Die mediale Funktion von statistischen Indizes

vorgelegt von
Martin H. Badicke
M.A.
geb. in Dresden

von der Fakultät I – Geistes- und Bildungswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Philosophie
- Dr. phil. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Walter Sendlmeier
Gutachterin:	Prof. Dr. Gisela Müller-Plath
Gutachter:	Prof. Dr. Norbert Bolz

Tag der wissenschaftlichen Aussprache: 4. Juli 2017

Berlin 2017

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich, Martin H. Badicke, die vorliegende Arbeit mit dem Titel „Die mediale Funktion von statistischen Indizes“ selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Ausführungen, die anderen veröffentlichten oder nicht veröffentlichten Schriften wörtlich oder sinngemäß entnommen wurden, habe ich kenntlich gemacht.

Die Arbeit hat in gleicher oder ähnlicher Fassung noch keiner anderen Prüfungsbehörde vorgelegen.

Ort, Datum: _____

Unterschrift: _____

"Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism."

Dave Barry

Abstract

The increasing number of indexes (composite indicators) in general and in media in particular seems to be a striking phenomenon concerning the communication of complex and multi-dimensional issues. Against the background of this communication activity the index generated by statistical methods is evidently a challenging object of research for media science and will be explored in this thesis. The objective of the analysis of the medial function of indexes will be pursued in two separate research directions, which reveal to be interdependent. The first one is concerned with the presence of indexes in media themselves. It will be analysed which indexes are treated and which function they fulfil in the given context. Initially detached, the second consideration pursues the question to what extent the index can be considered a (communication) medium in itself and which communication features are related to this aspect. The guiding approach of this thesis is an interdisciplinary method, where theories and instruments of different disciplines are combined within the frame of media science, after all philosophy, theory of science, history of science, sociology, informatics and statistics. An important finding of this interdisciplinary approach is that the medial function of indexes is strongly related to the structure of the measurement process in natural sciences and to its historical evolution. In the present thesis, this relation will be explored in depth and the actual prevailing conceptual ambiguities in the media (concerning indexes) will be illustrated in detail. Furthermore, on the basis of a detailed analysis of the mathematical-statistical construction methods of indexes a conceptual framework will be laid down in order to identify and evaluate factors, which influence the probability of communicational acceptance of indexes. Objectivity, reputation, transparency und trust reveal to be the important factors. Based on this theoretical discussion related to the construction and function of indexes as well as the basic conceptual framework and methods developed in the first chapters a quantitative content analysis for selected media has be conducted. For this reason, a specific innovative analytical tool has been developed by which also future content analysis might take advantage. The theoretical and empirical results show, that communication about indexes and by indexes is intrinsically imprecise, but allow to increase the probability to accept communication. In this sense the thesis show a third medial function of an index, which opens further possibilities of research.

Zusammenfassung

Ein auffälliges Phänomen im Zusammenhang mit der Kommunikation komplexer und mehrdimensionaler Sachverhalte ist die zunehmende Zahl von statistischen Indizes im Allgemeinen sowie in den Verbreitungsmedien im Besonderen. Vor dem Hintergrund dieser Kommunikationsleistung stellt der Index einen evident medienwissenschaftlichen Forschungsgegenstand dar, dem im Rahmen dieser Dissertation nachgegangen wird. Ziel der Arbeit ist die Analyse der medialen Funktion statistischer Indizes, die aus zwei Fragerichtungen jeweils einzeln sowie in gegenseitiger Abhängigkeit durchgeführt wird. Die erste Betrachtung konzentriert sich auf die Funktion von Indizes innerhalb der Verbreitungsmedien: Es wird untersucht, in welchem Kontext Indizes aufgegriffen werden und welche Funktionen sie dabei erfüllen. Zunächst losgelöst davon geht die zweite Betrachtung der Frage nach, ob Indizes selbst als (Kommunikations-)Medium betrachtet werden können und welche kommunikativen Eigenschaften damit einhergehen. Als Methode findet ein interdisziplinäres Vorgehen Verwendung, wobei die Theorien und Werkzeuge verschiedener Disziplinen innerhalb eines medienwissenschaftlichen Rahmens zusammengeführt werden. Dabei handelt es sich um Begriffe und Methoden aus der Philosophie, Wissenschaftstheorie, Wissenschaftsgeschichte, Soziologie, Informatik und Statistik. Eine wesentliche Erkenntnis dieser interdisziplinären Betrachtung ist es, dass die mediale Funktion von Indizes nicht unabhängig vom Bezug des Index auf den naturwissenschaftlichen Messprozess sowie dessen historischer Entwicklung ist. In der vorliegenden Arbeit wird dieser Bezug herausgearbeitet, Begrifflichkeiten werden präzisiert sowie aktuell vorherrschende terminologische Unklarheiten in den Verbreitungsmedien aufgezeigt. Zudem wird mit einer statistisch/mathematischen Betrachtung der Konstruktionsmethoden von Indizes eine Grundlage geschaffen, um Faktoren zu identifizieren und zu bewerten, die die Wahrscheinlichkeit für die Annahme einer Kommunikation von Indizes beeinflussen. Objektivität, Reputation, Transparenz und Vertrauen treten dabei als wichtige Faktoren hervor. Auf Basis der theoretischen Auseinandersetzung mit der Konstruktion und Funktion von Indizes sowie den dabei erarbeiteten Begrifflichkeiten und Methoden wird anschließend eine quantitative Inhaltsanalyse für ein ausgewähltes Verbreitungsmedium durchgeführt. Für die Inhaltsanalyse wurde ein neuartiges Analysetool entwickelt, von dem auch künftige inhaltsanalytische Arbeiten profitieren können. Die empirischen und theoretischen Ergebnisse der Analyse verdeutlichen, dass die Kommunikation mit und über Indizes unpräzise ist, jedoch erlauben es Indizes, die Wahrscheinlichkeit zur Annahme einer Kommunikation zu erhöhen. Daraus folgend erarbeitet die Dissertation eine dritte mediale Funktion von Indizes im instrumentellen Sinne, die weiteren Forschungen einen Anknüpfungspunkt bietet.

Inhaltsverzeichnis

Eidesstattliche Erklärung i

Abstract iii

Zusammenfassung iv

Inhaltsverzeichnis v

Abbildungsverzeichnis x

Tabellenverzeichnis xii

Abkürzungen xiv

1 Einleitung 1
- 1.1 Einleitung . . . 1
- 1.2 Relevanz . . . 12
- 1.3 Gegenstand . . . 13
- 1.4 Forschungsziel und Forschungsfragen . . . 14
- 1.5 Forschungsstand . . . 16
- 1.6 Methodisch-theoretisches Vorgehen . . . 18
- 1.7 Literatur und Material . . . 20
- 1.8 Aufbau der Arbeit . . . 21

2 Der Messprozess und die Bildung quantitativer Begriffe 26
- 2.1 Der Ursprung quantitativer Begriffe . . . 27
- 2.2 Der Messprozess nach Böhme . . . 29
  - 2.2.1 Quantifizierung . . . 31
  - 2.2.2 Metrisierung . . . 32
  - 2.2.3 Größe . . . 34
    - 2.2.3.1 Extensionale und intensionale Größen . . . 35
  - 2.2.4 Die Skalentheorie nach Stevens . . . 36
    - 2.2.4.1 Nominalskala . . . 39
    - 2.2.4.2 Ordinalskala . . . 40
    - 2.2.4.3 Intervallskala . . . 41
    - 2.2.4.4 Verhältnisskala und Absolutskala . . . 42
  - 2.2.5 Zahlen und das numerische System . . . 43
- 2.3 Bezug von Indikatoren und Indizes zum Messprozess . . . 45
  - 2.3.1 Indikatoren . . . 46

2.3.2 Indizes . . . 50

3 Evolution quantitativer Begriffe 56
3.1 Ursprünge des Messens . . . 57
3.1.1 Zahlen und Ziffern . . . 57
3.1.2 Die Ursprünge der Größenbegriffe . . . 59
3.2 Von Größenbegriffen zu Indikatoren und Indizes . . . 62
3.3 Zahlen in der Kommunikation . . . 69
3.3.1 Kommunikation von objektivem Wissen . . . 71
3.3.1.1 Der Wandel der Objektivität . . . 71
3.3.1.2 Die Ausdifferenzierung wissenschaftlicher Kommunikation 72
3.3.2 Objektivierungsleistung von Zahlen . . . 74
3.3.3 Zahlen und die Mobilisierung von Akzeptanz . . . 75

4 Die Konstruktion von Indizes 78
4.1 Funktionen und Probleme von Indizes . . . 79
4.2 Formalisierung des Indexproblems . . . 81
4.3 Schritte der Indexkonstruktion . . . 84
4.3.1 Theoretischer Rahmen . . . 87
4.3.2 Variablenselektion . . . 89
4.3.3 Datenqualitätsmanagement . . . 91
4.3.3.1 Fehlende Daten . . . 91
4.3.3.1.1 Arten von fehlenden Werten . . . 91
4.3.3.1.2 Fehlerursachen und ihre Konsequenzen . . . 93
4.3.3.1.3 Ausfallschema . . . 94
4.3.3.1.4 Quantifizierung des Datenausfalls . . . 94
4.3.3.2 Behandlung fehlender Daten . . . 95
4.3.3.2.1 Weitere Verfahren . . . 99
4.3.3.2.2 Fazit . . . 99
4.3.4 Prüfen der Datenstruktur . . . 100
4.3.4.1 Ebene der Indikatoren . . . 100
4.3.4.2 Ebene der Beobachtungen . . . 101
4.3.4.3 Modellerweiterung . . . 101
4.3.5 Die Aggregation der Daten . . . 103
4.3.5.1 Normalisierung der Daten . . . 103
4.3.5.2 Aggregationsform . . . 106
4.3.5.3 Gewichtung . . . 108
4.3.5.3.1 Faktorenanalyse . . . 112
4.3.5.3.2 Benefit-of-the-doubt-Gewichtung . . . 113
4.3.5.3.3 Unobserved components models . . . 118
4.3.5.3.4 Conjoint-Analyse . . . 121
4.3.5.3.5 Analytical hierarchy process . . . 123
4.3.5.3.6 Regressionsansatz . . . 126
4.3.5.3.7 Budget Allocation (Expertenmeinung) . . . 127
4.3.5.3.8 Öffentliche Meinung . . . 128
4.3.5.3.9 Entfernung zum Zielwert . . . 129
4.3.6 Prüfung der Konstruktion . . . 130

4.3.7 Kommunikation und Verbreitung . . . 133
4.3.7.1 Tabellen . . . 134
4.3.7.2 Visualisierung . . . 134
4.4 Beurteilung der Qualität eines Index . . . 147

5 Empirischer Teil: Inhaltsanalyse von Beiträgen mit Indexbezug 157
5.1 Vorgehen . . . 158
5.1.1 Explorative Datenanalyse . . . 159
5.1.2 Umsetzung der Automatisierung . . . 160
5.2 Untersuchungsinstrument . . . 162
5.2.1 Variablen und Variablenausprägung . . . 163
5.2.2 Beurteilung der Güte der Variablen . . . 165
5.3 Kriterien für die Selektion der Beiträge . . . 168
5.3.1 Beitrag mit Indexbezug . . . 168
5.4 Kriterien für die Bewertung der Beiträge . . . 169
5.4.1 Entwicklung der automatisch zu erhebenden Kriterien . . . 169
5.4.1.1 Direkte Kriterien . . . 169
5.4.1.1.1 Titel des Beitrags . . . 169
5.4.1.1.2 Datum . . . 170
5.4.1.1.3 Ressort . . . 170
5.4.1.1.4 Autor . . . 171
5.4.1.2 Indirekte Kriterien . . . 172
5.4.1.2.1 Wortanzahl . . . 172
5.4.1.2.1.1 Anzahl Wörter in der Überschrift . . . 173
5.4.1.2.1.2 Anzahl Wörter des Beitrages . . . 174
5.4.1.2.2 Anzahl Absätze . . . 175
5.4.1.2.3 Anzahl der Indexnennungen im Beitrag . . . 175
5.4.1.2.4 Satzlänge . . . 175
5.4.1.2.5 Nennung des Index im Titel . . . 177
5.4.1.2.6 Nennung des Index im ersten Absatz . . . 177
5.4.1.2.7 Nennung des Index im Beitrag . . . 178
5.4.1.2.8 Aktualität . . . 178
5.4.1.2.9 Anlass . . . 180
5.4.1.2.10 Auf Titelseite . . . 181
5.4.1.2.11 Illustration vorhanden . . . 181
5.4.1.2.12 Anzahl Zahlen . . . 181
5.4.1.3 Zusammenfassung der automatisierten Kriterien . . . 182
5.4.2 Entwicklung der nicht automatisch zu erhebenden Kriterien . . . 183
5.4.2.1 Indirekte Kriterien . . . 183
5.4.2.1.1 Genre . . . 183
5.4.2.1.2 Konflikt/Kontroverse/Problem . . . 185
5.4.2.1.3 Überraschung . . . 186
5.4.2.1.4 Zentralität . . . 186
5.4.2.1.5 Quellenangabe . . . 187
5.4.2.1.6 Vergleich . . . 187
5.4.2.1.7 Transparenz . . . 188
5.4.2.1.8 Argumentation . . . 188

5.4.2.2 Zusammenfassung der manuell zu erhebenden Kriterien . 188
5.4.3 Zusammenfassung der Bewertungskriterien 188
5.5 Wahl des Untersuchungsmediums 189
5.6 Selektion der Beiträge 192
5.6.1 Erstellung der Schlüsselwortliste 194
5.6.1.1 Abfrage der DBpedia 195
5.6.2 Auswahl zu analysierender Indizes 198
5.6.3 Aufgreifen der Beiträge 199
5.6.3.1 Die Suchmaske des FAZ-Archivs 202
5.6.3.2 Die Ergebnisse des FAZ-Archivs 203
5.7 Quantitative Inhaltsanalyse 204
5.7.1 Schema für die quantitative Inhaltsanalyse 204
5.7.2 Human Development Index 206
5.7.2.1 Inhaltsanalyse zum HDI 211
5.7.2.2 Der HDI im Kontext der Kommunikation 215
5.7.3 Happy Planet Index 219
5.7.3.1 Inhaltsanalyse zum HPI 224
5.7.3.2 Der HPI im Kontext der Kommunikation 228
5.7.4 OECD Better Life Index 231
5.7.4.1 Inhaltsanalyse zum BLI 237
5.7.4.2 Der BLI im Kontext der Kommunikation 242
5.7.5 GfK-Konsumklima-Index (KKI) 243
5.7.5.1 Inhaltsanalyse zum GfK-Index 245
5.7.5.2 Der GfK-Index im Kontext der Kommunikation 249
5.7.6 ifo-Index 251
5.7.6.1 Inhaltsanalyse zum ifo-Index 255
5.7.6.2 Der ifo-Index im Kontext der Kommunikation 264
5.7.7 Body-Mass-Index 265
5.7.7.1 Inhaltsanalyse Body-Mass-Index 267
5.7.7.2 Der Body-Mass-Index im Kontext der Kommunikation . 272
5.7.8 Michigan Consumer Sentiment Index 276
5.7.8.1 Inhaltsanalyse zum MCSI 279
5.7.8.2 Der MCSI im Kontext der Kommunikation 283
5.7.9 Deutscher Aktienindex 288
5.7.9.1 Inhaltsanalyse zum DAX 292
5.7.9.2 Der DAX im Kontext der Kommunikation 296
5.7.10 Korruptionswahrnehmungsindex 301
5.7.10.1 Inhaltsanalyse zum CPI 304
5.7.10.2 Der CPI im Kontext der Kommunikation 309
5.7.11 Gegenüberstellung und Zusammenfassung der Untersuchungseinheiten 311

6 Fazit **316**

Literaturverzeichnis **330**

A Tabellen **346**

B Diagramme **361**

C Abfragen **365**

D R-Code **366**

E Dokumentation **368**
E.1 Überblick über das R-Paket „analyseMedia“ . . . 368
E.2 Installation . . . 368
E.3 Anwendung der Module . . . 369
E.3.1 Selektion . . . 369
E.3.2 Datenabruf . . . 370
E.3.3 Datenverarbeitung und Erstellung deskriptiver Statistiken . . . 372
E.3.4 Stichprobenziehung . . . 373
E.4 Anpassung . . . 374
E.4.1 Hinzufügen eines neuen digitalen Mediums . . . 375
E.4.2 Anpassen einer R-Funktion . . . 375

Abbildungsverzeichnis

1.1 Unterscheidung des wissenschaftlichen „Messens“ 7

2.1 Messprozess nach Böhme 30
2.2 Skalen mit empirischer Grundoperation, mathematischer Gruppenstruktur und möglicher Statistik 38
2.3 Schematische Darstellung des indirekten Messens 47

4.1 Fehlerschemata 94
4.2 Zusammenhang zwischen Indikatoren und latentem Konstrukt 111
4.3 Restriktionsarten 116
4.4 Darstellung der Anteile 117
4.5 Beispiel für eine Tabelle mit dem Ranking von Ländern nach dem Happy Planet Index 135
4.6 Beispiel für eine Tabelle mit Grenzwerten 136
4.7 Poster zum OECD Better Life Index 137
4.8 Visualisierungstypen 144
4.7 Visualisierungstypen (Fortsetzung) 145

5.1 Module des R-Pakets mit Mapping zum Vorgehen 162
5.2 Beispiel für ein Liniendiagramm mit kumulierten Häufigkeiten 179
5.3 Yago-Graph 200
5.4 Schema für die Auswahl der zu analysierenden Indizes 200
5.5 Schema für die Analyse eines Index 208
5.6 Konstruktion des Human Development Index 208
5.7 Anzahl Beiträge zum Human Development Index 212
5.8 Aktualität der Beiträge zum Happy Planet Index 213
5.9 Anzahl Beiträge zum Happy Planet Index 224
5.10 Aktualität der Beiträge zum Happy Planet Index 225
5.11 Konzept zur Messung des Wohlbefindens 232
5.12 Anzahl Beiträge zum Better Life Index 238
5.13 Aktualität der Beiträge zum Better Life Index 239
5.14 Aktualität der Beiträge zum Gfk-Index 246
5.15 Anzahl Beiträge zum Gfk-Index 247
5.16 Aktualität der Beiträge zum ifo-Index 256
5.17 Anzahl Beiträge zum ifo-Index 257
5.18 Der ifo-Index im monatlichen Konjunkturbericht der FAZ 259
5.19 Anzahl Beiträge zum Body-Mass-Index 268
5.20 Aktualität der Beiträge zum MCSI 280

5.21 Anzahl Beiträge zum MCSI 281
5.22 Anzahl Beiträge zum DAX 293
5.23 Aktualität der Beiträge zum CPI 305
5.24 Anzahl Beiträge zum CPI 305

6.1 Schritte der Indexkonstruktion in der FAZ 321

B.1 Beispiel für ein Liniendiagramm 361
B.2 Beispiel für ein Streudiagramm 362
B.3 Beispiel für ein Balkendiagramm 362
B.4 Beispiel für ein gestapeltes Säulendiagramm 363
B.5 Beispiel für ein gestapeltes Radialsäulendiagramm 364

E.1 Dokumentation zum R-Paket „analyseMedia“ 378

Tabellenverzeichnis

2.1 Alternative Bezeichnungen für Index und Indikator 52

4.1 Vor- und Nachteile von Indizes 82
4.2 Vor- und Nachteile der Verfahren zur Prüfung der Datenstruktur 102
4.3 Vor- und Nachteile der Benefit-of-the-doubt-Gewichtung 118
4.4 Vor- und Nachteile des UC-Modells 121
4.5 Vor- und Nachteile der Conjoint-Analyse 123
4.6 Vergleichsmatrix mit drei Indikatoren 124
4.7 Relative Gewichte der drei Indikatoren 125
4.8 Vor- und Nachteile des AHP 126
4.9 Vor- und Nachteile des Regressionsansatzes 127
4.10 Vor- und Nachteile der Budget Allocation 128
4.11 Vor- und Nachteile der Öffentlichen Meinung 129
4.12 Vor- und Nachteile „Entfernung zum Zielwert“ 129
4.13 Visualisierungstypen für Indexwerte 137
4.14 Dimensionen und Kriterien des Data Quality Frameworks 148
4.15 Matrix zur Beurteilung der Konstruktions-Qualität eines Index 156

5.1 Variablentypen der Inhaltsanalyse 164
5.2 Tabelle für die Zuordnung der Ressorts von FAZ und FAS 171
5.3 Automatisch zu erhebende Kriterien für die Inhaltsanalyse 183
5.4 Manuell zu erhebende Kriterien für die Inhaltsanalyse 189
5.5 Beschreibung der numerische Kennzahlen 205
5.6 Ressort-Statistik zum Human Development Index 213
5.7 Numerische Kennzahlen zum Human Development Index 214
5.8 Ergebnisse der manuellen Inhaltsanalyse zum HDI 216
5.9 Ressort-Statistik zum Happy Planet Index 226
5.10 Numerische Kennzahlen zum Happy Planet Index 226
5.11 Ergebnisse der manuellen Inhaltsanalyse zum HPI 227
5.12 Ressort-Statistik zum Better Life Index 239
5.13 Numerische Kennzahlen zum Better Life Index 240
5.14 Ergebnisse der manuellen Inhaltsanalyse zum BLI 241
5.15 Numerische Kennzahlen zum GfK-Index 248
5.16 Ressort-Statistik zum GfK-Index 249
5.17 Ergebnisse der manuellen Inhaltsanalyse zum GfK-Index 252
5.18 Numerische Kennzahlen zum ifo-Index 259
5.19 Ressort-Statistik zum ifo-Index 260
5.20 Ergebnisse der manuellen Inhaltsanalyse zum ifo-Index (I) 261

5.21 Ergebnisse der manuellen Inhaltsanalyse zum ifo-Index (II) 262
5.22 Ergebnisse der manuellen Inhaltsanalyse zum ifo-Index (III) 263
5.23 Ressort-Statistik zum Body-Mass-Index 268
5.24 Numerische Kennzahlen zum Body-Mass-Index 269
5.25 Ergebnisse der manuellen Inhaltsanalyse zum BMI (I) 273
5.26 Ergebnisse der manuellen Inhaltsanalyse zum BMI (II) 274
5.27 Numerische Kennzahlen zum MCSI 281
5.28 Ressort-Statistik zum MCSI 282
5.29 Ergebnisse der manuellen Inhaltsanalyse zum MCSI 284
5.30 Ergebnisse der manuellen Inhaltsanalyse zum MCSI 285
5.31 Ergebnisse der manuellen Inhaltsanalyse zum MCSI 286
5.32 Anzahl Tage zwischen Nennung des DAX 292
5.33 Numerische Kennzahlen zum DAX 294
5.34 Ressort-Statistik zum DAX 294
5.35 Ergebnisse der manuellen Inhaltsanalyse zum DAX (I) 297
5.36 Ergebnisse der manuellen Inhaltsanalyse zum DAX (II) 298
5.37 Ergebnisse der manuellen Inhaltsanalyse zum DAX (III) 299
5.38 Numerische Kennzahlen zum CPI 306
5.39 Ressort-Statistik zum CPI 307
5.40 Ergebnisse der manuellen Inhaltsanalyse zum CPI 308
5.41 Ressort-Statistik je Index in Prozent 313
5.42 Mittelwerte je Index 315

A.1 Einträge in der deutschen Wikipedia zum Konzept oder Label „Index“ . . 346

E.1 Funktionen des R-Pakets „analyseMedia“ 376

Abkürzungen

AHP	**A**nalytic **H**ierarchy **P**rocesses
BAP	**B**udget **A**llocation **P**rocesses
BIP	**B**rutto **I**nlands **P**rodukt
BOD	**B**enefit **O**f The **D**oubt
CA	**C**onjoint **A**nalysis
FAS	**F**rankfurter **A**llgemeine **S**onntagszeitung
FAZ	**F**rankfurter **A**llgemeine **Z**eitung
MAR	**M**issing **A**t **R**andom
MCAR	**M**issing **C**ompletely **A**t **R**andom
MNAR	**M**issing **N**ot **A**t **R**andom
ML	**M**aximum **L**ikelihood
OAR	**O**bserved **A**t **R**andom
OECD	**O**rganisation for **E**conomic **C**o-operation and **D**evelopment
RDF	**R**esource **D**escription **F**ramework
TI	**T**ransparency **I**nternational
UCM	**U**nobserved **C**omponents **M**odels
UN	**U**nited **N**ations
UNDP	**U**nited **N**ations **D**evelopment **P**rogramme

Kapitel 1

Einleitung

1.1 Einleitung

Alle drei Jahre erhebt der *Happy Planet Index* den Anspruch, die Zufriedenheit und ökologische Effizienz eines Landes zu „messen“. Mit den Ergebnissen werden ausgewählte Nationen in eine Rangfolge gebracht und damit das subjektive und mehrdimensionale Konstrukt „Glück“ kommunizierbar gemacht.

Kann eine Aussage wie *Vietnam ist glücklicher als Deutschland* jedoch ohne persönliche Interaktion überhaupt akzeptiert werden? Kann sie als objektive und interpretationsfreie Beschreibung „der“ Realität angesehen werden? Existieren Möglichkeiten, um die Annahme einer solchen Kommunikation wahrscheinlicher zu machen?

Eines der auffälligsten Phänomene im Zusammenhang mit der Kommunikation von Komplexität und Mehrdimensionalität ist die zunehmende Zahl von Indizes im Allgemeinen sowie in den Verbreitungsmedien im Speziellen. Vor dem Hintergrund dieser Kommunikationsleistung stellt die Betrachtung des Index ein evident medienwissenschaftliches Problem dar.

Der Titel der vorliegenden Arbeit beinhaltet eine bewusst gewählte Ambiguität hinsichtlich der Hypothese über mediale Funktionen von Indizes. Dabei handelt es sich um eine Zweideutigkeit, die auch in der Analyse und Darstellung Niederschlag finden wird. So wird sich der Forschungsfrage zur *medialen Funktion* aus folgenden Betrachtungsweisen genähert:

1. Funktion des Index *in* den (Verbreitungs-) Medien

2. Funktion des Index *als* Medium

Diesen zwei Fragerichtungen soll sowohl separat als auch in gegenseitiger Abhängigkeit nachgegangen werden. Die erste Betrachtung konzentriert sich auf die Funktion von Indizes innerhalb der Verbreitungsmedien. Es ist zu untersuchen, in welchem Zusammenhang Indizes aufgegriffen werden und welche Funktionen sie an dieser Stelle erfüllen. Zunächst losgelöst davon geht die zweite Betrachtung der Frage nach, ob Indizes selbst als Medium betrachtet werden können und welche Rolle ihnen in diesem Sinne zukommt. Da bei der Diskussion, ob sie selbst *als Medium* zu verstehen sind, gewisse Überschneidungen mit den Eigenschaften von *auf Messung basierenden Zahlen* zu erwarten sind, wird die Betrachtung um diesen Punkt erweitert.

Aus beiden Betrachtungsweisen resultiert eine dritte – die Funktion des Index im instrumentellen Sinne. Hierbei geht es um die Annahme, dass *durch* den Index Systeme oder andere Medien beeinflusst werden. Beispielsweise beeinflusst der *Deutsche Aktienindex* die Märkte oder der *Body Mass Index* das individuelle Körperempfinden.

Die Forschungsfrage dieser Arbeit resultiert im weitesten Sinne aus der Beobachtung heraus, *alles messbar* oder *scheinbar messbar zu machen.* Eine Erscheinung der modernen zahlen- und datengetriebenen Welt ist die Zunahme der Produktion, Verarbeitung und des Konsums von Daten. „Big Data“ oder „Quantified Self“ sind nur zwei Beispiele, die im Zusammenhang mit dieser Entwicklung stehen. Trotz der damit verbundenen Komplexität und Mehrdimensionalität wird neben rein naturwissenschaftlichen Phänomenen auch versucht, sowohl intra- und interpsychologische als auch interindividuelle Prozesse (z.B. in Gesellschaft, Wirtschaft oder Soziologie) in Zahlen zu übersetzen. Der Körper wird mittels Fitnessarmbändern kontrolliert, die am Ende des Tages aufgrund der Schrittanzahl Bilanz ziehen; Kalorien können durch die Eingabe der Lebensmittel über den Tag gezählt werden, Partner können durch den Vergleich bestimmter Daten in den Profilen zueinander finden. Diverse Anwendungen für Smartphones und andere Geräte unterstützen dabei, zu messen, zu sammeln und zu vergleichen. Häufig werden diese Daten über (soziale) Netzwerke kommuniziert.

Der Autor dieser Arbeit wurde als Mitglied eines universitären Forschungsteams ebenfalls mit der Produktion und Verarbeitung von Daten konfrontiert. Im Forschungsprojekt

wurde der Versuch unternommen, den Begriff „Vertrauen“ auf objektive Weise und aus einer naturwissenschaftlichen Herangehensweise zu definieren und eine Unterscheidung nach *mehr* und *minder* möglich zu machen. Nach Prüfung verschiedener statistischer Methoden sollte dafür das Konstrukt eines statistischen *Index* näher untersucht werden. Dabei stand im Mittelpunkt, ob die Konstruktion eines Index dem wissenschaftlichen Anspruch des Projekts gerecht werden kann und inwieweit er die Kommunikation der Ergebnisse ermöglicht – insbesondere vor dem Hintergrund der eingangs aufgeworfenen Fragen zum *Happy Planet Index*.

Bei dem Besuch einer Konferenz zum Thema Vertrauen kam es zur Präsentation von zwei existierenden Ansätzen zur Konstruktion eines Vertrauensindex. Beide existieren seit mehreren Jahren parallel und erhalten mediale Beachtung. Jedoch ist anzunehmen, dass beide Indizes trotz der gemeinsamen Bezeichnung *Vertrauensindex* verschiedene Gegenstände abbilden. Schließlich sind die Auswahl der konstituierenden Merkmale sowie die Ergebnisse und Aussagekraft unterschiedlich. Das legt die Annahme nahe, dass die Merkmale den Gegenstand konstituieren.

Sollten beide Indizes verschiedene Gegenstände abbilden, ist für eine eindeutige Kommunikation zusätzliches Wissen über das jeweilige Indexkonzept notwendig. Andernfalls kann der Eindruck entstehen, beide operationalisieren den gleichen Begriff „Vertrauen“ oder den gleichen Gegenstand. Falls keine weiteren Informationen zum Indexkonzept vorliegen wird u.U. die Glaubwürdigkeit der Indizes in Frage gestellt. Bei unterschiedlichen Indexwerten stellt sich die Frage der Objektivität, bei gleichen Ergebnissen die der Redundanz.

Einer der Indizes wird von einer Agentur herausgegeben. Überraschend dabei erschien die Aussage auf der Konferenz, dass die wesentliche Funktion des Index gar nicht die Messung sei, sondern vielmehr die Medienwirksamkeit und die Möglichkeit, mit Zahlen eine „Story“ zu erzählen. Augenscheinlich handelte es sich dabei also um ein Instrument, um ein Verkaufspotential zu erhöhen und eine PR-Maßnahme mit Zahlen zu motivieren. Die Objektivität und Sinnhaftigkeit der Messung hingegen wurden kompromissbereit als nebensächlich und irrelevant abgetan, zumal die Konstruktionskriterien z.B. durch eine ständig wechselnde Auswahl an Indikatoren wissenschaftlich fragwürdig erschienen.

Genau an dieser Stelle – beim Vergleich des vom Autor der vorliegenden Arbeit vertretenen Indexkonzepts mit dem von der Agentur – wurden die verschiedenen Funktionen von

Indizes deutlich. Der vom Autor verfochtene wissenschaftliche Anspruch bedient sich des Index zur Darstellung und Kommunikation eines Konzepts im wissenschaftlichen Sinne. An die Konstruktion des Index werden also die gleichen Anforderungen gestellt wie an Messungen im klassischen Sinne. Die Selektion der konstituierenden Merkmale und die Konstruktionsmethoden richten sich nach dem wissenschaftlichen Anspruch der Objektivität und Reliabilität. Das bedeutet, dass die Ergebnisse der Konstruktion unabhängig von Ort, Raum und Person experimentell reproduzierbar sein sollten. Außerdem wird an das Konstruktionsmodell der Anspruch gestellt, dass es optimal ist – also aus der Vielzahl an möglichen Modellen das Konzept in optimaler Weise abbildet. Die Funktion des Index innerhalb der Verbreitungsmedien ist in dieser Hinsicht irrelevant.

Im Gegensatz dazu bedient sich der angesprochene Agentur-Ansatz der Funktion eines statistischen Index innerhalb der Verbreitungsmedien – unabhängig von seiner intrinsischen Information. Diese Vermutung wird gestärkt von der bereits weiter oben erwähnten Aussage, dass die (intrinsische) Funktion der Abbildung nebensächlich sei. Vermutlich ist die wesentliche Funktion in diesem Fall der „Hinweischarakter“. Indem sich der Index durch das Zeichen der Zahl von textuellen Informationen abhebt, signalisiert er: „Hier ist etwas wichtig, hier wurde verdichtet“. Dadurch suggeriert er Relevanz und ggf. auch Objektivität.

Die Erkenntnis der verschiedenen Anwendungen führt zu der Frage, welche Funktionen Indizes innerhalb der Verbreitungsmedien besitzen und ob noch weitere neben den o.g. existieren. Damit verbunden ist die Frage, ob ein Abstraktum wie „Vertrauen“ überhaupt messbar ist und anhand bestimmter Merkmale konkret dargestellt werden kann. Oder hat Heinrich von Kleist Recht, wenn er behauptet:

> „Das Glück kann nicht, wie ein mathematischer Lehrsatz bewiesen werden, es muß empfunden werden, wenn es da sein soll.“ (Kleist v., 1799/1997: S. 44f.)

Abstrakte, nicht-gegenständliche Begriffe, entziehen sich einer direkten Beobachtung. Können diese latenten Konstrukte dennoch mit Zahlen ausgedrückt werden? Welche Indikatoren können zur objektiven Beschreibung von Abstrakta beitragen und sie somit messbar machen?

Hier offenbart sich eine Schnittstelle zwischen Geistes- und Naturwissenschaften, wo Sprache allein nicht mehr objektiv genug zu sein scheint und Zahlen nicht genügen, um Bedeutung zu generieren. Dass der Index als Vermittler an diese Stelle treten kann, ist eine der Hypothesen, die diese Arbeit überprüfen will. Er vermag es, durch eine scheinbar objektive Zahl ein komplettes Konzept zu vertreten und eine ganze Datenfülle in einer Aussage zu komprimieren. Diffuse Annahmen über Begrifflichkeiten können dadurch auf einen Nenner gebracht und vergleichbar gemacht werden – so zumindest die Vermutung. Mit dieser Mittlerfunktion steht der Index in dieser Arbeit im Forschungsmittelpunkt. Die Frage nach seiner Funktion in den Verbreitungsmedien geht einher mit der Frage nach seiner Rolle als Kommunikationsmedium, die ebenfalls diskutiert werden soll.

Zunächst aber sollte in dieser Einleitung eine Annäherung an den Begriff Index selbst geschehen. Ist der Index in der Statistik definiert als eine Kennzahl, die das Ergebnis der Quantifizierung eines theoretischen Konzeptes anzeigt, so gibt es weitere gebräuchliche Indizes, die in den Medien repräsentiert sind. Ein häufiges Auftauchen des Begriffes, der im Übrigen etymologisch vom Lateinischen *indicere* – ansagen, melden oder *indicare* – anzeigen abgeleitet wird, ist in der Bedeutung des Index als eine Liste oder ein Verzeichnis festzustellen, wie zum Beispiel der Index der verbotenen Literatur. Auch in Büchern selbst stößt der Leser nicht selten auf einen Index. Dieser besteht meist aus Schlagwörtern, die mit einem Hinweis zu den Seitenzahlen, auf denen der jeweilige Begriff zu finden ist, versehen sind. Der Index ist in diesem Fall eine Art „Hinweisschild“ oder „Wegweiser“. In der Mathematik hingegen wird er beispielsweise zur Nummerierung unterschiedlicher Objekte herangezogen. Gleichbleibend in den Begrifflichkeiten ist jene Bedeutung, bestimmte Daten übersichtlich abzubilden.

In der Medizin ist *Index* passenderweise die Bezeichnung für den Zeigefinger – der Finger, der etwas anzeigt, indem er darauf deutet oder der erhobene Zeigefinger, der zum Innehalten ermahnt. Und nicht zuletzt ist eine vermutete Funktion über das Aufgreifen von Indizes in den Verbreitungsmedien jene, deutlich auf einen Sachverhalt hinzuweisen. Wie der Zeigefinger ragt die Kennzahl inmitten der Worte heraus und ist nicht nur ein „Stolperstein für das Auge“, sondern scheint auch der eindringliche Beweis von Wissenschaftlichkeit. Denn auch in der Semiotik bezeichnet der Index ein Zeichen, das direkt auf eine real-weltliche Erscheinung hinweist und mit dieser in räumlicher und zeitlicher Beziehung steht. Das ist eine Eigenschaft, die dem statistischen Index mehr oder

weniger ebenso zu eigen ist. Er macht es möglich, ausgewählte Referenten zeichenhaft darzustellen und auf deren Existenz hinzuweisen.

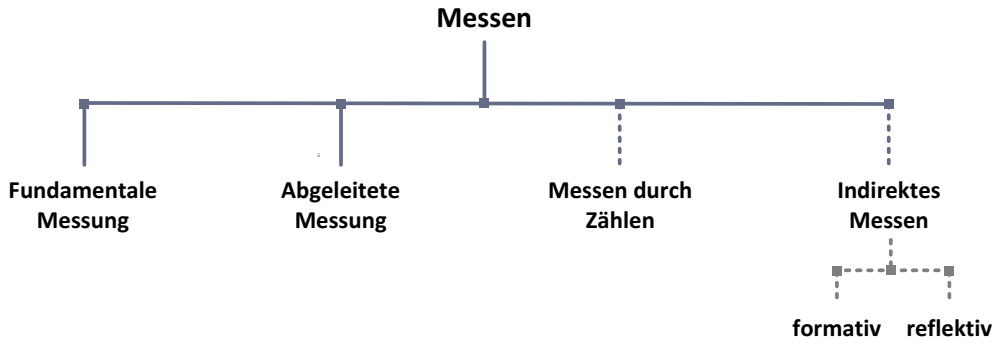
Dennoch weisen Indizes wie der *Happy Planet Index*, *Better Life Index*, *Big-Mac-Index* oder *Wolkenkratzerindex* nicht allein durch die Namensgebung auf deren Referenten hin und geben damit zunächst Zweifeln an ihrer Objektivität Raum. Die Bezeichnung signalisiert, dass es sich um eine quantitative Information handelt, jedoch lässt sich dem genannten Quantifizierungsobjekt keine eindeutige quantitative Struktur zuordnen. Entweder weil es sich um einen qualitativen Begriff handelt oder weil nicht unmissverständlich klar ist, um was es sich bei dem Gegenstandsbereich handelt. Zudem scheinen die Konzepte, die dahinter zu vermuten sind, eher oberflächlicher und labiler Natur zu sein. Was macht ein gutes Leben aus? Vermutlich ändern sich die maßgeblichen Kriterien zur Beantwortung der Frage in Abhängigkeit von Alter, Lebensstandard, Umfeld usw.. Kann dieses mehrdimensionale und äußerst volatile Konstrukt wirklich objektiv z.B. vom *Better Life Index* quantifiziert werden?

Eine ähnliche Frage kann bei dem mehrdimensionalen Konzept „Klimaerwärmung“ gestellt werden. Die Klimaerwärmung wird z.T. in den Medien mit einer Genauigkeit bis zur zweiten Nachkommastelle in Grad Celsius kommuniziert. Für den Informationsempfänger entsteht vermutlich der Eindruck, es handelt sich bei der Zahl um eine Messung. Tatsächlich ist jedoch auch diese Zahl als Index zu betrachten. Verschiedene Messpunkte für die Temperatur werden auf Basis eines Modells zusammengefasst und in einer Maßzahl kommuniziert. Auch hier ist die Frage zu stellen, wie objektiv *Anzahl* und *Ort* der Messpunkte sind und wie sensibel die resultierende Zahl auf Änderungen reagiert.

Bei einer Quantifizierung „über“ Indizes scheint immer auch ein Raum für Subjektivität und Willkür zu existieren. An welcher Stelle dieser Raum entsteht und welche Möglichkeiten bestehen, diesen zu kontrollieren, soll ebenfalls untersucht werden. Schließlich haben andere Disziplinen, allen voran die Physik, in harter und langwieriger Arbeit Methoden und Konzepte entwickelt, um Subjektivität und Willkür minimieren zu können. Wo genau der Unterschied zwischen den *ursprünglichen* Messungen z.B. in der Physik und den neueren *indikatorbasierten* Messsystemen liegt, zeigt die Messtheorie.

Aus dieser geht – historisch betrachtet – im Wesentlichen die Unterscheidung zwischen

Abbildung 1.1: Unterscheidung des wissenschaftlichen „Messens“ (nicht-durchgezogene Verbindungslinien signalisieren einen ungeklärten Status in der Fachliteratur)



Quelle: Eigene Darstellung

vier Messungen hervor. Eine Form der Messung ist die *„ursprüngliche“ Messung*. Ihre Eingrenzung wurde wesentlich von Campbell unter der Bezeichnung *fundamentales Messen* geprägt. Darunter ist die Zuordnung von Zahlen zu Objekten und Eigenschaften zu verstehen. Sie wird aus den „Naturgesetzen“ abgeleitet, welche auf das zu messende Phänomen einwirken. An dieser Stelle ergibt sich eine wichtige Eigenschaft der *fundamentalen Messung* – sie ist unabhängig von der Messung anderer Größen. Die Operationen des Vergleichs und der Kombination sind experimentell nachvollziehbar, was sie unabhängig von Raum, Zeit und Person reproduzierbar macht. So wird beispielsweise die *„Masse“* eines Körpers fundamental über das Verhalten des Objektes in einem reproduzierbaren Raum gemessen. Bei der Messung über eine Federwaage wird dafür die Dehnung der Feder benutzt. Dabei soll das Verhalten der *Dehnung* über das theoretisch eingeführte Konzept „Masse“ erklärt werden. Dessen Einführung verleiht den Vergleichs- und Kombinationsoperationen sowie den Zuordnungsregeln erst ihre theoretische Bedeutung. Die Verifikation letzterer erfolgt durch die Wirkung der Naturgesetze.

Eine weitere Art der Messung wurde von Campbell als *abgeleitete Messung* bezeichnet. Hierbei werden die Regeln für die Zuordnung nicht aus den Naturgesetzen abgeleitet, sondern aus einer bestehenden Theorie. Die konstituierenden Variablen der Theorie sind messbar oder wurden bereits gemessen, das zu messende Konstrukt jedoch nicht. So wird z.B. die *„Kraft“* mittels des zweiten Newtonschen Gesetzes über die *Masse* und die *Beschleunigung* bestimmt. Abgesehen von Fehlern durch die Messapparatur sind

abgeleitete Messungen ebenfalls reliabel, da ihre Bestandteile fundamental bestimmt werden und ihr Zusammenhang über die Naturgesetze determiniert ist.

Ob Zählen als Messung bezeichnet werden kann, ist nicht abschließend geklärt. Die von Coleman (1964: S. 71) geprägte Phrase *Messen durch Zählen* wird innerhalb der verschiedenen Wissenschaftsdisziplinen kontrovers diskutiert. Ausgangspunkt der Phrase ist die Soziologie, in der Quantifizierung oftmals auf bloßen Zählungen basiert. Die klassischen Messkriterien werden dabei lediglich implizit erfüllt, da die Operationen des Vergleichs und der Kombination auf einer Klasseneinteilung basieren (z.B. *Anteil Erwerbsloser eines Landes*). Die Grundlage der „Messung“ ist somit eine Theorie und nicht das Verhalten der Objekte.

Mit dem Anspruch, abstrakte Begriffe wie „Vertrauen“ oder „Glück“ zu quantifizieren, stoßen die *fundamentale* und abgeleitete Messung sowie das *Messen durch Zählen* an ihre Grenzen. Sie lassen derartige Messungen nicht zu, da die Begriffe nicht gegenständlich sind und auch kein bekannter Zusammenhang zu den Naturgesetzen besteht.

In Folge dessen hat sich in der Evolution des Messens die *indirekte Messung* herausgebildet. Hier wird das zu bestimmende Konzept nicht direkt, sondern indirekt über Indikatoren greifbar gemacht. Der Unterschied zum abgeleiteten Messen offenbart einen potentiellen Raum für Willkür. Kann der Zusammenhang der konstituierenden Variablen in der abgeleiteten Messung über Naturgesetze verifiziert werden, wird er bei der indirekten Messung definitorisch oder statistisch aus den Daten bestimmt.

Bei der Konstruktion von Indizes für die Quantifizierung eines Konzepts, z.B. „Glück“, wird allgemein von der Reduktion von Komplexität gesprochen. Oftmals handelt es sich um einen *mehrdimensionalen Merkmalsraum*, der über das Indexkonzept anhand eines meist eindimensionalen Maßes beschrieben werden soll. Der Vorteil eines solchen eindimensionalen Maßes ist Einfachheit und Übersichtlichkeit. Außerdem ermöglicht es den Vergleich von komplexen Sachverhalten z.B. in Form von Rangfolgen und trägt damit zur Reduktion der Informationsfülle bei. Die Objektivität des Index wird zum einen davon beeinflusst, wie die jeweiligen Merkmale des Index in den Zahlenraum überführt werden und zum anderen, wie diese unter dem Indexmodell selektiert und konkateniert werden.

Die Konstruktion eines Index kann zwei unterschiedlichen theoretischen Überlegungen folgen. Eine sieht die Indikatoren des Konzepts als Resultat, die andere als Ausgangspunkt. In beiden Varianten steht der resultierende Index für ein latentes Konstrukt. Lediglich die Wirkungsrichtung ist verschieden.

In ersterem Fall wird von *formativen Indikatoren* zur Konstruktion eines Index gesprochen. Das bedeutet, dass das latente Konstrukt als eine gewichtete Zusammenfassung seiner Indikatoren betrachtet werden kann. Bei formativen Modellen zur Konstruktion von Indizes werden die Indikatoren kausal als Ursache des latenten Konstruktes betrachtet. Ein formatives Modell zur Indexkonstruktion könnte z.B. zur Konstruktion eines Index zur Messung der Servicequalität eines Unternehmens benutzt werden. Verschiedene Merkmale (z.B. Bearbeitungsdauer, Kundenzufriedenheit, Fehleranzahl etc.) werden zu einer Grundaussage verdichtet. Die „Qualität“ wird über die selektierten Merkmale konstituiert. Wird ein Merkmal ausgetauscht oder eliminiert, ändert sich auch der Index.

Im zweiten Fall wird von *reflektiven Indikatoren* zur Konstruktion eines Index gesprochen. Hier wird das latente Konstrukt als (mathematische) Funktion seiner beobachteten Indikatoren modelliert. Dabei wird angenommen, dass sich das gesuchte latente Konstrukt in verschiedenen Indikatoren manifestiert und in diesen beobachtet werden kann. Ein Austauschen oder Eliminieren von Indikatoren verändert nicht das Konstrukt, sondern lediglich die Genauigkeit der „Messung“. Ein populäres Beispiel für ein solches Modell ist die Bestimmung des Konstrukts Intelligenz. Dieses entzieht sich einer direkten Beobachtung, kann aber über den Intelligenzquotienten und seine verschiedenen Merkmale „gemessen“ werden. Dabei wird angenommen, dass sich die Intelligenz in den beobachtbaren Merkmalen (z.B. Merkfähigkeit, numerische Kompetenz etc.) manifestiert. Der manifeste Anteil wird über das reflexive Modell zur Bestimmung des latenten Konstrukts *Intelligenz* benutzt.

Beim indirekten Messen sind die Zuordnungsregeln der Zahlen zu den Objekten arbiträr. Das heißt, dass sie aus den postulierten Zusammenhängen der messbaren Objekte des Konzepts und dem Konzept selbst abgeleitet werden. Bei der Beantwortung der Frage, ob es sich bei Indizes tatsächlich um eine Messung handelt und was sie eigentlich abbilden, ist die differenzierte Betrachtung nach der „Wirkungsrichtung“ sinnvoll. Wie später

aufgezeigt wird, lassen sich formative Indizes als reines Aggregationsproblem betrachten. Die Methoden der Konkatenation werden im Wesentlichen auf einer rein syntaktischen Ebene entwickelt und dem verdichteten Konzept fehlt per Definition ein Bezug zur Realität. Es existiert nicht losgelöst von den Indikatoren, sondern wird von diesen determiniert.

Jedoch tritt in der Geschichte des Messens eine solche terminologische Verwirrung, bei der sich Unklarheiten hinsichtlich des Messprozesses offenbaren, nicht zum ersten Mal auf. Denn bei der Betrachtung des Messprozesses geht es nicht ausschließlich um wissenschaftstheoretische Überlegungen. Vielmehr muss auch bei Indizes die traditionell erkenntnistheoretische Frage nach der Beziehung der Konstitution des Gegenstandes zu dem Erkenntnisvermögen gestellt werden. Teilweise wird dieser Frage ausgewichen, indem die Operationen innerhalb des numerischen Systems mit Begrifflichkeiten wie „Hyperrealität“ oder „Metameasurement“ von der Betrachtung ausgenommen werden.

Auf die kommunikativen Eigenschaften und Funktionen von Zahlen haben diese messtheoretischen Überlegungen und Erkenntnisse „mehr oder weniger“ Einfluss. Generell besitzen *auf Messung basierende Zahlen* ganz spezifische kommunikative Eigenschaften, die es unter bestimmten Voraussetzungen zulassen, sie als *symbolisch generalisiertes Kommunikationsmedium* zu bezeichnen. Die Zunahme ihrer Bedeutung hängt u.a. mit der Vermittlung objektiven Wissens, vornehmlich innerhalb der wissenschaftlichen Kommunikation, zusammen. Diese änderte sich im 18. und 19.Jahrhundert z.T. dahingehend, dass die Bedeutung der persönlichen Ansprache in Briefform zugunsten einer unpersönlichen Zeitschriftenform abnahm. Akzeptanz konnte weniger über persönliche Interaktion hergestellt werden. In diesem unpersönlichen Rahmen erhöhen Zahlen die Akzeptanz von Kommunikation, indem sie z.B. Objektivität signalisieren oder anzeigen, dass ihre Selektion bestimmte Voraussetzungen erfüllt. Außerdem wird der Anschein von Unpersönlichkeit und einer interpretationsfreien Beschreibung geweckt. Ähnlich wie bei Fotos entsteht bei Zahlen der Eindruck, sie stünden für die „Realität“ an sich. Zahlen, genau wie Fotos, sind schwieriger negierbar als Aussagen, da sie die Negation nicht direkt in sich tragen. Sie sind also tendenziell affirmativ.

Auf Messung basierende Zahlen signalisieren, dass ihre Selektion auf wissenschaftlichen Grundlagen basiert und nicht von persönlichen Ansichten oder Meinungen geprägt ist.

In Folge dessen wird sie der Umwelt zugerechnet und nicht den Beteiligten. Diese Objektivierungsleistung muss bei Indizes jedoch grundsätzlich in Frage gestellt werden. So handelt es sich bei der *Selektion* um Methoden, die subjektiv sein können. Darüber hinaus stellt sich die Frage, *welcher Gegenstand und welches Ziel dem Messprozess zugrunde liegen.* Diese Frage ist evident, da Zahlen als Ergebnis einer Größenmessung immer auf einen Referenten, d.h. auf die Außenwelt, verweisen. Das unterscheidet sie von Ziffern, die keinen externen Bezug besitzen und z.B. von der Mathematik verwendet werden. Insofern ist die Überzeugungskraft von Zahlen in hohem Maße abhängig von der Glaubwürdigkeit der Messverfahren und der Messapparatur sowie den Selektionsmethoden. Es geht konkret um die Reliabilität und Validität der Messung. Die Messergebnisse und das Messverfahren müssen wiederholbar sein und unabhängig von Raum, Zeit und Person zu den gleichen Ergebnissen führen.

Ein wichtiges Element für die Glaubwürdigkeit von Zahlen ist das Vertrauen in die Messung selbst. Nur wenn die Messung glaubwürdig ist, kann sie als tatsächliche Beschreibung der Außenwelt angesehen werden. Ist dies zutreffend, rücken die Verfahren und Methoden zur Herstellung der Zahlen in den Hintergrund. Das Vertrauen in die Messung ist z.T. auch auf außerepistemische Maßnahmen angewiesen. Dabei spielen Faktoren wie die Reputation des Urhebers oder die Transparenz der Messung eine Rolle.

Indizes, die nicht auf eine Messung im wissenschaftlichen Sinne zurückzuführen sind, teilen ggf. nur wenige oder gar keine der o.g. Eigenschaften. Auch eine entgegengesetzte Wirkung ist denkbar: Signalisieren die Indexwerte Subjektivität, Willkür oder eine interpretative Beschreibung bei der Selektion, wirkt sich das u.U. negativ auf die Glaubwürdigkeit und das Vertrauen in die Maßzahl aus. Zusammen mit Intransparenz kann die Akzeptanz für Kommunikation ggf. verringert werden. Die eingangs erwähnte angestrebte PR-Maßnahme zum Begriff „Vertrauen“ hätte in diesem Fall einen negativen Effekt. Und auch die Skepsis über die gesammelten Daten und deren Missbrauch für gerichtete Ziele wäre somit berechtigt.

Indizes werden zur Kommunikation von Komplexität und Mehrdimensionalität eingesetzt. Ihre mediale Funktion erscheint vielschichtig. Inwieweit die detaillierte Analyse der medialen Funktion eine wissenschaftliche Relevanz besitzt, wird im folgenden Abschnitt geklärt.

1.2 Relevanz

Indizes werden mit zunehmender Tendenz in den Verbreitungsmedien verwendet. Fast täglich erscheinen Meldungen zum DAX, wobei dem Großteil der Rezipienten die genaue Zusammensetzung der Zahl vermutlich nicht bekannt ist. Dies betrifft insbesondere Personen, die mathematisch weniger versiert sind. So kommt es, dass das theoretische Konzept hinter einem Index i.d.R. nicht angezweifelt wird. Doch sollte bei jedem Index zunächst die Frage aufgeworfen werden, was über und durch die betreffende Zahl eigentlich kommuniziert werden soll oder kann. Ist es überhaupt möglich, Abstrakta wie „Glück“ oder „Vertrauen“ wissenschaftlich zu messen und über einen Index abzubilden? Können abstrakte Begriffe in dem Sinne, dass sie für etwas Nichtgegenständliches stehen, überhaupt quantifiziert werden? Kann ein Index in diesem Fall die Mittlerfunktion an der Schnittstelle zwischen Geistes- und Naturwissenschaften übernehmen?

Mehrdimensionalität und Komplexität sind, wie bereits angesprochen, schwer zu kommunizieren und stellen die Medien damit vor eine große Herausforderung. Der statistische Index scheint in diesem Zusammenhang eine Lösung zu sein. Dennoch wird immer die Frage bestehen bleiben, wie objektiv die Konstruktion eines Index ist und ob er überhaupt in der Lage ist, „Realität“ abzubilden.

Die mediale Funktion von Indizes erscheint intuitiv vielschichtig. Sie werden eingesetzt, um Komplexität zu reduzieren und kommunizierbar zu machen, jedoch muss diese Funktion nicht zwangsläufig kongruent mit jener innerhalb der Verbreitungsmedien sein. Hinzu kommt, dass aus der Funktion als Medium sowie innerhalb der Verbreitungsmedien eine instrumentelle Funktion resultiert, indem z.B. Indizes auf andere Medien oder Systeme einwirken.

Zahlen und Messungen im Allgemeinen und Indizes im Speziellen sind in der Kommunikation allgegenwärtig. Umso relevanter ist es, allgemein zu klären, welchen Stellenwert sie darin besitzen sowie, ob und inwieweit Indizes als Messung angesehen werden können oder ob sie nicht vielmehr funktionalisiert werden. Gerade im PR-Bereich können durch den geschickten Einsatz von Indexwerten geschäftsrelevante Argumente geschaffen werden.

In Zeiten von *Big Data* entsteht der Eindruck, dass Argumente mit Zahlen wertvoller sind als ohne. Es erscheint nur das wahr zu sein, was mittels Zahlen belegt werden kann.

Der Mensch strebt sogar danach, seinen Körper mittels selbst gesammelter Daten messbar und vergleichbar zu machen. Das Sammeln von Daten ist nicht mehr Mittel zum Zweck, sondern mitunter ziellos und willkürlich. Die Menge der weltweit generierten Daten ist 2015 fast achtmal so groß wie fünf Jahre zuvor – Tendenz steigend. In diesem Zusammenhang kann der Index eine Vereinfachung darstellen, indem er ein mehrdimensionales System vereinfacht und unterschiedliche Daten in einem Wert zusammenfasst. Die bereits erwähnte „Klimaerwärmung“ ist ein Beispiel für das Bestreben, komplexe Sachverhalte mit Zahlen zu belegen und einfach kommunizierbar zu machen.

Die zunehmende Anzahl und Häufigkeit der Nennung von Indizes in den Verbreitungsmedien sowie ihre vielschichtige Funktion führen zu einer Relevanz, die die Untersuchung des Themas der vorliegenden Arbeit fruchtbar erscheinen lassen.

1.3 Gegenstand

Der Forschungsgegenstand der vorliegenden Arbeit ist die mediale Funktion von statistischen Indizes.

Ein statistischer Index wird dabei als Quantifizierung („Messung“) eines *mehrdimensionalen Merkmalsraums* angesehen. Die Zielgröße (das Indexkonzept) ist nicht direkt beobachtbar und messbar. Stattdessen wird sie formal betrachtet über eine *mathematische Funktion von Indikatoren* beschrieben. Der statistische Index wird stets mit Hilfe eines Modells gebildet und zeichnet sich in der Regel dadurch aus, dass der *mehrdimensionale Merkmalsraum* über das Indexkonzept anhand eines meist eindimensionalen Maßes beschrieben wird. Inhaltlich bedeutet das, dass Indizes Mehrdimensionalität reduzieren. Zudem lassen sie mindestens die Unterscheidung von *mehr* und *weniger* zu, teilweise jedoch auch eine kontinuierliche Abstufung.

Der Forschungsgegenstand umfasst die mediale Funktion der oben beschriebenen statistischen Indizes. Die mediale Funktion ist dabei in zweierlei Bereichen zu analysieren:

1. Index *als* Medium
2. Index *in* den (Verbreitungs-) Medien

Das bedeutet, dass einerseits das Konzept eines Index im Allgemeinen zur Diskussion steht, insbesondere im Bezug zum naturwissenschaftlichen Messprozess. Dabei wird im Vordergrund stehen, inwieweit er dazu beiträgt, Komplexität und Mehrdimensionalität zu kommunizieren sowie, welche kommunikativen Eigenschaften er generell besitzt und ob er als Medium angesehen werden kann. Über die Betrachtung der Konstruktionsmethoden wird die Objektivierungsleistung der Indizes untersucht. In diesem Zusammenhang werden generelle Eigenschaften von Zahlen, die auf Messung basieren, hinzugezogen, um Differenzen von Maßzahlen und Indexzahlen zu identifizieren.

Außerdem wird die Funktion von Indizes *innerhalb* der Verbreitungsmedien anhand der *Frankfurter Allgemeinen (Sonntags-) Zeitung* analysiert. Die *Frankfurter Allgemeine Zeitung* gehört zu einer der sieben überregionalen Tageszeitungen und ist im gesamten Bundesgebiet verbreitet. Neben *Die Welt* wird sie als Qualitätszeitung bezeichnet. Die FAZ gehört zur Verlagsgruppe *Frankfurter Allgemeine Zeitung/Frankfurter Societäts-Druckerei GmbH*. Sie hat einen Marktanteil von 3,0% .

Die Auswahl der Ausgaben der FAZ und FAS wird zeitlich vorerst nicht eingeschränkt. Konkreter Gegenstandsbereich sind hier alle Beiträge, die einen statistischen Index nennen oder sich auf einen beziehen.

Neben der Analyse der Funktion von Indizes (1) *als* Medium und (2) *in* den Verbreitungsmedien bleibt zu diskutieren, welche instrumentelle Funktion aus diesen beiden Betrachtungsweisen hervorgeht. Das bedeutet, inwieweit z.B. ein Index auf andere Systeme oder Medien einwirkt.

Der Gegenstandsbereich umfasst damit zusammengefasst die Ambiguität der *medialen Funktion* mit der Analyse der Funktion von Indizes *als* Medium und *innerhalb* der Medien. Dabei wird die Analyse der Funktion *innerhalb* der Medien am Beispiel der *Frankfurter Allgemeinen (Sonntags-) Zeitung* durchgeführt.

1.4 Forschungsziel und Forschungsfragen

Das ursprüngliche Ziel des Dissertationsprojektes bestand in der empirischen Analyse der medialen Funktion von statistischen Indizes innerhalb der Verbreitungsmedien. Früh

zeigte sich jedoch, dass dieser Rahmen ungeeignet für die Analyse war. Die uneinheitliche Verwendung und Betrachtung von Indizes in verschiedenen Wissenschaftsdisziplinen führte rasch zu der Notwendigkeit einer theoretischen Auseinandersetzung – vor allem medienwissenschaftlich, wissenschaftshistorisch, philosophisch und statistisch. Erst mit dieser Auseinandersetzung wird die empirische Analyse der medialen Funktion möglich.

Die Analyse der medialen Funktion folgt den drei bereits erwähnten Betrachtungsweisen des Gegenstandes. Indizes werden auf ihre kommunikativen Eigenschaften hin untersucht. Außerdem wird diskutiert, inwieweit sie als Medium Verwendung finden. Damit gehen folgende Fragen einher: *Werden Indizes verwendet, um Komplexität zu kommunizieren? Können Indizes als interpretationsfreie Beschreibung der Realität angesehen werden?* Oder *vermögen sie es sogar, die Akzeptanz von Kommunikation wahrscheinlicher zu machen?*

Vor dem detaillierten Betrachten dieser medialen Funktionen wirft der Index zusätzlich einige grundsätzliche Forschungsfragen auf, die sich auf den Messcharakter beziehen. *Ist ein Index tatsächlich als Messung zu betrachten und was bildet er eigentlich ab?* Damit geht die weitere Fragestellung einher, auf welche *Referenten* sich ein Index bezieht und welches *theoretische Konstrukt* hinterlegt ist. Bezüglich des theoretischen Konstrukts stellt sich wiederum die Frage: *Reduziert der Index, wie zunächst angenommen, Komplexität?* In der Regel wird über einen Index ein mehrdimensionales Konstrukt innerhalb einer Zahl verdichtet. In dieser Hinsicht wird oft von einer Reduktion der Komplexität gesprochen. Jedoch ist die Konstruktion von Indizes diesbezüglich z.T. mangelhaft. Die Beziehungen und Wechselwirkung zwischen den Einzelkomponenten werden nicht hinreichend beschrieben oder nur eine Teilmenge der einzubeziehenden Komponenten tatsächlich verwendet.

Darüber hinaus besteht das Ziel, die Funktion der Indizes innerhalb der Verbreitungsmedien anhand des Fallbeispiels der *FAZ* zu untersuchen. Diese Funktion ist sowohl separat sowie im Zusammenhang zu der Funktion des Index *als* Medium zu betrachten. Eine grundsätzliche Frage in diesem Kontext ist: *Werden Indizes in den Verbreitungsmedien eingesetzt, um Relevanz zu suggerieren?* Außerdem sollte geklärt werden: *Werden Indizes gegenüber fundamentalen oder abgeleiteten Messungen bevorzugt und wenn das der Fall ist, in welchen Situationen?* Dies ist insbesondere in Bezug zur höheren Komplexität des Indexkonstrukts zu bewerten. Letzteres leitet zur Frage über, welche Schritte der

Indexkonstruktion in den Beiträgen aufgegriffen werden. *Findet z.B. das Indexkonzept Beachtung oder ausschließlich der Indexwert? Wird das Wissen über das dahinterstehende Konzept erklärt oder als Wissen des Lesers vorausgesetzt? Und welchen Einfluss hat das auf die Kommunikation?*

Bei der Analyse der medialen Funktion statistischer Indizes besteht zudem der Anspruch, einem interdisziplinären Ansatz nachzugehen, der Natur- und Geisteswissenschaften vereint. So sollen bei diesem Vorgehen Methoden und Theorien verschiedener Disziplinen, insbesondere der Medienwissenschaften, der Statistik, Philosophie und Linguistik zusammengeführt werden.

Ein Modell, das aus den Naturwissenschaften entlehnt ist, soll den Geisteswissenschaften für den automatisierten Gebrauch zur Verfügung gestellt werden. Dieses geht aus einem weiteren zentralen Forschungsziel hervor, bei dem die *quantitative Inhaltsanalyse* für eine generelle Nutzung innerhalb der Medienwissenschaften auf Automatisierbarkeit geprüft wird. Dafür wird exemplarisch ein System entwickelt und bewertet. Input ist dabei ein beliebiger Textbeitrag, der innerhalb des Systems über verschiedene Funktionen zu quantitativen Kennzahlen verdichtet und ausgegeben wird. In der vorliegenden Arbeit richten sich die quantitativen Kennzahlen dabei nach der Analyse der Funktion von Indizes innerhalb der Verbreitungsmedien. Sie sind größtenteils generalisierbar sowie erweiterbar.

1.5 Forschungsstand

Trotz der regen Verwendung von statistischen Indizes in den Medien existieren bisher weitestgehend keine empirischen Studien zur medialen Funktion. Darunter würden z.B. die quantitativen Informationen zur Häufigkeit und Kontinuität ihrer Verwendung fallen. Aktuell bleibt der Beitrag statistischer Indizes innerhalb der Kommunikation theoretisch und empirisch unklar.

Es existieren theoretische Betrachtungen zu den kommunikativen Eigenschaften von Zahlen allgemein, dabei bleibt aber offen, inwieweit diese auf Indexzahlen angewendet werden können. Beispiele für diese Betrachtungen finden sich unter anderem in den Arbeiten von Theodore Porter oder Bettina Heintz. Porter untersucht die Attraktivität von

Zahlen und Messungen in der modernen Welt und diskutiert die Bedeutung von Objektivität innerhalb der kulturellen Entwicklung. Die Betrachtung von Heintz überschneidet sich teilweise mit der von Porter. Einer ihrer wesentlichen Beiträge ist die Diskussion von Zahlen als symbolisch generalisiertes Kommunikationsmedium im Luhmannschen Sinne.

Indizes werden innerhalb der Wissenschaft meist nur als mathematisch/statistisches Problem betrachtet und ihr Bezug zum naturwissenschaftlichen Messprozess von den verschiedenen Forschungsdisziplinen z.T. kontrovers diskutiert. Im Wesentlichen bleibt damit unklar, inwieweit die theoretischen Betrachtungen zu Zahlen im Allgemeinen überhaupt auf Indizes übertragbar sind.

Der Messprozess in den Naturwissenschaften wurde bereits erkenntnistheoretisch und wissenschaftstheoretisch ausgiebig untersucht. In der Psychologie findet sich ein analoger Verwendungsversuch des Messens, trotz fundamentaler Unterschiede. Auch die Einordnung neuartiger „Messtechniken“ wie die indikatorbasierten Messsysteme und dem Konzept der Indizes, z.B. in Politik- oder Sozialwissenschaften innerhalb der Messtheorie, ist nicht abschließend geklärt. Dazu gehört auch die zentrale Frage, ob statistische Indizes überhaupt als Messung bezeichnet werden können, sowie ob und auf welche Weise sie „Realität“ abbilden. Einen formalen Versuch unternehmen Besozzi und Zehnpfennig sowie Randolph.

Die mathematisch/statistische Betrachtung von Indizes lässt den Messcharakter, wie bereits angedeutet, weitestgehend außen vor. Die Literatur konzentriert sich hier vor allem auf statische Methoden zur Konstruktion und Validierung von Indizes. Die Ansätze gehen jedoch z.T. stark auseinander. Literatur aus der ersten Hälfte des 20. Jahrhunderts bezieht sich überwiegend auf Indizes als Verhältniszahlen und ist zum größten Teil im Bereich der Ökonomie einzuordnen. Vertreter dieser frühen Bewegung sind z.B. Fisher und Flaskämper, die sich auf Basis ihrer Anschauung jedoch äußerst uneinig sind, was sie in ihrer Arbeit explizit zum Ausdruck bringen. Bei neueren Betrachtungen, z.B. in Soziologie, Politik und Psychologie, kommen statistische Methoden für die Konstruktion von Indizes zum Einsatz. Eine vereinheitlichende Darstellung, die die Methoden der verschiedenen Disziplinen zusammenführt, fehlt weitgehend und wird am ehesten von Nardo u. a. versucht. Durch fehlende Standards existieren nach wie vor unterschiedliche Definitionen, Bezeichnungen und Vorstellungen von Indizes.

1.6 Methodisch-theoretisches Vorgehen

Für die Analyse der medialen Funktion von Indizes wird für die vorliegende Arbeit ein interdisziplinärer Ansatz gewählt, wobei die Theorien und Werkzeuge verschiedener Disziplinen zusammengeführt werden. Hauptsächlich handelt es sich dabei um die Medienwissenschaften, Philosophie, Wissenschaftsgeschichte, Soziologie, Informatik und Statistik. Dabei wird der naturwissenschaftliche Messprozess einer wissenschafts- und erkenntnistheoretischen Betrachtung unterzogen, und Quantifizierungsprodukte in Form von Zahlen werden auf ihre allgemeinen Eigenschaften innerhalb der Kommunikation hin untersucht. Ferner soll im Rahmen der medienwissenschaftlichen Betrachtung diskutiert werden, inwieweit Zahlen als symbolisch generalisierte Kommunikationsmedien nach der Systemtheorie von Niklas Luhmann verstanden werden können und inwiefern auch dem Index eine entsprechende Funktion eigen ist.

Über die Analyse von statistischen Indizes im Allgemeinen, ihre Einbettung in den Messprozess und ihre kommunikativen Eigenschaften wird die theoretische Grundlage für den empirischen Teil der Arbeit gelegt. Eine historische Betrachtung des Messens in den Naturwissenschaften sowie der Wirtschaft, Psychologie und Soziologie zeigt parallelen zu früheren wissenschaftlichen Herausforderungen und ermöglicht das Einordnen von Indizes in die Evolution quantitativer Begriffe. Dabei wird auch untersucht, inwieweit sich die Komplexität des Messprozesses und die Anforderungen an Messapparaturen im Laufe der Zeit verändert haben und welchen Zusammenhang diese Veränderungen hinsichtlich der Kommunikation aufweisen. Die geschichtliche Betrachtung in dieser Arbeit zielt darauf, die heutigen Trends zu erklären. Jedoch erhebt sie nicht den Anspruch auf Vollständigkeit. Vielmehr ist es ein Anliegen, punktuell Stationen der Entwicklung herauszugreifen, die den heutigen Umgang mit dem Messen und den daraus resultierenden Zahlen erklären. Ebenso ist die theoretische Betrachtung der kommunikativen Eigenschaften nur im engen Rahmen der statistischen Indizes im Medium Zeitung zu verstehen und nicht als allumfassende mediale Kommunikationstheorie.

Die Konstruktion von Indizes wird in Hinblick auf die statistischen Konstruktionsmethoden untersucht. Dabei werden verschiedene Methoden aufgezeigt, mit denen ein Index konstruiert werden kann. Außerdem ist es wichtig, zu erläutern, wie die Konstruktionsmethoden validiert und in ihrer Qualität beurteilt werden können. Erst so kann der Forschungsgegenstand Index begriffen und in seiner Funktion hinterfragt werden.

Innerhalb des empirischen Teils wird die mediale Funktion von Indizes anhand von Beiträgen der *Frankfurter Allgemeinen (Sonntags-) Zeitung* analysiert. Dabei wird ein innovativer Ansatz entwickelt, mit dem es u.a. möglich wird, der steigenden Zahl existierender Indizes gerecht zu werden. Über eine automatisierte Abfrage öffentlich zugänglicher Quellen kann dabei eine Liste von Indizes generiert und laufend aktuell gehalten werden. Gleichzeitig ermöglicht diese Auflistung das Einbeziehen unterschiedlicher Schreibweisen und z.T. auch verschiedener Sprachen. Die so erzeugte Liste bildet die Voraussetzung für die Selektion von Beiträgen, da sie sämtliche Aufgriffkriterien enthält. Sie ist die Grundlage für die empirische Datenerhebung.

Weiterhin gehört zu dem innovativen Ansatz dieser Arbeit eine automatisierte quantitative Inhaltsanalyse. Um die Ergebnisse der automatisierten Analyse bewerten zu können sowie den Aufwand überschaubar zu halten, wird sie anhand einer manuellen Stichprobe entwickelt. Sie stellt für medienwissenschaftliche Auswertungen eine Möglichkeit dar, diverse Beiträge aus verschiedensten Kanälen automatisiert quantitativ zu analysieren. Der Vorteil einer Automatisierung liegt u.a. darin, dass die Grundgesamtheit aller Beiträge einbezogen werden kann, was bei einer manuellen Analyse meist aus Gründen der Verhältnismäßigkeit nicht praktikabel ist. Außerdem ist auf Basis des automatisierten Ansatzes ein Monitoring möglich. So kann eine Selektion und deren interessierende quantitative Eigenschaften über einen ausgewählten Zeitraum verfolgt werden.

Für die quantitative Inhaltsanalyse wird im empirischen Teil eine Reihe aussagekräftiger Kriterien für die Untersuchung der medialen Funktion von Indizes erarbeitet. Sie bilden den Ausgangspunkt für die Generierung der quantitativen Aussagen über die selektierten Beiträge. Die automatisierte Inhaltsanalyse wird an einer manuellen motiviert. Das hat den Vorteil, dass die Güte und Zweckmäßigkeit der automatisiert erhobenen Kriterien beurteilt werden kann. Des Weiteren existieren zahlreiche Bewertungskriterien, die nur schwer oder gar nicht automatisierbar sind.

Da die automatische Lösung an der manuellen motiviert wird, findet an dieser Stelle die Stichprobentheorie Verwendung. Sie soll im empirischen Teil garantieren, dass der manuelle Aufwand in einem verhältnismäßigen Zusammenhang zu den Erkenntnissen steht. Eine Stichprobe ermöglicht trotz Reduzierung der zu analysierenden Beiträge Aussagen

über die Grundgesamtheit. Da jedoch nur eine Teilmenge an Beiträgen tatsächlich analysiert wird, finden generelle Aussagen unter einer quantifizierbaren Fehlerwahrscheinlichkeit statt.

Für die Aggregation und Visualisierung der automatisiert verarbeiteten Beiträge werden verschiedene Techniken der deskriptiven Statistik angewendet. Der Flut an Daten wird über eine Auswahl an – größtenteils vom Autor erstellten – Tabellen, Kennzahlen und Diagrammen gerecht geworden.

1.7 Literatur und Material

In Folge des interdisziplinären Ansatzes wird sich die vorliegende Arbeit an Literatur und Materialien aus verschiedenen Disziplinen bedienen. Für die Einordnung des Index in den naturwissenschaftlichen Messprozess wird auf die erkenntnis- und wissenschaftstheoretische Betrachtung von Gernot Böhme zurückgegriffen. In seiner Arbeit "Quantifizierung - Metrisierung: Versuch einer Unterscheidung erkenntnistheoretischer und wissenschaftstheoretischer Momente im Prozeß der Bildung von quantitativen Begriffen" ordnet er die verschiedenen Ansätze der Messtheorie und trennt erkenntnistheoretische von wissenschaftstheoretischen Ansätzen scharf. Damit gelingt ihm eine gesamtheitliche Betrachtung des Messens unter Einbeziehung unterschiedlicher Facetten. Später erforderten z.B. die in den Sozial- und Politikwissenschaften angewendeten indikatorbasierten Messansätze eine Erweiterung des Messprozesses von Böhme. In diesem Zusammenhang ist die Arbeit *Pragmatische Theorie der Indikatoren - Grundlagen einer methodischen Neuorientierung* von Rainer Randolph zu nennen, der ausgehend von Böhmes Ausführung das abgeleitete Messen über Indikatoren diskutiert.

Weniger formal, jedoch zu größeren Teilen wissenschaftstheoretisch, beschäftigen sich Besozzi und Zehnpfennig mit dem Bezug von Beobachtungssprache und theoretischer Sprache bei einer Quantifizierung über Indizes. In ihrem Beitrag "Methodologische Probleme der Index-Bildung" reduzieren sie den Index bei ihrer Betrachtung auf eine Funktion von Indikatoren und ermöglichen damit der vorliegenden Arbeit, die verschiedenen Ansätze zur Konstruktion von Indizes zusammenzuführen und zu verallgemeinern. Außerdem hilft ihre Darstellung bei der Beantwortung der Frage, ob Indizes überhaupt als Messung angesehen werden können.

Eine umfassende mathematisch/statistische Darstellung zur Konstruktion von Indizes bieten die Arbeiten *Tools for composite indicators building*; *Handbook on Constructing Composite Indicators: Methodology and User Guide* von NARDO U. A.. Sie stellt dabei sinnvolle und mögliche Methoden zusammen und bietet einen Leitfaden zur Konstruktion von Indizes sowie zur Beurteilung ihrer Qualität. Diese sollen der vorliegenden Arbeit als Ansatz dienen. Eine erkenntnistheoretische oder messtheoretische Auseinandersetzung fehlt dabei weitestgehend.

Für die medienwissenschaftliche Betrachtung von Zahlen allgemein werden Arbeiten von Theodore PORTER und Bettina HEINTZ einbezogen. Die Arbeit "Zahlen, Wissen, Objektivität: Wissenschaftssoziologische Perspektiven" von HEINTZ besteht u.a. in der Diskussion von Zahlen als symbolisch generalisierte Kommunikationsmedien im Sinne der Luhmannschen Systemtheorie sowie ihrer generellen Eigenschaften innerhalb der Kommunikation. Dafür bedient sie sich u.a. der Arbeiten von PORTER, der ebenfalls die Eigenschaften von Zahlen innerhalb der Kommunikation untersucht, jedoch nicht explizit in Verbindung zu Luhmann bringt. Als sein wesentliches Werk ist in diesem Zusammenhang *Trust in numbers* zu nennen. Die beiden Werke von PORTER und HEINTZ dienen der Diskussion der kommunikativen Eigenschaften der Indizes als Grundlage.

Für die quantitative Inhaltsanalyse im empirischen Teil der vorliegenden Dissertationsschrift wird *Medienresonanzanalyse eine Einführung in Theorie und Praxis* von RAUPP UND VOGELGESANG einbezogen. Darüber hinaus findet die Arbeit *The content analysis guidebook* zur Inhaltsanalyse von Kimberly NEUENDORF Beachtung. Die Entwicklung der Bewertungskriterien für die Inhaltsanalyse werden über diverse Autoren, insbesondere aus dem Bereich der Linguistik, gestützt. Die Inhaltsanalyse selbst basiert auf Beiträgen der *Frankfurter Allgemeinen (Sonntags) Zeitung*.

Für die Identifikation existierender Indizes dient die öffentlich zugängliche Online-Datenbank *DBpedia*. Die Inhalte von *DBpedia* werden fortwährend erweitert, womit auch zukünftig erscheinende Indizes abgedeckt sein sollten.

1.8 Aufbau der Arbeit

Es liegt nahe, die vorliegende Arbeit in zwei Hauptteile zu untergliedern – einen theoretischen und einen empirischen. In Anbetracht der Zusammenführung unterschiedlicher

Disziplinen unterteilt sich der theoretische Teil in weitere drei Blickwinkel: philosophisch, historisch/medienwissenschaftlich und statistisch.

Der erste Teil (Kapitel 2) der theoretischen Betrachtung beschäftigt sich weitestgehend philosophisch mit dem naturwissenschaftlichen Messprozess und der damit einhergehenden Quantifizierung der Welt. Es werden die Ursprünge von quantitativen Begriffen aufgezeigt und die verschiedenen messtheoretischen Ansätze seit der Antike zusammengefasst.

Danach wird der Prozess des Messens in seine einzelnen Komponenten zerlegt, Grundbegriffe werden geklärt sowie ontologische, erkenntnis- und wissenschaftstheoretische Aspekte herausgearbeitet. Als Basis dient die Darstellung des Messprozesses nach Gernot Böhme. In enger Abgrenzung werden Quantifizierung, Metrisierung, Größe und Zahl definiert sowie die Skalentheorie nach Stanley Stevens dargestellt.

Der Messprozesses nach Böhme wird danach um neuere Messkonzepte wie Indikatoren und Indizes erweitert. Im Wesentlichen wird die Frage beantwortet, inwieweit diese tatsächlich als Messungen zu interpretieren sind. Dafür wird ihre Stellung im Messprozess herausgearbeitet und ihre Eigenschaften sowie ihr Abbildungscharakter analysiert.

Der zweite Teil (Kapitel 3) beginnt mit einer historischen Betrachtung der Messung. Es werden die Ursprünge des Messens in der Menschheitsgeschichte herausgearbeitet sowie die Ausdifferenzierung von Zahlen, Ziffern und Größen verfolgt. Im Zentrum steht dabei die Frage, inwieweit die Zunahme der Komplexität in Wissenschaft und Gesellschaft zu neuen Anforderungen an die quantitativen Begriffe geführt hat. Vor diesem Hintergrund wird im anschließenden Abschnitt gezeigt, wie die Notwendigkeit, komplexe Sachverhalte quantifizieren zu können, zur Entwicklung von indikatorbasierten Messsystemen und einer neuen Form von Quantifizierung geführt hat.

Die historische Entwicklung der Quantifizierung ist anschließend die Grundlage für die Betrachtung von Zahlen innerhalb der Kommunikation. Ausgangspunkt ist die Rolle der Zahlen für die Kommunikation von objektivem Wissen. Da auch die wissenschaftliche Kommunikation zeitlichen Veränderungen unterliegt, wird auch diese Frage historisch eingebettet und eng an den Begriff der Objektivität gekoppelt.

Nach der Darstellung der Objektivierungsleistung von Zahlen werden ausgehend von den Überlegungen von Theodore Porter und Bettina Heintz die Eigenschaften von

Zahlen innerhalb der Kommunikation erarbeitet. Mit Bezug zur *Systemtheorie* von Niklas Luhmann wird die Frage verfolgt, inwieweit diese als symbolisch generalisiertes Kommunikationsmedium zu begreifen sind und inwieweit sie Akzeptanz mobilisieren.

Im dritten Teil (Kapitel 4) folgt die Betrachtung der Konstruktion von Indizes. Diese wird einer statistischen Sichtweise unterzogen, wobei Indizes zu Beginn einer philosophischen und formal-sprachlichen Betrachtung unterliegen. Es wird geklärt, welche messtheoretische Funktion Indizes besitzen und wie ihr Bezug zur „Realität“ formalsprachlich definiert werden kann.

Danach werden die verschiedenen Teilschritte der Konstruktion von Indizes vorgestellt. Die Kategorisierung in Teilschritte bietet dem empirischen Teil die Möglichkeit, differenzierte Aussagen über die Verwendung von Indizes in den Medien zu treffen. Dort wird der Frage nachgegangen, welche Aspekte der Indexkonstruktion in den Beiträgen Niederschlag finden.

Die Indexkonstruktion wird mit der Entwicklung des theoretischen Rahmens und Indexkonzepts eingeleitet. Hier geht es in erster Linie um den Zusammenhang zwischen dem Indexkonzept und den beobachtbaren Indikatoren.

Es folgt ein Abschnitt zur Variablenselektion, in dem aufgezeigt wird, welche Anforderungen sie zu erfüllen hat und welche Auswirkung verschiedene Entscheidungen auf den Index besitzen. Der nächste Schritt in der Konstruktion eines Index ist das Datenqualitätsmanagement. Innerhalb dieses Teils werden verschiedene statistische Verfahren zur Analyse der Datenqualität zusammengestellt. Im anschließenden Unterpunkt zur Datenstruktur handelt es sich um statistische Methoden, mit denen die Struktur hinter den Daten analysiert werden kann. Inhaltlich wird hier über statistische Methoden versucht, Aussagen über den Zusammenhang der Indikatoren aus den Daten selbst zu treffen.

Weiter geht es mit der Aggregation, der Art und Weise, wie verschiedene Indikatoren und Beobachtungen zu einem Index zusammengefasst werden. In diesem Schritt wird mathematisch der Zusammenhang zwischen Indexkonzept und beobachtbaren Variablen definiert. Anschließend folgt eine Beschreibung, wie die Konstruktion über statistische, intuitive und theoretische Tests geprüft werden kann.

Am Ende der Indexkonstruktion steht immer die Frage nach der adäquaten Visualisierung. So bildet auch hier die Betrachtung über geeignete Kommunikationsformen den letzten Punkt.

Nach der ausführlichen Darstellung der Konstruktion von Indizes folgt der Abschnitt zur Beurteilung der Qualität. In Anlehnung an Michela NARDO wird ein Schema entwickelt, mit dem die Qualität von Indizes bewertet werden kann.

Im empirischen Teil der vorliegenden Arbeit (Kapitel 5) wird die mediale Funktion von Indizes anhand der FAZ analysiert. Im ersten Schritt werden der Studienrahmen und das allgemeine Vorgehen entwickelt. Es folgt die Erläuterung der Methode einer explorativen Datenanalyse und deren Rechtfertigung sowie die Entwicklung des Untersuchungsinstruments.

Die Basis für die quantitative Datenerhebung der Beiträge der FAZ stellen die Kriterien der Selektion, Klassifikation und Bewertung dar, die nun umfassend behandelt werden. Sie werden gruppiert in automatisierbare und nicht automatisierbare Kriterien.

Zu den automatisierbaren Kriterien werden Definitionen und Anweisungen entwickelt, mit deren Hilfe quantitative Daten aus den Beiträgen maschinell erhoben werden können. Es findet eine Trennung nach direkten und indirekten Kriterien statt, wobei es sich bei den direkten Kriterien um Attribute eines Beitrags innerhalb der FAZ handelt. Indirekte Kriterien werden aus dem Inhalt der Beiträge abgeleitet. Der Abschnitt resultiert in einer Vorschrift zur Interpretation und Umwandlung von Beitragsdaten für die Weiterverarbeitung auf Basis von quantitativen Metadaten.

Weiterhin wird die Auswahl der FAZ für die quantitative Inhaltsanalyse begründet sowie die praktische Selektion der Beiträge erklärt. Dabei werden verschiedene Tools und öffentlich zugängliche Quellen einbezogen, die in der Selektion der Beiträge Verwendung finden. Dazu gehört u.a. das *DBpedia*-Projekt, das strukturierte enzyklopädische Daten zur Verfügung stellt. Dieses wird unter Nennung seiner Vor- und Nachteile beschrieben, außerdem wird auf die Abfragesprache SPARQL eingegangen. Der Unterpunkt mündet in einer Abfrage zur Erzeugung einer dynamischen Liste von Indizes.

Um den Aufwand der manuellen quantitativen Inhaltsanalyse überschaubar zu halten, wird im folgenden Abschnitt die Methode der Stichprobe eingeführt und eine Submenge an Indizes für die Inhaltsanalyse festgelegt.

Danach findet das Aufgreifen relevanter Beiträge statt, die in der quantitativen Inhaltsanalyse Verwendung finden. Die Indizes werden dann nach einem festen Schema analysiert, das zu Beginn der Inhaltsanalyse veranschaulicht wird. Es folgt eine einzelne theoretische und mathematisch/statistische Betrachtung eines jeden aufgegriffenen Index, in deren Zuge der Index nach bestimmten Kriterien beschrieben und bewertet wird. Auf jede theoretische Auseinandersetzung mit dem Index folgt jeweils eine empirische Inhaltsanalyse und ein Fazit. Dieses bezieht sich auf den entsprechenden Index im Kontext der Kommunikation.

Abgeschlossen wird der empirische Teil von einer vergleichenden Zusammenfassung aller Indizes. Die erhobenen Kriterien sprechen hier nicht mehr für einzelne Indizes, sondern werden im Kontext zu den Werten anderer Indizes betrachtet.

Kapitel 2

Der Messprozess und die Bildung quantitativer Begriffe

"A decision made by the numbers [...] has at least the appearance of being fair and impersonal."

Theodore M. Porter, 1995

Zahlenbasierte Darstellungen sind ein allgegenwärtiges Phänomen unserer datengetriebenen modernen westlichen Welt – ob in Politik, Wirtschaft oder Gesellschaft, ob in Medien oder Unternehmen. In vielen Lebensbereichen haben sich Rezipienten und Entscheidungsträger daran gewöhnt, Informationen auf Basis von Diagrammen, Infografiken oder Tabellen zu konsumieren (Passoth und Wehner, 2012: S. 7). Darüber hinaus folgen inzwischen viele dem Trend der Selbstvermessung (Quantified Self) und nehmen ihre eigenen Körper als Messobjekte wahr. Damit produzieren sie einen direkten Wettbewerb zwischen körpereigener Wahrnehmung und zahlenbasierter Darstellungen von Körperfunktionen, die dank modernsten Techniken unmittelbar abgerufen werden können. Was heute ein Trend ist, folgt im Grunde einer jahrtausendealten Erscheinung – der Darstellung und Kommunikation quantitativer Begriffe. Mit dem Aufkommen von Indizes hat diese mehr denn je die Berechtigung, erneut ausführlich diskutiert zu werden.

Das vorliegende Kapitel schafft einen Überblick über die Diskussion zur Bildung quantitativer Begriffe. Diese reicht historisch zurück bis zur aristotelischen Kategorienlehre.

Beginnend mit einer ontologischen Reflexion entwickelte sich der Diskurs über einen erkenntnistheoretischen bis hin zu einem wissenschaftstheoretischen. Diese Progression in drei Schritten soll im Folgenden skizziert werden.

Dabei liefert Gernot Böhme (1976) mit dem Werk "Quantifizierung - Metrisierung: Versuch einer Unterscheidung erkenntnistheoretischer und wissenschaftstheoretischer Momente im Prozeß der Bildung von quantitativen Begriffen" einen grundlegenden Beitrag zur systematischen Darstellung des naturwissenschaftlichen Messprozesses. Auf Basis seiner Arbeiten soll in diesem Kapitel der Messprozess in seine Teilprobleme zerlegt werden. Anschließend wird die Darstellung um das Phänomen der Indexzahlen erweitert. Zum Forschungszeitpunkt fehlt es in der hiesigen Literatur noch an einer erkenntnistheoretischen Auseinandersetzung mit dieser Art von quantitativen Begriffen, sodass die Ausarbeitung des Zusammenhangs von Indexzahl und naturwissenschaftlichem Messprozess einen ersten Grundstein dafür legen soll.

2.1 Der Ursprung quantitativer Begriffe

Der Begriff *Messung* ist problembehaftet und wird in der Alltagssprache oft unpräzise verwendet. Wie bereits in der Einleitung dieses Kapitels angemerkt wurde, hat die strukturierte Beschäftigung mit dem Prozess der Bildung quantitativer Begriffe eine weitreichende Geschichte. Fortschritte stellten sich in der Historie immer dann ein, wenn sich Hindernisse auftaten. Wie Böhme (1976: S. 210f.) zeigt, ergaben sich diese in der jüngeren Geschichte insbesondere aus dem Siegeszug der Naturwissenschaften und dem Aufkommen der Frage nach der Integrierbarkeit ihrer erfolgreichen Methoden wie Messen, Experimentieren und mathematischen Behandlungen in andere Disziplinen. Diese Frage war unter Fechner und Wundt zunächst auf das Fachgebiet der Psychologie beschränkt, dehnte sich jedoch schnell auf die Verhaltens-, Geistes-, und Gesellschaftswissenschaften aus. Die spezifischen Fragestellungen dieser Wissenschaften machten es jedoch mitunter schwer, die Methoden der Naturwissenschaften zu übernehmen. Die ausgiebige Diskussion, die seit dem Anstoß dieser Überlegung von v. Helmholtz und Campbell auftraten, führten neben den Fortschritten jedoch auch zu einer „terminologischen Verwirrung" (Pfanzagl, 1971) sowie, laut Böhme, zu einer Vernachlässigung relevanter Teile des Messprozesses:

> „Die Fortentwicklung von der ontologischen über die erkenntnistheoretische zur wissenschaftstheoretischen Reflexion ist auch nicht nur ein Fortschritt gewesen, sondern wurde zugleich mit Verlusten, wieder mit der Verdunkelung mancher Seiten des Problems erkauft.“ (Böhme, 1976: S. 210)

Der Ursprung der von Böhme angesprochenen Fortentwicklung kann im Rahmen der aristotelischen Kategorienlehre gesehen werden und ist rein ontologisch zu bewerten. Denn hier stand die Frage im Vordergrund, wie alles Seiende, was nicht Quantum ist, wie z.B. Relationen, Qualität oder Substanz trotzdem quantitativ zu denken sei. (ebd.: S. 209)

Später wurde das Thema u.a. bei Kant und v. Helmholtz erkenntnistheoretisch (Böhme, 1974; Helmholtz, 1879/1959) betrachtet und vorangetrieben. Gleichzeitig bildete sich jedoch auch erstmals eine rein wissenschaftstheoretische Betrachtung des Themas unter Fechner und Wundt innerhalb der psychologischen Forschung heraus (Böhme, 1976: S. 209f.). Diese Diskussion wurde im 20. Jahrhundert mit Beiträgen u.a. von Campbell, Stegmüller, Suppes und Zinnes sowie Stevens weiter vorangetrieben, führte neben den Fortschritten jedoch zu der von Böhme erwähnten „Verdunkelung“. Diese ist hauptsächlich darauf zurückzuführen, dass bei der Beschäftigung mit den quantitativen Begriffen nicht alle Bereiche des Messprozesses berücksichtigt wurden. Speziell geht es Böhme dabei um „die traditionell erkenntnistheoretische Frage nach der Beziehung der Konstitution des Gegenstandes zu den Erkenntnisvermögen“ (ebd.: S. 211).[1] Dieser Aspekt fehlt beispielsweise bei Stevens, der sich ausschließlich mit den Regeln beschäftigte, unter denen Zahlen den jeweiligen Phänomenen zugeordnet werden. Die Operationen im „Phänomenbereich“ lässt er komplett außer Acht (Stevens, 1959: S. 19ff.). Die Arbeiten von Suppes und Zinnes weisen einen ähnlichen Fokus auf. Sie reduzieren das Problem auf die numerische Darstellung eines empirisches Relationalsystems[2], welches für den Bereich der Phänomene vorausgesetzt wird (Suppes und Zinnes, 1962).

[1] Böhme ist jedoch ohne Zweifel klar, dass die Wissenschaftstheorie die Erkenntnistheorie „aufgehoben“ hat, da „die Möglichkeit von Wissenschaft weder auf das individuelle Erkenntnissubjekt noch überhaupt auf Erkenntnisvermögen heute sinnvoll gegründet werden kann, sondern vielmehr aus der Methodik kollektiv verfahrender Forschung begriffen werden muß“. (Böhme, 1976: S. 211)

[2] An empirical relational system is a relational system whose domain is a set of identifiable entities, such as weights, persons, attitude statements, or sounds. (Suppes und Zinnes, 1962: S. 10)

An dieser Stelle der Diskussion setzt Böhme ein und trennt erkenntnistheoretische und wissenschaftstheoretische Überlegungen. Er erweitert zudem den unvollständigen naturwissenschaftlichen Messprozess um fehlende Bestandteile. Für jede Messung – und im Speziellen für die Konstruktion von Indizes – ist dieser Schritt äußerst hilfreich, da er den normativen Wert entfaltet. (vgl. Abschnitt 2.2.1).

2.2 Der Messprozess nach Böhme

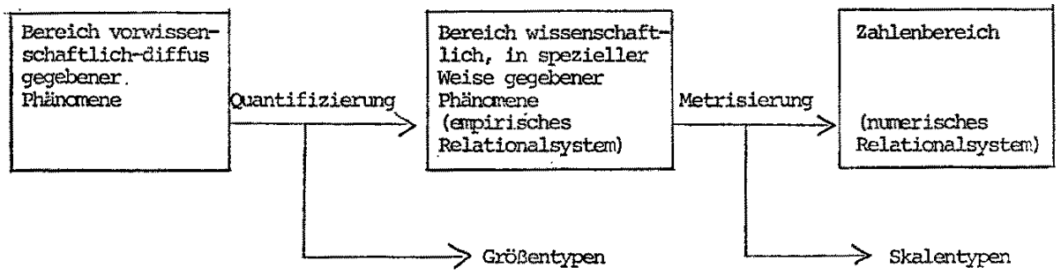
Den Ausgangspunkt Böhmes systematischer Darstellung des Messprozesses von 1976 bildet die Definition von Campbell. 1938 definierte dieser in seiner Schrift "Symposium: Measurement and Its Importance for Philosophy" *Messung* als Zuordnung von Zahlen zu Dingen oder Eigenschaften. Böhme kritisiert, dass innerhalb dieser Definition *Dinge* und *Eigenschaften* äußerst vage und vor allem unvollständig dargestellt sind. Er merkt an, dass z.B. auch Effekte, Ereignisse oder Empfindungen zahlenmäßig durch Messung bestimmt werden können und empfiehlt stattdessen die Verwendung des erweiterten Begriffs *Phänome* (Böhme, 1976: S. 12). Weitere Schwierigkeiten der Definition sieht Böhme in Campbells Verständnis der *Zuordnung* als eine Messung selbst, da Dinge an sich nicht gemessen werden können. So sei zwar die Aussage korrekt *Ich zähle die Vögel am Himmel*, jedoch laute sie im Zusammenhang einer Messung richtig *Ich messe die Anzahl der Vögel am Himmel.*

Böhme schlussfolgert, dass für die Messung von „Phänomenen" spezifische Prädikate (hier: Anzahl) entwickelt werden müssen, die den Zugang zu ihnen erst möglich machen. Mit dieser Schlussfolgerung kommt er zurück zu Campbells Definition und bemerkt die weitere Unvollständigkeit:

> „Wenn ich aber als Messung die Zuordnung der durch den Prozeß des Messens gewonnenen Zahlen zu solchen Prädikaten ansehe, so bemerke ich, daß ich sie dann schon als numerische Variable gedacht habe: Die Anzahl, die Länge, usw. ist eine Variable, die in diesem durch Messung bestimmten Fall den und den Wert annimmt." (ebd.: S. 12)

Auch Stegmüller unterschied bereits 1970 bei der Messung zwischen dem Prozess, durch den ein Prädikat als numerische Variable entwickelt wird (Anzahl, Länge etc.),

Abbildung 2.1: Messprozess nach Böhme



Quelle: Böhme (1976, S. 216)

und dem Prozess der konkreten Messung (Stegmüller, 1970). Er stellt damit fest, dass vor der Durchführung der eigentlichen empirischen Messung die begrifflichen Voraussetzungen geschaffen werden müssen. Die eigentliche Messung ist damit abhängig von der Bildung der numerischen Variable sowie transzendental dazu. Diese Unterscheidung ist eine wesentliche Erkenntnis, um den gesamten Messprozess hinreichend beschreiben und die wissenschaftstheoretischen Probleme von den erkenntnistheoretischen Fragen trennen zu können.

Der Messprozess, wie ihn Böhme in den Grundzügen 1976 definiert hat (S. 216), basiert weitestgehend auf dieser Unterscheidung von Begriffsbildung und konkreter Messung. Er ist schematisch in Abbildung 2.1 dargestellt. Die Inhalte der einzelnen Schritte des Messprozesses werden an dieser Stelle kurz erläutert und anschließend in Unterkapiteln ausführlicher diskutiert.

1. Der Bereich vorwissenschaftlich-diffus gegebener Phänomene ist als unsere lebensweltliche Umwelt zu begreifen. Über *Quantifizierungsverfahren* sollen diese Phänomene einem Zugriff zugänglich gemacht werden. Die Verfahren sind begriffsbildend und führen zu Größenbegriffen (z.B. Länge).

2. Das empirische Relationalsystem (strukturierter Phänomenbereich) ermöglicht einen wissenschaftlichen Zugriff auf die lebensweltlichen Phänomene.

3. Das numerische Relationalsystem (Zahlenbereich) wird über *Metrisierungsverfahren* geformt. Über diese findet das empirische Relationalsystem eine numerische Repräsentation. Die Abbildung wird über Skalen (vgl. Abschnitt 2.2.4) vorgenommen.

2.2.1 Quantifizierung

Der erste Teilschritt im Messprozess nach Böhme kann in die Erkenntnistheorie eingeordnet werden. Es geht um den Zugriff auf die diffusen lebensweltlichen Phänomene. Die hier entstehende Problematik ist, dass ein empirisch gegebener Bereich nicht selbstverständlich mathematisch repräsentiert werden kann:

> „Phänomenbereiche sind zunächst, d.h. vorwissenschaftlich, technisch oder lebensweltlich diffus und insbesondere nicht mit einer Struktur gegeben, die sich durch Größenbegriffe ausdrücken ließe.“ (Böhme, 1976: S. 215)

Der Phänomenbereich kann erst über Quantifizierungsverfahren zugänglich gemacht werden. Darunter fallen all jene Verfahren der Begriffsbildung, auf deren Basis Größenbegriffe gebildet werden. In der Sprache von Suppes und Zinnes geht es dabei um die Konstitution eines empirischen Relationalsystems. Dabei ist die Quantifizierung nicht das Entdecken einer an sich bestehenden quantitativen Struktur der Welt, sondern vielmehr ist sie als Teil der Organisation menschlichen Lebens zu begreifen. Sie gehorcht damit der Logik vom Mittel zum Zweck und ist auch auf Basis dieser Logik zu beurteilen. (Schlaudt, 2009: S. 8)

Außerdem wird „die Struktur des empirischen Relationalsystems [...] durch die Operationen geprägt, durch die wir uns einen Phänomenbereich in wissenschaftlicher Erfahrung zugänglich machen“ (Böhme, 1976: S. 217).

Das empirische Relationalsystem wird derart konstituiert, dass sein Begreifen in Größenbegriffen möglich wird. Dabei können u.U. verschiedene Quantifizierungsverfahren isomorphe empirische Relationalsysteme hervorbringen. Böhme spricht in solchen Fällen von denselben Größentypen. Außerdem ist anzumerken, dass der Begriff der *Größe* auch über den Folgeschritt der Metrisierung definiert wird. Denn ein Begriff stellt erst dann eine Größe dar, wenn er eine numerische Repräsentation als Vielfaches einer Einheit zulässt. (ebd.: S. 218)

Der Prozess der Quantifizierung und der Bildung eines empirischen Relationalsystems wird hinsichtlich der Konstruktion von Indizes in Abschnitt 4.3.1 genauer untersucht. Im Mittelpunkt stehen dabei die Operationen, welche die Struktur des empirischen Relationalsystems prägen. Denn die Spezifizierung des empirischen Zugangs bestimmt

den Unterschied zwischen dem wissenschaftlichen Gegenstand und dem zu messenden Phänomen, da es die Art der Spezifizierung prägt (Böhme, 1976: S. 217). Außerdem ist unter dem Aspekt der Validität das Quantifizierungsverfahren zu beachten:

> „Die Vergewisserung über das Quantifizierungsverfahren, das einem Größenbegriff zugrundeliegt, hat eine normative Funktion: Sie sichert die Validität von Messungen. Dieses Problem tritt in den Naturwissenschaften im allgemeinen nicht auf, weil man im Zuge des technischen Erkenntnisinteresses an den Naturphänomen ohnehin nur in ihrer quantifizierten Form interessiert ist. Dies ist anders in den Sozialwissenschaften, bei denen neben dem wissenschaftlich stilisierten Phänomen immer auch noch das Interesse an seiner lebensweltlichen Erscheinung erhalten bleibt (neben dem Intelligenzquotienten das Interesse an 'Intelligenz'). Dann aber wird die Frage relevant, ob man mit der Messung der jeweiligen Größe überhaupt mißt, was man wollte. Diese Frage kann nach unserer Analyse nur bedeuten, ob die Stilisierung des Phänomens zur Größe gerade seinen relevanten Aspekt herausbringt. Sie ist zu beantworten, wenn das Prinzip der Quantifizierung durchsichtig ist.“ (ebd.: S. 219f.)

Zur Erinnerung sei angemerkt, dass Böhme bewusst den Begriff (lebensweltliche) Phänomene benutzt, da er von der Messbarkeit alles empirisch Zugänglichem ausgeht. Er beschränkt das Messbare damit nicht nur auf Dinge und Eigenschaften, wie etwa Stevens und Campbell, sondern lässt auch Relationen, Qualitäten oder Substanzen zu. Dies ist notwendig, um die Theorie auch auf Messungen z.B. in den Sozialwissenschaften anwenden zu können.

2.2.2 Metrisierung

Der Prozess der Metrisierung ist laut Böhme der Wissenschaftstheorie zuzuordnen. Er bildet den Übergang von dem empirischen zum numerischen Relationalsystem. Das bedeutet, dass ein empirischer Bereich mit bestimmter Struktur innerhalb des Zahlenraums abgebildet wird. Es handelt sich dabei ausschließlich um ein Problem innerhalb

der quantitativen Begriffsbildung. Um dieses Teilproblem des Messprozesses zu behandeln, orientiert sich Böhme an der 1963 entwickelten "Basic measurement theory" von Suppes und Zinnes (1962) sowie an Stevens (1959).

Die Abbildung des empirischen Relationalsystems in den Zahlenraum wird über Skalen vollzogen. Die Operationen im Bereich der „Phänomene" spielen dabei keine Rolle mehr. Es geht einzig darum, die empirischen Relationen und den im Prozess der Quantifizierung gebildeten Begriff angemessen im Zahlenraum zu repräsentieren.

Die Metrisierung, also die Konstruktion einer numerischen Variable aus einem empirischen Prädikat, führt zur allgemeinen Skalentheorie. Erste formale Darstellungen der Maßskalen gehen auf den Amerikaner Stanley Stevens[3] zurück. In seiner Theorie der Maßskalen von 1946 veröffentlichte er vier Skalentypen (vgl. Abbildung 2.2) und diskutierte die zulässigen mathematischen Operationen. Seine Arbeit verdeutlicht, dass über bestimmte empirische Relationalsysteme bestimmte Skalen (numerische Funktionen) erstellt werden können, die hinsichtlich bestimmter Transformationen eindeutig sind. (vgl. Abschnitt 2.2.4)

Suppes und Zinnes (1962: S. 27ff.) führen zwei fundamentale Probleme in diesem Zusammenhang an: Das Repräsentationsproblem und das Eindeutigkeitsproblem. Das Repräsentationsproblem thematisiert die Notwendigkeit einer isomorphen oder homomorphen Abbildung. Das bedeutet, dass eine umkehrbar eindeutige (und bedeutungsgleiche) Abbildung zwischen den Teilen des empirischen Bereichs und des numerischen Bereichs existieren muss. Das zweite Problem thematisiert die Eindeutigkeit. Die Frage ist dabei, welche mathematischen Transformationen eine Eindeutigkeit der Abbildung zulassen.

Ohne die Erfüllung der Eindeutigkeit erscheint eine Messung nicht sinnvoll. Durch Messungen soll erreicht werden, dass eine Struktur des empirischen Bereichs durch Beziehungen zwischen Zahlen ausgedrückt werden kann (Böhme, 1976: S. 214). Bei einer nicht strukturgleichen Abbildung sind Aussagen über die numerische Abbildung nicht eindeutig auf das empirische System übertragbar, was Falschaussagen zur Folge haben kann.

[3] Hintergrund für seine Forschungsarbeit war u.a. sein Anspruch, Empfindungen „messbar" zu machen. Dieses Vorhaben wurde vorerst von der Fachwelt belächelt.

Ein weiteres Problem entsteht bei der Weiterverarbeitung der gewonnenen Daten (im numerischen System). So bestimmt die Transformationsregel (z.B. der verwendete Skalentyp) die möglichen statistisch/mathematischen Methoden. Funktionen oder statistische Modelle, die auf nicht eindeutige oder unpassende Daten angewendet werden, können zu fehlerhaften Aussagen führen. So ist z.B. die Mittelwertberechnung über eine Ordinalskala[4] nur bedingt sinnvoll, da u.U. Ergebnisse möglich sind, die im empirischen Relationalsystem nicht existieren. In der Praxis wird dieses Problem aufgrund der Zweckmäßigkeit oftmals akzeptiert.

2.2.3 Größe

Größen gehen, wie bereits angemerkt, aus empirischen Relationalsystemen hervor. Sie beschreiben die „lebensweltlichen Phänomenen“ unter dem quantitativen Aspekt und bilden damit bestimmte Relationalsysteme. Es handelt sich also bei einer Größe um quantitative Eigenschaften oder Zustände von „Phänomenen“.

Größen werden über das Produkt von Zahlenwert und *Einheit* dargestellt. Die Einheit ist dabei eine grundsätzlich willkürliche Festlegung einer Bezugsgröße der gleichen Art[5]. Bei der Messung einer Distanz ist eine mögliche Einheit z.B. der Meter. Jedoch sollte die Größe gegenüber Transformationen der Einheit invariant sein: So sollte beispielsweise das Größensystem *Länge* unter der Einheit *Meter* dieselben Aussagen treffen wie unter der Einheit *Kilometer*. (Böhme, 1976: S. 217ff.)

Prinzipiell ist ein Größenbegriff nicht an eine bestimmte Messmethode gebunden. So kann z.B. die *Länge* entweder über die *Dauer* gemessen werden, die das Licht von A nach B benötigt, oder über einen Zollstock. Beide Messmethoden ermöglichen valide Messergebnisse für die Größe *Länge*.

Die Größenverhältnisse von Gegenständen können über Zahlen ausgedrückt werden. Jedoch können die Zahlen, im Gegensatz zu Ziffern, nicht losgelöst von der Größe betrachtet werden.

[4] Eine beispielhafte Ordinalskala ist: 1 = groß; 2 = eher groß; 3= eher klein; 4 = klein.

[5] D.h., eine Masse kann z.B. nur mit einer Masse, eine elektrische Ladung nur mit einer elektrischen Ladung verglichen werden (vgl. Eulitz, Scheuermann und Thier (1968: S. 205f.)).

Im Folgenden werden extensive und intensive Größen unterschieden. Damit soll deutlich werden, dass die statistisch/mathematische Weiterverarbeitung der Messergebnisse durch die Größe beeinflusst wird.

2.2.3.1 Extensionale und intensionale Größen

Größen beschreiben Eigenschaften oder Zustände „lebensweltlicher Phänomene“, wie im vorhergehenden Abschnitt aufgezeigt wurde. Sie werden über das Vielfache einer Einheit dargestellt. Bei genauerer Betrachtung fällt jedoch auf, dass gewisse Größen unterschiedlich auf Änderungen reagieren. Dieser Umstand führt zu der Diskussion über extensionale und intensionale Größen oder über die von Leibniz als Kongruenz und Äquipollenz bezeichnete Größenschätzung (Leibniz, 1696/1996: S. 207ff.).

Eine Größe kann auf zwei Weisen gegeben sein: entweder über die Extension der formalen Anschauung oder über die Intensität der Empfindung, mit der ein empirischer Gegenstand wahrgenommen wird. Extensive Größen haben ihren Ursprung in der Form der Anschauung, intensive Größen in dem Mehr oder Weniger der „Empfindung“. Ihre Unterscheidung ist maßgeblich für die statistisch/mathematische Weiterverwendung der Messergebnisse. (Böhme, 1974)

Der Unterschied zwischen extensiv und intensiv soll anhand eines zur Hälfte mit Wasser gefüllten Glases exemplifiziert werden. Die Temperatur des halbvollen Wasserglases verändert sich nicht, wenn Wasser mit derselben Temperatur hinzugegeben wird. Die Größe *Temperatur* bleibt trotz der Veränderung der Größe des Systems unverändert.

Zustandsgrößen, die durch Größenänderungen des Systems gleich bleiben, werden intensive Größen genannt (hier: *Temperatur*). Wird beim Auffüllen des Glases jedoch statt der *Temperatur* das *Volumen* gemessen, würde das volle Glas Wasser sehr wohl einen unterschiedlichen Messwert erhalten. Zustandsgrößen, die sich durch Größenänderungen des Systems verändern, werden extensive Größen genannt (hier: *Volumen*). Erstes nannte Leibniz Größenschätzung durch Äquipollenz, letzteres durch Kongruenz (Leibniz, 1696/1996: S. 207ff.).[6]

[6] Für Leibniz findet die Bildung einer Größe grundsätzlich durch die Wiederholung eines Grundmaßes statt. Die Wiederholung könne dabei auf zwei verschiedene Weisen stattfinden: eine formelle und eine virtuelle. Bei ersterer wird das Subjekt, von dem die Kraft ausgeht, tatsächlich wiederholt. Die Größe wird aus dem Grundmaß zusammengesetzt. Ein Teil der zu schätzenden Größe wird mit dem Grundmaß in Beziehung gebracht. Leibniz spricht hier von Kongruenz (5 Meter ist gleich 5 mal 1 Meter). Bei der

Für den Teilschritt der Metrisierung (Abbildung 2.1) ist die Unterscheidung von extensiv und intensiv entscheidend. Extensive Größen sind a priori gegeben. Sie sind unabhängig von der „Erfahrung“. Die Form der Anschauung erlaubt eine Metrisierung in Form einer Verhältnisskala. Intensive Größen dagegen verhalten sich anders. So diskutiert Böhme, ob diese überhaupt als Größe bezeichnet werden sollten (Böhme, 1974), da sie keine Vielheit oder Teile ihrer selbst enthalten. Sie sind vielmehr durch eine Einheit gegeben, die durch Apprehension[7] definiert wird.

Die Erkenntnis der möglichen Zustände von Größen wird später von Stevens bei der Entwicklung der Skalentheorie aufgegriffen werden. Intensive Größen sind anders zu metrisieren als extensive und ermöglichen auch nicht die Anwendung aller statistischen Methoden. Die inadäquate Metrisierung einer Größe kann unzutreffende und falsche Aussagen zur Folge haben.

2.2.4 Die Skalentheorie nach Stevens

Für die Überführung des empirischen Relationalsystems in den Zahlenraum werden in der Regel sogenannte Skalen verwendet (vgl. Abschnitt 2.2.2). Stevens (1946) entwickelte mit seiner Schrift “On the Theory of Scales of Measurement” den Grundstein dafür. Als Rahmen für seine Skalentheorie entwickelte er eine neue Definition von Messung:[8]

> „Measurement, in the broadest sense, is defined as the assignement of numerals to objects or events according to rules.“ (ebd.: S. 677)

Das Neuartige an der Definition war, dass die Zuordnung von Ziffern zu Dingen oder Ereignissen *unter Regeln* stattfindet. Mit dieser Erweiterung der Definition versuchte er der Kritik an seinem Vorhaben, Empfindungen messen zu wollen entgegen zu treten (Schlaudt, 2009: S. 241).

Stevens wählt in seiner Definition bewusst den Begriff *Ziffer*, der für ihn zwei Bedeutungen hat (vgl. Stevens (1946)). Mal steht er für die „physische Tintenspur auf dem

virtuellen Wiederholung, bei der Leibniz von Äquipollenz spricht, wird nur im Grundmaß gedacht, es ist nicht wirklich existent in der Größe. Diese kann nicht aus dem Grundmaß zusammengesetzt werden, d.h. sie kann nicht mit dem Grundmaß kongruiert werden (5 Grad Celsius ist nicht gleich fünf mal 1 Grad Celsius). Das Grundmaß findet sich bei der virtuellen Wiederholung damit nicht in der Größe wieder, sondern vielmehr in der Wirkung oder Ursache. Die virtuellen Teile bezeichnet er daher als Grade. (Leibniz, 1696/1996: S. 207ff.)

[7] Geistige Erschließung und Bewusstwerdung einer Erscheinung.

[8] Nach dem o.g. Messprozess genauer gesagt Metrisierung.

Papier“ (Schlaudt, 2009: S. 242), mal für die rein logische Beziehung, für die eine Ziffer stehen kann. Ohne die erste Bedeutung wäre seine Theorie nicht haltbar (vgl. Abschnitt 2.2.4.1), da z.B. die Rückennummer eines Fußballspielers sonst nicht als Teil einer Skala betrachtet werden könnte und keine booleschen Aussagen zur Gruppenzugehörigkeit unter der Theorie zulässig wären.

Für einige Kritiker war Stevens Skalentheorie lediglich eine Rechtfertigung der Psychometrie, bei der nicht messbare (qualitative) „Phänomene“ durch Skalen in den Schein einer naturwissenschaftlichen Messung gehüllt werden. Tatsächlich hat die Skalentheorie entscheidend auf die weitere Entwicklung der Messtheorie eingewirkt.

Entscheidend für die Formulierung der Skalentheorie war die Klärung der Beziehung zwischen *Größe*, *Maßgröße* und *Einheit* (Berkeley, 1710/2004). Durch die Messung wird die *Größe* in Verbindung zu einer *Maßzahl* gebracht. Sie hängt jedoch nicht nur von der *Maßzahl* ab, sondern auch von der Größe der gewählten *Einheit*. Dieser Zusammenhang von *Maßzahl*, *Größe* und *Einheit* ist natürlich nachteilig für eine Messung. Wie Stevens jedoch bemerkte, ist die *Maßzahl* dem Verhältnis von *Größe* und *Einheit* eindeutig verbunden, da die *Maßzahl* die Häufigkeit der *Einheit* angibt. Da die Bedeutung der *Maßzahl* relativ zur Einheit ist, muss deren Verhältnis definiert werden. Damit wird die Relativität per Definition absolut und die Skalentheorie ermöglicht die Definition dieses Verhältnisses.

Werden mehrere Elemente des gleichen empirischen Bereichs über Skalen dem numerischen Bereich zugeordnet, bleiben die Verhältnisse unabhängig von den zugeordneten Zahlen erhalten – zumindest wenn die Zuordnung nach festgelegten Regeln vorgenommen wird (vgl. die o.g. Definition von Messung nach Stevens). So hat es keinen Einfluss auf das Verhältnis, wenn z.B. einer Gruppe von 10 Personen die Zahlen 10, 5 oder 2 zugeschrieben werden, solange die gleiche Einheit auf eine zweite Gruppe mit 20 Personen angewendet wird (hier: 20, 10 oder 4). Das Verhältnis 10/20, 5/10 und 2/4 bleibt in allen Fällen gleich.

Schwierigkeiten treten jedoch auf, sobald die Verhältnisse der Größen nicht direkt zugänglich sind und nicht linear beschrieben werden können. Sobald ein Messwert über beliebige Funktionen transformiert werden kann, entstehen Probleme zwischen der abgebildeten und der eigentlichen Realität. Die Verhältnisse sind nicht mehr strukturtreu.

Abbildung 2.2: Skalen mit empirischer Grundoperation, mathematischer Gruppenstruktur und möglicher Statistik

Scale	Basic Empirical Operations	Mathematical Group Structure	Permissible Statistics (invariantive)
NOMINAL	Determination of equality	*Permutation group* $x' = f(x)$ $f(x)$ means any one-to-one substitution	Number of cases Mode Contingency correlation
ORDINAL	Determination of greater or less	*Isotonic group* $x' = f(x)$ $f(x)$ means any monotonic increasing function	Median Percentiles
INTERVAL	Determination of equality of intervals or differences	*General linear group* $x' = ax + b$	Mean Standard deviation Rank-order correlation Product-moment correlation
RATIO	Determination of equality of ratios	*Similarity group* $x' = ax$	Coefficient of variation

Quelle: STEVENS (1946: S. 679)

STEVENS diskutiert in seiner Skalentheorie vier Skalentypen hinsichtlich ihrer mathematischen Eigenschaften. Sie werden in Abbildung 2.2 zusammengefasst und in den folgenden Unterabschnitten näher beschrieben. In Kapitel 4 wird darauf zurückgekommen, um die Eigenschaften und Werte von Indexzahlen beurteilen zu können.

Skalen sind möglich, weil isomorphe Abbildungen zwischen Gegenständen und Eigenschaften von Ziffernfolgen bestehen. Die Art der anzuwendenden Skala hängt in der Regel von der Natur der zu messenden Dinge und der Wahl der Verfahrensweise ab. Die von STEVENS definierten Skalentypen sind kumulativ zu lesen, was bedeutet, dass auf eine Skala auch die Operationen aller vorhergehenden Skalen anwendbar sind. Bei der Wahl der Intervallskala ist es also auch möglich, die Gleichheit der Intervalle zu bestimmen.

Die Metrisierung ist neben der Quantifizierung eines der Schlüsselelemente im Messprozess von BÖHME. Für die Glaubwürdigkeit der „Messung“ ist es wichtig, dass dieser Schritt transparent nachvollzogen und reproduziert werden kann. In Abschnitt 3.3 und 5.7 werden diesbezüglich Unterschiede zwischen wissenschaftlicher Kommunikation und Kommunikation eines Verbreitungsmediums wie der Frankfurter Allgemeinen Zeitung zum Vorschein treten.

Um die Qualität von Messergebnissen bewerten zu können, werden im Folgenden die vier Skalen nach STEVENS vorgestellt. Da sich seit STEVENS die Skalentheorie weiterentwickelt hat, wird diese Darstellung im Anschluss um eine zusätzliche Skala erweitert.

2.2.4.1 Nominalskala

Die Nominalskala bietet in der Anwendung die meisten Freiheiten im Vergleich zu den anderen Skalentypen. STEVENS selbst lässt offen, ob eine Abbildung auf den numerischen Raum unter Verwendung dieser Skala überhaupt als Messung bezeichnet werden sollte.

Über die Nominalskala werden empirische Prädikate über eine reine Zuweisungsregel dem Zahlen- oder Ziffernraum zugeschrieben. Dabei wäre es kein Unterschied, würden statt der Zahlen Symbole oder Buchstaben verwendet werden. Zwei Verwendungsarten ergeben sich für die Nominalskala:

1. **Individuelle Unterscheidung:** Durch das Zuschreiben von Zahlen zu empirischen Phänomenen (z.B. das Durchnummerieren von Fußballspielern) können Unterscheidungen getroffen werden, über die wiederum boolesche Aussagen möglich sind, wie z.B. A und B sind gleich oder ungleich.
2. **Differenzierung von Gruppenzugehörigkeiten:** Über das Zuweisen derselben Zahl oder Ziffer können Einheiten einer Gruppe identifiziert und unterschieden werden (z.B. Zuschreiben von Wohnungen zum jeweiligen Postleitzahlenbereich).

Ein paarweises Vertauschen der zugeschriebenen Zahlen unter der allgemeinen Substitutionsgruppe- oder Permutationsgruppe lässt die Skalenform invariant. Es spielt für die Unterscheidung z.B. keine Rolle, ob die Gemeinde A die Postleitzahl der Gemeinde B hat oder umgedreht (zulässige Transformationen: $f(x) = f(x') \Rightarrow x = x'$).

Die aus den Freiheiten der Verwendung entstehenden Ungenauigkeiten der Skala schränken, verglichen mit den anderen Skalentypen, die statistisch/mathematische Weiterverarbeitung der Ergebnisse stark ein. Aus den genannten Beispielen ergeben sich für (1) das Zählen von Spielern mit einer bestimmten Rückennummer, für (2) können zusätzliche Gruppenstatistiken erstellt werden. Dazu gehört z.B. der Modalwert, d.h. der am häufigsten auftretende Wert sowie Minima und Maxima, d.h. der am seltensten

und häufigsten vorkommende Wert. Praktisch ist es damit z.B. möglich, die Anzahl der Wohnungen verschiedener Postleitzahlenbereiche zu vergleichen. Außerdem können einfache Kontingenzmethoden für Hypothesenprüfungen zur Verteilung über die Gruppen angewendet werden. (Stevens, 1946: S. 678f.)

2.2.4.2 Ordinalskala

Eine Ordinalskala kann für die Abbildung auf einen Zahlenraum benutzt werden, sobald die Elemente des empirischen Bereichs in einer Rangfolge stehen. Das bedeutet, dass sie zu einer Aussage über größer, gleich oder kleiner fähig sein müssen (z.B. *Person A* ist größer als *Person B*).

Bei der Überführung einer empirischen Struktur in den numerischen Bereich unter Verwendung einer Ordinalskala sind alle Transformationen zulässig, die die Rangordnung nicht verändern. Abbildungen die unter Transformationen ordnungserhaltend sind, werden als isoton bezeichnet (zulässige Transformationen: $x < x' \Rightarrow g(x) < g(x')$).

Abbildungen unter der Ordinalskala ermöglichen mehr Aussagen über die empirischen Relationen als unter der Nominalskala (vgl. Abschnitt 2.2.4.1). Außerdem lassen sich zusätzliche statistische Methoden auf die unter der Ordinalskala gewonnenen Daten anwenden.

In der Praxis wird für ordinal-skalierte Daten oftmals auch der Mittelwert und die Standardabweichung berechnet. Diese Praxis wird kontrovers diskutiert. Grundsätzlich verbieten ordinal-skalierte Daten solche Berechnungen, da die Abstände zwischen den Merkmalen nicht zwingend linear sind[9]. Die pragmatische Begründung ist jedoch, dass die Verwendung fruchtbare Erkenntnisse liefert und eine Analyse der Daten andernfalls gar nicht oder nur schwierig möglich wäre.

Auch bei der Konstruktion von statistischen Indizes findet die Methode ordinal-skalierte Daten über den Mittelwert zu verdichten Anwendung. Die Aussagen aus solchen Berechnung sind achtsam zu verwenden. Es sind Situationen möglich, in denen Mittelwerte berechnet werden, die im empirischen Raum nicht existieren. Beispielsweise kann ein Mittelwert angewendet auf die Ordinalskala *1 = groß, 2= eher groß, 3 = eher klein und*

[9]Bei der Ordinalskala ist es möglich, dass der Abstand einzelner Skalenelemente nicht-linear ist. Es sind lediglich zuverlässige Aussagen über das Mehr und Minder möglich. Die Intervalle zwischen den einzelnen Skalenpunkten wirken sich direkt auf den Mittelwert aus.

4 = klein den Wert 2.5 ergeben. Dieser existiert jedoch nicht auf der Skala. Trotzdem lässt sich u.U. die Aussage treffen, dass im Mittel eine Ausprägung zwischen eher groß und eher klein gegeben ist. Jedoch sind solche Aussagen abhängig von den Intervallen der Skala. Problematisch wird z.B. die Mittelwertberechnung auf die o.g. Skala, wenn die zugeordneten Zahlen jeweils *1,2,3* und *10* betragen. U.U. wird der Mittelwert dann zwischen *3* und *10* liegen. Dieses Verhalten leitet über zu der Frage, ob der Abstand zwischen dem ersten und zweiten Skalenpunkt in der Realität tatsächlich um ein Vielfaches geringer ist als zwischen dem dritten und vierten. Sind die Abstände zwischen den Intervallen bekannt und interpretierbar, ist die Intervallskala (vgl. Abschnitt 2.2.4.3) anzuwenden.

Wie bereits angedeutet lässt die Ordinalskala, im Vergleich zur Nominalskala, die Verwendung von zusätzlichen statistischen Methoden zu. Darunter fällt u.a. die Korrelationsanalyse. Mit ihrer Hilfe können z.B. „Wenn-dann-Beziehungen“ aufgedeckt werden. Kausalbeziehungen lassen sich damit jedoch nicht beweisen. (Stevens, 1946: S. 679)

2.2.4.3 Intervallskala

Abbildungen unter der Intervallskala führen zu quantitativen Daten im eigentlichen Sinne. Im Gegensatz zu der Ordinalskala (vgl. Abschnitt 2.2.4.2) sind bei der Intervallskala die Abstände zwischen den Skalenelementen exakt definiert. Einschränkungen bei der Weiterverwendung der Daten ergeben sich durch den fehlenden Nullpunkt und aus dem Umstand, dass Verhältnisse nicht direkt abgebildet werden. Die Einschränkungen können anhand der Temperaturskala Grad-Celsius verdeutlicht werden: Der Nullpunkt dieser Skala wurde willkürlich festgelegt, was im Vergleich zur Kelvin-Temperaturskala deutlich wird. Celsius und Kelvin nutzen beide dieselbe Quantifizierungfunktion. Sie bestimmen die Temperatur über den Volumenzuwachs. Ihre Werte sind bei gleicher Temperatur jedoch verschieden. Dies liegt an der Festlegung des Nullpunktes, der bei der Kelvin-Temperaturskala im „absoluten Nullpunkt“ liegt. Durch Addition einer Konstanten können beide Temperaturskalen ineinander umgerechnet werden (Kelvin = Grad Celsius + 273,15 K). Zulässige Transformationen sind: $h(x) = ax + b$ mit $a \neq 0$.

Direkte Aussagen über die Verhältnisse sind unter der Intervallskala nicht zulässig. Es kann z.B. nicht die Aussage getroffen werden: *es ist doppelt so warm wie gestern*, wenn heute 20 Grad Celsius gemessen wurden und gestern 10 (vgl. Abschnitt 2.2.3.1). Ein

Ausweg bietet der Vergleich von Verhältnissen oder das Verhältnis von Verhältnissen. Der Vergleich von Verhältnissen ist reine Datenverarbeitung und hebt das Skalenniveau von Intervall-Niveau auf das Verhältnis-Niveau (vgl. Abschnitt 2.2.4.4). Bezogen auf das eben angeführte Beispiel ist die Aussage zulässig: *die Temperaturänderung von gestern auf heute ist doppelt so hoch, wie die von vorgestern zu heute* (heute-gestern / heute-vorgestern = 2).

Auf intervallskalierte Daten können alle statistischen Methoden angewendet werden, die keinen Nullpunkt erfordern. Dabei sind beispielsweise Durchschnittswerte sinnvoll, sämtliche Multiplikationsoperationen jedoch nicht. (Stevens, 1946: S. 679)

2.2.4.4 Verhältnisskala und Absolutskala

Die Verhältnisskala ist eine metrische Skala – laut Skalentheorie nach Stevens die mit dem höchsten Skalenniveau. Die Verwendung der Verhältnisskala erfordert einen natürlichen Nullpunkt sowie Operationen zur Bestimmung aller vier Relationen: Gleichheit, Rangfolge, Gleichheit der Intervalle und Gleichheit der Verhältnisse.

Nur bei der Verhältnisskala sind Multiplikation und Division erlaubt, d.h. es können Verhältnisse beliebiger Art $x = y * z$ gebildet werden. Außerdem können die Werte der Skala durch Multiplikation einer Konstanten beliebig transformiert werden. Zulässige Transformationen sind: $h(x) = ax$ mit $a \neq 0$. (ebd.: S. 679f.)

Neben der Intervall- und Verhältnisskala wurde eine weitere Kardinalskala eingeführt, bei der die Abstände mess- und interpretierbar sind. Die Absolutskala wird in Stevens Schrift zwar nicht explizit genannt, da sie Bezug zu dem von Campbell geprägten Begriff Messen durch Zählenäufweist, wird die Skala an dieser Stelle zur Vollständigkeit aufgenommen. Die Absolutskala erfüllt dieselben Voraussetzungen wie die Verhältnisskala. Zusätzlich ist ihre Maßeinheit natürlich gegeben, wobei die Skaleneinheit nicht frei wählbar ist (zulässige Transformationen sind: $h(x) = x$). Ein Beispiel für diese Skala ist die Bevölkerungszahl eines Stadt mit dem Nullpunkt "keine Einwohner". (Saint-Mont, 2011: S. 30)

Alle fünf Skalentypen führen zu einer Abbildung des empirischen Bereichs in den Zahlenbereich und entfalten in diesem Zusammenhang Relevanz für die Konstruktion von Indizes. Die Methoden der Konstruktion stehen dabei in direkter Abhängigkeitsbeziehung zur angewendeten Skala.

2.2.5 Zahlen und das numerische System

Werden aus Relationen im numerischen Raum Aussagen auf die „Umwelt“ übertragen, ist das entsprechende Wissen über die Größe und deren Messung notwendig, denn erst mit der Kenntnis von Metrisierung und Quantifizierung können valide Aussagen über die Umwelt getroffen werden. Durch Mathematik und Statistik können aus Zahlen neue Zahlen erzeugt werden – völlig ohne Bezug zu deren Bedeutung. Die Glaubwürdigkeit von Messergebnissen hängt dann nicht nur vom Messprozess ab, sondern auch von den verwendeten mathematischen und statistischen Methoden.

Die Grundlage des numerischen Relationalsystems wird von Zahlen gebildet. Auf die Relationen zwischen den Zahlen lassen sich wiederum statistische Verfahren anwenden, wodurch sich aus den Zahlen neue Zahlen entstehen. Der notwendige Prozess verläuft auf einer rein syntaktischen Ebene, ist also entkoppelt von der Bedeutung – ähnlich der Deduktion eines formalen Beweises (vgl. Vollmer (2003)). Auf der syntaktischen Ebene stehen Zahlen nicht nur für verschiedenste „Phänomene“, sondern potentiell auch für andere Zahlen. Durch Aggregation, Transformation oder Kombination kann der Raum an Vergleichsmöglichkeiten vergrößert werden, z.B. über die Erstellung von Ranglisten. (Heintz, 2007: S. 79)

Die syntaktische Ebene hat Dedekind (1893) bei seiner Definition von Zahlen noch nicht berücksichtigt. In *Was sind und was sollen die Zahlen?* geht er wie folgt auf die Bedeutung von Zahlen ein:

> „Die Zahlen sind freie Schöpfungen des menschlichen Geistes, sie dienen als Mittel, um die Verschiedenheit der Dinge leichter und schärfer aufzufassen.“ (S. VII,VIII)

Die Aussage von Dedekind bezieht sich ausschließlich auf die Abbildung von „Phänomenen“ in den numerischen Raum und muss in Hinblick auf die moderne (reine) Mathematik ausgebaut werden. Bei der Abbildung von „Phänomenen“ beziehen sich Zahlen

auf einen externen Referenten. Differenzen lassen dabei ihre Unterscheidung zu. In der Mathematik sind die Bezüge jedoch rein interner Natur. Es existiert kein Verweis auf einen externen Referenten. Die Mathematik operiert nicht mehr mit Zahlen, sondern genauer gesprochen mit Ziffern und Zeichen (vgl. Frege (1884/1990)).

Die Mathematik in Form eines geschlossenen Systems, welches Objekte und Regeln selbst erzeugt, ist eine Entwicklung des 19. Jhd. (Heintz, 2007: S. 71ff.). Bei der Betrachtung von Zahlen muss damit neben dem Messprozess auch immer die syntaktische Ebene einbezogen werden, da innerhalb dieser u.U. Aggregationen, Transformationen oder Kombinationen angewendet wurden. Die Glaubwürdigkeit einer Zahl hängt nicht nur von der Transparenz der angewendeten Messverfahren ab, sondern auch von den angewendeten statistisch/mathematischen Verfahren. Diese können, müssen aber nicht selbstreferenziell sein.

Die oben genannten Eigenschaften zeigen den wesentlichen Unterschied zwischen Ziffern und Zahlen: Zahlen, die auf Messungen basieren, verweisen auf einen externen Referenten. Ihre Bedeutung kommt nur über die Referenz zustande. Ziffern (als Zeichen) hingegen sind selbstreferenziell und besitzen keinen externen Bezug.

Um beurteilen zu können, welche statistisch/mathematischen Verfahren auf Zahlen angewendet werden dürfen, ist das Wissen über den Quantifizierungs- und Metrisierungsprozess notwendig. Zahlen allein kann nicht angesehen werden, welche Verfahren zulässig sind. Wie oben bereits angedeutet, kann auf eine Reihe von Zahlen immer der Mittelwert gebildet werden. Jedoch ist diese Berechnung nicht immer zielführend. Ob dieses Verfahren der Aggregation effektiv ist, lässt sich erst unter dem Wissen der angewendeten Skala oder der Metrisierung feststellen (vgl. 2.2.4). In diesem Zusammenhang lohnt die Betrachtung, ob und ggf. inwieweit Journalisten und Autoren in ihren Beiträgen eigenständige Aussagen oder Schlussfolgerungen auf Basis von Zahlen formulieren und ob diese adäquat sind.

In Abschnitt 3.1.1 wird das Thema *Zahlen* in den historischen Kontext eingeordnet.

2.3 Bezug von Indikatoren und Indizes zum Messprozess

Im oben dargestellten Messprozess nach Böhme wurden die verschiedenen Messarten der messtheoretischen Literatur deutlich: (1) *fundamentales Messen*, (2) *abgeleitetes Messen* und (3) *Messen durch Zählen*. Die drei Prozesse sollen an dieser Stelle zusammengefasst dargestellt werden.

Die beiden Begrifflichkeiten fundamentales und *abgeleitetes Messen* gehen auf Campell (1928) zurück. Unter *fundamentalem Messen* versteht man die Zuordnung von Zahlen zu Objekten und Eigenschaften. Diese wird aus den „Naturgesetzen“ abgeleitet, welche auf das zu messende Phänomen einwirken. Daraus ergibt sich eine wichtige Eigenschaft der fundamentalen Messung – sie ist unabhängig von der Messung anderer Größen. Die hier auftretenden Operationen des Vergleichs und der Kombination sind experimentell nachvollziehbar, was sie unabhängig von Raum, Zeit und Person reproduzierbar macht. So wird beispielsweise die *„Masse“* eines Körpers fundamental über das Verhalten des Objektes in einem reproduzierbaren Raum gemessen. Bei der Messung über eine Federwaage wird dafür die Dehnung der Feder benutzt. Dabei soll das Verhalten der *Dehnung* über das theoretisch eingeführte Konzept „Masse“ erklärt werden. Dessen Einführung verleiht den Vergleichs- und Kombinationsoperationen sowie den Zuordnungsregeln erst ihre theoretische Bedeutung. Die Verifikation letzterer erfolgt durch die Wirkung der Naturgesetze. (Besozzi und Zehnpfennig, 1976: S. 11)

Beim *abgeleiteten Messen* werden die Regeln für die Zuordnung nicht aus den Naturgesetzen abgeleitet, sondern aus einer bestehenden Theorie. Die konstituierenden Variablen der Theorie sind messbar oder wurden bereits gemessen, das zu messende Konstrukt jedoch nicht. So wird z.B. die *„Kraft“* mittels des zweiten Newtonschen Gesetzes $F = m*a$ über die *Masse* m und die *Beschleunigung* a bestimmt. Abgesehen von Fehlern durch die Messapparatur sind abgeleitete Messungen ebenfalls reliabel, da ihre Bestandteile fundamental bestimmt werden und ihr Zusammenhang über die Naturgesetze determiniert ist.

Coleman (1964) erweiterte die Begrifflichkeiten von Campell um die des *Messens durch Zählen* (S. 71). Eine Phrase, die innerhalb der verschiedenen Wissenschaftsdisziplinen kontrovers diskutiert wird. Sein Ausgangspunkt dabei ist die Soziologie, in der

Messen oftmals auf bloßen Zählungen basiert. Über simple Quotienten werden z.B. Anteile einer Ethnie an der Gesamtbevölkerung berechnet. So basiert beispielsweise der Anteil an Christen einer Region auf der Zählung aller Personen des Gebiets, die ins Verhältnis gesetzt werden zu der Anzahl der dort lebenden Christen. Hier werden die o.g. klassischen Messkriterien lediglich implizit erfüllt, da die Operationen des Vergleichs und der Kombination auf einer Klasseneinteilung (z.B. Religionszugehörigkeit) basieren. Die Grundlage ist somit eine Theorie und nicht das Verhalten der Objekte. Trotz dieses Unterschieds ist bei einer Messung durch Zählen auch eine Isomorphie vom numerischen System und den Objekten gegeben. Außerdem ist auch die Bedingung der Eindeutigkeit der Zuordnung erfüllt. Allerdings erfolgt die Validierung der Zuordnung nicht über die Messoperationen selbst, sondern über die Überprüfung der Theorie.

In Abgrenzung zu den dargestellten Messarten (1) bis (3) wird im Folgenden der Versuch unternommen, den Zusammenhang von Indikatoren und Indizes zum Messprozess nach Böhme zu erarbeiten.

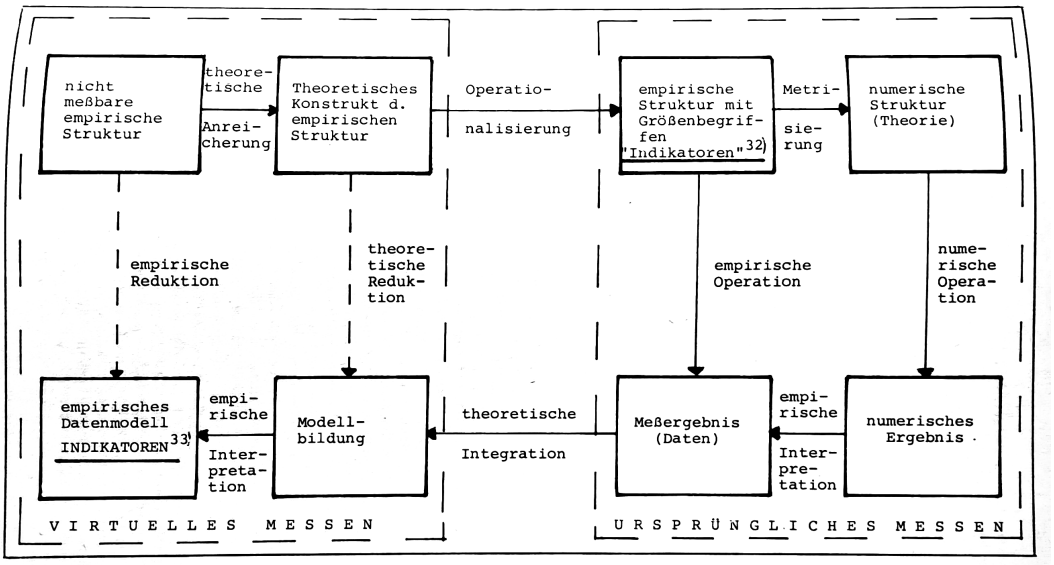
2.3.1 Indikatoren

Der Messprozess nach Böhme (vgl. Abschnitt 2.2) basiert auf Quantifizierungsverfahren, die es ermöglichen, „Phänomene“ über Größenbegriffe in ein empirisches Relationalsystem zu überführen. Damit geht der Prozess von beobachtbaren, messbaren oder realisierbaren „Phänomenen“ aus. Doch wie geht man mit einem theoretischen Konstrukt um, zu dem kein unmittelbares empirisches Korrelat gefunden werden und außerdem die verwendeten Terme nicht der Beobachtungs- und Messsprache zugeordnet werden können, sondern ausschließlich dem Vokabular der Theorie?[10]

Aus diesem Messproblem geht das *indirekte Messen* hervor. Es handelt sich dabei um einen Prozess der wissenschaftlichen Informationsgewinnung (Werner, 1975a: S. 56f.). Abbildung 2.3 fasst den gesamten Prozess nach Randolph zusammen. Der Bereich des *ursprünglichen Messens* entspricht dabei dem Bereich rechts von der Quantifizierung im Messprozesses nach Böhme (vgl. Abschnitt 2.2). Randolph betrachtet indirektes

[10] Für einen zusammenfassenden Überblick über Beobachtungssprache und theoretische Sprache mit Bezug auf Carnap sei auf Besozzi und Zehnpfennig (1976: S. 14-20) verwiesen.

Abbildung 2.3: Schematische Darstellung des indirekten Messens



Quelle: Randolph (1979, S. 27)

Messen als zwei verbundene Prozesse: Der ursprünglichen Messung[11] ist als Reflexion ein virtueller Ablauf gegenübergestellt.[12] Die virtuelle Messung macht das nicht direkt beobachtbare „Phänomen“ über Indikatoren dem Prozess der ursprünglichen Messung zugänglich. Dies sei sinnvoll, wenn das theoretische Konstrukt empirisch gehaltvoll, sein Basisbereich jedoch nicht messbar ist. (Randolph, 1979: S. 28)

Indikatoren können in diesem Messprozess laut Randolph als Instrumente für das indirekte Messen angesehen werden. Sie treten an der Schnittstelle zwischen virtueller und ursprünglicher Messung in Erscheinung (vgl. Abb. 2.3). An diesem Punkt sieht Randolph Ähnlichkeiten zur Erfindung von Teleskop und Mikroskop in der Physik, da diese Erfindungen zum Übergang von direkter zu indirekter Beobachtung führten und damit eine gewisse Analogie zum methodischen Instrument der Indikatormessung in den Sozialwissenschaften darstellen. Jedoch machen „operationalisierende Indikatoren [...] lediglich einen Schritt im Prozess des indirekten Messens aus; bildlich gesprochen also eine Linse, aber kein Mikroskop“ (ebd.: S. 30). Die Indikatoren sind in diesem Zusammenhang die quantitative Struktur der nicht messbaren Struktur und entfalten damit nach Randolph (ebd.: S. 30) zwei Funktionen oder Wirkungsweisen:

[11] Randolph definiert das ursprüngliche Messen als die einfachste Art der Messung, bei der den Intensitäten eines Merkmals reelle Zahlen direkt zugeordnet werden können (z.B. Länge Gewicht). Die Relationen der Ausprägungen des Merkmals werden im numerischen Raum abgebildet (Randolph, 1979: S. 22).

[12] Von dieser virtuellen Messung aus betrachtet, erscheint die ursprüngliche Messung als indirekt.

1. **Operationalisierung:** Indikatoren operationalisieren theoretische Konstrukte, d.h. sie überführen sie in messbare Begriffe. Sie verbinden rein theoretische Terme mit empirischen.

2. **Modellfunktion:** Für nicht unmittelbar messbare und beobachtbare empirische Sachverhalte stellen Indikatoren eine Modellfunktion dar.

Hinsichtlich der Begrifflichkeit der Indikatoren ist darauf zu achten, dass sie im Prozess der indirekten Messung an zwei Stellen in Erscheinung treten – an der Schnittstelle zwischen virtuellem und ursprünglichem Messen, zusätzlich jedoch auch als das Ergebnis der indirekten Messung. Im Vergleich zu Messungen in den klassischen Naturwissenschaften steht die indirekte Messung vor einigen Herausforderungen. Die Genauigkeit und Zuverlässigkeit der Messungen in den Naturwissenschaften orientiert sich an der ausnahmslosen Gültigkeit der Naturgesetze. Bei Indikatoren und der indirekten Messung ist dies i.d.R. nicht der Fall. Wie Randolph zeigt, orientieren sich Indikatoren nur bedingt an dem Ideal der klassischen Naturwissenschaften und den damit verbundenen Anforderungen. Er führt die Ursache u.a. darauf zurück, dass die „idealen Bedingungen indirekten Messens im realen Meßprozeß stets nur in mehr oder weniger großer Annäherung gegeben sein können [...]“ (Randolph, 1979: S. 60). Das zeige sich insbesondere in der Sozialwissenschaft. Dort ist der Mensch der Untersuchungsgegenstand und es ist in Folge dessen nicht ohne Weiteres möglich, wie z.B. in der Physik, ein Experiment mehrmals durchzuführen. Wird ein Proband zum wiederholten Mal der gleichen Situation ausgesetzt oder mit der gleichen Frage konfrontiert, ist er nicht mehr unvoreingenommen. Diese Konditionierung führt zu dem Problem der statistischen Unabhängigkeit.

Bei der Wahl von Indikatoren als quantitative Annäherung an das theoretische Konstrukt ist Willkür zu erwarten, da oftmals verschiedene Größen und Proxyvariablen als „Substitute“ in Betracht kommen. Randolph (ebd.: S. 60f.) nennt in diesem Zusammenhang einige Beispiele, die zu Problemen bei der indirekten Messung führen:

- Es fehlen Theorien und Gesetzmäßigkeiten für gültige indirekte Messungen.
- Es besteht die Gefahr, dass Bereiche unbeachtet bleiben, die bisher nicht quantifiziert wurden.
- Messinstrument und Messobjekt könnten sich gegenseitig beeinflussen.

An die Stelle von Gesetzmäßigkeiten für eine „gültige“ indirekte Messungen treten verschiedene Ansätze der Indikatorenbildung. Werner nennt in Zusammenhang zu Politik und Gesellschaft folgende: a) normativer Ansatz, b) Performanzansatz, c) systemwissenschaftlicher Ansatz und d) subjektiver Ansatz. Im Folgenden werden die Ansätze nach Werner (1975a: S. 63-92) grob zusammengefasst:

a) Der normative Ansatz ist laut Werner der beherrschende der Anfangsbewegung sozialer Indikatoren. Konzeptioneller Ausgangspunkt sind Normen und Interessen, die im Wesentlichen auf Regierungsebene deklariert werden. Auf diese Weise sollen Indikatoren wertgebundene Zustandsanalysen liefern. Ihre Auswahl, Definition und Erhebung wird durch regierungsamtliche oder ihr nahestehende Interessen geprägt. Wegen der Bindung an Interessen ist dieser Ansatz äußerst kritisch zu hinterfragen, da die Ergebnisse ggf. nicht objektiv sind.[13]

b) Der Performanzansatz zielt auf die Perzeption und Evaluation von Maßnahmen ab. So geht es z.B. nicht um eine allgemeine Einschätzung der Entwicklung, sondern um die Prüfung der Beurteilung des Erfolgs bestimmter Maßnahmen.

c) Der systemwissenschaftliche Ansatz behandelt die Zusammenfassung mehrerer Indikatoren zu einem System, wobei die Indikatoren alle einem Gegenstandsbereich entstammen. Dieser wird als mehrdimensional aufgefasst und daher nur über ein System von Indikatoren adäquat wiedergegeben.

d) Der Untersuchungsgegenstand des subjektiven Ansatzes ist das Individuum mit seinen Verhaltensweisen, Einstellungen und Satisfaktionen. Die Erhebungsfelder sind z.B. individuelle Ziele, Wünsche, Hoffnungen, Urteile oder Zufriedenheiten.

Indikatoren bieten eine Möglichkeit, nicht direkt beobachtbare Objekte zu quantifizieren. Die Verbindung zwischen dem zu messenden latenten Konstrukt und dem beobachtbaren Objekt wird über die Theorie hergestellt und resultiert in einem Indikator. Jedoch existieren auch theoretische Konstrukte, die über einen einzigen Indikator nicht hinreichend genau abgebildet werden können. Die Verwendung von mehr als einem Indikator zur Bestimmung eines theoretischen Konstrukts führt zum Indexproblem.

[13] Ein formales Beispiel für die Bildung eines normativen Index bietet Formel 4.10.

2.3.2 Indizes

In der Literatur herrscht Uneinigkeit, ob Indizes und Indikatoren als Messung betrachten werden sollten und welche Bedeutung sie überhaupt besitzen. Besozzi und Zehnpfennig (1976) schreiben dazu einleitend:

> „Das Problem der Index-Bildung kennzeichnet eine in den Sozialwissenschaften häufig anzutreffende Situation, in der man nicht ganz genau weiß, was man eigentlich mißt. Diese Unwissenheit bezieht sich sowohl auf die inhaltliche Bedeutung als auch auf die quantitativen Zusammenhänge des zu messenden Konzepts und anderen, die Messung beeinflussenden Variablen.“ (S. 9)

Besozzi und Zehnpfennig meinen mit dem Begriff *Messen* eine indirekte Messung oder nach dem Vokabular von Torgerson (1958) *Messen by fiat*. Damit werden Indizes im Messprozess genauso behandelt wie Indikatoren. Beim indirekten Messen sind die Zuordnungsregeln der Zahlen zu den Objekten arbiträr. Das bedeutet, dass sie aus den postulierten Zusammenhängen der messbaren Objekte des Konzepts und dem Konzept selbst abgeleitet werden. Der Zusammenhang zwischen den beobachtbaren Objekten und dem theoretischen Konzept und der damit verbundenen nicht beobachtbaren latenten Variablen wird also aus der Theorie abgeleitet. Der Unterschied von Indizes besteht darin, dass sie mindestens zwei Indikatoren enthalten und in dieser Hinsicht eine höhere Komplexität und Dimensionalität aufweisen.

Es ist umstritten, ob solche Zuordnungsregeln als Messung bezeichnet werden sollten (vgl. Besozzi und Zehnpfennig (1976: S. 11)). Schnell (2005) grenzt die Indexbildung klar von Messungen ab:

> „Sowohl die Indexbildung als auch die Skalierungsverfahren sind Auswertungsverfahren, keine Datenerhebungs- oder Messverfahren.“ (S. 158)

Sehr eingeschränkt wird das Problem der Indexbildung von Flaskämper (1928) behandelt, indem er Indizes auf eine Sonderform des statistischen Vergleichs reduziert (S. VI).[14] Beiden gemein ist, dass die Konstruktion von Indizes außerhalb des Messprozesses eingeordnet wird, d.h. auf den Abbildungen im numerischen Raum.

[14] Vgl. auch Fisher (1923) und Ralph (2015).

Wie Besozzi und Zehnpfennig (1976: S. 9) zeigen, mangelt es bis heute grundsätzlich an einer einheitlichen Systematisierung der Indexkonstruktion.[15] Die Diskussion zum messtheoretischen Status der Konstruktion von Indizes sei unzureichend, was eine rein forschungstechnische Behandlung des Problems verbiete. Sicherlich wird dieser Zustand z.T. von den unterschiedlichen Zielstellung der Forschungsdisziplinen beeinflusst. Insbesondere in den Sozial- und Politikwissenschaften steht die Messung theoretischer Konzepte bei der Indexkonstruktion im Vordergrund. Nicht selten existiert dabei ein wesentlicher Unterschied zwischen Theorie und Praxis. Indizes werden ad hoc gebildet, ohne dass der Zusammenhang von Konzept und Indikator innerhalb einer Theorie spezifiziert wurde. Die theoretische Relevanz der Konzepte wird jedoch über ihre ursprüngliche Bedeutung abgeleitet, ohne dass sie über eine konzeptuelle Analyse untersucht wurde. Dieser Umstand bietet viel Raum für Diskussionen und verhindert z.T. die korrekte Interpretation der Indexwerte (vgl. Kapitel 5).

Besozzi und Zehnpfennig (ebd.: S. 13) zeigen weiter, dass neben dem ungeklärten Status von Indizes innerhalb des wissenschaftlichen Messprozesses erschwerend hinzukommt, dass zum Terminus *Index* in der Literatur zahlreiche alternative Bezeichnungen verwendet werden. In der Test- und Skalierungstheorie wird u.a. vom *Gesamt-Score* gesprochen. Dieser basiert auf *Indikatoren*, die jedoch oft als *Item* bezeichnet werden. Der Begriff Index ist in diesem Fall eine mehr oder weniger willkürliche Zuordnung von Beobachtungen zu einer latenten Variablen (dem Konzept). Teilweise werden *Index* und *Indikator* auch alternativ verwendet (Lazarsfeld und Rosenberg, 1956). Manchmal wird statt dem Begriff *Indikator* auch *Variable* benutzt. In der Ökonomie sind bisweilen auch die Begriffe *Indexziffern*, *Indexzahlen* sowie *Kompositindikatoren* gebräuchlich. Das Beispiel der Klimaerwärmung als mehrdimensionales Konstrukt und damit Index verdeutlicht eine weitere Schwierigkeit zu den alternativen Schreibweisen – in der Praxis wird z.T. in der Bezeichnung erst gar nicht kenntlich, dass es sich bei der Maßzahl um eine Zusammenfassung mehrerer Dimensionen oder ein theoretisches Konzept handelt.

In der Wissenschaftsliteratur lassen sich für Indizes und Indikatoren derzeit die in Tabelle 2.1 zusammengestellten Termini beobachten.

[15] Der aktuelle Stand der Systematisierung wird in Kapitel 4 dargelegt. Für einen Einblick in die Heterogenität der Konstruktionsmethoden von Indizes sei auf Bonjean, Hill und MacLemore (1967) verwiesen.

Tabelle 2.1: Alternative Bezeichnungen für Index und Indikator

Konzept	Alternative Bezeichnung
Index	Indexziffer
	Indexzahl
	Kompositindikator
	Indikator
	Gesamt-Score
	Kennzahl
Indikator	Item
	Variable

Quelle: Eigene Darstellung

In der vorliegenden Arbeit werden quantifizierte theoretische Konzepte mit dem Terminus *Index* bezeichnet. Unter dem Begriff Index wird dabei ein („Mess-“) Modell der folgenden allgemeinen Form verstanden

$$I_k = f(x_1, x_2, ..., x_k). \tag{2.1}$$

Dabei sind $x_1, x_2, ..., x_k$ die Indikatoren, mit denen die Variable X bestimmt werden soll. Sie sind die Elemente des Index I_k. Formel 2.1 verdeutlicht damit den Zusammenhang der beobachtbaren Indikatoren und dem theoretischen Konzept. Mit der Begrifflichkeit von Carnap (1956) gesprochen, wird damit eine Verbindung von Beobachtungssprache und theoretischer Sprache hergestellt.[16]

An dieser Stelle wird die Auffassung von Werner (1975b) geteilt und das Problem der Indexkonstruktion als Reduktion eines manifesten, n-dimensionalen Eigenschaftsraums zu einer oder mindestens *n-1* Dimensionen angesehen:

> „Die Konstruktion eines Index kann aufgefasst werden als die Reduktion einen n-dimensionalen Merkmalsraumes R^n in einen Raum R^m, wobei $m < n$. Meist ist $m = 1$ [...]“. (S. 1)

Dem folgend, muss in Formel 2.1 immer $k \geq 2$ sein. Das bedeutet, dass ein Index nur in einer Situation konstruiert werden kann, in der keine Variable existiert, die das zu bestimmende Konzept konsistent genauer „messen“ kann als jede mögliche Kombination der Indikatoren-Menge $x_1, x_2, ..., x_k$. (Besozzi und Zehnpfennig, 1976: S. 12)

[16] Für einen zusammenfassenden Überblick von Beobachtungssprache und theoretischer Sprache bezüglich der Indexkonstruktion sei auf Lazarsfeld und Rosenberg (1956: S. 14ff.) verwiesen.

Aus der formalen Darstellung folgt, dass Indizes dann Anwendung finden, wenn ein Konstrukt oder theoretischer Begriff operationalisiert werden soll, welches nicht über einen einzelnen Indikator abgebildet werden kann. Grund kann entweder sein, dass ein einzelner Indikator das Konstrukt nicht hinreichend genau beschreibt oder die Theorie verschiedene Dimensionen anspricht. (Schnell, 2005: S. 158)

Das Ziel des Index ist es, ein mehrdimensionales (latentes) Konstrukt anhand eines meist eindimensionalen Maßes zu beschreiben. Als Beispiel für solch ein Konstrukt nennt Zeisel (1970: S. 78ff.) z.B. die Veränderungen in den Preisen der Waren und Dienstleistungen, die über den Verbraucherpreisindex quantifiziert werden. Ausgewählte Produkte werden in einem virtuellen Warenkorb zusammengestellt und fließen auf diese Weise als Indikatoren in den Index ein. Beispiele für latente Konstrukte sind die Messung von „Intelligenz“ über den *Intelligenzquotienten* oder die Bestimmung des „Glücks“ einer Nation über den *Happy Planet Index*. Hierbei spricht die Theorie verschiedene Dimensionen an[17], postuliert jedoch eine einzige gemeinsame latente Variable.

Der Vorteil eines solchen eindimensionalen Maßes ist Einfachheit und Übersichtlichkeit. Außerdem ermöglicht es den Vergleich von komplexen Sachverhalten z.B. in Form von Rangfolgen (Polzin, 2007: S. 70) und trägt damit zur Reduktion von Komplexität und Informationsfülle bei (Werner, 1975a: S. 151ff.).

Laut Werner (1975b) handelt es sich bei Indizes um ein reines Aggregationsproblem, bei dem n Größen oder Indikatoren zu einer oder mindestens $n\text{-}1$ Dimension zusammengefasst werden. Die Erzeugung von Zahlen aus Zahlen (z.B. über Aggregation) findet oft auf einer rein syntaktischen Ebene statt. Das bedeutet, dass die Konstruktion von Indizes dann weitgehend unabhängig von den erhobenen Daten und ihrer Bedeutung ist und grundsätzlich immer durchgeführt werden kann. Die entstehenden Zahlen können prinzipiell für alles stehen – für verschiedene „Phänomene“, aber auch für andere Zahlen (Heintz, 2007: S. 79). Power (2004) bildet diesbezüglich den Begriff „metameasurement“, Vollmer (2003) den Begriff „Hyperrealität“ von Zahlen. Aussagen die dem Index entnommen werden, haben definitionsgemäß keinen Bezug zur lebensweltlichen Umwelt, sondern entspringen einzig dem Vokabular einer Theorie. In diesem Zusammenhang entsteht schnell die Frage, was und ob der Index eigentlich misst, quantifiziert oder abbildet. Was den *Intelligenzquotienten* betrifft, so lautet die genaue Antwort, die

[17] Bei der Intelligenz z.B. Merkfähigkeit, Einfallsreichtum, zahlengebundenes Denken etc..

Leistung in einer Anzahl von Testsituationen. „In der Regel mißt der Index [aber] noch etwas Zusätzliches, das indirekt mit dem, was er direkt mißt, verbunden ist, etwas, das dann oft das eigentliche Objekt des Index ist“ (Zeisel, 1970: S. 82). Hier offenbart sich ein Unterschied zu den Naturwissenschaften. Denn diese sind im Zuge des technischen Erkenntnisinteresses nur am Messobjekt in seiner quantifizierten Form interessiert. Hingegen hat beim Intelligenzquotienten die Psychologie neben der wissenschaftlich stilisierten Form von Intelligenz – dem Intelligenzquotienten – auch das Interesse an seiner lebensweltlichen Erscheinung- der Intelligenz (vgl. Abschnitt 2.2.1).

Zeisel (ebd.) formuliert den Gegenstandsbereich des Index in Folge dessen deutlich weiter, als es der wissenschaftlich stilisierte Begriff zulassen würde:

> „Im allgemeinen kann man sagen, daß ein Index all das mißt, was irgendwie mit dem Gegenstand, der das direkte Objekt der Messung ist, zusammenhängt.“ (S. 82)

Das Zitat verdeutlicht, wie diffus Indexkonzepte sein können. Insbesondere wenn die Beziehungen zu anderen Objekten „irgendwie“ definiert sind. Sicherlich muss auch in Frage gestellt werden, ob der Begriff „Messung“ bei solchen Konzepten sinnvoll ist. Indizes sehen sich aktuell der großen Herausforderung gegenüber, die Beziehung der einzelnen Merkmale zum Indexkonzept wissenschaftlich zu fundieren und angemessen zu kommunizieren. Sobald diffuse Begrifflichkeiten oder Gefühle wie „Glück“, „Liebe“ oder „Vertrauen“ über Indizes operationalisiert werden sollen, muss der Bezug der Indikatoren zum Indexkonzept hinreichend beschrieben werden. Jede Person hat vermutlich eine andere Definition und Auffassung von z.B. „Glück“. Wird dieser Begriff dennoch über eine Auswahl von Indikatoren operationalisiert, muss davon ausgegangen werden, dass die Merkmale einen neuen Gegenstand konstituieren, der mehr oder weniger mit der individuellen Auffassung und Semantik von „Glück“ gemeinsam haben. Eine transparente Darstellung des Indexkonzepts kann zumindest sicherstellen, dass über das gleiche gesprochen wird. Andernfalls wird z.B. das operationalisierte Konzept „Glück“ durch den Rezipienten womöglich mit dem Begriff selbst assoziiert.

Die Verbindung der Indikatoren mit dem Indexkonzept folgt zwei verschiedenen theoretischen Überlegungen. Eine sieht Indikatoren als Resultat, die andere als Ausgangspunkt:

> „Der Komplex, den ein Index zu beschreiben versucht, kann entweder aus einer Vielzahl von Einzelheiten bestehen [...], oder aus vielen Dimensionen einer einzelnen Einheit.“ (Zeisel, 1970: S. 78)

In beiden beschriebenen Varianten steht der resultierende Index für ein nicht direkt beobachtbares Konstrukt. Lediglich die Wirkungsrichtung ist verschieden. Beim ersten Fall kann von *formativen Indikatoren* ausgegangen werden. Das bedeutet, dass das latente Konstrukt als eine gewichtete Zusammenfassung seiner Indikatoren betrachtet werden kann. Bei formativen Modellen zur Konstruktion von Indizes werden die Indikatoren kausal als Ursache des latenten Konstruktes betrachtet. Daraus folgt, dass ein Austauschen von Indikatoren zu Veränderung des Konstruktes führt.

Im zweiten Fall kann von *reflektiven Indikatoren* ausgegangen werden. Das bedeutet, dass das latente Konstrukt als Funktion seiner beobachteten Indikatoren modelliert wird. Dabei wird davon ausgegangen, dass sich das gesuchte latente Konstrukt in verschiedenen Indikatoren manifestiert. Über statistische Verfahren wird diese Manifestation zurückgerechnet auf das latente Konstrukt. Bei reflektiven Modellen wird das latente Konstrukt als Ursache für die beobachteten Indikatoren angesehen. Ein Verändern des Konstrukts führt folglich zu einer Veränderung in allen Indikatoren.[18]

Zusammenfassend wird festgestellt, dass Indizes nicht zweifelsfrei als Messung angesehen werden dürfen. Ihr Anspruch ist die Quantifizierung von theoretischen Konzepten. Die Aussagen auf Basis von Indizes beziehen sich dabei nicht immer nur auf den stilisierten Gegenstand des Index, sondern gehen z.T. über das theoretische Konzept hinaus. Die verschiedenen Entscheidungen, die bei der Konstruktion eines Index getroffen werden, sind z.T. willkürlich (z.B. Auswahl oder Zusammenfassung der Indikatoren), u.a. weil es an einer einheitlichen Systematisierung fehlt. Der Bezug zwischen Indexkonzept und Beobachtungswelt wird über eine Theorie hergestellt. Sie legt den Zusammenhang zwischen Indikatoren und Indexkonzept fest. Die Interpretation der resultierenden Indexwerte hängt damit sowohl von der Theorie als auch von den Indikatoren ab.

Eine umfassende Darstellung der Indexkonstruktion liefert Kapitel 4.

[18] Weiterführende Informationen zur Erfassung von latenten Konstrukten mit Hilfe von formativen und reflektiven Messmodellen bietet Christophersen und Grape (2006).

Kapitel 3

Evolution quantitativer Begriffe

≪Misura ciò che è misurabile,
e rendi misurabile ciò che non lo è.≫

Galileo Galilei

Das Messen ist aus unserer alltäglichen Welt nicht mehr wegzudenken. Neben rein naturwissenschaftlichen Phänomenen werden aktuell, sowohl intra- und interpsychologische als auch interindividuelle Prozesse (z.B. in Gesellschaft, Wirtschaft oder Soziologie) in Zahlen übersetzt. Bereits zur Zeit der frühen Hochkulturen hat der Mensch begonnen, die unübersichtliche Vielfalt der Dinge, Bewegungen und Ereignisse seiner Umwelt über Zahlen zu ordnen und so leichter verarbeitbar zu machen. Die Messung der Zeit strukturiert unseren Tagesablauf, die Messung der Geschwindigkeit bestimmt, wie schnell wir von A nach B kommen und die Messung des Gewichts, wie viel wir für unser Nahrungsmittel bezahlen müssen – um nur wenige Beispiele zur Anschauung herauszugreifen.

Keine Industrie kann im Wettbewerb des Weltmarktes ohne Messung bestehen. Laut Haustein (2001) und Randolph (1979) sind Maß und Gewicht dabei zwingende Voraussetzungen für die Steigerung der Wirtschaftsleistung und der Produktivität. Das geordnete Maß- und Gewichtswesen senkt die Komplexität der vielgestaltigen Wechselbeziehungen zwischen Völkern, Ländern und Staaten und bildet die Grundlage für „jede wirtschaftliche Transaktion, für wissenschaftliche Tätigkeiten, künstlerische Arbeiten, Verkehr, Rechtsprechung, Finanzwesen und Internet“ (Haustein, 2001: S. IX).

Mit der Zunahme der Komplexität des zu messenden Gegenstandes und dem Anspruch, neben Quantitäten auch Qualitäten zu messen, steigt gleichzeitig die Herausforderung an die Messapparatur. Der Anspruch von Galilei, alles zu messen, was messbar ist, und alles andere messbar zu machen, ist in Zeiten von Big Data aktueller denn je. Doch die Entwicklung der Messung reicht weitaus länger in die Vergangenheit zurück und Errungenschaften wie die Standardisierung von Maßen und Gewichten mussten hart erkämpft und diskutiert werden.

3.1 Ursprünge des Messens

3.1.1 Zahlen und Ziffern

In Abschnitt 2.2 wurde Böhmes Unterscheidungsversuch von erkenntnistheoretischen und wissenschaftstheoretischen Momenten im Prozess der Bildung quantitativer Begriffe ausführlich dargestellt. Das letzte Element des Prozesses bildet das numerische Relationalsystem mit seinen Zahlen. Die Ursprünge dafür gehen zurück bis zur Jungsteinzeit. Damals entstanden die ersten Zahlen aus dem Bedürfnis heraus, verschiedene Mengen von gleichartigen Dingen miteinander vergleichen zu können. Damit konnten z.B. Tiere, Früchte, Feuersteine etc. auf die Mitglieder einer Horde verteilt werden. (Trapp, 1998: S. 149)

Das Bedürfnis für das Zählen ist dem Menschen laut Stanislav Dehaene (1999) sogar angeboren. In seinem Buch *Der Zahlensinn oder warum wir rechnen können* beschreibt er die Existenz eines „Zahlensinns“. Dieser existiere unabhängig von erlerntem kulturellen Wissen. Dehaene zeigt über Grenzwertbetrachtungen, bei denen zwei marginal verschiedene Mengen miteinander verglichen werden sollen, dass der Unterschied auch ohne bewusstes Zählen auf ein Element genau festgestellt werden kann.[1] Ebenfalls belegt er empirisch, dass Menschen ohne bewusstes Zählen im Mittel die tatsächliche Menge treffen. Die Abweichung von der exakten Anzahl folgt einer Normalverteilung. Jedoch bleibt die Frage offen, ob der von Dehaene beschriebene Zahlensinn der Ausgangspunkt für die Entwicklung des abstrakten Zahlenverständnisses der Kulturen darstellte.

Erste Formen von Zahlzeichen wurden später in der Altsteinzeit über Kerben in Knochen oder Strichen in Felswänden nachgewiesen. Sie bilden damit die ältesten Spuren

[1] Die Untersuchungseinheiten bestanden aus Kindern und Tieren.

menschlicher Schrift. Danach belegen Dokumente aus den Hochkulturen an Nil, Euphrat und Tigris mit ihren Zahlensystemen den Beginn der Arithmetik. Beispielsweise finden sich auf der Keule des Königs Narmer (ca. 3000 v.Chr.) Zahlen wie 10.000, 100.000 usw., die auf der Addition aneinandergereihter Hieroglyphen beruhen. Mathematische Operationen wie Multiplikation, Division und Brüche waren dieser Hochkultur ebenfalls bekannt. (Ebbinghaus, 1992: S. 9)

Die Babylonier arbeiteten mit Keilsymbolen auf Tontäfelchen für die Darstellung von Zahlzeichen und trieben damit die Entwicklung zum aktuellen numerischen System maßgeblich voran. Mit ihren vielseitigen Rechenverfahren und der formalen Darstellung von Rechenvorschriften und Rechentafeln war ihr Einfluss auf Arithmetik und Algebra maßgeblich (Neugebauer, 1935).

Zwischen den Babyloniern und den Griechen setzte sich jedoch ein Unterschied im Zahlensystem durch. Die Babylonier arbeiteten mit einem Sexagesimal-System, die Griechen mit einem dekadischen. Der Unterschied besteht in der unterschiedlichen Basis von 6 bzw. 10. Heutzutage äußert sich dieser Unterschied z.B. innerhalb der Zeit (60 Minuten = 1h) und dem Gewicht (1000g = 1kg). (Trapp, 1998: S. 149ff.)

Die Pythagoreer gaben der Zahl dann philosophische Bedeutung. In ihrer Weltanschauung ist das Universum durch Zahlen und Verhältnisse charakterisiert, was den direkten Bezug vom System lebensweltlicher Phänomene auf das numerische System bedeutet. Euclides (3. Jh. v. Chr./2005) definiert die Zahlen in seinen Büchern *Die Elemente: Bücher I - XIII* von 325 v.Chr. als eine aus Einheiten zusammengesetzte Menge, wobei er die Einheit als dasjenige bezeichnet, wonach jedes Ding als Eines genannt wird (VII S. 1, 2). Die Einheit besteht damit nicht aus Einheiten. Sie bildet die Grundlage des Zählens und den Ursprung der Zahl. Aristoteles sieht das ähnlich. Für ihn ist das, was in diskrete Teile zerlegbar ist, Vielheit und die begrenzte Vielheit ist für ihn die Zahl. In diesem Zusammenhang wird allerdings deutlich, dass es für die Griechen nur die natürlichen Zahlen gab (Ebbinghaus, 1992: S. 11f.).

Dennoch waren in der Antike neben den natürlichen Zahlen auch rationale (Brüche, Verhältnisse) und gewisse irrationale Zahlen wie die Kreiszahl und Quadratwurzel bekannt. Die letzten beiden stellten jedoch eine eigenständige Lehre der kommensurablen und inkommensurablen Proportionen und eine Erweiterung der natürlichen Zahlen dar (Wussing, Alten und Wesemüller-Kock, 2008). Proportionen wurden auf diese

Weise bis ins 17. Jhd. als Zahlen behandelt. Später schränkte Isaac Newton (1643-1727) die Definition von Zahlen ein. Er sah die Zahl als abstraktes Verhältnis einer Quantität zu einer anderen derselben Art (vgl. Hess (2010: S. 14)). Letzteres fungiert dabei als Einheit.

Campbell (1928) führte den Begriff des *Fundamental Measurements* ein. Hierbei repräsentieren die numerischen Beziehungen zwischen gewissen Zahlen bestimmte natürliche Beziehungen von Dingen und Größen. Scott und Suppes sehen das ähnlich. Bei ihnen ist einem empirischen relationalen System ein numerisches System mit Zahlen gegenübergestellt. Beide Systeme beinhalten eine Menge von empirischen bzw. numerischen Objekten sowie Relationen auf diese Mengen. Damit wird eine Abbildung von Objekten und Relationen des empirischen Systems auf das numerische System strukturerhaltend ermöglicht (Homomorphismus) (Krantz, Luce und Suppes, 1989).

Eine wegweisende Schrift zu den Zahlen wurde 1912 von Dedekind veröffentlicht. In *Stetigkeit und irrationale Zahlen* heißt es:

> "Die bisher übliche Einführung der irrationalen Zahlen knüpft nämlich geradezu an den Begriff der extensiven Größen an – welcher aber selbst nirgends streng definirt wird – und erklärt die Zahl als das Resultat der Messung einer solchen Größe durch eine zweite gleichartige [eine Einheit]". (S. 11)

Die Kritik Dedekinds ist im historischen Kontext zu betrachten. Damals war es üblich, reelle Zahlen per se als Maßzahlen zu betrachten. Zahl und Messung standen zu dieser Zeit damit in engem Zusammenhang.

Fortschritte in der Definition der Zahlen wurden im 19. Jhd. von der Mathematik forciert. Über die steigende Abstraktion und die Betrachtung der Mathematik als System ohne externen Bezug wurden Begriffe eingeführt, die rein „fiktiven" Charakter hatten. In diesem Zusammenhang sind die imaginären Zahlen zu nennen. Ohne systematische Herleitung wurden diese bis weit ins 19. Jhd. hinein verwendet. (Heintz, 2000: S. 263)

3.1.2 Die Ursprünge der Größenbegriffe

Die Ursprünge des numerischen Systems gehen, wie oben bereits angemerkt, zurück bis zur Neu- und Altsteinzeit und sind eng verbunden mit der Zählung von Dingen. Mit

dem Aufkommen der Kultur entwickelte sich parallel zum System der Zahlen das der Größe und Einheit und damit auch das System des Messens.

Messen hat als Wesensbestandteil von Kultur bereits innerhalb der Urvölker und den primitiven Naturvölkern der Erde Niederschlag gefunden. Den Mittelpunkt bildeten die drei Größen *Länge*, *Zeit* und *Gewicht*. Besonders Tauschhandel und Wasserbau förderte dabei den Fortschritt (Alberti, 1957: S. 1). Probleme bereiteten anfangs neben der Genauigkeit besonders die Wiederholbarkeit und Unabhängigkeit der Messung. So konnte der Fall eintreten, dass das Gewicht eines Lebensmittels in verschiedenen Dörfern unterschiedlich war oder sich von einem auf den anderen Tag änderte.

Für die Bestimmung der *Zeit* nutzte der Mensch anfangs beispielsweise natürliche Einheiten wie die Dauer, die der Mond für einen Umlauf um die Erde benötigt oder die Dauer zwischen Saat und Ernte. Es handelte sich damit um nicht fest definierte Einheiten, da sie veränderlich waren, d.h. von variablen Faktoren abhingen. (ebd.)

Die Messung von *Länge* und *Gewicht* unterlag anfangs ebenfalls einer solchen Abhängigkeit und den damit verbundenen Problemen der Reproduzierbarkeit. Häufig wurden Einheiten benutzt, die vom Körper oder der eigenen Leistungsfähigkeit abgeleitet wurden.

Messungen waren damit ungenau und nur bedingt reproduzierbar. Einheiten wie z.B. Elle oder Fuß bleiben zwar bei ein und derselben Person weitestgehend gleich, jedoch treten spätestens beim Vergleich mit Personen abweichender Fuß- und Ellengröße Probleme auf. Harald Witthöft (1998) übersetzt dazu Agricola:

> "Mehrere Völker haben zwar derartige Maße zuerst nach Breite und Länge der menschlichen Gliedmaße festgelegt und sie mit denselben Wörtern wie die Gliedmaßen selbst bezeichnet; sie sind jedoch unterschiedlich, weil die Glieder an einem Menschen selten die gleichen wie an einem anderen sind.
>
> So war das olympische Stadion länger als die andren Stadien Griechenlands, obwohl jedes einzelne 600 Fuß maß. Doch jenes hat Herakles mit seinen Füßen gemessen gehabt, diese andre.
>
> Es ist eine bekannte Tatsache, daß die Spanne von der Länge eines jeden von der Natur wohl– gebildeten menschlichen Körpers 1/24 beträgt, der Fuß 1/6, die Elle 1/4. Pythagoras kannte diese Verhältnisse der Glieder

> zueinander; er schloß daher von der Länge des Fußes, die er im Stadion von Olympia beobachtete, auf die Gestalt des Herakles.“ (S. 11)

Bei Menschen, die nicht dieser Körperdefinition entsprechen, ist dieses Verhältnis jedoch anders und das Ableiten der Größe fehlerbehaftet. Die Probleme der Relativität waren auch den „Alten“ bekannt. Die Einheit wurde deswegen teilweise auf externe Gegenstände wie Korb oder Kanne ausgedehnt, was in gewisser Weise einen Versuch der Standardisierung der Einheit darstellt. (Alberti, 1957: S. 7)

Jedoch waren diese Versuche noch weit von einem Standard- oder Einheitsmaß entfernt. Dieses ist notwendig, um die Messung reproduzierbar und unabhängig von Ort, Zeit und Person zu machen. Aussagen wie z.B. *3 Wegstunden entfernt* konnten auch bei Bezug zu derselben Person nie eine genaue Zeit abbilden, da die Vergleichsgröße nicht feststand.

Den Versuch einer Standardisierung der Größe findet Krickeberg (1914: S. 724) bei den nordwestkalifornischen Hupa. Hier wurde das individuelle Körpermaß mit Hilfe von Dentalien bekannter Länge „normiert“. Dafür wurden Striche auf die Innenfläche des linken Vorderarmes tätowiert, mit denen dann Längenmessungen durchgeführt werden konnten. Es kann dabei von einer eingeschränkten Reproduzierbarkeit und Unabhängigkeit von Ort, Zeit und Person ausgegangen werden.

Ein anderes Beispiel für den Versuch einer Standardisierung kam von den Ureinwohnern Brasiliens. Diese maßen die Entfernung über die Anzahl von Bächen, die zwischen Anfangs- und Endpunkt lagen (Alberti, 1957: S. 10).

Trotz der eklatanten Schwäche der o.g. sehr frühen Ansätze einer Standardisierung sind sie als Fortschritt zu werten. Sie machten die Abbildung vom empirischen Raum auf den numerischen Raum „objektiver“. Außerdem bleibt anzumerken, dass selbst zur aktuellen Zeit noch mit diversen „Verlegensheitsmaßen“ gemessen wird. Die Strecke von Daumen zu Zeigefinger wird für kurze Längenmessungen benutzt oder unsere Schrittlänge für längere Strecken. Die Einheiten Fuß oder Faustgroß sind dabei jedoch als Schätzung zu bezeichnen. (ebd.: S. 7)

Zwischen den frühen Naturvölkern und dem Abendland wurde die Weiterentwicklung der Maße durch die Mystik verhindert oder zumindest verzögert (ebd.: S. 10). Schnee berichtet in diesem Zusammenhang z.B. von Eingeborenen Neuguineas, die wegen ihres

Glaubens Entfernungen falsch einschätzten. Sie waren der Überzeugung, dass die ausfahrenden Schiffe von Dämonen festgehalten würden, die auf bestimmten Küstenfelsen lebten. Diese Felsen lagen höher als der Rest der Küste, weswegen sie bei der Ausfahrt länger am Horizont gesehen wurden. Daraus ergab sich für die Eingeborenen die Täuschung, dass sie von den Felsen festgehalten würden. Flüstern bei der Ausfahrt sollte verhindern, dass die Dämonen aufgeweckt werden.

Als ein weiteres Beispiel beschreibt Seligmann (1910) einen Ritus im Pandschab. Dort wurden Pferde von den Ohren bis zur Schwanzspitze mit einem Baumwollfaden vermessen, um festzustellen, ob das Pferd verhext ist. Die Wollschnur wurde abgeschnitten, um einen Käseteig gewickelt und ins Feuer geworfen. Brannte die Schnur, bevor der Teig gekocht war, war das Pferd verhext. (S. 266)

Die mystische Betrachtung änderte sich mit einer umfassenderen Beobachtung der Umwelt. Alberti (1957) konstatiert, dass bereits in den alten Kulturvölkern von feinen ausdifferenzierten Messmethoden ausgegangen werden kann.

Ein entscheidender Schritt für die Standardisierung der Messung lässt sich auf das Zweistromland zwischen Euphrat und Tigris zurückführen. Die dortigen Kulturen waren von Wasserbau geprägt und ein ausgeklügeltes Messsystem war für viele Prozesse Voraussetzung für die dort ansässigen Sumerer und Chaldäer. So ist es nicht verwunderlich, dass die Fortschritte speziell dort erzielt wurden. Einer davon hat bis heute maßgebliche Bedeutung: Die Einteilung des Tages in 24 Stunden und der Stunde in 60 Minuten.

Vom Zweistromland breiteten sich die Erfahrungen über Babylon nach Ägypten, Phönizien und Griechenland aus. Ohne Erkenntnisse zum standardisierten Messsystem wären wohl die hervorragenden Ingenieurbauten in z.B. Ägypten (Pyramide des Königs Cheops um 2570 v.Chr.) oder den Reichen der Inkas und Azteken nicht möglich gewesen. (ebd.)

3.2 Von Größenbegriffen zu Indikatoren und Indizes

Mit dem Zuwachs der Fragestellungen in Wissenschaft und Gesellschaft steigt auch die Anforderung an die quantitativen Begriffe. Wie Burzan (2014: S. 1029) ausführt, sind Forschungsfragen und Hypothesen nicht selten innerhalb der quantitativen empirischen Forschung zu allgemein und mehrdeutig formuliert, um sie direkt über Größen messen zu

können. Bei der Operationalisierung der komplexen Sachverhalte spielen Indikatoren als *Anzeiger* über eine indirekte Messung eine entscheidende Rolle (vgl. Abschnitt 2.3.1). Die erste theoretische Auseinandersetzung zur Indikatorbildung und -verwendung zeigt sich in den 1920ern und 1930ern im Rahmen der gesamtwirtschaftlichen Beobachtungen und Planungen (Randolph, 1979: S. 32). Oskar Morgenstern (1956) fasst die Gründe für die neuen Anforderungen an die Messapparatur für die Wirtschaftswissenschaft wie folgt zusammen:

> "Je weiter [...] eine Wissenschaft fortschreitet, desto weniger offensichtlich und desto weniger leicht unmittelbar zugänglich werden die sich in den einzelnen Teilkomponenten auszudrückenden Erscheinungen und Daten, die von Bedeutung sind. Tatsächlich muß man sich selbst in der Wirtschaftswissenschaft in zunehmenden Maße mit Beziehungen und Begriffen befassen, die jenseits der Vorstellungswelt des durchschnittlichen Laien liegen." (S. 229)

Die neuen Forschungsfragen führen dazu, dass zu einem gegebenen theoretischen Konstrukt kein unmittelbares empirisches Korrelat zugeordnet werden kann. Die verwendeten Terme sind nicht mehr der Beobachtungs- und Messsprache zuzuordnen, sondern gehören ausschließlich zum Vokabular einer Theorie. Indikatoren übernehmen an dieser Stelle die Operationalisierung der theoretischen Konstrukte, d.h. die Überführung in messbare Begriffe. (Randolph, 1979: S. 26ff.)

Wie in Abschnitt 3.2 erwähnt, wurde die Frage der Indikatoren anfangs insbesondere von der Ökonomie aufgeworfen und behandelt. Die gesamtwirtschaftlichen Indikatoren entsprangen vorrangig der Wirtschaftsstatistik und Konjunkturforschung. Beispielhaft sei die 1938 vom *National Bureau of Economic Research* der Vereinigten Staaten entwickelte „Indikatorbatterie" zur Beschreibung und Analyse des komplexen Phänomens „economic process" genannt. Insgesamt wurden 21 Indikatoren ausgewählt, die die konjunkturelle Entwicklung hinreichend beschreiben sollten. Danach erhielten Indikatoren Einzug in betriebswirtschaftliche Betrachtungen und Entscheidungen, z.B. in Form von Input- und Output-Indikatoren, die den Betriebszustand sowie das Betriebsgeschehen in zusammengefasster Form beschreiben und messen sollten. (ebd.: S. 32ff.)

Soziale und politische Indikatoren forcierten die Entwicklung weiter:

> "Wegen der angewachsenen Steigerung von Kompliziertheit und Komplexität soll die Steuerungskapazität des politischen Systems [...] angepaßt werden, d.h. erweitert werden. Inhalt eines großen Teils der Ansätze der Sozialindikatorenforschung ist es, Werte und Ziele einer Gesellschaft abzubilden, um die soziale Entwicklung zu diagnostizieren [und zu] prognostizieren [...]." (Randolph, 1979: S. 36-37)

Einen wesentlichen Beitrag zur Indikatorbewegung leistete Raymond Bauer (1966) mit *Social indicators*. Dieses Werk steht in Beziehung zu dem intensiven Verteidigungs- und Raumfahrtprogramm der USA und dem daraus resultierenden großen Bedarf an Personal. Die Auswirkungen sowie die ökonomischen und sozialen Voraussetzungen der erhöhten Nachfrage wurden 1962 über die Gründung des *Council on Space Efforts and Society* untersucht und ein Jahr später vom *Commitee in Space* der *American Academy of Arts and Science* weitergeführt. Gemeinsam mit dem *MIT* wurde die Reihe *Technology, Space and Society* herausgegeben, in der auch die eben genannte Pionierarbeit von Bauer erschien. (Randolph, 1979: S. 37)

In anderen Ländern wurden später von verschiedenen Institutionen Indikatoren, z.B. zur Messung der Lebensqualität, entwickelt. Die Ansätze waren jedoch sehr unterschiedlich. D.h., die Zahl und inhaltliche Ausrichtung der Indikatoren sowie ihre Gewichtung variierte.

Der Versuch, die Gesellschaft zu vermessen, hat zu einer umfangreichen Fachliteratur geführt. Die verschiedenen Perspektiven und Überschneidungen verhinderten allerdings einen tragfähigen Konsens in der Sozialindikatorenforschung (vgl. dazu ebd.: S. 37ff.). So ist es nicht verwunderlich, dass zahlreiche kritische Meinungen zu dem „Quantifizierungswahn" existieren:

> "Eine Ursache für das Faszinosum und den Mißbrauch der Quantität ist die Tatsache, daß die Gesellschaft immer komplexer, quantitativ vielfältiger wird und der zahlenmäßige Ausdruck als ein Rettungsanker für ihre Durchschaubarkeit erscheint [...] Überall, wo Einsicht in eine differenzierte, widersprüchliche Situation und ihre Ursachen fehlt, da stellen sich Prozentsätze oder Durchschnitte ein, die Wahrheitsgehalte vorgaukeln". (Haustein, 2001: S. 6)

Die Quantifizierung von Konzepten und komplexen Begriffen birgt die Gefahr von Subjektivität, z.B. durch spezifische Interessen. Dies wird besonders deutlich bei sogenannten normativen Indikatoren. Diese enthalten Interessen, die in der Regel auf Regierungsebene deklariert sind. Die Indikatoren spiegeln damit die Interessen und Normen der Regierung wider. Es kann nicht von Objektivität im wissenschaftlichen Sinne ausgegangen werden, zumal diese Art der Indikatoren anfällig für Manipulationen ist. (Randolph, 1979: S. 62)

Mit der Bildung von Indikatoren wurde dem Bedürfnis verschiedener Wissenschaftsdisziplinen nachgegangen, Messungen oder Quantifizierungen durchzuführen, wo naturwissenschaftliche Methoden an ihre Grenzen stoßen:

> "Die unbestreitbaren Erfolge unserer klassischen Wissenschaftstradition beruhen darauf, daß die theoretische Betrachtungsweise sich auf solche Strukturen und Systeme beschränkt hat, die von minimaler logischer Komplexität waren und deren strukturelle Komplikation radikal reduzierbar war. Das derart gesammelte Wissen genügt aber unseren heutigen Bedürfnissen nicht mehr im geringsten. [...] Was dringend benötigt wird, sind wissenschaftliche Methoden mit denen das Verhalten von Systemen von hoher logischer Komplexität und entsprechender struktureller Komplikation studiert werden kann." (Günther, 1968: S. 337)

Vermutlich mündet das von Haustein im Zitat oben angesprochene Problem des „Mißbrauchs der Quanität" auch im Siegeszug der Naturwissenschaften und dem Aufkommen der Frage nach der Integrierbarkeit ihrer erfolgreichen Methoden wie Messen, Experimentieren und mathematischen Behandlungen in andere Disziplinen. Hauptproblem bei der Integrierbarkeit ist die höhere Komplexität der Forschungsfragen.

Indikatoren und Indizes vermögen es, nicht beobachtbare bzw. komplexe und mehrdimensionale Konzepte zu quantifizieren, ohne den naturwissenschaftlichen Anforderungen einer Messung nachkommen zu müssen. Wissenschaftliche Betrachtungen zu Indizes entwickelten sich ungefähr zeitgleich mit denen von Indikatoren. Das Ziel von Indizes ist in der Regel die Reduktion von Komplexität, indem mehrere Dimensionen zu einer einzigen Maßzahl zusammengefasst werden. In diesem Zusammenhang können Indizes sicherlich

als logische Konsequenz von Indikatoren angesehen werden. Indikatoren sind Ergebnis einer *indirekten Messung* und i.d.R. eindimensional (vgl. Abschnitt 2.3.1).

In der Literatur werden die Begriffe Indikatoren und Indizes z.T. auch synonym verwendet, in der vorliegenden Arbeit wird sich davon jedoch distanziert und an Indikatoren die Bedingung von Eindimensionalität geknüpft sowie die Operationalisierbarkeit über lediglich eine Variable oder ein Item. Bezogen auf den in Kapitel 2 dargestellten Messprozess, wird hierbei eine quantitative Struktur unter einmaliger Metrisierung in den numerischen Raum überführt. Sollen mehrdimensionale Konzepte operationalisiert werden, können verschiedene Indikatoren Verwendung finden, wobei der Index mindestens zwei Indikatoren unter einem Modell zusammenfasst. Die Verbindung der Indikatoren mit dem Indexkonzept basiert damit auf einer Theorie.

Bei der Konstruktion von Indizes für die Quantifizierung eines Konzepts, z.B. „Glück“, wird allgemein von der Reduktion von Komplexität gesprochen. Damit kann der Index als Versuch angesehen werden, den von Günther im Zitat oben angesprochenen Bedürfnissen zu genügen. Das Konzept des Index beschreibt i.d.R. einen mehrdimensionalen Merkmalsraum anhand eines meist eindimensionalen Maßes. Außerdem ermöglicht es den Vergleich von komplexen Sachverhalten z.B. in Form von Rangfolgen und trägt damit zur Reduktion der Informationsfülle bei. Wie oben angedeutet, enthält das Konzept des Index jedoch z.T. Subjektivität. So wird die Objektivität des Index zum einen davon beeinflusst, wie die jeweiligen Merkmale des Index in den Zahlenraum überführt werden, und zum anderen, wie diese unter dem Indexmodell selektiert und konkateniert werden. (vgl. Kapitel 4)

Eine zentrale Eigenschaft des Indexes ist damit die Informationsverdichtung und einfache Abbildung komplexer Sachverhalte. Flaskämper (1928: S. VI) spricht bei Indizes von einem „Spezialproblem der Logik des statistischen Vergleichs“. So können komplexe Sachverhalte wie die Lebenserhaltungskosten über den Verbraucherpreisindex zeitlich oder regional über eine einzige Maßzahl miteinander verglichen werden. Analog kann die „Intelligenz“ von Personen über den Intelligenzquotienten verglichen werden. (Zeisel, 1970: S. 82)

Bereits vor dem 1. Weltkrieg existierten Indizes zu Phänomenen des sozialen Lebens, die z.B. das Preisniveau in ihrer zeitlichen Entwicklung abbildeten (Flaskämper, 1928:

S. I). Danach verbreiteten sich Indizes in den 1920ern in der Ökonomie im Zusammenhang mit der Inflation und den konjunkturellen Bewegungen. Es wurden zahlreiche Konstruktionsmethoden bezüglich der Wahl des Basisjahres, der Gewichtung und der Art der Zusammenfassung (additiv, multiplikativ) erprobt (Werner, 1975a: S. 152).

Die steigende Verwendung von Indizes zu Zeiten der Inflation führte zu einer umfangreichen Literatur zu Technik und Anwendung. Eine umfassende theoretische Leistung lieferte in diesem Zusammenhang 1922 Irving Fisher mit dem Werk *The making of index numbers: a study of their varieties, tests, and reliability.* Sein Ziel war die Darstellung von allen möglichen Indexformeln. Ausgangspunkt sind sechs Grundtypen, die auf verschieden definierten Komponenten und unterschiedlichen Summierungsvorschriften beruhen. Mit zahlreichen Möglichkeiten der Wahl des Basisjahres, der Gewichtung, der Verkettung und des „Kreuzens“ von Indizes erarbeitet er eine Zusammenstellung von 134 Indexformeln. Dazu entwickelt er Teststatistiken, mit denen die Angemessenheit einer Formel zu einem gegebenen Material geprüft werden kann. Flaskämper (1928) äußerte sich kritisch gegenüber dieser Darstellung. Er wirft Fisher Formalismus vor und bezweifelt, dass das Indexproblem mit der Entwicklung von Formeln gelöst werden könne. Seiner Ansicht nach geht es eher um das Begreifen der logischen Grundstruktur eines Index. Außerdem sei es sinnvoller, die inhaltlichen Voraussetzungen des Materials zu prüfen, statt zahlreiche Tests anzuwenden. Es geht also um die Frage, bis zu welchem Maß die Konstruktion von Indizes auf einer rein formalen und syntaktischen Ebene vorgenommen werden kann.

Auch die Definition von Indizes, Indexzahlen oder Indexziffern wurde zu dieser Zeit kontrovers diskutiert. Nach Flaskämper (ebd.) sollen Indexzahlen gleichartige und einander koordinierte statistische Größen miteinander vergleichen. Sie seien logische Instrumente des statistischen Vergleichs und ähneln den Verhältniszahlen. Als Beispiel nennt er:

> “Wenn wir z. B. von einer Großhandelsindexziffer oder von einer Aktienindexziffer sprechen, so meinen wir damit eine Zahl, die angibt, wie hoch das Niveau der Großhandelspreise oder der Aktienkurse zu einem bestimmten Zeitpunkte ist, verglichen mit dem entsprechenden Niveau zu einer anderen, als Ausgang gewählten Zeit, wobei der Wert der letzteren gleich 1 oder gleich

> 100 gesetzt wird. Oder wenn bei Reallohnvergleichen der Reallohn von London gleich 100 gesetzt wird und darauf die Reallöhne anderer Städte bezogen werden, so sprechen wir auch hier von Indexziffer.“ (S. 1)

Dabei wird deutlich, dass die Vorstellung von Flaskämper ganz der Tradition der Preisindizes nach Paasche und Laspeyres folgend, von einer Normierung über ein Basisjahr geprägt ist.[2] Auch können für Flaskämper einfache Erscheinungen, wie der Preis einer einzelnen Ware, im Zeitvergleich als Index betrachtet werden. Den Unterschied zwischen einfachen und komplexen Phänomenen macht er über die Begrifflichkeiten Einzel- und Generalindex kenntlich.

Weigel (1921b) definiert Indexzahlen breiter. Für ihn müssen Indexzahlen (1) absolute Zahlen sein, die (2) durch „Reihenverschmelzung“ zustande gekommen sind (S. 459).

> “Indexziffern werden in der Statistik berechnet, um prägnante, schlagwortartige Ausdrücke für Erscheinungen zu gewinnen, die zwar äußerlich – etwa ihrer Benennung nach – einheitliche Tatsachen sind, in Wirklichkeit aber die Summe einer mehr oder weniger großen Zahl von Einzelerscheinungen darstellen, die sämtlich berücksichtigt werden müssen, wenn man zu einem zusammenfassenden Urteil über die Gesamterscheinung kommen will, die aber ohne weiteres eine solche zusammenfassende Beurteilung entweder überhaupt nicht oder wenigstens sehr schwer gestatten.“ (Weigel, 1921a: S. 135)

Weigel (ebd.) bemängelt, dass trotz der regen Nutzung von Indizes die theoretische Auseinandersetzung mangelhaft sei (S. 128). Eine Kritik, die auch in der aktuellen Literatur Bestand hat. Weitere Beiträge zu Indizes in dieser Zeit liefern u.a. Bortkiewicz, Gini, Haberler, Hermberg und Walsh.[3] Die theoretische Auseinandersetzung und Anwendung bezog sich dabei nach wie vor hauptsächlich auf den ökonomischen Bereich. Jedoch lassen sich die Fragestellungen auch auf andere Forschungsdisziplinen übertragen. Nach der Inflation lagen die entscheidenden Treiber für Indizes im Zusammenhang zur

[2] Für weiterführende Informationen zum Preisindex nach Paasche und Laspeyres sei auf Fahrmeir (1997: S. 551f.) verwiesen.

[3] Vgl. Bortkiewicz (1923, 1924), Haberler (1927), Hermberg (1930) und Walsh (1921). Für weitere Autoren wird auf die bedeutende Diskussion von Frisch (1936) verwiesen. Er erwähnt Allen, Bortkiewicz, Bowley, Gini, Haberler, Keynes, Konüs, Pigou sowie Staehle und diskutiert sie ausführlich.

modernen Konjunkturstatistik in den Gebieten des sozialen und wirtschaftlichen Lebens (Flaskämper, 1928: S. I).

Danach erweiterten insbesondere die ökonomischen, politischen und sozialen Indizes die formale Darstellung und theoretische Auseinandersetzung (vgl. dazu Nardo u. a. (2005, 2008), Pickel (2013), Pickel und Pickel (2012), Ralph (2015), Werner (1975a) und Werner (1975b)). Herausgehoben sei hierbei die umfassende Darstellung zur Konstruktion von Indizes von Nardo u. a. (2008), die von der OECD publiziert wurde. In den allgemeinen Lehrbüchern der Statistik werden Indizes verhältnismäßig knapp behandelt (vgl. dazu Anderson (1954), Bourier (2013), Fahrmeir (1997), Holland und Scharnbacher (2006), Kellerer (1960), Pfanzagl (1966), Schnell (2005) und Wagenführ (1967)).

Die aktuelle Verbreitung von Indizes wird ausführlich in Kapitel 5 behandelt, ihre Konstruktion in Kapitel 4.

3.3 Zahlen in der Kommunikation

Seit dem Aufkommen der Naturwissenschaften und der wissenschaftstheoretischen Auseinandersetzung mit dem Messprozess erhalten Zahlen[4] innerhalb der Kommunikation eine steigende Bedeutung (vgl. Heintz (2007)). Wegen ihrer spezifischen Eigenschaften behandelt Theodore Porter (1995) sie in seinem Buch *Trust in numbers* sogar als Kommunikationsmedium:

> "Numbers, graphs, and formulas [are] first of all strategies of communication. [...] Reliance on numbers and quantitative manipulation minimizes the need for intimate knowledge and personal trust." (S. viii-ix)

Die Objektivität von Zahlen erhöhe die Akzeptanz für Kommunikation und komme immer dann zum Einsatz, wenn persönliches Vertrauen fehle oder wenn ein Konsens nicht mehr über persönliche Interaktion hergestellt werden könne (vgl. Porter (ebd.: S. 213ff.),Porter (1992: S. 643)).

Bettina Heintz (2007) betrachtet die Thesen von Porter in Bezug auf die Systemtheorie von Luhmann und modifiziert Porters Thesen dahingehend, dass Zahlen –

[4] Gemeint sind Zahlen, die durch Messen einer Größe entstehen.

ähnlich wie nach Luhmann z.B. Liebe oder Geld – als symbolisch generalisiertes Kommunikationsmedium betrachtet werden können (S. 65). Unter symbolisch generalisierten Kommunikationsmedien sind jene Kommunikationsanweisungen zu verstehen, die es unter einer an sich unwahrscheinlichen Kommunikation ermöglichen, Akzeptanz herzustellen. Damit helfen sie bei der Lösung des Problems der dreifachen Unwahrscheinlichkeit: Verstehen, Erreichen von Empfänger und Kommunikationserfolg (Luhmann, 1981).

Die Sprache wird dabei als das primäre Medium betrachtet. Ihre Entwicklung erst ermöglicht Kommunikation, die an das Verstehen gekoppelt ist. Auf dem Medium Sprache basieren die weiteren Verbreitungsmedien, die den Kreis der Empfänger erweitern. Beispielhaft seien hier als Verbreitungsmedien Bücher und Zeitungen genannt. Aus dem Dazwischenschalten eines vermittelnden Mediums resultiert ein Versatz von Mitteilung und Verstehen (im Vergleich z.B. zu mündlicher Interaktion[5]), was weitere Verfahren erfordert, um die Annahme der Kommunikation zu sichern. An diese Stelle treten nun nach Luhmann die symbolisch generalisierten Kommunikationsmedien. Sie erhöhen die Akzeptanz von Kommunikation, indem sie signalisieren, dass ihre Selektion bestimmten Voraussetzungen folgt (Luhmann, 1997: S. 321). Sie definieren sich daher nicht als Sachverhalte, „sondern [als] 'semantische Einrichtungen', die auf Sachverhalte hinweisen, und ihnen Kausalität zuschreiben“ (Heintz, 2007: S. 366). Das Kommunikationsmedium *Liebe* ist somit z.B. eine Kommunikationsanweisung, mit der Liebe ausgedrückt werden kann. Es handelt sich dabei nicht um das Gefühl an sich (ebd.: S. 366). Genauso ist das Kommunikationsmedium *Wahrheit* zu begreifen. Es signalisiert, dass die Selektion einer Information unter bestimmten (anerkannten) Voraussetzungen durchgeführt wurde und nicht auf bloßer Willkür basiert. *Wahrheit* ist damit in diesem Zusammenhang keine Eigenschaft von Sätzen.

Inwieweit Zahlen auf Selektionen basieren und was sie darüber hinaus als symbolisch generalisiertes Kommunikationsmedium auszeichnet, wird in den folgenden Unterabschnitten zusammengefasst (vgl. Heintz (ebd.)). Folgende Eigenschaften einer auf Messung basierenden Zahl bilden dabei die Grundlage der Überlegung:

- Sie basiert auf einer Selektion, die bestimmte Voraussetzung erfüllt.

[5] Mündliche Interaktion erhöht die Bereitschaft, Mitteilungen zu akzeptieren und Konsens zu signalisieren (vgl. dazu Goffman (1982)).

- Sie weckt den Anschein von Unpersönlichkeit, d.h. einer interpretationsfreien Beschreibung.
- Sie weist nicht über das Beschriebene hinaus.
- Sie enthält Informationen in hochverdichteter Form.
- Sie ist tendenziell affirmativ (beinhaltet im Gegensatz zu einem Satz nicht automatisch die Möglichkeit einer Negation).

3.3.1 Kommunikation von objektivem Wissen

3.3.1.1 Der Wandel der Objektivität

Analog zur geschichtlichen Entwicklung der Messung wandelte sich auch die Kommunikation von objektivem Wissen. Entscheidende Veränderungen zeichneten sich in der wissenschaftlichen Kommunikation im Verlauf des 19. Jhd. ab, z.B. in Form von Journalaufsätzen.

Wissenschaftshistorisch änderte sich vor allem die Art und Weise, wie objektives Wissen garantiert wurde. Das heutige Verständnis von objektiv als unabhängig von persönlichen Einflüssen und das Gleichsetzen von Objektivität und Wahrheit ist relativ spät im 19. Jhd. entstanden. Die Ursprünge dafür liegen im 16. Jhd. und dem Aufkommen der modernen Wissenschaften als eigenständiges Funktionssystem. Als grundlegendes Erkenntnisinstrument galten seither Experiment und Empirie. Bücher und Autoritäten wurden durch eigene Beobachtungen ersetzt. Nur empirisch Erfahrbares und intersubjektiv Überprüfbares konnte sich als wissenschaftliche Tatsache etablieren. (Heintz, 2007: S. 67f.)

Das Problem in diesem Zusammenhang lag in der Abhängigkeit von Raum, Zeit und Beobachter. Beobachtungen ließen sich nicht einfach auf andere Orte und andere Zeiten übertragen. Auch konnten sie zwischen verschiedenen Beobachtern variieren. Die Lösung bestand in der Eliminierung der Subjektivität und Körperlichkeit. Die Natur sollte quasi nicht mehr durch Personen sprechen, sondern direkt für sich selbst. (ebd.: S. 67f.) Dieser Wunsch ist der Ausgangspunkt für den Siegeszug der „mechanical objectivity“, wie sie von Daston und Galison (1992) bezeichnet wurde (S. 82).

Über die Verbreitung von Beobachtungs- und Messgeräten und strengen methodischen Regeln sollte eine Unabhängigkeit von menschlichen Sinnen und ihrer Fehleranfälligkeit erreicht werden. Die Körperwahrnehmung wurde durch Apparate ersetzt, durch die die Natur frei von jeglicher Subjektivität für sich selbst sprechen konnte. (Heintz, 2007: S. 68)

Im Auftrag der *Royal Society* führte beispielsweise Robert Hooke regelmäßig Wetterbeobachtungen durch und entwickelte dafür spezielle meteorologische Messgeräte. Diese sollten als „künstliche Organe“ die Fehleranfälligkeit der Sinneswahrnehmung ausgleichen. Später entstand aus dem Projekt der Vorläufer einer automatischen Wetterstation. (Schramm, Schwarte und Lazardzig, 2006: S. 70f.)

3.3.1.2 Die Ausdifferenzierung wissenschaftlicher Kommunikation

Die Objektivität der Messung hing nicht nur von dem Beobachtungs- oder Messinstrument ab. Für die Überführung in eine wissenschaftliche Tatsache musste sie vielmehr von der unabhängigen Fachgemeinschaft anerkannt und akzeptiert werden. Die Akzeptanz einer wissenschaftlichen Leistung unterlag in der Geschichte ebenfalls einem Wandel. In der Frühphase wurden wissenschaftliche Tatsachen durch Anwesenheit von sozial anerkannten Zeugen geschaffen. Experimentell erzeugte Fakten und Beobachtungen erhielten Glaubwürdigkeit und Autorität durch die Bestätigung von namentlich genannten Zeugen, die die Ergebnisse mit ihrer Unterschrift bestätigten (Wöllenstein, 2008: S. 8ff.). Auch in der bereits angesprochenen 1660 gegründeten *Royal Society* wurde Objektivität anfangs über diese Mechanismen der Zeugenschaft hergestellt. Bei den Zusammenkünften der Gesellschaft wurden Experimente vorgeführt und diskutiert. Zusätzlich konnten Beobachtungen reproduziert werden, wenn Beobachter der Gesellschaft ihre Objekte übersandten. (vgl. Gläser (2006) und Schramm, Schwarte und Lazardzig (2006))

Mehrmals wurde jedoch die wissenschaftliche Professionalität der Gesellschaft z.B. durch die Figur des adligen *Virtuoso* und privaten Finanziers in Frage gestellt (Berndt und Fulda, 2012: S. 199ff.) Die Tradition der Zeugenschaft wandelte sich im 18. und 19. Jhd.. An die Stelle der Zeugen traten wissenschaftliche Qualifikationen und das Befolgen bestimmter anerkannter Forschungstechniken. In dieser Phase änderte sich auch die Präsentationsform und Verbreitung der Ergebnisse. (Gläser, 2006: S. 232ff.)

Bis Ende des 18. Jhd. war die wissenschaftliche Kommunikation häufig formal geprägt: durch die Briefform. Dies macht sich durch die direkte Ansprache des Lesers und die persönlichen Elementen bemerkbar (Stichweh, 1984: S. 401ff.). Die Glaubwürdigkeit der Argumente war an die Person des Autors gebunden. Über die Form des Briefes versuchte der Autor, den Leser persönlich zu überzeugen und sich selbst als Vermittler von Wahrheit einzusetzen. Das änderte sich im Verlauf des 19. Jhd., als wissenschaftliche Publikationen in Form des Zeitschriftenaufsatzes zunahmen. Sie richteten sich an ein anonymes Publikum und avancierten zum Standard der wissenschaftlichen Kommunikation. (Gläser, 2006: S. 235f.)

Mit dem Zeitschriftenaufsatz verschwanden die persönlichen Elemente. Die Mitteilungen wurden anonymer, indem sowohl Autor als auch Adressat in den Hintergrund rückten. Auch die Sprache wurde unpersönlicher und exakter. Damit wurde ein grundsätzlich neuer Stil geprägt, der keine Simulation eines Gesprächs mehr darstellte, sondern durch seine Form ausdrückt, dass ein Sachverhalt wiedergegeben wird und keine persönlichen Urteile oder Meinungen. Der Anspruch auf Wahrheit wird hierbei über die Beachtung wissenschaftlicher Standards bekräftigt. (Daston, 1991: S. 381ff.)

Der Wechsel der Form brachte neue Anforderungen mit sich:

> "Mit der veränderten Kommunikationsform versagen Strategien, die auf sozialer Glaubwürdigkeit und direktem Kontakt beruhen wie im Falle der 'Genthemen-Wissenschaft'". (Heintz, 2007: S. 69)

Wie Heintz (ebd.) zeigt, waren neue Überzeugungsstrategien notwendig, um die Akzeptanz der wissenschaftlichen Aussagen wahrscheinlich zu machen. Strategien, die ohne persönliches Vertrauen und gemeinsame Wahrnehmungen auskommen (S. 69). Indem sich die Form der Mitteilung an strikten Regeln orientierte, wurde die Rolle der o.g. Zeugen ersetzt. Dazu gehörte u.a., dass:

- Die Mitteilung ein exaktes Nachvollziehen ermöglicht.
- Die Begrifflichkeiten klar definiert sind.
- Die Gedankengänge lückenlos nachvollziehbar sind.
- Die Ergebnisse unabhängig von Raum, Zeit und Person reproduzierbar sind.

Über die Standardisierung von Messung (vgl. dazu Abschnitt 3.1.2) und Kommunikation wurde letztere anschlussfähig. Über diverse Standardmaße (z.B. Gewicht, Zeit etc.) konnte weltweit problemlos und eindeutig kommuniziert werden.

Laut Heintz (2007) ist der Wechsel von persönlicher Anrede zu entpersonalisierten Beschreibungen typisch für die Operationsweise von symbolisch generalisierten Kommunikationsmedien (S. 71). Wissenschaftliche Kommunikation bringt sie dabei in Verbindung mit dem von Luhmann beschriebenen Kommunikationsmedium *Wahrheit*. Dieses signalisiert, dass eine Aussage auf wissenschaftlichen Grundlagen basiert und damit nicht nur von persönlichen Ansichten oder Meinungen geprägt ist (Luhmann, 1992: S. 173ff.). Die Selektion der Information werde dabei der Umwelt und nicht den Beteiligten zugerechnet. Da wissenschaftliche Aussagen offenkundig Sachverhalte beschreiben, die unter anerkannten methodischen Verfahren zustande gekommen sind, erhöht sich laut Heintz die Akzeptanzwahrscheinlichkeit. Sie werden mit Objektivität assoziiert, da sie einen externen Sachverhalt wiedergeben und keine persönliche Meinung. (Heintz, 2007: S. 71)

3.3.2 Objektivierungsleistung von Zahlen

Aus Abschnitt 2.2.4 geht hervor, dass Zahlen grundlegend von Ziffern zu unterscheiden sind. Zahlen verweisen als Ergebnis einer Größenmessung immer auf einen Referenten, d.h. auf die „Außenwelt“. Ziffern hingegen haben keinen externen Bezug (vgl. Frege (1884/1990)). Insofern ist die Überzeugungskraft von Zahlen in hohem Maße abhängig von der Glaubwürdigkeit der Messverfahren (Quantifizierung, Metrisierung) und der Messapparatur (Porter, 1995: S. 213f.). Es geht hierbei um die Reliabilität und Validität der Messung. Die Messergebnisse müssen wiederholbar und unabhängig von Raum, Zeit und Person zu den gleichen Ergebnisse führen.

Wie in Kapitel 2 deutlich wurde, ist die Vergleichbarkeit von „Phänomenen“ nicht intrinsisch, sondern basiert auf den Prozessen der Quantifizierung und Metrisierung. In diesem Zusammenhang ist es erstaunlich, dass Zahlen oftmals ohne Zweifel hingenommen werden:

> "Zahlen, die nicht strittig sind, werden nicht als selektive Beschreibungen einer zugrunde liegenden Wirklichkeit angesehen, sondern mit dieser selbst gleichgesetzt." (Heintz, 2007: S. 75)

Dieser Umstand ist darauf zurückzuführen, dass einer mechanischen Messung, die auf Messgeräten und quantitativen Verfahren basiert, eine größere Zuverlässigkeit zugeschrieben wird als direkten Beobachtungen. Qualitative Einschätzungen sind nur unter Einschränkung explizierbar und selten standardisierbar. Bei Messungen wird hingegen davon ausgegangen, dass sie unabhängig von Ort, Zeit und Person wiederholt zu den gleichen Ergebnissen führen. (vgl. Heintz (2007), Porter (1995) und Schlaudt (2009))

Die Teilschritte des Messprozesses (Kapitel 2) zeigen jedoch, dass Messergebnisse nicht zwingend auf den Bereich der Phänomene übertragen werden können. Ein ungenaues Messgerät, z.B. in Form einer ungünstig gestellten Frage innerhalb eines Fragebogens oder durch die Wahl einer unpassenden Skala, kann Artefakte zur Folge haben. Jedem Schritt des Messprozesses muss daher höchste Aufmerksamkeit und Präzision geschenkt werden.

Ein wichtiges Element für die Glaubwürdigkeit von Zahlen ist das Vertrauen in die Messung selbst. Nur wenn die Messung glaubwürdig ist, kann sie als tatsächliche Beschreibung der Außenwelt angesehen werden. Auf dieser Basis rücken die Verfahren und Methoden zur Herstellung der Zahlen in den Hintergrund. (Porter, 1995: S. 213f.)

In Zeitungsbeiträgen, die in dieser Arbeit im Fokus der Betrachtung stehen, werden die Verfahren zur Herstellung von Indexzahlen in der Regel nicht genannt (vgl. Abschnitt 5.7). Das Vertrauen in die Zahlen wird daher weitestgehend auf außerepistemische Maßnahmen angewiesen sein, z.B. über Monopolisierung und Netzwerkbildung (vgl. Jones und Dugdale (2001)).

3.3.3 Zahlen und die Mobilisierung von Akzeptanz

Indem Zahlen signalisieren, dass ihre Erstellung auf eine spezifische Auswahl rückführbar ist, erhöhen sie laut Luhmann die Wahrscheinlichkeit der Annahme einer Kommunikation:

> "Man kann eine zugemutete Kommunikation annehmen, wenn man weiß, daß ihre Auswahl bestimmten Bedingungen gehorcht; und zugleich kann derjenige, der eine Zumutung mitteilt, durch Beachtung dieser Bedingungen die Annahmewahrscheinlichkeit erhöhen und sich selbst damit zur Kommunikation ermutigen." (Luhmann, 1997: S. 321)

Zahlen signalisieren Objektivität und werden in diesem Zusammenhang oft mit Wahrheit gleichgesetzt. Indem sie scheinbar von subjektiven Elementen befreit sind, entsteht der Eindruck, dass die Natur für sich selbst sprechen kann. Zahlen werden nicht mehr als Abbildung betrachtet, sondern mit der Realität gleichgesetzt – ähnlich wie bei fotografischen Bildern. Unterstützend wirkt der Umstand, dass Zahlen und Bilder, im Gegensatz zur Sprache, nicht über das Beschriebene hinausweisen. Bilder scheinen die Realität abzubilden, die außerhalb von ihnen liegt, ohne ihre Bildhaftigkeit zu thematisieren. Darin liegt ihre Objektivierungsleistung (Heintz, 2007: S. 79).

Scheinbar existiert kein Unterschied zwischen der Realität und dem Bild. Tatsächlich handelt es sich jedoch nur um eine Selektion, die eine Vielzahl von Informationen ignoriert. Ähnlich verhalten sich Zahlen. In Form von Indizes sind Zahlen hoch aggregierte Konstrukte. Die zugrundeliegenden Messmethoden und Rohdaten sind schwierig nachzuvollziehen. Dafür müssen genaue Informationen über das Indexkonzept und die Methoden der Quantifizierung und Metrisierung bekannt sein. Die Komplexität und Kontingenz der Konstruktion verschwindet hinter den Zahlen und lässt damit auch das Problem der Manipulierbarkeit mitschwingen.

Bilder und Zahlen teilen eine weitere Gemeinsamkeit. Im Gegensatz zu sprachlichen Aussagen lassen sie sich schwieriger negieren. In der Sprache ist immer eine Ja-/Nein-Fassung enthalten. Damit beinhaltet sie automatisch die Möglichkeit der Negation. Im Fall der Zahlen ist die Negation dagegen deutlich aufwändiger. Hier muss eine Alternativfassung aktiv erarbeitet werden, z.B. über die Reproduktion der Zahlen. Da dafür ein umfassendes Verständnis vom Messprozess sowie technische Apparate erforderlich sind, ist jedoch davon auszugehen, dass solch eine Reproduktionsleistung selten von Lesern einer Tageszeitung übernommen werden kann. Anders als in der Wissenschaft beruht in solchen Fällen die Akzeptanz von Zahlen eher auf einem „Einverständnis", als auf „Verständnis" (ebd.: S. 79). Außerhalb der Wissenschaften ist daher zu erwarten, dass die Zurechnung von Objektivität zu Zahlen potentiell instabiler und kontroverser ist:

> "Die Zurechnung von Objektivität impliziert folglich ein doppeltes Vertrauen: einerseits in die Zuverlässigkeit der Messverfahren, über die die Rohdaten gewonnen wurden, und andererseits in die Zuverlässigkeit der verwendeten statistischen Verfahren und deren korrekte Anwendung. Nur unter dieser Bedingung kann Quantifizierung als Kommunikationsmedium fungieren und die Kommunikation zwischen Personen erleichtern, die sich persönlich nicht kennen."[6] (Heintz, 2007: S. 79)

Im empirischen Teil der vorliegenden Arbeit (Kapitel 5) wird untersucht, von welchen Voraussetzungen die Zurechnung von Objektivität zu statistischen Indizes abhängt.

[6] Nach der Darstellung des Messprozess von Böhme ist von dreifachem Vertrauen auszugehen (Quantifizierung, Metrisierung, statistische Verfahren) (vgl. Abschnitt 2.2.1).

Kapitel 4

Die Konstruktion von Indizes

"Index numbers could never simply be observed; they normally involved extensive data collection and often difficult or at least tedious calculations. Their credibility required that they be calculated, even if from bad data, and it has never been acceptable to adjust a number on the basis of judgment alone, however expert."

Theodore M. Porter, 1995

In Kapitel 3 wurde gezeigt, dass sich Indizes parallel zur steigenden Komplexität in Wissenschaft und Gesellschaft entwickelten. In den 1920er Jahren erhielt ihre Verbreitung innerhalb der Ökonomie im Rahmen der Inflation einen starken Impuls und dehnte sich von da aus u.a. auf Politik und Gesellschaft aus. Von Anfang an griffen dabei die Medien auf diese Art der Quantifizierungen zurück (vgl. Kapitel 5). Da im Verlauf der Arbeit der Fokus auf dem Medium Zeitung liegt, hier nur ein kleiner Ausblick auf den Umfang der Indexzahlen: Allein in der Wikipedia befinden sich fast 200 Einträge zu dem Konzept *Index* (vgl. Tabelle A.1). Auch Bandura behandelt in seiner Studie von 2008 178 Indizes. Dabei geht er u.a. auf Attribute wie Erstellungsjahr, Urheber, Umfang und Häufigkeit der Veröffentlichung ein.

Warum erfreuen sich Indizes solcher Beliebtheit? Wie bereits angesprochen wurde, liegt die Annahme nahe, dass sie (scheinbar) Komplexität oder Mehrdimensionalität reduzieren und daher besonders gerne in schwer fassbaren Bereichen von Wirtschaft, Gesellschaft, Entwicklung oder Umwelt genutzt werden. Um sich dieser Frage hinreichend

zu nähern und eine Basis für die anschließende Artikelanalyse zu schaffen, widmet sich dieses Kapitel den vielfältigen Funktionen und möglichen Problemen von Indizes, die sich im Folgenden aus den Konstruktionsmethoden ergeben. Aus diesem Grund wird in diesem Kapitel die Methodologie der Indexkonstruktion problematisiert, indem einzelne Konstruktionsschritte und die damit einhergehende Qualitätsdimension aufgeführt werden. Diese dienen im empirischen Teil der vorliegenden Arbeit als Grundlage der Bewertung der Indizes.

4.1 Funktionen und Probleme von Indizes

Wie in Kapitel 2 dargelegt wurde, verlangen bestimmte Fragestellungen die Operationalisierung von Konzepten oder theoretischen Begriffen, die sich aus mehr als einem Merkmal zusammensetzen. Dies kann der Fall sein, wenn ein beobachtbares (eindimensionales) Merkmal das Konzept nicht hinreichend genau abbildet oder die Theorie mehrere Dimensionen anspricht. Eine Lösung für dieses Problem stellen Indizes dar. Mindestens zwei Dimensionen werden dabei über Indikatoren i.d.R. zu einer Maßzahl zusammengefasst. (Schnell, 2005: S. 158)

Das Interpretieren einer einzelnen Indexzahl scheint häufig einfacher als eine ganzheitliche Betrachtung verschiedener Indikatoren. So werden Indizes unter unterschiedlichsten Fragestellungen genutzt, um komplexe Sachverhalte zu vereinfachen und kommunizierbar zu machen. Werden sie z.B. in der Wirtschaft z.T. für die automatisierte Steuerung von Prozessen herangezogen (DAX[1]), stoßen sie in sozialen oder politischen Bereichen Diskussionen an (vgl. Pickel (2013) und Randolph (1979)) oder lenken den gesellschaftlichen Diskurs auf einen Brennpunkt.

Indizes werden stets mit Hilfe eines Modells gebildet und haben in der Regel mehrere Dimensionen, die sie vereinen. Die Zielgröße ist nicht direkt beobachtbar und messbar. Indexzahlen ermöglichen es, Trends komplexer Zusammenhänge aufzuzeigen und dadurch Veränderungen zu beobachten. Aus dieser Eigenschaft ergibt sich nach Saisana, Saltelli und Tarantola (2005) neben der Möglichkeit, komplexe, mehrdimensionale Realitäten[2] abzubilden eine Vielzahl an weiteren hilfreichen Funktionen und Vorteilen

[1] Z.B. indem durch Über- oder Unterschreiten definierter Grenzwerte Kauf- oder Verkaufsanweisungen ausgelöst werden.

[2] Es bleibt hier offen, welche „Realität“ gemeint ist. Im Folgenden wird dem Zusammenhang von „Realität“ und Index in aller Allgemeinheit nachgegangen.

in der Nutzung von Indizes. Sie ermöglichen nicht nur eine leichtere Interpretation einer großen Anzahl an Indikatoren, sondern reduzieren gleichsam die sichtbare Menge dieser, ohne dabei die zugrunde liegenden Informationen zu verlieren. Das gestatte eine Nutzung der größtmöglichen Informationsmenge unter einem gegebenen Limit an Dimensionen.

Bezüglich der medialen Analyse sind zudem die Eigenschaften der (Index-) Zahlen als symbolisch generalisiertes Kommunikationsmediums relevant (vgl. Abschnitt 3.3). In diesem Zusammenhang sind sie als jene Kommunikationsanweisungen zu verstehen, die es unter einer an sich unwahrscheinlichen Kommunikation ermöglichen, Akzeptanz herzustellen. Sie kommen zum Einsatz, wenn persönliches Vertrauen fehlt oder wenn ein Konsens nicht mehr über persönliche Interaktion hergestellt werden kann. (vgl. Abschnitt 3.3)

Der Index hilft auf diese Weise beim Anreichern einer Fragestellung mit quantitativen Informationen. In der Folge vereinfacht er außerdem die Kommunikation über einen bestimmten Sachverhalt, indem er eine Verständigungsbasis schafft und bei der Formulierung von aussagekräftigen Aussagen unterstützt. Erst dadurch wird ein Vergleich komplexer Dimensionen in der öffentlichen oder wissenschaftlichen Diskussion ermöglicht.

Neben der Vielzahl an Möglichkeiten, einen Index „gewinnbringend“ einzusetzen, können aufgrund der vermeintlich reduzierten Komplexität jedoch auch Probleme entstehen (Lippe und Kladroba, 2004). In Tabelle 4.1 werden die im Folgenden ausformulierten Vor- und Nachteile zusammengefasst.

Indizes basieren neben der zugrunde liegenden Messung auf mathematischen oder statistischen Modellen, die z.T. willkürliche oder nicht-objektive Elemente enthalten können (vgl. Abschnitt 4.3.5.2). Der allgemein anerkannte Anspruch an eine Messung, objektiv und wiederholbar zu sein, ist bei der Konstruktion von Indizes nicht gesichert. Dementsprechend existiert eine breite Diskussion über Sinn und Unsinn von Indizes (hier „composite indicators“):

> “[. . .] it is hard to imagine that debate on the use of composite indicators will ever be settled [. . .] official statisticians may tend to resent composite indicators, whereby a lot of work in data collection and editing is „wasted“ or „hidden“ behind a single number of dubious significance. On the other hand, the temptation of stakeholders and practitioners to summarise complex and

> sometime elusive processes (e.g. sustainability, single market policy, etc.) into a single figure to benchmark country performance for policy consumption seems likewise irresistible.“ (Saisana, Saltelli und Tarantola, 2005)

In der Nutzung von Indizes können weitere Schwierigkeiten auftauchen und den Vorteil dadurch schmälern. So scheint die vermeintlich leichtere Interpretation komplexer Daten mithilfe eines Index zunächst vorteilhaft, obwohl sie gleichermaßen eine Fehlerquelle darstellt und zu vereinfachten Aussagen über einen komplexen Sachverhalt einlädt. Wenn das theoretische Konstrukt anders quantifiziert wird, als die traditionelle Bedeutung (Semantik) des Begriffs erforderlich machen würde, können Fehlinterpretationen, falsche Schlussfolgerungen oder gar Fehlentscheidungen die Folge sein.

So führt beispielsweise die Benennung des *Happy Planet Index* zu Missverständnissen. Der Begriff *Happy Planet* erweckt in manchen Betrachtern den Eindruck, es handle sich um eine Maßzahl für „Glück“. Umso überraschter sind sie, wenn sie die USA auf einem der letzten Plätze im weltweiten Ranking wiederfinden. Ursache für dieses Missverständnis ist sicherlich die irreführende Benennung des Index. In Wirklichkeit steht er eher als Maß für die ökologische Effizienz, mit der die Zufriedenheit eines Landes erzeugt wird (vgl. Abschnitt 5.7.3).

Weitere Probleme können aus einer inadäquaten Konstruktion des Index (z.B. fehlerhafte Gewichtung) oder einer ungeeigneten Messapparatur (z.B. Skala) resultieren. Neben einer fehlerhaften ist auch eine manipulative Konstruktion denkbar, z.B. wenn die Auswahl und Gewichtung der einzelnen Indikatoren spezifischen Interessen folgt (vgl. dazu Abschnitt 2.3.1). Auch das bewusste Auslassen kompliziert zu messender Dimensionen führt eine einseitige und damit nicht ausreichende Betrachtung eines Sachverhalts mit sich. Um die Aufmerksamkeit auf diese speziellen Fehlerquellen zu lenken, soll nach einer formal-sprachlichen Darstellung des Indexproblems in Abschnitt 4.3 die Konstruktion von Indizes anhand der einzelnen Teilschritte nachvollziehbar gemacht werden.

4.2 Formalisierung des Indexproblems

Der Begriff Index wird an dieser Stelle nach der Definition vonBesozzi und Zehnpfennig (1976) als eine Funktion von Indikatoren in der folgenden allgemeinen Form

Tabelle 4.1: Vor- und Nachteile von Indizes

Vorteile von Indizes:	**Nachteile von Indizes:**
• Kann komplexe, mehrdimensionale Realitäten zur Unterstützung von Entscheidungsträgern zusammenfassen.	• Eine inadäquate Konstruktion kann zu fehlerhaften und irreführenden Aussagen führen.
• Leichter interpretierbar als Vielzahl von Indikatoren.	• Lädt zu vereinfachten Aussagen ein.
• Reduziert die Menge sichtbarer Indikatoren, ohne die darunter liegende Information zu verlieren.	• Signalisiert ggf. Messbarkeit, wo keine ist.
• Ermöglicht die Nutzung von mehr Informationen unter einem gegebenen Limit.	• Kann missbraucht werden, um gewünschte Entscheidungen durchzusetzen, wenn die Konstruktion intransparent oder fachlich unsauber ist.
• Lenkt die Aufmerksamkeit auf spezielle Fragestellung.	• Die Gewichtung und Auswahl von Indikatoren kann subjektiv motiviert sein.
• Erleichtert die Kommunikation mit Öffentlichkeit (Bürger, Medien usw.).	• Kann Mängel in speziellen Dimensionen verschleiern und die Reaktion auf Probleme verzögern, wenn die Indexkonstruktion intransparent ist.
• Hilft bei der Formulierung von Aussagen.	• Kann zu einseitiger Betrachtung führen, wenn schwer zu messende Dimensionen ignoriert werden.
• Ermöglicht den Vergleich komplexer Dimensionen.	

Quelle: In Anlehnung an Nardo u. a. (2008: S. 13f.)

angesehen (vgl. auch Lazarsfeld (1959) und Lazarsfeld und Rosenberg (1956))

$$I_k = f(x_1, x_2, ..., x_k). \tag{4.1}$$

Die Variablen $x_1, x_2, ..., x_k$ fungieren dabei als Indikatoren für die zu bestimmende Variable X und sind die Elemente des Index I_k. Formel 4.1 beschreibt damit den Zusammenhang von theoretischem Konstrukt und beobachtbaren Indikatoren.

Die Bedeutung der Indikatoren zur Bestimmung der Variable X kann wie folgt formalisiert werden

$$x_i = g_i(X, Z_1, Z_2, ..., Z_n). \tag{4.2}$$

Dabei sind $Z_j (j = 1, 2, ..., n)$ eine Reihe von Variablen, welche die Bestimmung von X über den Index I_k beeinflussen.

Für die Konstruktion eines Index werden die beiden Funktionen f und g sowie deren

Elemente so definiert, dass der Index I_k eine Funktion der zu bestimmenden Variable X plus einem Störterm ist

$$I_k = h(X) + \varepsilon_{I_k}. \tag{4.3}$$

Dabei wird davon ausgegangen, dass

$$\varepsilon_{I_k} < \varepsilon_{I_{k-1}}. \tag{4.4}$$

Der Fehlerterm ε sammelt alle zufallsbedingten Schwankungen. Sein Erwartungswert $E(\varepsilon)$ ist gleich Null. Es wird angenommen, dass sich die zufallsbedingten Schwankungen mit steigender Anzahl an Beobachtungen aufheben und der Fehler[3] insgesamt abnimmt. In Bedingung 4.4 wird der Zusammenhang zwischen der Anzahl der Freiheitsgrade des Index (diese werden über die Anzahl der Indikatoren k bestimmt) und der Größe des Fehlers zusammengefasst. Es wird ersichtlich, dass die Genauigkeit die benötigte Anzahl an Elementen bestimmt. Optimaler Weise sollte die Anzahl der Elemente so lange erhöht werden, bis eine Erhöhung um eine Einheit keinen wesentlichen Einfluss mehr auf den Fehler von I_k hat, d.h. bis Bedingung 4.4 nicht mehr erfüllt ist.

Indizes müssen definitionsgemäß immer mindestens zwei Elemente besitzen (vgl. Abschnitt 2.3.2). Für die obere Darstellung folgt demnach die zusätzliche Bedingung, dass $k \geq 2$ sein muss. Inhaltlich impliziert diese Bedingung, dass Indizes nur dann konstruiert werden, wenn keine Variable x_0 definiert werden kann, die alleinig konsistent genauere Bestimmung von X liefert als alle Kombinationsmöglichkeiten der Menge an Indikatoren $x_1, x_2, ..., x_k$.

Für die Konstruktion von Indizes sind vier Entscheidungen nötig. In Abschnitt 4.3 wird genauer darauf zurückgekommen. Um den Bezug zur formalen Darstellung zu klären, seien sie hier im Groben zusammengefasst.

Die ersten Entscheidung findet innerhalb der Definition des theoretischen Rahmens (vgl. Abschnitt 4.3.1) statt. Hier geht es um die Definition des Zusammenhangs zwischen dem theoretischen Konzept und den beobachtbaren Variablen, d.h. um die Entwicklung eines Modells.

[3] Der Fehler bezieht sich dabei auf die Summe von systematischen und zufallsbedingten Abweichungen, die bei der Bestimmung des Konzepts auftreten.

Eine weitere Entscheidung wird im Rahmen der Variablenselektion (vgl. Abschnitt 4.3.2) getroffen. Hier geht es um die Auswahl der k beobachtbaren Variablen zur Bestimmung des theoretischen Konzepts, d.h. der Variable X. Meist kann die Variable X neben den k beobachtbaren Variablen über eine größere Anzahl m $(m > k)$ funktional verbunden werden. Daraus folgt, dass Formel 4.1 nur eine Untermenge der Möglichkeiten definiert. Formal geschrieben beschreibt sie nur eine Untermenge

$$A_{I_k}(x_1, x_2, ..., x_k) \tag{4.5}$$

aus der Menge

$$B_X(x_1, x_2, ..., x_k, ..., x_m) \tag{4.6}$$

Die Menge 4.6 beinhaltet alle m Variablen, die für die Bestimmung von X auf Basis der Theorie möglich sind. Für die Identifizierung einer optimalen Untermenge von k Indikatoren ist die Entwicklung von Kriterien erforderlich. Die Entscheidung über die Menge und Auswahl der Indikatoren wirkt sich, wie bereits angemerkt, auf den Fehler aus.

Die dritte Entscheidung bezieht sich auf das Gleichungssystem 4.2. Sie wird in Abschnitt 4.3.4 thematisiert. Hier geht es um die Analyse der Struktur der Indikatoren und der Lösung des Gleichungssystems zur Verbindung von theoretischem Konzept und beobachtbaren Variablen.

Der letzte Schritt besteht in der Entscheidung über eine geeignete Kombination der Indikatoren zu einem n-Dimensionalen Index (wobei meist $n=1$). Es müssen Fragen zur Normalisierung, Gewichtung und Aggregation geklärt werden. Aus allen Möglichkeiten ist dabei jene zu wählen, die den Fehler minimiert sowie theoretisch interpretierbare Ergebnisse liefert.

4.3 Schritte der Indexkonstruktion

Einige Indizes sehen sich in Wissenschaft und Öffentlichkeit einer großen Skepsis gegenüber. Nicht selten liegt dies an ihrer Konstruktion, Funktion und der fehlenden

Transparenz. Während manche Indizes als PR-Werkzeuge[4] abgetan werden, unterliegen andere dem Verdacht, ihr alleiniger Zweck läge im Untermauern von bereits existierenden Behauptungen[5]. Um derartige Risiken zu vermeiden, ist bei der Entwicklung und Konstruktion von Indizes auf größtmögliche Transparenz zu achten, die wie folgt geschaffen werden kann:

1. Die Qualität von Indizes hängt nicht nur von der Methodologie, sondern auch von der Genauigkeit des theoretischen Rahmen ab. So ist es a) entscheidend, dass der Konstrukteur des Index das zu messende Phänomen möglichst klar definiert und Ungenauigkeiten in der Begrifflichkeit vermeidet. Außerdem muss b) die Datengrundlage „sauber“ und objektiv[6] sein. Jede Verzerrung in den Daten wirkt sich unmittelbar auf die Qualität des Index aus.

2. Es existiert eine Vielzahl ausführlich diskutierter und anerkannter statistischer Methoden für die Konstruktion von Indizes, die wann immer es möglich ist, zum Einsatz kommen sollten, um auf heuristische und willkürliche Konstruktionselemente zu verzichten. Durch den Nachweis statistischer Methoden ist ein Index transparenter und leichter zu rechtfertigen.

3. Fehlende Transparenz bei der Veröffentlichung von Indexwerten führt in erster Linie zu Misstrauen und kann ferner Fehlinterpretationen verursachen, was u.U. dem Ruf schädigen kann. Das vordergründige Ziel des Index ist zwar die Reduktion von Komplexität, dennoch ist das Angebot eines Drill-down[7] auf sämtliche Ebenen (die nicht dem Datenschutz widersprechen) unumgänglich. Auch die Nachvollziehbarkeit der Erstellung des Indexwertes ist ausschlaggebend, um bei negativen Veränderungen Skepsis zu vermeiden.

Der Aufbau dieses Kapitels folgt Nardo u. a. (2005). Es handelt sich bei diesem Werk um eine Publikation der Europäischen Kommission. Sie lässt sich als zentrales Werk innerhalb einer ganzen Reihe von Publikationen der Europäischen Union und der OECD zum Thema Indizes ansehen (u.a. Nardo u. a. (2005, 2008), Saisana und Saltelli

[4] Z.B. *GPRA-Vertrauensindex* oder *Edelman Trust Barometer*.
[5] Diese Vermutung wird besonders bei hoch-korrelierten Indizes geweckt.
[6] Zur objektiven Wissensvermittlung vgl. auch Abschnitt 3.3.1.
[7] Gemeint ist die Betrachtung der Daten auf einer niedrigeren Hierarchie-Stufe.

(2012), Saisana und Tarantola (2002) und Tarantola u. a. (2006)).[8] In der Politik werden mit einer Vielzahl von Indizes komplexe Zusammenhänge kommuniziert und überschaubar gemacht.[9] Da es in der Wissenschaft an einer umfassenden und einheitlichen Darstellung des Indexproblems fehlt, kann die Menge der Publikationen der Europäischen Kommission und der OECD als Versuch angesehen werden, diesen Mangel selbständig (außerhalb des Wissenschaftsbetriebs) zu beheben. Die verschiedenen Wissenschaftsdisziplinen müssen sich somit der Kritik stellen, ihrer Aufgabe – eine wissenschaftliche Auseinandersetzung mit dem Indexproblem – unzureichend gerecht geworden zu sein. Publikationen unter politischen Organisationen sind deutlich schwieriger von Interessen zu trennen und sollten insbesondere vor diesem Hintergrund analysiert und bewertet werden.

Zum aktuellen Zeitpunkt ist die Darstellung von Nardo u. a. (2005), zumindest aus rein methodologischer Sicht, die umfassendste. Jedoch bezieht sich die Darstellung einseitig auf Ländervergleiche und ist überwiegend statistisch/mathematisch geprägt. Sie wird in diesem Kapitel verallgemeinert und erweitert. Ebenfalls werden die Erkenntnisse aus Kapitel 2 und 3 Berücksichtigung finden, indem z.B. die Konstruktionsmethoden von wissenschaftlichen Messungen abgegrenzt sowie auf ihre kommunikativen Eigenschaften hin untersucht werden. Die in Teilaspekte getrennte Indexkonstruktion wird an dieser Stelle kurz zusammengefasst und im Verlauf des Kapitels ausführlich beschrieben, bewertet und problematisiert.

Theoretischer Rahmen (Abschnitt 4.3.1): Der theoretische Rahmen stellt die Basis für die Definition und Kombination der Indikatorvariablen dar. Das zu bestimmende mehrdimensionale „Phänomen“ muss im Vorfeld klar abgegrenzt und eindeutig definiert werden.

Variablenselektion (Abschnitt 4.3.2): Die genutzten Daten müssen hinsichtlich des zu bestimmenden „Phänomens“ objektiv, belastbar und reproduzierbar sein, wobei u.U. der Einbezug von Proxy-Variablen möglich ist.

[8] Teilweise werden die Werke auch in wissenschaftlichen Zeitschriften unter dem Namen der Europäischen Kommission veröffentlicht.

[9] Z.B. Korruptionsindex, Better Life Index, Human Development Index; für weitere Beispiele sei auf Bandura (2008) verwiesen.

Datenqualitätsmanagement (Abschnitt 4.3.3): Sämtliche statistische Analysen sind abhängig von der Qualität der Daten. Um diese zu gewährleisten, müssen folgende Fragen geklärt werden: Wie wird mit fehlenden Daten umgegangen? Werden Ausreißer einbezogen? Sind die Variablen reliabel? Wird der Umgang mit diesen Fällen nicht sorgfältig behandelt, kann es zu Verzerrungen kommen.

Prüfen der Datenstruktur (Abschnitt 4.3.4): Statistische Analysen decken in der Regel Strukturen auf, die unter den jeweiligen Daten liegen. So erleichtern sie z.B. die Bildung von Item-Batterien, identifizieren redundante Indikatoren und offerieren letztendlich Vorschläge der Zusammenfassung.

Aggregation (Abschnitt 4.3.5): Bestimmte Erkenntnisse, die über die Struktur gewonnen werden, führen u.U. zu notwendigen Anpassungen in den Daten. Ggf. müssen Ausreißer ausgeschlossen oder Skalen standardisiert werden. Anschließend folgt die Wahl einer mathematischen Gleichung für die Konstruktion des Index, wobei ebenfalls statistische Analysen unterstützen, z.B. um die Art und Höhe der Gewichtung festzulegen.

Prüfung (Abschnitt 4.3.6): Eine Prüfung des Index auf Sensitivität, Reliabilität, Validität, Aussagekraft und seine Fähigkeit der Generierung von neuen Informationen ist Voraussetzung vor der Veröffentlichung. In diesem Zusammenhang ist eine Untersuchung der Korrelation zu anderen Indikatoren und Indizes ratsam.

Reporting (Abschnitt 4.3.7): Auch die Darstellungsweise des Index sollte im Vorfeld diskutiert werden, da sie die Interpretation der Werte zu beeinflussen vermag. Die Wahl einer adäquaten Visualisierungstechnik ist daher relevant.

4.3.1 Theoretischer Rahmen

Am Anfang der Indexkonstruktion steht die Entwicklung eines exakten theoretischen Rahmens, der das Fundament des Index bildet. In dieser Definition ist vordergründig, mit welchen Indikatoren das Konzept operationalisiert wird. Eine Modifikation dieser Liste ist in späteren Analyseschritten durchaus möglich. Die Notwendigkeit kann eintreten, wenn sich etwa herausstellt, dass zwei Indikatoren denselben Beitrag zur Bestimmung des Konzepts leisten. Entscheidend bei der Zusammenstellung der Indikatoren ist, dass:

> „This process should ideally be based on what is desirable to measure and not on which indicators are available.“ (Nardo u. a., 2008: S. 22)

Als Ziel sollte also, wie in Kapitel 2 bereits beschrieben, stets das zu operationalisierende Konzept im Fokus stehen und weniger die verfügbaren Indikatoren. Andernfalls wird über die zur Verfügung stehenden Indikatoren ein Konzept konstituiert, was die aufgeworfene Fragestellung womöglich kaum oder verzerrt beantwortet. Auch hier ist die Transparenz ein entscheidender Faktor. So dienen folgende vier Punkte nach Nardo u. a. (ebd.: S. 22) als Leitfaden:

- **Konzeptbeschreibung:** Mittels der Konzeptbeschreibung kann sich ein Leser Klarheit über das zu bestimmende Konzept verschaffen. Daher sollte sich das Konzept stets auf den theoretischen Rahmen beziehen und erklären, wie die verschiedenen Indikatoren oder Subindizes zusammenhängen. Der Better Life Index der OECD z.B. basiert auf der Idee, dass Wohlstand über elf Aspekte abgebildet werden kann (z.B. Wohnverhältnisse, Einkommen). Über diese elf Aspekte wird Wohlstand operationalisiert und über entsprechende Skalen metrisiert. Mittels dieser Operationalisierung wird eine klare Verbindung zwischen theoretischem Rahmen und der Struktur des Index hergestellt. (vgl. Schnell (2005: S. 109f.))

- **Bestimmung der Subindizes:** Mehrdimensionale Konzepte können über verschiedene Subindizes abgebildet werden. Dies erleichtert u.U. das Verständnis vom Index und, z.B. über eine Drill-down-Funktionalität, auch die Analysen. Dabei ist eine statistische Unabhängigkeit der Subindizes entscheidend. Zudem ist eine theoretische und empirische Rechtfertigung für die Subindizes Voraussetzung. So kann z.B. der Fall eintreten, dass unter Anwendung einer Faktorenanalyse (Abschnitt 4.3.5.3.1) die Unabhängigkeit gewährleistet wird, jedoch das Ergebnis unzureichend interpretierbar ist. Das ist denkbar, wenn z.B. Indikatoren statistisch einem Subindex zugeschrieben werden, der theoretisch nicht zu begründen ist.

- **Auswahlkriterien:** Damit sie bei der Indikatorenselektion als Leitfaden fungieren können, muss zuvor die Auswahl der Indikatoren definiert werden. Dabei sollten die Kriterien vorrangig das zu operationalisierende Konzept beschreiben, z.B. über eine Klassifikation „Output-Input-Prozess“. Bei Indizes, die sich laut Definition auf Output-Leistung beziehen, sollten nur Indikatoren einbezogen werden, die

Output-Faktoren messen. So kann ein Innovationsindex einen Indikator *Anzahl von neuen Produkten oder Dienstleistungen* einbeziehen, sollte jedoch F&E-Ausgaben ausschließen. Der erste Indikator misst einen Output und kann damit direkt mit Leistung in Verbindung gebracht werden. Der zweite Indikator hingegen misst einen Inputfaktor, sodass er nur indirekt in die Leistung eingeht.[10] (vgl. SCHNELL, 2005: S. 109f.)

- **Hypothesen:** Mithilfe der theoretischen Überlegungen des Indexkonstrukts werden Hypothesen formuliert, die bei abschließender Indexprüfung als Validierung dienen. Mögliche Aussagen sind z.B. *Land A hat eine höhere Varianz als Land B* oder *A ist im Ranking über B*.

4.3.2 Variablenselektion

Die Konstruktionsregeln und die Indikatorvariablen des Indexkonzepts sind die entscheidenden Faktoren für die Qualität der Indexwerte. Im Wesentlichen geht es darum, das zu bestimmende theoretische Konzept beobachtbar zu machen, unter Anwendung einer indirekte Messung und verschiedener Indikatoren (vgl. Kapitel 2). Der Anspruch an die Qualität der Indikatorenvariablen ist hoch anzusetzen (ebd.: S. 111). Die Selektion einer Variablen orientiert sich idealerweise an der Eigenschaft, Bedeutung, analytischen Verlässlichkeit, Aktualität und Verfügbarkeit (NARDO u. a., 2008: S. 23).

Die Variablenauswahl wird unter Einbezug des theoretischen Rahmens getroffen. Wie in der Formalisierung des Indexproblems deutlich wurde (vgl. Abschnitt 4.2), existiert jedoch keine fest definierte Menge und Auswahl an Indikatoren. Vielmehr wird die abschließend getroffene Wahl nur eine, im Idealfall optimale, Submenge aller möglichen Indikatoren besitzen. Dieser Schritt ist z.T. subjektiv, da es an allgemeingültigen und wissenschaftlichen Regeln zur Variablenselektion fehlt. Ferner können nicht verfügbare Daten oder Variablen die Qualität der Indexwerte negativ beeinflussen. Dieser Herausforderung sehen sich insbesondere internationale Indizes gegenüber.[11] Aufgrund der Knappheit von quantitativen, gesicherten Daten („hard data“), die über alle Länder

[10] Unter anderen Umständen würde der Quotient *F&E-Ausgaben je neues Produkt* sicherlich sinnvoll sein. Jedoch nicht in einem Index, der Innovationsleistung misst, sondern eher in einem, der z.B. Produktivität misst.

[11] Vgl. dazu den *Happy Planet Index* in Abschnitt 5.7.3.

verfügbar sind, muss nebenbei z.T. auf ungesicherte Erkenntnisse („soft data“) wie Meinungen, Erwartungen aus Umfragen oder politische Bewertungen zugegriffen werden (Petersen, 2004). Diese Datenknappheit führt nicht selten zu Veränderungen der Variablenselektion über verschiedene Erhebungswellen hinweg, was die Vergleichbarkeit der Wellen mitunter stark beeinflusst. (Nardo u. a., 2008: S. 23)

Proxy-Variablen können einen Ausweg aus der Datenknappheit, verursacht durch nicht zur Verfügung stehende oder nicht erhebbare Variablen, darstellen. Sie fungieren als Annäherung an die gewünschte Größe. Beispielhaft sei hier die Größe *Anzahl an Berufstätigen, die einen Computer nutzen*, genannt. Falls diese Zahl nicht zur Verfügung steht, ist stattdessen der Rückgriff auf die *Zahl der Berufstätigen, die Zugriff auf einen Computer haben*, denkbar. Bei der Verwendung derartiger „weicher Daten“ („soft data“) ist große Sorgfalt geboten und in jedem Fall ist die Genauigkeit dieser Variablen durch statistische Maßnahmen, wie Korrelationsanalysen oder Sensitivitätsanalysen, zu prüfen. Außerdem ist eine Untersuchung der Abhängigkeiten der Proxy-Variablen von externen Einflüssen, die u.U. keine Auswirkung auf die gesuchte Zielgröße haben, notwendig.

Auch der Variablentyp (Input-, Output- oder Prozessvariable) spielt bei der Auswahl eine Rolle. Handelt es sich laut Theorie bei dem Index um eine Output-Größe, so ist auch die Indikatorvariable entsprechend zu wählen.

Ein weiteres relevantes Kriterium ist die Skalierung (vgl. Abschnitt 2.2.4). In manchen Fällen ist eine Normierung der gewählten Variablen nötig, um sie mit anderen Populationen oder anders skalierten Variablen vergleichbar zu machen. Bei einem Ländervergleich wäre dies durch eine Normierung auf das Bruttoinlandsprodukt (BIP) des jeweiligen Landes denkbar – beispielsweise in der Form *Anteil Bildungsausgaben am BIP*.

Die Konstruktion eines Index ist mit großem finanziellen Aufwand verbunden, weshalb aus Gründen der Ersparnis oftmals Zugeständnisse in der Datenerhebung gemacht werden. Insbesondere internationale Indizes produzieren bei adäquater Datenerhebung immense Kosten, was je nach Budget zu Kompromissen bei der Datenqualität führt. Nicht selten wird auf kostenfrei zugängliche Daten ausgewichen, wie z.B. die der Wold Value Study[12]. Diese pragmatische Herangehensweise ist so lange akzeptabel, wie maximale

[12]Die Wold Value Study (WVS) wurde seit 1981 über sechs Wellen hinweg in einer Vielzahl von Ländern durchgeführt. Die Ergebnisse und Dokumentation der Erhebungen können unter `www.worldvaluessurvey.org` (besucht am 30. Sep. 2015) heruntergeladen werden.

Transparenz geboten ist. Sicherlich ist es aus Gründen der Vergleichbarkeit nicht optimal, dass z.B. der *Happy Planet Index* seine Indikatoren zwischen verschiedenen Wellen durch Abweichungen in der Fragestellung verändert (vgl. Abschnitt 5.7.3). Jedoch ist diese Anpassung dokumentiert und für den Nutzer nachvollziehbar dargestellt. Das macht je nach Anspruch einen individuellen Umgang mit den Indexwerten möglich.

4.3.3 Datenqualitätsmanagement

4.3.3.1 Fehlende Daten

Wie bereits angesprochen basieren Indizes auf Ergebnissen einer „Messung“ oder zumindest einer Datenerhebung. Potentiell stellen fehlende Daten bei der Indexkonstruktion ein Problem dar, da sie direkten Einfluss auf die Qualität des Index haben. Werden Daten nicht in einem kontrollierten oder experimentellen Umfeld gewonnen, besteht bei der Erhebung die Gefahr fehlender Werte und damit von Datenlücken. Diese Lücken, die z.B. durch Auslassen von Antworten durch die Befragten entstehen, führen bei der Analyse der Daten und der späteren Aggregation zum Index zu Verzerrungen und in ihrer Konsequenz zu Fehlschlüssen. (Nardo u. a., 2005: S. 35)

4.3.3.1.1 Arten von fehlenden Werten Fehlende Werte werden durch unterschiedliche Formen der Nichtantwort verursacht, die jeweils eine adäquate Methode der Behandlung erfordern. Insgesamt lassen sich, wie folgt, sechs Arten der Nichtantwort („Nonresponse“) unterscheiden. (Albers, 2009: S. 119-120)

- **Beim Unit-Nonresponse** handelt es sich um die Verweigerung der Teilnahme des Befragten an der gesamten Untersuchung. Denkt man an die traditionelle Umfrage, äußert sich das z.B. in dem Ausbleiben des Rücklaufs des Fragebogens. Das führt zu einem Sinken der Rücklaufquote. Problematisch hierbei ist in erster Linie das Fehlen der Information über den Grund der Nichtantwort.
- **Bei Item-Nonresponse** hingegen sind lediglich einzelne Antworten auf einen Teil der Fragen nicht vorhanden. Somit existiert zwar ein Datensatz des Befragten, jedoch ist dieser unvollständig.

- **Der Wave-Nonresponse** ist eine Spezialform, bei der Unit- und Item-Nonresponse gleichzeitig auftreten, was etwa bei Längsschnittuntersuchungen (ein Befragter wird z.B. jedes Jahr erneut zum selben Thema befragt) vorkommt. Befragte antworten dabei z.B. in Folgeuntersuchgen nicht mehr oder lassen eine Untersuchungswelle aus oder beantworten nicht alle Fragen.

- **Drop-Outs** beziehen sich ebenfalls auf Paneluntersuchungen und entstehen, wenn z.B. einzelne Untersuchungseinheiten im Verlauf ausscheiden und für Folgeuntersuchungen nicht mehr zur Verfügung stehen.

- **Beim Observation-Nonresponse** handelt es sich um fehlende Antworten einzelner Probanden eines definierten Clusters. Dieses kann z.B. eine Familie sein, bei der in diesem Fall ein Familienmitglied zum Untersuchungszeitpunkt nicht erfasst werden kann.

- **Beim Missing by Design** handelt es sich um eine gezielte Nichterhebung von Daten, die meist durch eine Filterfrage zustande kommt. Fragen zur Arbeitszufriedenheit sind etwa nur sinnvoll, wenn der Befragte einer Beschäftigung nachgeht. Durch die Bestätigung der Filterfrage *Gehen Sie einer Beschäftigung* nach schließt sich die Frage zur Zufriedenheit an. Das Fehlen der Daten ist in diesem Fall also einkalkuliert.

Bis auf den letzten Fall *Missing by Design* können nach Albers (2009: S. 120) die unterschiedlichsten Gründe zum Fehlen der Antworten führen. Der Proband ist beispielsweise nicht erreichbar; bei der Beantwortung werden einzelne Fragen übersehen; alle Antwortmöglichkeiten auf eine Frage werden als nicht zutreffend erachtet; der Befragte möchte nicht antworten oder kennt die Antwort schlichtweg nicht. Darüber hinaus können Datenfehler durch Tippfehler, unzulässige Antworten, Übertragungsfehler oder Missverständnisse entstehen. Die Ursachen können damit sowohl innerhalb als auch außerhalb des Messprozesses liegen.

Ein weiteres Problem besteht darin, dass es nicht nachvollziehbar ist, ob die Ursache der fehlenden Werte auf relevante Informationen verweist. Der Fall würde dann eintreten, wenn z.B. ein Unternehmen eine Frage zur betrieblichen Gesundheitsförderung deswegen unbeantwortet lässt, weil es der Öffentlichkeit verheimlichen möchte, dass es diese nicht betreibt. Eher irrelevant wäre die Ursache, wenn z.B. der Umfragebogen auf dem Postweg

verloren geht, da die fehlenden Daten dann nicht mit den abgefragten Variablen in Verbindung stehen. (Albers, 2009: S. 120)

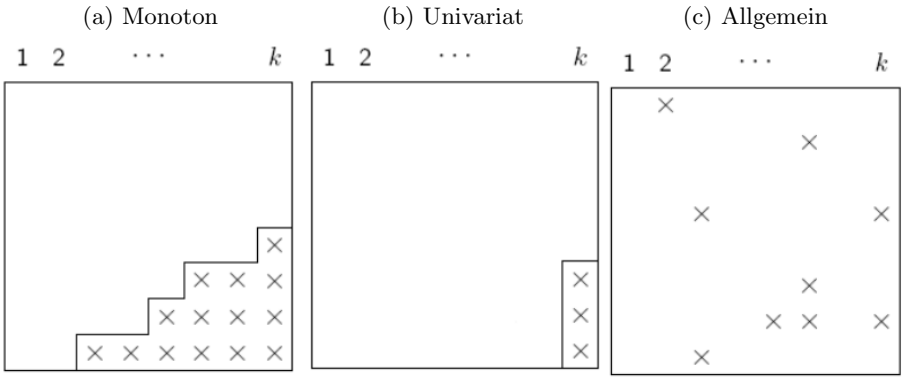
4.3.3.1.2 Fehlerursachen und ihre Konsequenzen Wie eben beispielhaft angemerkt wurde, kann die Ursache für einen Datenausfall mit dem Untersuchungsinhalt in Verbindung stehen oder rein zufällig erfolgen. Zur Unterscheidung der denkbaren Ursachen fehlender Werte existieren nach Albers (ebd.: S. 120f.) drei Bewertungen:

- **Missing Completely At Random (MCAR):** Der Datenausfall ist zufällig bedingt, sodass die beobachteten Daten keine Informationen über einen Fehlendmechanismus beinhalten (z.B. der Fragebogen geht auf dem Postweg verloren).
- **Missing At Random (MAR):** Der Datenausfall ist nur bedingt zufällig, d.h. dass die beobachteten Informationen den Grund für das Fehlen der Daten enthalten. Wurde etwa in einem Fragebogen eine bestimmte Antwort nicht erbracht, ist die Ursache für die Nichtantwort in einer der beantworteten Fragen enthalten.
- **Missing Not At Random (MNAR):** Der Datenausfall ist nicht zufällig und verzerrt die Ergebnisse. Die Antwortwahrscheinlichkeit hängt von den fehlenden Daten ab. Dies kann sich darin äußern, dass Befragte mit speziellen Charakteristika (z.B. Einkommen kleiner 800 EUR) bei bestimmten Fragen zum Verweigern der Antwort tendieren, z.B. *Wie viel Geld geben Sie im Monat für Erholung aus.*

Auf die Spezialform Observed At Random (OAR) wird an dieser Stelle nicht weiter eingegangen. Im Gegensatz zu MAR, bei der die Antworten zu einer Variablen unabhängig von der Ausprägung dieser Variablen sind, sind bei OAR die Antworten zu einer Variablen unabhängig von der Ausprägung der anderen Variablen (vgl. Rubin (1976: S. 584)).

Der Datenausfall bei MCAR ist – abgesehen davon, dass er eine ungenauere Analyse verursacht – als unproblematisch anzusehen. Er verursacht keinerlei Verzerrungen oder Verfälschungen am Ergebnis. Das trifft sowohl auf Unit- als auch auf Item-Nonresponse zu. Anders verhält sich der MNAR-Ausfall, der sehr wohl zu Verzerrungen führen und somit Ergebnisse erzeugen kann, die nicht die „Realität“ abbilden. (Albers, 2009: S. 121)

Abbildung 4.1: Fehlerschemata



Quelle: In Anlehnung an Albers (2009: S. 122)

4.3.3.1.3 Ausfallschema Die Beschreibung des Datenausfalls erfolgt in den drei folgenden Schemata:

a) Das monotone Schema setzt eine Ordnung der Items voraus, sodass wenn ein Item nicht vorhanden ist, das nachfolgende ebenso fehlt (z.B. durch einen Drop-Out).

b) Beim univariaten Schema treten die fehlenden Werte unspezifisch in einem Item oder einer Gruppe von Items auf, wobei eine oder mehrere Beobachtungen fehlen können.

c) Beim allgemeinen Schema ist das Fehlen von Daten in der Matrix zufällig verteilt.

Das Ausfallschema kann Rückschlüsse auf die Ursachen der Nichtantwort ermöglichen. So ist unter dem unvariaten Schema die Prüfung eines MCAR-Ausfalls möglich. Dafür erfolgt zunächst eine Trennung in vollständige und unvollständige Datensätze, woraufhin getrennt nach Teildatensatz die Verteilung für die vorhandenen Variablen ermittelt wird. Anschließend wird die Verteilung der Variablen über die Teildatensätze mittels t-Test oder Maximum-Likelihood-Quotienten-Test verglichen. Liegen keine signifikanten Abweichungen vor, ist ein MCAR naheliegend. Die Überprüfung der anderen Schemata ist deutlich komplizierter und z.T. unmöglich. (Albers, 2009: S. 121f.)

4.3.3.1.4 Quantifizierung des Datenausfalls Für eine größtmögliche Transparenz eines Index sind einige Maßzahlen zur Quantifizierung des Datenausfalls sinnvoll.

Denkbar sind dabei etwa die Rücklaufquote, das Verhältnis von vollständigen zu unvollständigen Datensätzen oder das Verhältnis von leeren zu allen Datenfeldern (vgl. Bankhofer (1995: S. 31)).

4.3.3.2 Behandlung fehlender Daten

Im Umgang mit fehlenden Daten existieren drei anerkannte Methoden, die im Folgenden kurz mit Vorgehen und möglichen Problemen erläutert werden: 1) Beobachtung löschen, 2) Singuläre Imputation und 3) Multiple Imputation.

1) Beobachtung löschen: Bei der Methode des Löschens werden die fehlenden Werte bei der Auswertung einfach vernachlässigt. Diese Vorgehensweise ignoriert potentielle systematische Unterschiede zwischen vollständigen und unvollständigen Beobachtungen. Dadurch können im Zweifelsfall verzerrte Schätzungen eintreten. Es wird zwischen zwei Varianten unterschieden:

- **Es werden nur vollständige Datensätze genutzt** (Complete Case Analysis): Es handelt sich dabei um das simpelste und ein universell anwendbares Vorgehen, bei dem alle unvollständigen Datensätze ignoriert oder gelöscht werden, sodass nur vollständige Datensätze in die Auswertung eingehen. Dieses Verfahren ist ausschließlich für den MCAR-Ausfall anwendbar, da andernfalls mit verzerrten Schätzern und Ergebnissen zu rechnen ist. Am besten ist es für das unvariate Ausfallschema geeignet. Doch problematisch ist bei der Anwendung, dass es zu einer schnellen Reduktion des Datensatzes kommt, da schon eine einzige fehlende Antwort den Ausschluss des gesamten Datensatzes erzeugt. (Albers, 2009: S. 123)
- **Es werden verfügbare Datensätze genutzt (Available Case Analysis):** Dieses Vorgehen bezieht alle verfügbaren Datensätze zwecks Ermittlung der Verteilungsmaße der einzelnen Variablen ein. Bei anderen Maßen, wie der Kovarianz zwischen verschiedenen Variablen, werden hingegen wieder nur die vollständigen Datensätze benutzt. Das verhindert einen Teil des Informationsverlustes, indem unvollständige Datensätze nicht komplett ausgeschlossen werden. Als nachteilig erweist sich hier jedoch der Fakt, dass die Maße auf unterschiedlichen Stichproben beruhen können, wodurch sie nicht mehr vergleichbar sind. (ebd.: S. 123f.)

2) Singuläre Imputation: Die Imputation (engl. unterstellen) beschreibt das Vorgehen, fehlende Werte durch Schätzungen zu ersetzen, was in der Regel mit vielen Unsicherheiten behaftet ist. Im Gegensatz zu der noch zu beschreibenden Methode der multiplen Imputation findet diese Unsicherheit bei der singulären Imputation keine Berücksichtigung. Für die singuläre Imputation existiert eine Reihe an populären Ad-hoc-Verfahren:

- **Expertenschätzung:** Ein Experte wird dazu berufen, die entstandenen Lücken in der Datentabelle zu füllen. Ausgewählt wird dieser aufgrund seines Erfahrungsschatzes auf dem Gebiet der untersuchten Materie. Eine intersubjektive Überprüfbarkeit ist bei dieser Methode jedoch nur bedingt möglich und erfordert eine ausführliche Dokumentation der imputierten Werte. Hier muss das abgegebene Urteil des Experten eindeutig nachvollziehbar sein. Bei großen Datensätzen ist der hohe Aufwand, den diese Methode mit sich bringt, weder handhabbar noch wirtschaftlich vertretbar. (Albers, 2009: S. 124)
- **Mittelwertergänzung:** Diese Methode beinhaltet das Ersetzen der fehlenden Werte durch das arithmetische Mittel der verfügbaren Werte oder alternativ durch den Median oder Modus. Da es sich um ein Verfahren mit wenig Rechenaufwand handelt, ist es einfach in der Durchführung. Zwar findet dabei keine Veränderung des Mittelwertes (oder Modus und Median) statt, jedoch werden die weiteren Verteilungsmaße bei Anwendung des Verfahrens verzerrt, außerdem die Verhältnisse der Variablen verändert. (ebd.: S. 124f.)
- **Regressionsimputation:** Ein Regressionsmodell, das über die vorhandenen Datensätze gebildet wird, dient bei dieser Methode als Basis zur Bestimmung der fehlenden Werte. Die imputierten Werte liegen per Definition auf der Regressionsgeraden. Um diesem Fall entgegenzuwirken, kann den Imputationswerten ein Störterm hinzugerechnet werden. (ebd.: S. 125)

Die oben genannten drei Methoden können bei einem MCAR-Ausfall eingesetzt werden. Jedoch ist Vorsicht angebracht, da der errechnete Standardfehler mit den imputierten Werten gemessen an den wahren Werten unterschätzt wird und weitere Test-Statistiken überschätzt werden. (ebd.: S. 125-128)

- **Predictive Mean Matching:** Dieses Verfahren ist eine Erweiterung der Regressionsimputation. In Anschluss daran wird nach jenem real beobachteten Wert gesucht, der dem errechneten Imputationswert am nächsten kommt. Der gefundene Wert wird dann anstelle des aus dem Regressionsmodell bestimmten Werts für die Imputation benutzt. Damit wird sichergestellt, dass nur real beobachtete Werte im Datensatz vorkommen.

- **Historische Ergänzung und Interpolation:** Dabei handelt es sich um ein Verfahren, das bei Längsschnittuntersuchungen angewendet werden kann. Falls ein fehlender Wert bereits in einer vorhergehenden Untersuchungswelle bestimmt wurde, kann dieser historische Wert für eine Imputation genutzt werden. U.U. wird der historische Wert um einen Trend korrigiert. Bei mehr als zwei verfügbaren Zeitreihen kann ggf. der fehlende Wert über eine lineare Interpolation bestimmt werden. Z.B. wenn bei drei Untersuchungswellen die mittlere Welle einen fehlenden Wert aufweist.

- **Dummy Variable Adjustment:** Bei diesem Verfahren wird ein Regressionsmodell bestimmt. Die fehlenden Werte der unabhängigen Variable des Regressionsmodells werden durch eine Konstante, z.B. Null oder das beobachtete Mittel ersetzt. Eine Dummyvariable des Modells zeigt an, ob ein fehlender Wert vorlag oder nicht. Dieses Verfahren generiert verzerrte Schätzer.

- **Hot-Deck-Verfahren:** Dieses Verfahren existiert in zwei Varianten. Bei der *Random Within Cell* Variante werden sämtliche Datensätze in sinnvolle Imputationsklassen getrennt. Innerhalb einer Klasse werden dann die fehlenden Werte mittels einer Zufallsauswahl durch verfügbare Werte derselben Klasse ersetzt. Die zweite Variante *Nearest Neighbour* ist deterministisch. Die gesamte Datentabelle wird in vollständige und unvollständige Datensätze aufgeteilt. Mittels eines Distanzmaßes wird dann einem unvollständigen Datensatz der ähnlichste vollständige Datensatz zugeordnet und der fehlende Wert durch den des nächsten Nachbars ersetzt. (Toutenburg und Heumann, 2008: S. 287)

- **Cold-Deck-Verfahren:** Dieses Verfahren ähnelt dem Hot-Deck-Verfahren und kann bei Längsschnittuntersuchungen angewendet werden. Dabei kommen jedoch nur Datensätze aus früheren Erhebungen als „nächster Nachbar“ in Frage. (ebd.: S. 287)

- **Doppeln:** Hierbei wird für jeden gelöschten unvollständigen Datensatz zufällig ein vollständiger dupliziert.

Neben den Ad-hoc-Verfahren existieren noch modellbasierte Verfahren. Dabei handelt es sich um Methoden zur Bestimmung von Parametern, wie Mittelwerten, Varianzen, Kovarianzen etc. über Maximum-Likelihood-Modelle (ML-Modell) oder Bayes-Schätzungen. (Toutenburg und Heumann, 2008: S. 288)

Ein Beispiel für ein ML-Modell ist der Expectation-Maximization-Algorithmus (EM-Algorithmus). Der erste (Iterations-)Schritt gleicht dem einer normalen Imputation. Die Parameter und Kovarianzmatrizen werden auf Basis der verfügbaren Daten geschätzt. Mit ihrer Hilfe kann auf die fehlenden Werte geschlossen werden. Danach werden neue Parameterwerte auf Basis der beobachteten und imputierten Werte berechnet („maximization"-Schritt). Mit den neu erhaltenen Werten wird dann im „expectation"-Schritt weitergerechnet. Die Schleife wird solange ausgeführt, bis sich die neuen Parameterwerte nicht mehr ändern. Der Algorithmus ist konvergiert.

Die Bayesianische Datenvermehrung (gehört zu Gruppe der Markov-Chain-Monte-Carlo-Methode) ähnelt dem EM-Algorithmus. Der Unterschied besteht darin, dass bei letzterem mit Zufallswerten gearbeitet wird. Der imputierte Wert wird um einen Zufallsterm erweitert und für die Parameter, wie Mittelwert und Kovarianz, werden je Imputationsschleife zufällig Werte aus einer posteriori Verteilung gezogen. (Nardo u. a., 2008: S. 55f.)

3) Multiple Imputation: Wie bereits angemerkt, findet bei den Methoden der singulären Imputation die Unsicherheit keine Berücksichtigung. Anders ist dies bei der multiplen Imputation. Pro leerem Datenfeld wird nicht ein einzelner Wert, sondern $m > 1$ Werte eingesetzt. Die Anzahl m ist dabei willkürlich, liegt aber in der Praxis zwischen drei und zehn. Durch dieses Vorgehen erhält man m vollständige Datentabellen. Eine Zelle gibt es also in verschiedenen Varianten. Die verschiedenen Datentabellen können dann einzeln mit Standardmethoden analysiert werden. Die resultierenden m Maßzahlen (und Statistiken) werden danach über alle m gemittelt. (Molenberghs u. a., 2014: S. 1307f.)

Die Wahl der Imputationsmethoden ist frei. Grundsätzlich kann eine Vielzahl der oben beschriebenen Methoden angewendet werden. In der Regel werden jedoch Monte-Carlo-Techniken bevorzugt. Bei einem MAR-Ausfall liefert die multiple Imputation meist verlässliche Ergebnisse. (Nardo u. a., 2008: S. 58-62)

4.3.3.2.1 Weitere Verfahren Weitere Verfahren, auf die an dieser Stelle nicht detailliert eingegangen wird:

- Sample-Selection-Modelle, die einen MNAR-Ausfall behandeln sollen. Hintergrund ist, dass der Ausfall über ein Modell erklärt und eingerechnet wird. Das Verfahren ist umstritten.
- Gewichtungsverfahren, bei denen vollständige Datensätze ein stärkeres Gewicht bekommen als unvollständige. Eine Spezialform ist Resampling (z.B. Bootstrapping), bei dem aus allen Beobachtungen zufällig neue gezogen werden. Der Datensatz wird also vergrößert.

4.3.3.2.2 Fazit Die Idee der Imputation als Lösung ist attraktiv und gefährlich zugleich. Sie wiegt den Anwender in der Sicherheit, die Daten seien vollständig. Jedoch existieren Situationen, in denen der Datenausfall nicht ausreichend gering ist, und die Standardschätzer nicht zu erwartungsgemäßen Ergebnissen führen. Das heißt, dass sie bei der Anwendung auf die imputierten Werte nicht dem wahren Wert entsprechen. Es liegt dann eine Verzerrung vor, die im systematischen Fehler des Schätzers ausgedrückt wird.

Generell bietet die Imputation die Möglichkeit den Fehler bei der Bestimmung des Index I_k in Formel 4.1 zu reduzieren. Jedoch ist je nach Umstand auch gegenteiliges möglich.

Die Unsicherheit der imputierten Werte sollten über Varianzschätzer abgebildet werden. Dies ermöglicht die Berücksichtigung der Effekte der Imputation in der Analyse. Die singuläre Imputation ist jedoch bekannt dafür, dass sie die Varianz unterschätzt, weil sie die Unsicherheit der Imputation nur teilweise reflektiert. Anders ist es bei der multiplen Imputation, bei der für einen fehlenden Wert mehrere imputierten Werte existieren. Diese Methode kann die Unsicherheit effektiver reflektieren.

Kein Imputationsmodell ist ohne Annahme. Daher sollten die Ergebnisse der Imputation genauestens auf ihre statistischen Eigenschaften geprüft werden. (Nardo u. a., 2008: S. 24f.)

4.3.4 Prüfen der Datenstruktur

Vor der Konstruktion eines Index sollte die Struktur hinter den Daten analysiert und verstanden werden. Dafür stellt die Statistik eine Reihe von multivariaten Verfahren zur Verfügung. Es kommen z.B. Hauptkomponentenanalyse, Faktorenanalyse, Cronbachs Alpha und die Clusteranalyse in Frage.

Der Analyst prüft und entscheidet mittels der Daten, ob die im theoretischen Schritt (siehe Abschnitt 4.3.1) definierte Struktur tatsächlich zutreffend ist und ob die ausgewählten Indikatoren das zu messende Phänomen numerisch abbilden können.

Die multivariaten Methoden können auf zwei Ebenen angewendet werden: Auf der Ebene der Indikatoren und Variablen sowie auf Ebene der Beobachtungen. Im Folgenden soll ein grober Überblick über die genannten Verfahren gegeben werden. Danach wird ausführlich auf die Hauptkomponenten- und die Faktorenanalyse eingegangen.

4.3.4.1 Ebene der Indikatoren

Bei der Frage, auf welche Weise verschiedene Indikatoren sich in Abhängigkeit zueinander ändern und wie sie zusammenhängen, kann die Hauptkomponentenanalyse angewendet werden. Dafür werden korrelierte Variablen in ein neues Set von unkorrelierten Variablen transformiert. Die Kovarianzmatrix oder die Korrelationsmatrix der Variablen dient als Ausgangspunkt für die Transformation. Die Faktorenanalyse kann diese Frage auch beantworten. Sie ist ähnlich zur Hauptkomponentenanalyse, basiert jedoch auf einem bestimmten statistischen Modell. (ebd.: S. 25f.)

Die Korrelation eines Sets an Indikatoren kann außerdem über die Berechnung von Cronbachs Alpha analysiert werden. Cronbachs Alpha ist ein Schätzer für die interne Konsistenz einer Skala. (Hartig, Frey und Jude, 2008: S. 124f.)

Die Hauptkomponenten- / Faktorenanalyse und Cronbachs Alpha liefern wertvolle Informationen über die Struktur der Daten. Allerdings sollten beide Methoden nicht angewendet werden, wenn die Zahl der Beobachtungen im Vergleich zur Anzahl der Indikatoren klein ist. Die Ergebnisse würden keine bekannten statistischen Eigenschaften besitzen.

4.3.4.2 Ebene der Beobachtungen

Die Clusteranalyse ermöglicht es, viele Beobachtungen in handhabbare Datensets zu klassifizieren. Die Klassen werden dabei auf Basis der Ähnlichkeit zwischen verschiedenen Beobachtungseinheiten ermittelt. Die Clusteranalyse findet NARDO u. a. (2008: S. 26) folgend bei der Bildung von Indizes in diesen Bereichen Einsatz:

- Aggregation (Klassifikation) von Beobachtungen.
- Wirkungsanalyse von theoretischen Entscheidungen.
- Reduktion von Komplexität ohne auf einzelne Indikatoren verzichten zu müssen.
- Festlegung von Imputationsgruppen (unter Reduktion der Varianz der imputierten Werte).

Bei einer großen Zahl an Indikatoren oder wenn einige Indikatoren nicht zur identifizierten Clusterstruktur beitragen, können diskrete und kontinuierliche Modelle angewendet werden. So wird z.B. bei einer „Tandem Analyse“ zuerst eine Hauptkomponenten-/Faktorenanalyse und im zweiten Schritt eine Clusteranalyse durchgeführt. Jedoch ist Vorsicht angebracht, da die Hauptkomponenten-/Faktorenanalyse u.U. die wahre Clusterstruktur verdeckt.

Multivariate Verfahren helfen dabei, die Eignung der Daten zu beurteilen und liefern das Verständnis für die Auswirkungen der methodologischen Entscheidungen bei der Konstruktion des Index. So können z.B. Untergruppen von Indikatoren gefunden werden, die dann über Gewichtung und Aggregation entscheiden.

4.3.4.3 Modellerweiterung

Es existieren verschiedene Methoden, um Hauptkomponenten-/Faktorenanalyse und Clusteranalyse zu verbinden. Bei der „Factorial k-means Analysis“ wird die k-means

Tabelle 4.2: Vor- und Nachteile der Verfahren zur Prüfung der Datenstruktur

Verfahren	Vorteile	Nachteile
Hauptkomponenten-tenanalyse/ Faktorenanalyse	• Zusammenfassung von Indikatoren unter Beibehaltung der größtmöglichen Varianz. • Größte Ladungen werden den Indikatoren mit der größten Varianz in den Originaldaten zugeschrieben (maximales Differenzierungspotential).	• Die Korrelationen entsprechen nicht unbedingt dem wahren Einfluss der Indikatoren auf das zu bestimmende latente „Phänomen“. • Empfindlich bei Änderungen der Basisdaten (Revision, Erweiterung [z.B. Hinzufügen von Beobachtungen] etc.). • Empfindlich bei Ausreißern (diese produzieren eine große Varianz). • Empfindlich bei kleinen Stichproben. • Reduktion des Beitrags von Indikatoren, die wenig mit anderen Indikatoren zusammenhängen.
Cronbachs Alpha	• Messung der internen Konsistenz einer Skala (wie gut beschreiben die Indikatoren das mehrdimensionale Konzept). • Zusammenfassung ähnlicher Objekte.	• Die Korrelationen entsprechen nicht unbedingt dem wahren Einfluss der Indikatoren auf das zu bestimmende latente „Phänomen“. • Nur aussagekräftig, wenn der Index über eine einzelne Skala gemessen wird (z.B. Summe der Indikatoren).
Clusteranalyse	• Alternative Methode, um Beobachtungen zu gruppieren. • Bietet Einblicke in die Datenstruktur.	• Ausschließlich beschreibendes Verfahren. • U.U. intransparent, wenn die methodologischen Entscheidungen nicht klar motiviert und beschrieben wurden.

Quelle: Nardo u. a. (2008: S. 26f.)

Clusteranalyse mit Aspekten der Faktoren- und Hauptkomponentenanalyse verbunden. Ein diskretes Clustermodell wird zusammen mit einem kontinuierlichen Faktorenmodell simultan an die Daten angepasst. Ziel ist die beste Partition von Objekten zu identifizieren, die bzgl. der Methode der kleinsten Quadrate über die besten orthogonalen Linearkombinationen der Variablen (Faktoren) beschrieben werden können. Der Anwendungsbereich dieser Methode ist vielseitig, da er zwei Ziele gleichzeitig verfolgt: die Datenreduktion auf Ebene der Beobachtungen und die Synthese auf Ebene der Indikatoren. Die Methode hat einen schnell alternierenden Kleinste-Quadrate-Algorithmus. Sie dient als Alternative zu der o.g. „Tandem-Analyse“. (Nardo u. a., 2005: S. 34)

4.3.5 Die Aggregation der Daten

4.3.5.1 Normalisierung der Daten

Bevor Indikatoren aggregiert werden können, ist eine Normalisierung notwendig. Zumindest dann, wenn die Indikatoren in verschiedenen Maßeinheiten vorliegen, was die Regel ist. Eine Normierung ist beispielsweise beim Vergleich der Preisvariation von Milch und Autos hilfreich. Milchpreise schwanken im Cent-Bereich, während der Kaufreis für Autos im Mittel im 1000-Euro-Bereich schwankt. Ein weiteres allgemeines Beispiel für die Notwendigkeit einer Normierung ist die Verwendung zweier Indikatoren, bei der ein Indikator über eine Likert-Skala[13] mit fünf Ausprägungen metrisiert wird, der andere mit einer Likert-Skala mit nur vier Ausprägungen. Noch deutlicher wird das Beispiel, wenn ein Indikator über eine metrische und der andere über eine ordinale Skala bestimmt wird (vgl. 2.2.4). Ein Beispiel für die Anwendung einer Normalisierung bietet der *Happy Planet Index* in Abschnitt 5.7.3. Für die Normalisierung bieten sich verschiedene Techniken an:

1. Die Rangbildung ist die einfachste Form der Normalisierung. Die Methode ist robust gegenüber Ausreißern, ermöglicht jedoch nur den relativen Vergleich von

[13] Eine mögliche Likert-Skala mit fünf Elementen: *trifft zu (1), trifft eher zu (2), teils-teils (3), trifft eher nicht zu (4), trifft nicht zu (5).* Für weitere Informationen zur Likert-Skala wird auf Bortz (2006: S. 224) verwiesen.

Merkmalsträgern. Dies liegt daran, dass die absoluten Werte durch die Rangbildung verloren gehen. (Jacobs, Smith und Goddard, 2004: S. 36f.)

$$\text{Indik}_{\text{n}} = Rang(X_{\text{n}}) \tag{4.7}$$

2. Bei der Standardisierung werden die Indikatoren zu einer einheitlichen Skala mit dem Mittelwert 0 und der Standardabweichung 1 transformiert. Dieses Verfahren ist anfällig gegenüber Extremwerten. So haben Indikatoren mit Extremwerten einen größeren Einfluss auf den Index. Ein besonders gutes Ergebnis eines Indikators hätte einen größeren Einfluss als mehrere mit einem durchschnittlichen Wert. Mit dem Streichen des besten und schlechtesten Werts kann dieser Effekt korrigiert werden. Auch in Frage kommt – je nach gewünschtem Einfluss - eine unterschiedliche Gewichtung der Indikatoren. (Kessler, 2007: S. 65f.)

$$\text{Indik}_{\text{n}} = \frac{X_{\text{n}} - \bar{x}}{s} \tag{4.8}$$

3. Die Min-/Max-Normalisierung führt zu einem identischen Wertebereich von [0, 1] bei allen Indikatoren. Dafür wird von dem individuellen Wert der minimale Wert abgezogen und durch den Wertebereich des Indikators geteilt. Das Verfahren ist nicht robust gegenüber Extremwerte. Diese können den transformierten Indikator verzerren. Auf der anderen Seite kann die Min-/Max-Normalisierung den Bereich von Indikatoren eines sehr schmalen Intervalls vergrößern und damit seinen Einfluss auf den Index erhöhen - mehr als das z.B. bei der Standardisierung der Fall ist. (Jacobs, Smith und Goddard, 2004: S. 37f.)

$$\text{Indik}_{\text{n}} = \frac{X_{\text{n}} - min(X)}{max(X) - min(X)} \tag{4.9}$$

4. Der Abstand zu einem Referenzwert misst die relative Position eines Indikators zu einem Referenzpunkt. Ein Referenzpunkt ist z.B. ein spezieller Zielwert in einem bestimmten Zeitintervall.[14] Alternativ kann der Referenzwert auch der Durchschnitt aller Werte sein und den Wert 1 zugeschrieben bekommen. Die anderen Werte würden dann im Abstand zum Durchschnitt betrachtet werden. Normalisierte Werte größer als 1 liegen dann z.B. über dem Durchschnitt. Optional kann

[14] Ein Beispiel ist die Aufnahme politischer Ziele. Für weiterführende Informationen vgl. Abschnitt 2.3.1.

der normalisierte Wert mit 100 multipliziert werden. (Jacobs, Smith und Goddard, 2004: S. 38)

$$\text{Indik}_\text{n} = \frac{X_\text{n}}{X_\text{ref}} \tag{4.10}$$

5. Die prozentuale Abweichung zum Bestwert ist eine leichte Abwandlung des *Abstandes zu einem Referenzwert*. Anstelle des Referenzwertes wird der Maximalwert eingesetzt und der Quotient mit 100 multipliziert. Inhaltlich bedeuten die errechneten Werte, wie weit ein Wert n eines Indikators prozentual vom Bestwert des Indikators entfernt ist.

$$\text{Indik}_\text{n} = (\frac{X_\text{n}}{max(X)}) * 100 \tag{4.11}$$

6. Bei der kategorialen Skalierung wird jedem Indikatorwert ein Scorewert zugeschrieben. Die Kategorien können dabei numerisch (1,2,3, ...) oder qualitativ („sehr gut“, „gut“, ...) sein. Oftmals werden die Scores auf Basis der Perzentile der Verteilung der Indikatorwerte vergeben. Z.B. könnten die 10% besten Werte einen Score von 100 bekommen, Werte zwischen dem 80%- und 90%-Perzentil den Score 90 usw. Problematisch dabei ist, dass ein Großteil der Verteilungsinformation über die Kategorisierung verloren geht. Außerdem kann es u.U. sein, dass bei geringer Varianz innerhalb der Werte die Kategorisierung unabhängig von der Verteilung erzwungen wird. Eine Möglichkeit, dieses Problem zu beheben, besteht darin, die Grenzen der Kategorien abhängig von der Verteilung der Ursprungswerte zu bestimmen. (Nardo u. a., 2005: S. 49f.)

   Beispiel:

$$\text{Indik}_\text{n} = \begin{cases} 0 & \text{falls } X_\text{n} < \text{P}^{33} \\ 50 & \text{falls } X_\text{n} \leq \text{P}^{66} \\ 100 & \text{falls } \text{P}^{66} \leq X_\text{n} \end{cases} \tag{4.12}$$

7. Indikatoren unter und über dem Durchschnitt werden derart transformiert, dass Werte um den Mittelwert die 0 erhalten sowie Werte unter und über dem Durchschnitt 1 und -1. Diese Art der Methode ist einfach und robust gegenüber Ausreißern. Jedoch werden die Willkürlichkeit der Schwellenwerte und das Fehlen der

Information zu den absoluten Werten kritisiert. Wenn Wert A 10% und Wert B 400% über dem Durchschnitt liegen, erscheinen beide unter dem Wert 1 = über dem Durchschnitt. (Nardo u. a., 2008: S. 86)

$$\text{Indik}_\text{n} = \begin{cases} 1 & \text{falls } \frac{X_\text{n}}{\bar{x}} > (1+p) \\ 0 & \text{falls } (1-p) \leq \frac{X_\text{n}}{\bar{x}} \leq (1+p) \\ -1 & \text{falls } \frac{X_\text{n}}{\bar{x}} < (1-p) \end{cases} \tag{4.13}$$

p... ist ein willkürlicher Wert um den Mittelwert

Die Wahl einer passenden Normalisierung muss sorgfältig getroffen werden. Ggf. muss die Anpassung der Skala einbezogen und auf besonders „schiefe“ Verteilungen geachtet werden. Bei der Wahl der Normalisierungsmethode sind sowohl die Eigenschaften der Daten einzubeziehen als auch die Ziele des Index. Über Robustheitstests kann der Einfluss auf die Ergebnisse geprüft werden.

4.3.5.2 Aggregationsform

Die Gewichtung und Aggregationsform zum Zusammenfassen der verschiedenen Indikatoren und Merkmale kann prinzipiell auf zwei verschiedene Arten durchgeführt werden: *formalanalytisch* oder *normativ* (Werner, 1975a: S. 160). Im Wesentlichen existieren vier Gewichtungsarten für diesen Schritt (Schnell, 2005: S. 161ff.).

Über Benchmarks kann gezeigt werden, dass die Gewichtung einen maßgeblichen Einfluss auf den Index und z.B. die darin enthaltenen Rangreihenfolge der Merkmalsträger hat (Nardo u. a., 2008: S. 31). Bevor näher auf die Festlegung der Gewichtungsfaktoren eingegangen wird, werden im Folgenden die vier verschiedenen Methoden der Zusammenfassung vorgestellt:

- **Bei additiven Indizes** werden die einzelnen Merkmale additiv verknüpft. Bei dieser Variante geht man davon aus, dass die verschiedenen Merkmale voneinander unabhängig auf die Zieldimension wirken. Die Indikatoren können sich gegenseitig kompensieren. Ein besonders hoher Wert einer Dimension kann einen besonders niedrigen einer anderen ausgleichen. Diese Methode bietet sich an, wenn alle Indikatoren die gleiche Maßeinheit verwenden. (Schnell, 2005: S. 163)

- **Bei multiplikativen Indizes** werden die einzelnen Merkmale multiplikativ verknüpft. Falls einer der Merkmale den Wert 0 besitzt, kann keiner der anderen Indikatoren diesen Wert kompensieren. Der Wert der Zieldimension wird ebenfalls 0 sein. Diese Methode bietet sich an, wenn die Indikatoren sich zu einem bestimmten Maß nicht kompensieren dürfen. Bei der geometrischen Aggregation werden höhere Indikatorwerte mit einem höheren Indexwert belohnt. (Schnell, 2005: S. 163)

- **Gewichtete additive Indizes** werden auf Basis der unterschiedlichen Wichtigkeit der Indikatoren in Bezug auf die Zieldimension gebildet. Dabei werden die Merkmale mit dem jeweiligen Gewicht multipliziert und danach additiv zur Zieldimension verknüpft. Die Indikatoren gehen damit proportional zu ihren Gewichten in den Index ein. Die Gewichte können dabei auf verschiedene Arten gewonnen werden. (ebd.: S. 164)

Lineare und geometrische Aggregation, d.h. (gewichtete) additive und multiplikative Indizes, ermöglichen bis zu einem gewissen Grad, dass sich Indikatoren gegenseitig kompensieren. Ein schlechter Wert in einem Indikator kann durch einen guten Wert eines anderen Indikators ausgeglichen werden. Dieses Verhalten widerspricht u.U. der Motivation hinter den Gewichten, die der Wichtigkeit des Indikators entsprechen. In der linearen Aggregation ist der Kompensationseffekt konstant. In der geometrischen ist er für geringe Werte jedoch geringer. Dies führt zu einigen Folgeüberlegungen, wie Indexwerte am leichtesten verändert werden können.

Bei einem geringen Wert eines bestimmten Indikators wird bei der geometrischen Aggregation ein deutlich höherer Wert eines anderen Indikators benötigt, um den Indexwert zu verbessern. Beim Vergleich der linearen und geometrischen Aggregation schneiden Beobachtungseinheiten mit geringen Werten in der linearen Aggregation besser ab als bei der geometrischen. Auf der anderen Seite haben unter der geometrischen Aggregation kleine Änderungen bei niedrigen Werten einen viel größeren Einfluss auf den Index als bei hohen Werten. Ein Merkmalsträger mit einem niedrigen Indexwert wird sich also auf die Verbesserung von Bereichen oder Sektoren konzentrieren, die niedrige Werte aufweisen, da dadurch eine größere Verbesserung im Index zu erwarten ist.

Um sicherzustellen, dass die Gewichtung als Maß für die Wichtigkeit auch bei der Aggregation erhalten bleibt, können Techniken angewendet werden, die das gegenseitige Kompensieren von Indikatoren verbieten:

- **Bei der Mehrkriterien-Aggregation** wird eine nicht-kompensatorische Aggregationslogik festgelegt. Damit können sich Indikatoren nicht gegenseitig kompensieren. Diese Methode wird gewöhnlich bei der Zusammenfassung von sehr unterschiedlichen Dimensionen benutzt, z.B. bei der Aggregation von physikalischen, sozialen und ökonomischen Daten. Falls der theoretische Rahmen bei diesen drei Dimensionen verbietet, dass z.B. die Zunahme der ökonomischen Leistung nicht die soziale Zersplitterung ausgleichen können soll, ist weder lineare noch geometrische Aggregation passend. Der Ansatz über mehrere Kriterien bei der Aggregation kann dieses Problem lösen, z.B. indem die Aggregation über verschiedene Zielwerte durchgeführt wird. Die Indexwerte wären dann ordinal skaliert. (Nardo u. a., 2005: S. 76f.)

4.3.5.3 Gewichtung

Viele Indizes basieren auf einer Gleichgewichtung. Das bedeutet jedoch nicht, dass keine Gewichtung angewendet wurde, sondern vielmehr, dass alle Indikatoren das gleiche Gewicht besitzen. Inhaltlich wird also davon ausgegangen, dass alle Indikatoren den gleichen Beitrag für den Index leisten. Bei der Nutzung von Subindizes ist darauf zu achten, dass bei einer unterschiedlichen Anzahl von Indikatoren je Subindex keine Gleichgewichtung mehr vorliegt. Bei der weiteren Aggregation der Subindizes zu dem Gesamtindex erhalten Indikatoren von Subindizes mit mehr Indikatoren ein geringeres Gewicht als Indikatoren von Subindizes mit wenigen Indikatoren. Eine nicht balancierte Indexstruktur ist die Folge.

Eine Korrelation zwischen verschiedenen Indikatoren stellt ein weiteres Problem für die Gleichgewichtung dar. Existiert zwischen zwei Indikatoren eine hohe Korrelation ist das ggf. ein Anzeichen, dass sie dasselbe „Phänomen“ bestimmen. Gehen die beiden Indikatoren dann zusammen mit allen anderen Indikatoren gleichgewichtet in den Index ein, wird das durch die zwei Indikatoren bestimmte „Phänomen“ mit doppeltem Gewicht in den Index eingehen. Die Korrelation zwischen den Indikatoren kann, je nach Skalenniveau, z.B. über den Korrelationskoeffizient nach Pearson oder Spearman berechnet

werden.[15] Lösungen sind der Ausschluss oder eine geringere Gewichtung hoch korrelierter Indikatoren.

Die Entscheidung über den Umgang mit korrelierten Indikatoren sollte jedoch nicht ausschließlich statistisch, sondern auch inhaltlich getroffen werden. Eine geringe Korrelation zwischen Indikatoren ist normal und muss nicht zwangsläufig für eine Doppelzählung oder Redundanz sprechen.

Bei der Betrachtung von Zeitreihen ist es vorteilhaft, wenn sich die Gewichte über die Zeit nicht ändern. Dies erhöht Transparenz, Vergleichbarkeit und Objektivität. Nichtsdestotrotz existieren auch Indizes, bei denen die Gewichte von Welle zu Welle angepasst werden, z.B. wenn der Indexwert Prioritäten oder Ziele enthält. (Jacobs, Smith und Goddard, 2004: S. 44-48)

Für die Berechnung der Gewichte je nach Aggregationsform kommen verschiedene Methoden in Frage:

- Gleichgewichtung
- Expertenmeinung (4.3.5.3.7)
- Hauptkomponenten-/Faktorenanalyse (4.3.5.3.1)
- Benefit of the doubt (4.3.5.3.2)
- Unobserved components models (4.3.5.3.3)
- Conjoint analysis (4.3.5.3.4)
- Analytic hierarchy processes (4.3.5.3.5)
- Regressionsansatz (4.3.5.3.6)
- Budget allocation processes (4.3.5.3.7)
- Öffentliche Meinung (4.3.5.3.8)
- Entfernung zum Zielwert (4.3.5.3.9)

[15] Für weiterführende Informationen zu den genannten Zusammenhangmaßen sei auf Fahrmeir (1997: S. 135ff.) verwiesen.

Die Gewichte der Indikatoren müssen nicht zwangsläufig über statistische Methoden hergeleitet werden. Nicht selten werden Experten eingesetzt, um die Gewichte inhaltlich oder heuristisch festzulegen. Derartige Gewichte sind z.T. leichter kommunizier- und erklärbar (vgl. Abschnitt 4.3.5.3.7).

Im Folgenden wird auf die verbreitete Form der gewichteten additiven Indizes näher eingegangen.

Methoden für die Gewichtung additiver Indizes

Die Festlegung der Gewichtung additiver Indizes kann über verschiedene Methoden durchgeführt werden. Eine Methode basiert auf Expertenwissen. Dabei wird die Gewichtung aufgrund von theoretischen Überlegungen oder durch Heuristiken vorgenommen. Bei dieser Methode sollte die Anzahl der Indikatoren eher gering sein, da dieses Vorgehen einen hohen kognitiven Aufwand bedeutet.

Ein anderes Verfahren basiert auf einem multiplen Regressionsmodell. Hierbei wird die Zieldimension über eine Expertenschätzung gebildet. Die Gewichte der Indikatoren werden dann in Abhängigkeit dieser Zielvariablen bestimmt. Für dieses Verfahren ist damit eine externe Information notwendig, die z.T. subjektiv sind. Die Faktorenanalyse hingegen ist ein Verfahren, das keine externen Informationen voraussetzt. Sie benötigt jedoch korrelierende Indikatoren. Die Gewichte der Faktorenanalyse stellen die Stärke des Zusammenhangs zu der Zielgröße dar.

Wirkungsanalyse

Die Wirkung der verschiedenen Gewichtungsarten und Methoden kann u.a. über die Berechnung der durchschnittlichen absoluten Differenzen der Rangzuweisung, basierend auf den Indexwerten zu den Merkmalsträgern, bestimmt werden. Damit kann z.B. die Sensitivität der unterschiedlichen Methoden quantifiziert werden. Dazu werden die Merkmalsträger jeweils für die Referenzmethode $Rank_{\text{ref}}(I_{\mathbf{c}})$ und die Alternativmethode $Rank(I_{\mathbf{c}})$ in eine Rangfolge c = 1,..., M gebracht und verglichen.

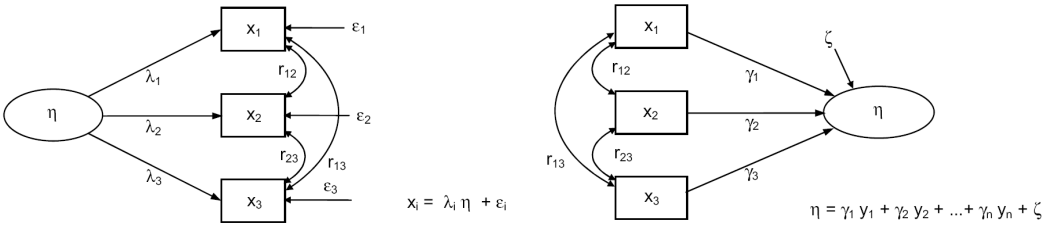
Zusammenhang zwischen Indikatoren und Indexkonstrukt

Zu Beginn der Analyse sollte die Beziehung der Indikatoren zum Indexkonzept geklärt werden. Dabei existieren zwei Perspektiven des Zusammenhangs. Sie unterscheiden sich

Abbildung 4.2: Zusammenhang zwischen Indikatoren und latentem Konstrukt

(a) Latentes Konstrukt mit reflektiven Indikatoren

(b) Latentes Konstrukt mit formativen Indikatoren



(a) η: latente Variable; λ: Faktorladung; x: reflektiver Indikator; ε: Messfehler auf Indikatorebene; τ: Korrelation zwischen den Indikatoren
(b) η: latente Variable; λ: Gewicht; y: formativer Indikator; ζ: Messfehler auf Ebene der latenten Variable τ: Korrelation zwischen den Indikatoren

Quelle: Christophersen und Grape (2006)

bezüglich ihrer Kausalitätsrichtung: Wirken die Indikatoren auf den Index oder wirkt der Index auf die Indikatoren? (Schlittgen, 2009: S. 447)

Bei der Annahme, dass die Zielgröße unabhängig von seinen Indikatoren existiert und sich die Zielgröße lediglich in den Indikatoren manifestiert, wird von einem reflexiven Modell ausgegangen. Die Kausalitätsrichtung wirkt also vom latenten Konstrukt zu den Indikatoren. Die latente Variable erklärt dabei die beobachteten Variablen. Ändert sich z.B. das Konstrukt, wird diese Änderung in allen beobachteten Variablen sichtbar. Das reflexive Modell geht davon aus, dass das latente Konstrukt „real“ und unabhängig von der „Messung“ existiert. Die Indikatoren in diesem Modell sind grundsätzlich austauschbar, da das Hinzufügen oder Eliminieren eines Indikators keine Auswirkung auf den Inhalt des latenten Konstrukts hat. (ebd.: S. 261f.)

Ergibt sich das Konstrukt jedoch erst durch die Indikatoren, spricht man von einem formativen Modell. Die Kausalitätsrichtung verläuft von den Indikatoren zu dem Konstrukt. Das Konstrukt wird also durch seine Indikatoren vollständig determiniert. Eine Änderung der Indikatoren führt in der Regel zu einer Änderung des Konstrukts, wobei sich Änderungen der Indikatoren auch gegenseitig kompensieren können. Die Indikatoren sind in dem formativen Modell, anders als im reflexiven, nicht austauschbar, da andernfalls das Konstrukt verändert würde. (ebd.: S. 447f.)

4.3.5.3.1 Faktorenanalyse Der Begriff Faktorenanalyse bezeichnet nicht ein bestimmtes statistisches Verfahren, sondern ist ein Sammelbegriff für viele, z.T. sehr verschiedene Techniken (Brachinger und Ost, 1996: S. 639). Dabei handelt es sich um multivariate Verfahren, die zur Analyse der Beziehung von quantitativen Variablen eingesetzt werden. Der Vorteil der Faktorenanalyse ist dabei, dass sie die Beziehung eines ganzen Merkmalskomplexes ganzheitlich beschreiben kann (Weber, 1974: S. 16). Damit ist sie z.B. einzelnen Korrelationskoeffizienten überlegen, die nur die Beziehung von zwei Variablen beschreiben. Die Faktorenanalyse teilt sich in die zwei Bereiche: die explorative und konfirmatorische Faktorenanalyse. Die explorative Faktorenanalyse stellt ein modell- und hypothesengenerierendes Verfahren dar. Es geht also um das Aufdecken von Strukturen und Zusammenhängen zwischen den Variablen. Dabei kann über die korrelative Beziehung der Variablen abgeleitet werden, ob sich verschiedene Variablen zu wenigen Faktoren zusammenfassen lassen.

Im Gegensatz dazu stellt die konfirmatorische Faktorenanalyse ein strukturprüfendes Verfahren dar. Hier wird z.B. auf Basis theoretischer Überlegungen von einem bestimmten Zusammenhang zwischen den beobachteten Variablen und den dahinterstehenden Faktoren ausgegangen.

Explorative Faktorenanalyse

Die Techniken der Faktorenanalyse decken Strukturen in den Daten auf. Über die Identifikation von Faktoren wird im Idealfall eine Datenreduktion erzielt. Zu den Techniken zählen u.a. die Hauptkomponentenanalyse, Hauptfaktorenanalyse, Image-Analyse, Maximum-Likelihood-Faktorenanalyse, Kanonische Faktorenanalyse, Alpha-Faktorenanalyse und die Maximum-Determinanten-Lösung. Im Lösungsalgorithmus zur Bestimmung der Ladungsmatrix liegt ihr Unterschied. Dieser beschreibt den Zusammenhang zwischen den Indikatoren und den dahinterstehenden Faktoren. Die verschiedenen Techniken lassen sich in zwei Gruppen teilen und unterscheiden sich hinsichtlich ihrer Interpretation über die o.g. Wirkungsrichtung:

- **Bei der Hauptkomponentenanalyse** geht es im Wesentlichen um das Zusammenfassen von Daten. Die Zielsetzung ist Datenreduktion. Dafür wird nach wenigen sogenannten Hauptkomponenten gesucht, die es ermöglichen, die Varianz der verschiedenen Indikatoren hinreichend zu approximieren. Damit stehen bei der

Hauptkomponentenanalyse die Varianzaspekte im Vordergrund. Der Zusammenhang zwischen einer Hauptkomponente und den Indikatoren wird über Ladungen beschrieben. Die Hauptkomponenten werden über diese Ladungen interpretiert, da sie quasi einen Sammelbegriff für die ihnen zugeordneten hoch ladenden Indikatoren darstellen. In der Regel werden mindestens zwei Indikatoren für die Bildung einer Hauptkomponente einbezogen. Grundsätzlich geht die Hauptkomponentenanalyse nicht von gemeinsamen und spezifischen latenten Faktoren aus. Es ist eher davon auszugehen, dass die Indikatoren auf die Hauptkomponente wirken. Daher wird die Hauptkomponentenanalyse oftmals auch nicht unter die Verfahren der Faktorenanalyse eingereiht, sondern gesondert beschrieben. Die Hauptkomponentenanalyse stellt im Wesentlichen ein rein mathematisches Verfahren dar. (vgl. Kessler (2007: S. 22-37))

- **Bei der (Haupt-) Faktorenanalyse** steht vor allem die Kovarianz- und die Korrelationsstruktur der Indikatoren im Vordergrund. Unterstellt wird dabei eine kleine Anzahl nicht beobachtbarer (latenter) Variablen, von denen angenommen wird, dass sie die Kovarianz- und Korrelationsstruktur der Indikatoren erklären. Sie werden als gemeinsame Faktoren bezeichnet und stellen quasi die Ursache für hohe Ladungen dar. (Härdle, 2007: S. 185ff.)

Die Faktorenanalyse ermöglicht neben der Reduktion von Daten (durch die Identifikation von Faktoren oder Hauptkomponenten) auch die Reduktion der Datenerhebung. Ist z.B. die Menge an Indikatoren zu groß und soll reduziert werden, kann über die Höhe der Ladungen eine Reduktionsregel festgelegt werden. Diese kann darin bestehen, dass Indikatoren ausgeschlossen werden, deren Exklusion keinen wesentlichen Einfluss auf den Faktor hat. Es werden also nur die Indikatoren in die Bildung des Faktors einbezogen, die die Variation aller beobachteten Variablen hinreichend abbilden.

4.3.5.3.2 Benefit-of-the-doubt-Gewichtung Bei der Konstruktion eines Index sieht sich der Konstrukteur mehreren Fragen gegenüber, die die Ergebnisse des Index maßgeblich beeinflussen: Welche Indikatoren werden zur Bestimmung des Konzepts benutzt; wie werden die Daten normalisiert und aggregiert; wie wird mit fehlenden Daten oder schlechter Datenqualität umgegangen? Die Beantwortung dieser Fragen beeinflusst jeweils das Ranking der Indexwerte und verursacht somit eine mögliche Angriffsfläche

und die Frage, warum sich gerade dafür entschieden wurde. Bei einem Ländervergleich liegt es als Land mit schlechten Ergebnissen z.B. nahe, die Gewichtung des Index zu kritisieren. Insbesondere dann, wenn dieses Land unter einer anderen Gewichtung besser abgeschnitten hätte. (Nardo u. a., 2008: S. 92-94)

Ausgangspunkt

Ausgangspunkt der Benefit-of-the-doubt-Gewichtung ist die kontroverse Diskussion über die Gewichtung von Indikatoren bei der Zusammenführung zu einem Index. In der Praxis ist es oft nicht zweifelsfrei möglich, eine bestimmte Gewichtung zu begründen und wissenschaftlich zu motivieren. In solch einem Fall ist damit zu rechnen, dass sich z.B. Länder im unteren Bereich eines Indexrankings kritisch über die Gewichtung äußern und an der Glaubwürdigkeit der Werte zweifeln.

Insbesondere beim Vergleich von z.B. politischen Maßnahmen einzelner Länder kann nicht immer unterstellt werden, dass die einzelnen Maßnahmen in allen Ländern das gleiche Gewicht besitzen. Ein Land könnte etwa einen ganz speziellen Wirtschaftsbereich fördern, der im Gesamtindex nur mit einem geringen Gewicht Berücksichtigung findet. Ist es gerechtfertigt, dass dieses Land schlechter rankt als andere, nur weil es andere Prioritäten setzt?

Dieser Zweifel soll durch die Benefit-of-the-doub-Gewichtung beseitigt werden. Dafür erhält jedes Land eine individuelle Gewichtung. Und zwar die beste Gewichtung, die es für seine spezifischen Indikatorwerte gibt. Der daraus berechnete Score wird dann ins Verhältnis zu dem Score des besten Landes der Studie gesetzt und daraus der Indexwert berechnet. Jede andere Gewichtung würde also zu einer Verschlechterung des Ranges führen. (Cherchye u. a., 2007: S. 1-3)

Berechnung der individuellen Gewichtung

Bei der Gewichtung handelt es sich um eine gewichtete additive Zusammenfassung:

$$I_{\mathrm{c}} = \sum_{i=1}^{m} w_{\mathrm{c,i}} * y^{n}_{\mathrm{c,i}} \tag{4.14}$$

I_{c} . . . Index für Kategorie c

$w_{\mathrm{c,i}}$ . . . Gewicht für Kategorie c und Indikator i (i=1,. . . ,m)

$y^{n}_{\mathrm{c,i}}$ . . . (ggf. nomalisierter) Wert für Kategorie c und Indikator i (i=1,. . . ,m)

Es werden also alle Indikatoren zu dem Index zusammengefasst. Dabei hat jede Kategorie (z.B. jedes Land) und jeder Indikator eine individuelle Gewichtung.

Der Indexwert selbst besteht dann in einer Art Benchmark. Der Wert der Kategorie wird ins Verhältnis zu dem Wert der besten Kategorie unter derselben Gewichtung w gesetzt:

$$I_c = \frac{\text{Performance Wert}}{\text{Benchmark Wert}} = \frac{\sum_{i=1}^{m} w_{\text{c,i}} * y_{\text{c,i}}}{\sum_{i=1}^{m} w_{\text{c,i}} * y_{\text{i}}^{\text{B}}} \tag{4.15}$$

Danach gilt es das Problem zu lösen, wie die individuelle Gewichtung je Kategorie und Indikator bestimmt werden kann, um sowohl den *Performance Wert* als auch den *Benchmark Wert* zu berechnen. Dazu sollte erneut das Ziel der Benefit-of-the-doubt-Gewichtung einbezogen werden: Es geht darum, jeder Kategorie die Gewichtung zuzuschreiben, mit der es den bestmöglichen Indexwert erreichen kann. Es handelt sich also um ein Maximierungsproblem:

$$I_c = \max_{w_{\text{c,i}}} \frac{\sum_{i=1}^{m} w_{\text{c,i}} * y_{\text{c,i}}}{y_{\text{j,i}} \in \max_{\text{Länder der Studie}} \sum_{i=1}^{m} w_{\text{c,i}} * y_{\text{j,i}}} \tag{4.16}$$

Es werden folglich die Gewichte für jeden einzelnen Indikator so bestimmt, dass eine spezifische Kategorie im Vergleich zur besten Kategorie unter dieser Gewichtung den bestmöglichen Wert erhält. Diese Maximierungsrechnung wird für jedes Element (z.B. Land) individuell durchgeführt.

Damit ist die Festlegung der Gewichtung über jeden Zweifel erhaben. Über die Maximierung wird eine Gewichtung festgelegt, die jedes Land die bestmögliche Position im Ranking zusichert. Jede andere Gewichtung würde es absteigen lassen.

Über die Benefit-of-the-doubt-Gewichtung wird die Anforderung konstanter Gewichte fallen gelassen und jedem Land oder jeder Kategorie eine individuelle Gewichtung ermöglicht. Daraus entstehende Probleme können über Restriktionen bei der Maximierung eingeschränkt werden. (Cherchye u. a., 2007: S. 4-7)

Definition von Maximierungs-Restriktionen

Durch das oben dargestellte Maximierungsverfahren können unerwünschte Ergebnisse auftreten. So ist denkbar, dass verschiedene Indikatoren für eine spezifische Kategorie ein Gewicht von Null erhalten und keinen Einfluss auf den Index besitzen. Ebenfalls ist – bei der Betrachtung von Ländern als Kategorie – denkbar, dass sich ein spezifisches

Abbildung 4.3: Restriktionsarten

Absolute Sub-indicator share restrictions
$\alpha_i \leq w_{j,i} y_{j,i} \leq \beta_i$
Ordinal Sub-indicator share restrictions
$w_{j,6}\, y_{j,6} \leq w_{j,5}\, y_{j,5} \leq w_{j,2}\, y_{j,2} \leq w_{j,3}\, y_{j,3} \leq w_{j,1}\, y_{j,1} \leq w_{j,7}\, y_{j,7} \leq w_{j,4}\, y_{j,4}$
Relative Sub-indicator share restrictions
$\alpha_i \leq \frac{w_{j,i}\, y_{j,i}}{w_{j,k}\, y_{j,k}} \leq \beta_i$
Proportional Sub-indicator share restrictions
$\alpha_i \leq \frac{w_i y_{j,i}}{\sum_{i=1}^{m} w_i y_{j,i}} \leq \beta_i$
Restrictions pertaining to category shares
$\alpha \leq \frac{\sum_{i \in Sa} w_{j,i}\, y_{j,i}}{\sum_{i=1}^{m} w_{j,i}\, y_{j,i}} \leq \beta$

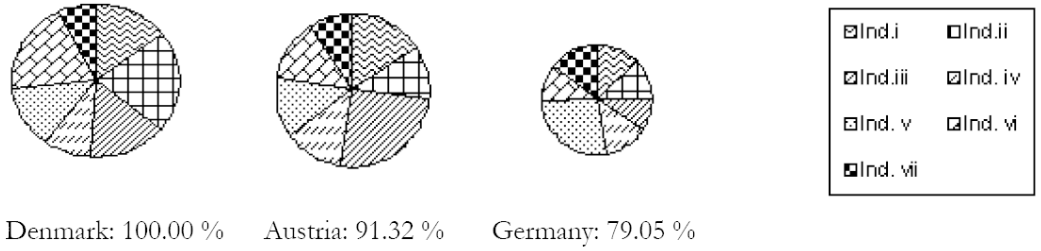
Quelle: Cherchye u. a. (2007)

Land gezielt auf einen Indikator konzertiert, der bei allen anderen Ländern schwach ausgeprägt ist. Ein gutes Ranking in der Tabelle und sehr ungleiche Gewichte wären die Folge. Um dieses Verhalten zu verhindern, können Einschränkungen bei der o.g. Maximierung getroffen werden. Dabei ist davon auszugehen, dass es Experten leichter fällt, grundsätzliche Einschränkungen zu definieren, als z.B. Entscheidungen auf Ebene der individuellen Gewichte zu treffen.

Ein Ansatz für die Definition von Einschränkungen ist das Festlegen des minimalen oder maximalen Beitrags eines „Sub-Indikatoranteils“ zum Gesamtindex. Dieser Sub-Indikatoranteil ist das Produkt des ursprünglichen Indikatorwerts $y_{c,i}$ und dem ihm zugeschriebenen Gewicht $w_{c,i}$. Der daraus resultierende Indexwert kann dann z.B. als die Summe der einzelnen Sub-Indikatoranteile interpretiert werden. Restriktionen auf dieser Ebene haben den Vorteil, dass sie unabhängig von der Maßeinheit sind. Auch sind die Einschränkungen leichter interpretierbar als z.B. Einschränkungen auf Ebene der individuellen Gewichte. (Cherchye u. a., 2007: S. 8-11)

Eine Aufstellung der wichtigsten Restriktionsarten findet sich in Abbildung 4.3. Die letzten beiden Restriktionen sind sogenannte *Pie-Charts Restriktionen*. Sie beziehen sich auf den Beitrag einen Sub-Indikatoranteils zum Gesamtindexwert. Bei der Festlegung der Restriktionen können die Experten also mit einfachen Prozentangaben arbeiten.

Abbildung 4.4: Darstellung der Anteile



Quelle: Cherchye u. a. (2007)

Soll ein Indikator am Indexwert einen Beitrag von 10 bis 15 Prozent besitzen, ist bei der *Proportional Sub-indicator share restriction* z.B. $= 10$ und $= 15$. Die Experten können damit auch nach einer Budget-Logik vorgehen, indem sie von 100% ausgehen und jeder Dimension einen Anteil zuschreiben. Die o.g. Maximierungsgleichung muss dann mit den bestimmten Restriktionen erweitert werden. Eine mögliche Darstellung der Anteile zeigt Abbildung 4.4. Dabei stellen die schraffierten Flächen den zugeordneten Beitrag und die Größe der Torte den absoluten Indexwert in Prozent dar.

Hintergrund und Bezug zur Datenhüllenanalyse

Die Datenhüllenanalyse (auch Data Envelopment Analyse (DEA) genannt) stellt den Ausgangspunkt für die Benefit-of-the-doubt Analyse dar. Es handelt sich bei der DEA um eine Technik zur Effizienz-Analyse aus dem Bereich des *Operations Research*. Ziel ist die vergleichende Messung der Effizienz von Organisationseinheiten und Entscheidungseinheiten. Die Effizienz der Entscheidungseinheiten wird relativ zueinander bestimmt. Entscheidungseinheiten können verschiedenste Objekte sein, die durch Inputs wie z.B. Kosten, Arbeitsaufwand und Outputs wie z.B. Umsatz charakterisiert werden. Entscheidungseinheiten können z.B. Universitäten oder Bankfilialen sein. Schnell wurde die Ähnlichkeit der DEA-Problemstellung zu der der Indexkonstruktion erkannt und das Konzept unter dem Begriff der Benefit-of-doubt-Analyse auf Indizes erweitert.

Die DEA wird auf Charnes, Cooper und Rhodes (1978) zurückgeführt. Die ursprüngliche Fragestellung der DEA untersucht, wie die Effizienz der Entscheidungseinheiten bestimmt werden kann, wenn die Anzahl von Inputs und Outputs beobachtet wurde, jedoch keine zuverlässigen Informationen zum Preis verfügbar sind und niemand über Informationen zur Produktions- oder Kostenfunktion verfügt. Mit Hilfe der DEA wird

Tabelle 4.3: Vor- und Nachteile der Benefit-of-the-doubt-Gewichtung

Vorteile:	**Nachteile:**
• Sensibel gegenüber individuellen Präferenzen, da die Gewichte individuell über die beobachtete Informationen definiert sind. • Eine Rangliste auf Basis der Indexwerte ist unabhängig von theoretischen Annahmen bzgl. der Gewichte. • Kein Diskussionspotential zu Benachteiligungen von Populationen bei der Festlegung der Gewichtung.	• Gewichte werden spezifisch für jede Population berechnet. Damit ist kein Vergleich zwischen den Populationen möglich. • Ggf. sind Einschränkungen hinsichtlich der Gewichte notwendig, wenn alle Populationen einen Indexwert von 1 erhalten. In diesem Fall besteht die Möglichkeit, dass nicht immer Lösungen zur Maximierung existieren. • Ggf. verschiedene Gewichtungsschemata je Population. • Gewichtung widerspricht ggf. Expertenmeinungen. • Der Merkmalsträger mit der besten Leistung wird seine Fortschritte nicht im Index wiederfinden.

Quelle: In Anlehnung an Nardo u. a. (2005: S. 62)

die relative Effizienz der Entscheidungseinheiten gemessen, da diese innerhalb der Gruppe als Vergleichsmaßstab dienen. Eine Produktions- oder Kostenfunktion ist nicht notwendig, jedoch müssen die Entscheidungseinheiten die gleichen oder zumindest ähnliche Inputs und Outputs besitzen.

Hinsichtlich der Indizes kann das DEA Problem derart umgedeutet werden, dass z.B. die Bedeutung einer politischen Maßnahme für ein Land unbekannt ist. Wird diese Maßnahme über einen Indikator gemessen, ist ihr Gewicht bei der Verdichtung zu einem Index unbekannt. Über den Effizienzansatz der DEA kann die Effizienz der Maßnahme jedoch im Bezug zu einem anderen Land bestimmt werden. (Cherchye u. a., 2007: S. 4-7)

4.3.5.3.3 Unobserved components models (UCM) Diese Art von Modellen kann verwendet werden, um die Gewichtung von Indikatoren festzulegen. Die Gewichte werden in den UC-Modellen über eine Maximum-Likelihood-Schätzung bestimmt. Die Idee dahinter ist, dass die Indikatoren von einer unbeobachteten Variablen und einem Fehlerterm abhängen. Über die Schätzung der unbeobachteten Komponente kann die Beziehung zwischen dem Index und seinen Komponenten beleuchtet werden. Die aus dem

Modell heraus bestimmten Gewichte minimieren den Fehler des Index. Das Vorgehen ähnelt einer Regression.

Vorgehen

Für die Anwendung des UC-Modells ist je nach Spezifikation eine Mindestanzahl an Indikatoren notwendig. Bei der Berechnung eines Index, der z.B. die Innovation eines Landes misst, sind mindestens drei Indikatoren notwendig. Andernfalls ist das Modell nicht exakt identifiziert.

Im Folgenden wird das Modell für mehrere Indikatoren und Kategorien vorgestellt. Dabei bestehen die beobachteten Daten aus mehreren Indikatoren $q = 1, \ldots, Q_c$ und mehreren Kategorien $c = 1, \ldots, M_q$. $\text{Indik}_{c,q}$ ist der beobachtete Wert des Indikators q von Kategorie c. α_q und β_q sind unbekannte Parameter zur Verbindung der Indikatoren $\text{Indik}_{c,q}$ und des unbeobachteten „Phänomens“ ph_c. Dabei besteht die Verbindung in Form einer linearen Funktion mit einem zusätzlichen Fehlerterm $\varepsilon_{c,q}$. Der Fehlerterm bindet zwei Arten von Fehlern. Erstens den Fehler der Messung in den verschiedenen Kategorien, z.B. über falsche Messgeräte. Zweitens den Fehler der Indikatoren.

$$\text{Indik}_{c,q} = \alpha_q + \beta_q * ph_c + \varepsilon_{c,q} \tag{4.17}$$

Der Fehlerterm $\varepsilon_{c,q}$ hat den Mittelwert Null und eine konstante Varianz für einen Indikator über die verschiedenen Kategorien (aber eine unterschiedliche Varianz zwischen den Indikatoren). Außerdem wird die Annahme getroffen, dass die Kovarianz ebenfalls Null ist.

$$E(\varepsilon_{c,q}^2) = \sigma_q^2 \tag{4.18}$$

$$E(\varepsilon_{c,q}\varepsilon_{i,h}) = 0 \text{ für } c \neq i \text{ oder } q \neq h \tag{4.19}$$

Die Fehler sind also unabhängig von den Indikatoren. Diese Behauptung baut auf der Idee auf, dass alle Indikatoren einen unabhängigen und einzigartigen Beitrag zum Index leisten. Ein Indikator darf also nicht etwas messen, was bereits ein anderer Indikator misst. Diese Voraussetzung kann unter einer Modellerweiterung fallen gelassen werden. Jedoch wird das Modell damit komplizierter. Um das Modell weiter zu vereinfachen, wird die Annahme getroffen, ph_c sei eine Zufallsvariable mit Mittelwert Null und einer

einheitlichen Varianz. Zusätzlich wird davon ausgegangen, dass die Indikatoren auf einen Wertebereich zwischen 0 und 1 skaliert wurden. Die Annahme, dass ph_c und $\varepsilon_{c,q}$ gemeinsam normalverteilt sind vereinfacht die Schätzung von ph_c für Kategorie c. Dafür wird der Mittelwert der bedingten Verteilung der unbeobachteten Komponente genutzt:

$$E(ph_c \mid \text{Indik}_{c,1}, \ldots, \text{Indik}_{c,Q_c}) = \sum_{q=1}^{Q_c} w_{c,q} * \frac{\text{Indik}_{c,q} - \alpha_q}{\beta_q} \tag{4.20}$$

Die Gewichte sind gegeben durch:

$$w_{c,q} = \frac{\frac{1}{\sigma_q^2}}{1 + \sum_{q=1}^{Q_c} \frac{1}{\sigma_q^2}} \tag{4.21}$$

Durch die Gewichtsfunktion $w_{c,q}$ wird eine geringere Genauigkeit von Indikator q zu einem geringeren Gewicht führen. Die Gewichtsfunktion $w_{c,q}$ ist eine abnehmende Funktion bezüglich der Varianz von Indikator q.

Die Gewichte hängen im Zähler von der Varianz des Indikators q ab und im Nenner von der Summe der Varianzen aller Indikatoren. Falls eine Kategorie c nicht für alle Indikatoren Werte vorweist, kann die Anzahl der summierten Indikatoren im Nenner über die Kategorien variieren. Das Gewicht $w_{c,q}$ hängt über diesen Umstand also auch von Kategorie c ab. Dies kann nicht vergleichbare Indexwerte zur Folge haben. Sind die Indikatoren jedoch über alle Kategorien gleich, ist das Gewicht $w_{c,q}$ unabhängig von den Kategorien.

Die Varianz der bedingten Verteilung ist gegeben durch:

$$var(ph_c \mid \text{Indik}_{c,1}, \ldots, \text{Indik}_{c,Q_c}) = \frac{1}{1 + \sum_{q=1}^{Q_c} \frac{1}{\sigma_q^2}} \tag{4.22}$$

Die Varianz kann als ein Maß für die Genauigkeit des Index betrachtet werden. Über sie können auch Konfidenzintervalle berechnet werden. Die Varianz ist über die Anzahl der Indikatoren für jede Kategorie abnehmend und zunehmend in Bezug auf den Störterm für jeden Indikator.

Über die Annahme der Normalverteilung von ph_c und $\varepsilon_{c,q}$ wurde das Modell vereinfacht. Die zu schätzenden Parameter sind α_q, β_q und σ_q^2. Daher sind mindestens drei Indikatoren je Kategorie für ein exakt spezifiziertes Modell notwendig. Die Likelihood-Funktion

Tabelle 4.4: Vor- und Nachteile des UC-Modells

Vorteile:	**Nachteile:**
• Die Gewichte sind unabhängig von Ad-hoc-Restriktionen.	• Reliabilität und Robustheit der Ergebnisse ist Abhängig vom Umfang der Daten. • Hoch korrelierte Indikatoren führen zu Identifikationsproblemen. • Reagiert sensitiv auf Ausreißer. (Gewichte sind abnehmende Funktion der Varianz der Subindikatoren.)

Quelle: In Anlehnung an Nardo u. a. (2005: S. 66)

der beobachteten Daten (Formel 4.21) wird hinsichtlich der unbekannten Parameter α_q, β_q und σ_q^2 und ihren zurück in die Formeln 4.20 und 4.21 eingesetzten geschätzten Werten maximiert, um den Index und die Gewichte zu bestimmen. (Nardo u. a., 2005: S. 64-66)

4.3.5.3.4 Conjoint-Analyse Die Conjoint-Analyse (auch Verbundmessung) wurde in der Psychologie entwickelt und stellt ein multivariates Verfahren dar, was insbesondere im Bereich Marketing Verwendung findet (Gustafsson, 2007: S. 3). Dabei wird als erstes die Bedeutung eines (ggf. fiktiven) Gutes gemessen. Im zweiten Schritt wird über das dekompositionelle Verfahren der Anteil einer einzelnen Komponente am Gesamtnutzen bestimmt. Die einzelnen Komponenten und Eigenschaften des Gutes erhalten bestimmte Bedeutungsgewichte, mit denen ein möglichst allgemeingültiges Gesamt-Präferenzurteil der Verbraucher über ein Gut abgeleitet werden kann. In der Automobilindustrie ist etwa die Information relevant, welche Rolle die Farbe, Leistung, Ausstattung oder Größe auf den Autokauf eines Verbrauchers hat. Je nach Präferenz kann der optimale Preis für ein Auto mit bestimmten Eigenschaften definiert werden. (Baier, 2009: S. 3ff.)

Diese Art der Analyse kann auf Indizes übertragen werden. In der CA werden Respondenten mit verschiedenen Szenarien konfrontiert. Diese können z.B. aus einem Satz an unterschiedlichen Indikatorwerten bestehen. Die Respondenten bringen die verschiedenen Kombinationen an Indikatorwerten in eine Reihenfolge. Die durch das Ranking definierte Präferenz der Respondenten wird dann über die CA in die einzelnen Komponenten, d.h. Indikatoren zerlegt. (Nardo u. a., 2008: S. 98f.)

Zur Bestimmung der Gewichtung der Subindikatoren des *Happy Planet Index*[16] könnten folgende Szenarien präsentiert werden:

1. Hohe Lebenserwartung, hohe Zufriedenheit, hoher ökologischer Fußabdruck
2. Geringe Lebenserwartung, hohe Zufriedenheit, geringer ökologischer Fußabdruck
3. Hohe Lebenserwartung, geringe Zufriedenheit, geringer ökologischer Fußabdruck

Die CA könnte als Ergebnis die Gewichte (Präferenzen) der Komponenten Lebenserwartung, Zufriedenheit und ökologischer Fußabdruck ableiten.

Obwohl diese Methode eine statistische Analyse für die Bewertung der Daten benutzt, hängt die Gewichtung der Indikatoren von den Antworten der Personen ab. So werden z.B. Experten oder Politiker nach ihren Präferenzen befragt, wobei ihnen verschiedene Szenarien an Indikatorkombinationen präsentiert werden. Die Werte der Indikatoren werden über die verschiedenen Szenarien und zwischen den Respondenten variiert. Eine Präferenzfunktion wird über die Informationen der verschiedenen Szenarien geschätzt. Die Wahrscheinlichkeit der Präferenz kann nach Nardo u. a. (2008: S. 99) als Funktion der verschiedenen Indikatorwerte geschätzt werden, die in den Szenarien definiert wurden:

$$pref_c = P(\text{Indik}_{1C}, \text{Indik}_{2C}, \ldots, \text{Indik}_{QC}) \tag{4.23}$$

Dabei ist Indik_{qc} der Wert oder das Level von Indikator $q = 1, \ldots, Q$ für Kategorie $c = 1, \ldots, M$. Nachdem die Wahrscheinlichkeit über ein *Choice Modell* (z.B. Probit, Multinomiales Logit etc.) geschätzt wurde, können die Ableitungen in Bezug auf die Indikatoren der Präferenzfunktion als Gewichte zur Aggregation der Indikatoren zum Index genutzt werden:

$$I_c = \sum_{q=1}^{Q} \frac{\partial P}{\partial \text{Indik}_{qc}} \text{Indik}_{qc} \tag{4.24}$$

Das Differential der Funktion P wird am Indifferenzpunkt zweier alternativer Zustände berechnet. Im Falle eines bestimmten Indikators wird der Grenznutzen oder die Grenzrate der Substitution bestimmt. $\frac{\partial P}{\partial I_{qc}}$ ist die Indikatorgewichtung und zeigt an, wie sich die Präferenz mit der Anpassung des Indikators ändert. Damit ergibt sich die Anforderung, dass schlechte Werte einer Dimension durch eine andere ausgeglichen werden

[16] Für weiterführende Informationen zum *Happy Planet Index* wird auf Abschnitt 5.7.3 verweisen.

Tabelle 4.5: Vor- und Nachteile der Conjoint-Analyse

Vorteile:	**Nachteile:**
• Gewichtung wird über Abwägungen bestimmt. • Die Bewertung der Gemeinschaft wird einbezogen.	• Unterstellt Kompensierbarkeit der Indikatoren. • Erfordert eine Nutzenfunktion. • Hängt von der Auswahl der Respondenten ab (die die Gewichtung durch ihre Präferenzen bestimmen).

Quelle: In Anlehnung an Nardo u. a. (2008: S. 101)

können. Dieser Umstand verdient Aufmerksamkeit, da es ggf. inhaltlich nicht haltbar ist, dass Dimensionen und Indikatoren sich gegenseitig kompensieren können. (Nardo u. a., 2005: S. 71f.)

4.3.5.3.5 Analytical hierarchy process (AHP) Der Analytic Hierarchy Process (auch Analytische Hierarchieprozess (AHP) genannt) wurde von dem Mathematiker Thomsa L. Saaty in den 1970ern entwickelt und ist eine Methode um Entscheidungsprozesse zu unterstützen. (Alonso und Lamata, 2006: S. 445)

Der AHP gehört zur präskriptiven Entscheidungstheorie. Ähnlich wie die Nutzwertanalyse bietet er eine Entscheidungshilfe, um komplexe Entscheidungen zu vereinfachen und rationaler zu treffen. Dafür wird das Grundproblem in seine Bestandteile zerlegt und in eine Hierarchie gebracht. Qualitative und quantitative Elemente können dabei berücksichtigt werden. In dem Prozess werden Meinungen systematisch abgeleitet, indem Paare von Kriterien miteinander verglichen werden. (Lütters und Staudacher, 2008: S. 45f.)

Die Anwendung der AHP auf die Indexkonstruktion führt zu Einschränkungen: Erstens muss es theoretisch möglich sein, dass sich Indikatoren gegenseitig kompensieren können. Das bedeutet, dass ein schlechter Indikatorwert durch einen besonders guten Wert eines anderen Indikators ausgeglichen werden kann. Zweitens steht die durch den AHP bestimmte Gewichtung nicht für die Wichtigkeit der Indikatoren. Das bedeutet, dass die Gewichtung nicht die Wichtigkeit der einzelnen Indikatoren widerspiegelt, das gesuchte Phänomen zu erklären.

Tabelle 4.6: Vergleichsmatrix mit drei Indikatoren

Indikator/Attribut	A	B	C
A	1	$\frac{1}{2}$	$\frac{1}{4}$
B	2	1	$\frac{1}{3}$
C	4	3	1

Vorgehen

Die Grundidee des AHP besteht im Paarvergleich von Indikatoren (im AHP-Vokabular: Attribute). Es werden zwei Indikatoren ausgewählt und die Frage gestellt: Welche der beiden ist wichtiger und wie viel wichtiger? Die Stärke der Präferenz wird über eine semantische Ordinalskala von 1 bis 9 gemessen. Damit wird die gleiche Messgröße über alle Messpaare sichergestellt. Eine Präferenz von 1 bedeutet Gleichheit der beiden Indikatoren und 9, dass einer der Indikatoren neunmal wichtiger ist als der andere. Auf diese Weise werden alle Indikatoren miteinander verglichen und bewertet, zumindest dort, wo die Wahrnehmung eine eindeutige Unterscheidung zulässt.

Mögliche Bewertungsskala:

1. gleiche Bedeutung
2.
3. etwas größere Bedeutung
4.
5. deutlich größere Bedeutung
6.
7. sehr viel größere Bedeutung
8.
9. absolut dominierend

Als Ergebnis des Vergleichs ergibt sich eine Matrix, in der alle Indikatoren miteinander gekreuzt werden und der jeweilige Präferenzwert ($A_{ii} = 1, A_{ij} = 1/A_{ji}$) zugeordnet wird (vgl. Tabelle 4.6). Beispielsweise ist Indikator B zweimal wichtiger als Indikator A und folglich Indikator A nur halb so wichtig wie Indikator B. Die Entscheidungen zur Präferenz der jeweiligen Indikatorpaare wird als Präferenzwert ($A_{ii} = 1, A_{ij} = 1/A_{ji}$) in die Matrix eingetragen. Er wird für jeden Befragten erstellt. (Nardo u. a., 2008: S. 96-98)

Tabelle 4.7: Relative Gewichte der drei Indikatoren

Indikator/Attribut	A	B	C	Prioritäten
A	1	$\frac{1}{2}$	$\frac{1}{4}$	13,6%
B	2	1	$\frac{1}{3}$	23,8%
C	4	3	1	62,5%

Berechnung der Gewichte

Aus der oben dargestellten Matrix sind der Eigenvektor und der maximale Eigenwert zu berechnen. Mit ihrer Hilfe können die relativen Gewichte für die einzelnen Indikatoren berechnet werden. Ein Vorteil dieser Methode ist, dass über die Berechnung der Eigenwerte die Konsistenz der Vergleichsmatrix geprüft werden kann. Für das genannte Beispiel aus Tabelle 4.6 ergeben sich die in Tabelle 4.7 genannten Gewichte.

Berücksichtigung von Inkonsistenzen

Oftmals sind die Entscheidungen der Befragten bezüglich der Wichtigkeit von Indikatoren nicht konsistent. Wenn z.B. ein Befragter angibt, dass A sehr viel wichtiger als B ist, und B wichtiger als C ist und dann C wiederum wichtiger als A ist, liegt ein Konsistenzproblem vor. Denn C kann auf Basis der beiden ersten Antworten nicht wichtiger als A sein. Die Aussagen sind dann weniger vertrauenswürdig. Trotzdem verdienen sie Berücksichtigung, da Inkonsistenz in der Natur des Menschen liegt. Es muss vielmehr darum gehen, den Grad der Inkonsistenz zu bestimmen. Nur so können die Gewichte in der öffentlichen Kommunikation gerechtfertigt werden.

Der AHP toleriert Inkonsistenz über Redundanz: Für eine $Q \times Q$-Matrix reichen *Q-1* Vergleiche um die Gewichtung der Q-Indikatoren zu berechnen. Im AHP werden $\frac{Q(Q-1)}{2}$ Vergleiche durchgeführt. Die Redundanz der Werte ist einerseits rechenintensiv, andererseits hat sie die nützliche Eigenschaft, dass analog zu einer Schätzung der Durchschnitt über wiederholte Beobachtungen berechnet wird. Die Folge sind Gewichte, die weniger anfällig gegenüber Fehlentscheidungen sind. Darüber hinaus ermöglicht die Redundanz die Berechnung des Fehlers, der durch eine Fehlentscheidung entsteht - der Inkonsistenz-Quotient. Dabei wird davon ausgegangen, dass geringe Quotienten die Gewichtung nicht wesentlich beeinflussen (0.1 bis 0.2 werden als Grenze genannt). (Nardo u. a., 2008: S. 98)

Tabelle 4.8: Vor- und Nachteile des AHP

Vorteile:	**Nachteile:**
• Sowohl qualitative als auch quantitative Daten können verwendet werden.	• Ggf. höherer Rechenaufwand, da die Methode eine hohe Anzahl an Paarvergleichen durchführt.
• Erhöhung der Transparenz des Index.	• Die Ergebnisse hängen von der Auswahl der Bewerter ab.

Quelle: In Anlehnung an Nardo u. a. (2005: S. 70)

4.3.5.3.6 Regressionsansatz Ein lineares Regressionsmodell kann etwas über die Beziehung von einer großen Anzahl von Indikatoren $\text{Indik}_{1c}, \text{Indik}_{2c}, \ldots, \text{Indik}_{Qc}$ und einer einzelnen Output-Variablen Y_c sagen. Das Regressionsmodell wird entweder zur Überprüfung von bereits getroffenen Gewichtungsentscheidungen oder für die Berechnung von Gewichten genutzt. Bei letzterem müssen Experten den Indexwert Y_c schätzen. Dieser wird dann in Abhängigkeit zu den einzelnen Indikatoren gebracht. Über das Regressionsmodell wird dann ihre Gewichtung β_Q berechnet. Dieses Vorgehen hängt stark von der Expertenmeinung ab.

Daraufhin wird ein (üblicherweise) lineares multiples Regressionsmodell geschätzt, um die relativen Gewichte der Indikatoren zu erhalten:

$$\widehat{Y}_c = \widehat{\alpha} + \widehat{\beta}_1 I_{1c} + \ldots + \widehat{\beta}_Q I_{Qc} \tag{4.25}$$

Dabei ist $\widehat{Y}_c$, $c = 1, \ldots, M$, das Konzept, das die Indikatoren beschreiben sollen, α ist eine geschätzte Konstante und β_1 bis β_Q sind die Regressionskoeffizienten (Gewichte) der zugehörigen Indikatoren $I_1, I_2, \ldots, I_Q$. Das Regressionsmodell geht von der Annahme eines linearen Zusammenhangs und der Unabhängigkeit der erklärenden Variablen untereinander aus. Falls diese Variablen korreliert sind, werden die Schätzer eine höhere Varianz besitzen. Das bedeutet, dass die Schätzer instabil und Hypothesentests damit weniger aussagekräftig sind. Bei dem Extremfall der Kollinearität der Regressoren ist das Modell nicht mehr eindeutig. Das ist jedoch höchst hypothetisch, da in diesem Fall ein einziger Indikator für die Bildung des Index ausreichen würde und damit die Bildung eines Index überflüssig wäre.

Neben der Berechnung von Gewichten kann der Regressionsansatz bei der Prüfung und

Tabelle 4.9: Vor- und Nachteile des Regressionsansatzes

Vorteile:	**Nachteile:**
• Methode, die auch ohne Korrelation der Indikatoren untereinander angewendet werden kann. • Beinhaltet keine Manipulation der Gewichte durch Restriktionen. • Kann zur Aktualisierung oder Validierung bereits angewendeter Gewichte genutzt werden. • Kann zur Bestimmung von Gewichten genutzt werden.	• Liefert nicht belastbare Ergebnisse bei hoch korrelierten Indikatoren (Alternative: Faktorenanalyse). • Erfordert größere Datenmengen, um Schätzer mit bekannten statistischen Eigenschaften zu erhalten.

Quelle: In Anlehnung an Nardo u. a. (2005: S. 64)

Justierung der Gewichte helfen. Darüber hinaus können über das Modell z.B. politische Maßnahmen interpretiert werden, indem es den relativen Einfluss jeder politischen Maßnahme auf die Zielgröße quantifiziert. (Saisana und Tarantola, 2002: S. 11)

4.3.5.3.7 Budget Allocation (Expertenmeinung) Bei der Budget Allocation (BAL) werden Experten zu der Wichtigkeit von Indikatoren befragt. Sie erhalten ein Budget von N Punkten und müssen dieses über die Gesamtanzahl der Indikatoren verteilen. Indikatoren die hervorgehoben werden sollen, erhalten mehr Budget. Die BAL beinhaltet vier Schritte:

1. Auswählen der Experten
2. Verteilung des Budgets auf die Indikatoren durch die Experten
3. Berechnung der Gewichte
4. Iteration der BAL bis zur Konvergenz (optional)

Bei der Auswahl der Experten kommt es darauf an, dass sie ein großes Spektrum an Erfahrung, Wissen und Kritik mitbringen. Nur so kann eine ausgewogene Gewichtung gefunden werden. Besondere Aufmerksamkeit sollte darüber hinaus die Identifikation der Populationen bekommen, aus der die Experten ausgewählt werden. So können Experten aus speziellen Regionen, in denen z.B. die Umweltverschmutzung sehr stark ist, dem

Tabelle 4.10: Vor- und Nachteile der Budget Allocation

Vorteile:	**Nachteile:**
• Gewichtung wird auf Basis von Experten festgelegt (nicht technisch). • Expertenmeinungen erhöhen eher die Legitimität von Indizes, da vor der Festlegung eine Diskussion zu Gewichtung steht, in deren Folge ein Konsens steht.	• Reliabilität der Gewichte: Gewichte können spezielle (lokale) Bedingungen widerspiegeln, wodurch Expertenmeinungen nicht ohne Weiteres auf ein anderes Gebiet übertragen werden können. • Bei zu vielen Indikatoren kann der kognitive Aufwand zu anspruchsvoll sein. Die Abhängigkeit zwischen den Indikatoren kann zu komplex werden und damit zu inkonsistenten Meinungen führen. • Die Gewichtung steht u.U. nicht für die Wichtigkeit der Indikatoren, sondern z.B. für die Dringlichkeit, mit der ein bestimmter Bereich behandelt werden sollte.

Quelle: In Anlehnung an Nardo u. a. (2005: S. 67)

Thema Umwelt eine ganz andere Priorität zuweisen als Experten aus einer sehr sauberen Gegend. (Saisana und Tarantola, 2002: S. 18)

4.3.5.3.8 Öffentliche Meinung Anstatt eines Experten kann auch die Öffentlichkeit für die Gewichtung der Indikatoren eines Index befragt werden. Ein Beispiel dafür ist die Internetseite des OECD Better Life Index (OECD, 2015a). Dort erhält jeder Besucher die Möglichkeit, die elf Indikatoren des Index nach den eigenen Präferenzen zu gewichten. Danach wird der Indexwert jedes Landes auf Basis dieser Gewichtung berechnet. Auf diese Weise konnten bis jetzt (2015) 60.000 Besucher-Gewichtungen zusammengetragen werden. Probleme hinsichtlich der Repräsentativität müssen natürlich bei Veröffentlichungen berücksichtigt werden. Dennoch zeigt das Beispiel, wie die Allgemeinheit an der Gewichtung von Indikatoren beteiligt werden kann. Das scheint insbesondere dann sinnvoll, wenn es um die Messung von Phänomenen geht, die durch die öffentliche Meinung beeinflusst werden, von ihr abhängen oder durch sie definiert sind.

Ein weiterer Vorteil dieser Gewichtungsart ist neben den geringen Kosten die relativ leichte Verfügbarkeit. Bei der Befragung der Öffentlichkeit ist es schwieriger, nach der Budgetierung von 100 Punkten zu fragen als nach der Wichtigkeit einzelner Probleme, Maßnahmen etc..

Tabelle 4.11: Vor- und Nachteile der Öffentlichen Meinung

Vorteile:	**Nachteile:**
• Befasst sich mit Fragen der öffentlichen Agenda. • Ermöglicht allen Beteiligten, ihre Präferenzen auszudrücken.	• Impliziert die Messung von „Bedenken“. • Bei zu vielen Indikatoren kann es zu Inkonsistenz kommen.

Quelle: In Anlehnung an Nardo u. a. (2005: S. 67f.)

Tabelle 4.12: Vor- und Nachteile „Entfernung zum Zielwert“

Vorteile:	**Nachteile:**
• Leicht nachvollziehbare Gewichtungsmethode.	• Bei internationalen Vergleichen sind Zielwerte nicht immer verfügbar. • Die Vorteile einer Maßnahme sollten nicht an den aktuellen Zielen beurteilt werden.

Quelle: In Anlehnung an Nardo u. a. (2005: S. 68)

Ähnlich wie bei der Expertenbefragung kann diese Art der Gewichtungsfindung nur angewendet werden, wenn die Anzahl der Indikatoren nicht zu groß ist. Andernfalls ist der kognitive Aufwand zu hoch und es ist mit Inkonsistenz zu rechnen. (Saisana und Tarantola, 2002: S. 18f.)

4.3.5.3.9 Entfernung zum Zielwert Statt der Gewichtung von Indikatoren kann auch die Notwendigkeit oder Dringlichkeit (z.B. für eine politische Maßnahme) bestimmt werden. Der Abstand von aktuellen Indikatorwerten zu definierten Zielwerten erlaubt die Quantifizierung der Dringlichkeit. Ist der Abstand zum definierten Zielwert besonders hoch, ist ein Eingreifen notwendiger, als es bei leichten Abständen der Fall ist. Diese Art der Gewichtung wird realisiert durch die Division von Indikatorwert und Zielwert. Beide werden in derselben Einheit bestimmt. Der resultierende dimensionslose Parameter kann über die simple Mittelwertbildung über alle Indikatoren zum Index verdichtet werden.

Als Zielwerte kommen mehrere Möglichkeiten in Betracht. Zielwerte können politisch motiviert sein oder Überlegungen zur Nachhaltigkeit folgen. Ferner resultieren sie etwa aus der Vermeidung von Schäden. Alternativ können auch reale Orientierungswerte, wie bestimmte Länder- oder Jahreswerte, dienen. (Nardo u. a., 2005: S. 68)

4.3.6 Prüfung der Konstruktion

Bei der Konstruktion eines Index werden viele Entscheidungen getroffen, die u.U. die Ergebnisse des Index maßgeblich beeinflussen. Dazu gehört z.B. die Wahl der Indikatoren, die Normalisierungsmethode (Abschnitt 4.3.5.1), die Aggregationsmethode (Abschnitt 4.3.5.2) und die Gewichtungsart (Abschnitt 4.3.5.3). Der Einfluss der verschiedenen Entscheidungen und die daraus resultierende Verteilung der Indexwerte ist daher etwa durch Simulation zu testen.

Sensitivitätsanalyse

Mit der Sensitivitätsanalyse können Parameter und Indikatoren identifiziert werden, die das Ergebnis der Indexkonstruktion signifikant beeinflussen, z.B. die Varianz der Indexwerte. Die gleichzeitige Veränderung aller Indikatoren stellt sicher, dass der Einfluss der Parameter untereinander und deren gemeinsame Wirkung auf das Ergebnis der Indexkonstruktion berücksichtigt werden kann. Eine Möglichkeit zur Durchführung einer solchen Analyse ist die *Monte-Carlo-Analyse* (MC). Es wird eine Annahme über die Verteilung der Parameterwerte getroffen und zufällig eine Anzahl an Parameterkombinationen ausgewählt. Aus den resultierenden Indikatorensätzen und den dazugehörigen Indexwerten können z.B. mit Hilfe einer multiplen linearen Regression die Indikatoren bestimmt werden, die das Ergebnis signifikant beeinflussen. (Tarantola u. a., 2006: S. 1138f.)

Unsicherheitsanalyse

Die Unsicherheitsanalyse untersucht, wie sich die Unsicherheit der Eingangsfaktoren über die Struktur des Index auf die Indexwerte niederschlägt. Je nach Entscheidung (Wahl der Indikatoren, Gewichtungsart etc.) verändert sich ggf. die Unsicherheit und der Fehler der Indexwerte. In der Regel wird zu jedem berechneten Indexwert ein Vertrauensintervall angegeben. Beim *Happy Planet Index* nimmt ein Land einen bestimmten Rang in der Länderliste ein (vgl. Abschnitt 5.7.3). Beim Einbeziehen der Unsicherheit wird ein Intervall angegeben: z.B. mit einer Wahrscheinlichkeit von 95%, liegt der Rang des Landes zwischen Platz x und y. Je größer die Unsicherheit ist, desto größer ist der Abstand von x und y. Ziel ist es, den Abstand zwischen x und y klein zu halten. (ebd.: S. 1136f.)

Auch für diese Analyse kann die Monte-Carlo-Simulation benutzt werden (vgl. Saisana, Saltelli und Tarantola (2005)). Unter einer angenommenen Unsicherheitsverteilung zufällig variierter Werte wird die Berechnung n Mal wiederholt. Die Verteilung der n Ergebnisse lässt einen Rückschluss auf die Auswirkung der Unsicherheit der Eingangsdaten und Parameter auf die Ergebnisse der Indexkonstruktion zu. (ebd.: S. 310)

Folgende Schritte sind bei der Untersuchung der Unsicherheit denkbar:

- Aufnahme und Ausschluss von Indikatoren.
- Modellierung des Datenfehlers über bekannte Informationen der Varianzschätzung.
- Verwendung alternativer Methoden, z.B. singuläre oder multiple Imputation.
- Verwendung alternativer Normalisierungsmethoden (Standardisierung, Rangbildung etc.).
- Verwendung alternativer Gewichtungsmethoden.
- Verwendung alternativer Aggregationsmethoden.
- Verwendung alternativer Gewichte.

Zusammenfassung

Die Unsicherheitsanalyse wird in der Regel öfter angewendet als die Sensitivitätsanalyse. Jedoch bietet die iterative Verwendung von beiden im Prozess der Indexkonstruktion Potential, um die Indexstruktur zu verbessern, da sich auch die Sensitivität auf die Unsicherheit auswirkt. Die Ergebnisse der beiden Analysen werden in Streudiagrammen dargestellt - mit den Indexwerten auf der horizontalen Achse und den verschiedenen Parametern der Unsicherheit auf der vertikalen Achse. Sie helfen, die Muster in den Input-Output-Beziehungen aufzudecken. (Saisana und Tarantola, 2002: S. 56ff.)

Außerdem können die Ergebnisse der verschiedenen Simulationen über die Berechnung der durchschnittlichen absoluten Differenzen der Rangzuweisung von den Indexwerten der Merkmalsträger verglichen werden. Dazu werden die Merkmalsträger jeweils für die Referenzmethode $Rank_{\text{ref}}(I)$ und die Alternativmethode $Rank(I)$ in eine Rangfolge $c = 1, \ldots, M$ gebracht und verglichen. (Seipelt, 2010: S. 7)

Über die beiden Analysen kann geprüft werden, wie robust der Index aus statistischer Sicht ist. Dennoch ist es keine Garantie für den Index, denn der theoretische Unterbau ist nach wie vor das wichtigste Element. Die statistischen Analysen helfen dabei, die theoretischen Überlegungen erneut in Frage zu stellen und zu überprüfen. Retrospektiv (und u.U. Out-of-sample) sollte sichergestellt werden, dass das theoretisch hergeleitete Model gut auf die Daten passt.

Weitere Tests

Neben der Analyse der Sensitivität und Unsicherheit sind die Reliabilität, Aussagekraft und (Diskriminante-) Validität zu untersuchen und mit dem theoretischen Rahmen abzugleichen (Früh, Wünsch und Klopp, 2004: 529f.):

- **Die Reliabilität** äußert sich in der Forderung nach zeitlicher Stabilität des Index und der Annahme einer Homogenität äquivalenter Messungen. Das bedeutet z.B., dass der gleiche (psychische, soziale) Stimulus auf der Aggregatebene bei verschiedenen statistisch vergleichbaren Personen in gleichen Kontexten zu gleichen Werten des Index führt.
- **Bei der Validität** (Face- oder Konstruktvalidität) gilt es zu prüfen, ob das Messinstrument auf eine extreme, systematische Variation der konstituierenden Bedingungen theoriegemäß reagiert und dabei die Skalenendpunkte auch erreicht. Außerdem sind die Konstruktionsprinzipien auf praktisch unmögliche, aber konstruktimmanent mögliche Situationen zu prüfen. Theoretische Überlegungen sollten über das Konstrukt empirisch bestätigt werden.
- **Bei der diskriminanten Validität** geht es um die Frage, ob das Instrument tatsächlich etwas anderes misst als bereits existierende, ähnliche Konstrukte. Falls ein ähnliches Konstrukt existiert, ist zu prüfen, ob die Operationalisierung des Konstrukts die Differenzen abbilden kann.
- **Als Letztes ist die Aussagekraft** des Index zu prüfen, also die Fähigkeit, neue Informationen zu generieren.

4.3.7 Kommunikation und Verbreitung

Die Darstellungsform von Indexwerten hat große Bedeutung. Indizes sollten eine klare und eindeutige Aussage transportieren und es dem Nutzer ermöglichen, die Informationen schnell zu verarbeiten. Dafür bieten sich zwei Darstellungsformen an: Diagramme und Tabellen. Tabellen ermöglichen die Darstellung der kompletten Information. Die transportierte Aussage kann damit aber u.U. schwerer aufgenommen werden als beispielsweise bei geeigneten Diagrammen. Je nach Index kann eine Kombination von beiden Darstellungsformen vorteilhaft sein. Dem Betrachter wird die Möglichkeit geboten, entweder bei der durch das Diagramm geformten Grundaussage zu verweilen oder über die Tabelle tiefer in die Informationen einzusteigen.

Die Darstellung der Indexwerte kann sich auf verschiedene Ebenen beziehen. Die höchste Aggregationsstufe bilden dabei die Indexwerte selbst (Ebene des Index). Diese stellen jedoch z.T. nur den Ausgangspunkt einer Analyse dar. Sie reduzieren die Komplexität der darin enthaltenen Informationen und ermöglichen eine einfache und übersichtliche Darstellung vielschichtiger Sachverhalte. Um den Entscheidungsträgern und Nutzern des Index eine feingranulare Analyse zu ermöglichen, sind neben den Indexwerten selbst auch die Einzelwerte aufzubereiten. Über einen „Drill-down“ hat jeder Betrachter dann die Möglichkeit von dem Indexwert tiefer in die Informationen einzusteigen, indem er sich beispielsweise einen bestimmten Indikatorwert anschaut (Detailebene). Das schafft Transparenz und ermöglicht dem Betrachter eigene Ursachenforschung. Die Verwendung von Farbskalen und geeigneten Diagrammen kann die Übersichtlichkeit der Ergebnisse erhöhen.

In Gestalt der *International Business Communication Standards* (IBCS) existieren Vorschläge, um Diagramme und Tabellen hinsichtlich ihrer inhaltlichen Konzeption, der visuellen Wahrnehmbarkeit und ihrer semantischen Notation standardisiert zu gestalten. Die Vorschläge stehen zur freien Nutzung unter einer CC-Lizenz[17]. Sie werden in diesem Abschnitt Beachtung finden. Außerdem existieren diverse Versuche, die Vielzahl an existierenden Diagrammtypen zu klassifizieren. (vgl. Abela (2006), Rahlf (2014), Ribecca (2015) und Schwabish (2014))

[17] Bei der Creative-Commons-Lizenz handelt es sich um verschiedene Standard-Lizenzverträge, mit denen ein Autor der Öffentlichkeit Nutzungsrechte an seinen Werken einräumen kann.

Für die Visualisierung und Darstellung von Daten existiert eine Vielzahl von Möglichkeiten[18]. An dieser Stelle finden nur jene für die Darstellung von Indexwerten (sowie Subindizes und Indikatoren) relevanten Beachtung.

4.3.7.1 Tabellen

Die Tabelle ist die einfachste Art der Präsentation. Die Indexwerte werden getrennt nach einem Attribut oder einer Kategorie (z.B. Höhe des BIP je Land) in einer Tabelle dargestellt. In der Regel werden die Werte dabei absteigend sortiert und ggf. mit der Information zum Wert der Vorperiode erweitert. Dies kann z.B. simpel über ein Pfeilsystem (hoch, horizontal, runter) umgesetzt werden, wodurch kenntlich gemacht wird, ob der Wert im Vergleich zur Vorperiode gestiegen, gefallen oder gleich geblieben ist. Neben absoluten Werten werden in der Tabellenform auch oft Informationen zum Rang dargestellt.

Die Tabelle bietet ein umfassendes Bild der Ergebnisse, ist jedoch meist sehr detailliert und visuell nicht ansprechend. Ein Beispiel findet sich in Abbildung 4.5.

Tabelle mit Grenzwerten

Um die Übersichtlichkeit zu steigern, ist das Benutzen von Grenzwerten in Verbindung mit einer individuellen Einfärbung eine Alternative zur o.g. Tabelle. Tabelle 4.6 bietet dafür ein Beispiel. Es wurden fünf Bänder auf Basis der Indexwerte gebildet. Je nach Indexwert wird ein Land in eines dieser Bänder einsortiert.

4.3.7.2 Visualisierung

Für die Visualisierung von Indexwerten steht eine Vielzahl an Diagrammtypen zur Verfügung. Bei der Wahl sollte die Zielgruppe beachtet werden. Mit Linien- und Säulendiagramme ist sicherlich eine Vielzahl der Zeitungskonsumenten bereits in Kontakt gekommen. Das Lesen und Interpretieren dieser Diagramme sollte ohne größere Probleme möglich sein. Einige Zusammenhänge lassen sich jedoch ggf. über andere Diagrammtypen leichter oder platzsparender transportieren. In dieser Hinsicht ist Vorsicht angebracht.

[18] Einen sehr umfangreichen Überblick über Visualisierungsmöglichkeiten getrennt nach Funktion bietet Ribecca (2015).

Abbildung 4.5: Beispiel für eine Tabelle mit dem Ranking von Ländern nach dem Happy Planet Index

Staat	Rang	Happy Planet Index
Albanien	18	54,1
Norwegen	29	51,4
Schweiz	34	50,3
Moldawien	40	48
Vereinigtes Königreich	41	47,9
Deutschland	46	47,2
Österreich	48	47,1
Frankreich	50	46,5
Italien	51	46,4

Quelle: Wikipedia (o.J.)

Mit komplexen Diagrammen, welche auf den ersten Blick nicht ohne tiefgreifendes statistisches Wissen interpretierbar sind, wird riskiert, dass die gewünschte Aussage nicht vermittelt wird. Nur wenige Personen außerhalb der Statistik können z.B. etwas mit einem Boxplot-Diagramm (Abbildung 4.8c) anfangen, geschweige denn mit Begriffen wie Interquartilsabstand etc.. Einfache Diagrammtypen mit weniger Informationen sind daher gegenüber einem breiten Publikum zu bevorzugen.

In letzter Zeit haben sich die Visualisierungsmöglichkeiten deutlich erweitert. Einen größerer Entwicklungsschub gab es durch die Digitalisierung und die Steigerung der Interaktivität.[19] Nachteil der interaktiven Lösungen ist die Abhängigkeit von Endgeräten, wie Tablett, PC oder Smartphone. Der *OECD Better Life Index* nutzt auf seiner Webpräsenz großzügig die Möglichkeiten der Interaktivität und ermöglicht damit u.a. die Übersichtlichkeit einer Vielzahl von Indikatoren aufrecht zu halten (vgl. OECD (2015a)). Offline haben sich für die Darstellung von Indexwerten neben einfachen Darstellungen

[19] Beispielhaft sei die frei-nutzbare interaktive Diagrammsammlung *Google Charts* von Google genannt (Google, o.J.).

Abbildung 4.6: Beispiel für eine Tabelle mit Grenzwerten

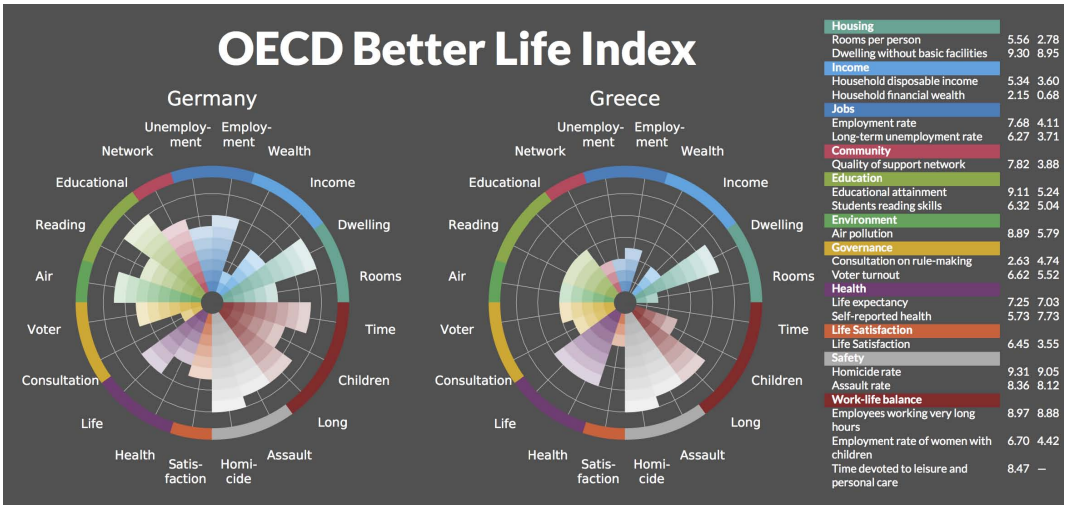
	score	Well below average (under 20)	Below average (20-40)	Average (40-60)	Above average (60-80)	Well above average (over 80)
Finland						
TAI	74				X	
Patents	19	X				
Royalties	46			X		
Internet	86					X
Tech exports	63				X	
Telephones	100					X
Electricity	100					X
Schooling	82					X
University	100					X
Japan						
TAI	70				X	
Patents	100					X
Royalties	24		X			
Internet	21		X			
Tech exports	100					X
Telephones	100					X
Electricity	100					X
Schooling	78				X	
University	36		X			

Quelle: Nardo u. a. (2008: 38)

wie Säulen- oder Liniendiagrammen auch inhaltlich leicht interpretierbare und illustrative Design-Lösungen durchgesetzt. Abbildung 4.7 zeigt ein Beispiel für solch eine umfassendere Lösung am Beispiel des *OECD Better Life Index*.

Für die visuelle Darstellung von Indexwerten steht in der Regel der Vergleich von Werten im Vordergrund (vgl. Abschnitt 3.2). In Betracht kommt dabei z.B. der Vergleich von Ländern oder Zeitperioden. Eine weitere Betrachtungsweise ist die Darstellung der Verteilung. Eine mögliche Fragestellung in diesem Zusammenhang ist, *wie verteilen sich die Indexwerte innerhalb einer Population*. Solche und weitere Fragen können auf Ebene des Index gestellt werden oder auf einer detaillierteren Ebene, die z.B. über Subindizes oder Indikatoren definiert ist. Tabelle 4.13 fasst mögliche Visualisierungstypen zusammen. Ihnen werden die zugehörigen Funktionen zugeordnet und eine Empfehlung abgegeben, ob sie auf Index- und/oder Detail-Ebene angewendet werden sollten. Die Visualisierungstypen der Tabelle werden im Folgenden kurz zusammengefasst. Abbildung 4.8 zeigt jeweils ein stark vereinfachtes Beispiel für den beschriebenen Visualisierungstyp.

Abbildung 4.7: Poster zum OECD Better Life Index



Quelle: Rahlf (2014: S. 586)

Tabelle 4.13: Visualisierungstypen für Indexwerte

Visualisierungstyp	Funktion				Ebene	
	Werte-vergleich	Zeitver-gleich	Ver-hältniss	Vertei-lung	Index	Detail
Balkendiagramm	x	x			x	
Bevölkerungspyramide	x				x	
Boxplot	x			x		x
Blasendiagramm	x	x	x	x	x	
Bubble Map			x		x	
Choroplethenkarte	x				x	
Flächendiagramm		x			x	
Gestapelte Fläche		x				x
Gestapelte Säule			x			x
Gruppierte Säule	x			x	x	x
Heatmap	x				x	x
Histogramm	x	x		x	x	
Liniendiagramm	x	x			x	x
Radarchart	x					x
Radial-Säulendiagramm	x	x	x			x
Streudiagramm				x	x	

Quelle: Eigene Darstellung

Liniendiagramm

Liniendiagramme (Abbildung 4.7m) werden eingesetzt, um quantitative Daten über ein kontinuierliches Intervall oder die Zeit darzustellen. Bei Bedarf können Gruppen über verschiedene Linien vergleichend gegenübergestellt werden.

Sollen Trends oder zeitliche Zusammenhänge von Indexwerten gezeigt werden, wird in der Regel das Liniendiagramm benutzt. Auf der horizontalen Achse wird die zeitliche Dimension abgetragen, auf der vertikalen Achse die Indexwerte. Neben dem Vergleich von Gruppen kann es u.U. auch sinnvoll sein, Subindexwerte oder Indikatoren in das Diagramm als zusätzliche Linien aufzunehmen. Hinsichtlich der Übersichtlichkeit wird empfohlen, die Anzahl der Linien überschaubar zu halten. Verschiedene Farben je Subindex, Indikator oder Gruppe können die Übersichtlichkeit erhöhen.

Als Werte können entweder die absoluten Werte, die absolute Veränderung in Prozent oder transformierte Werte abgetragen werden. Als Transformation bietet sich der Bezug z.B. zu einem Basisjahr an. In einem bestimmten Jahr wird der Wert 100 definiert und alle anderen Werte werden relativ zu diesem transformiert. Analog kann die Veränderung zum Vorjahr transformiert werden. Abbildung B.1 zeigt ein Beispiel für die Anwendung eines Liniendiagramms zur Abtragung von Indexwerten.

Streudiagramm

Ein Streudiagramm (Abbildung 4.7q) bietet die Möglichkeit, zwei Metriken und Variablen gleichzeitig für verschiedene Populationen oder Kategorien darzustellen. Dafür werden in einem kartesischen Koordinatensystem die Werte der beiden Variablen jeweils auf einer Achse abgetragen. Über diese Darstellung lassen sich z.B. leicht Korrelationen zwischen den beiden Variablen feststellen. Daraus ergibt sich für die Darstellung von Indexwerten ein großer Vorteil. Soll sowohl die Entwicklung als auch die absoluten Werte eines Indexwertes über eine Gruppe verglichen werden, bietet sich diese Art des Diagramm sehr gut an. Abbildung B.2 bietet ein praktisches Beispiel dieses Diagrammtyps für die Visualisierung von Indexwerten. Auf der vertikalen Achse wurden die absoluten Werte des Index je Land und auf der horizontalen Achse die durchschnittliche Entwicklung der Werte zur Vorperiode abgetragen. Dadurch ergeben sich im Diagramm vier leicht interpretierbare Quadranten: Anführer, Angreifer, Verfolger, Anhänger. Die

Anführer werden darüber charakterisiert, dass ihr absoluter Indexwert über dem Durchschnitt liegt und er sich im Vergleich zur Vorperiode verbessert hat. Die Angreifer zeichnen sich über einen Indexwert unterhalb des Durchschnitts aus, jedoch haben sie sich gegenüber der Vorperiode verbessert. Die Verfolger verfügen über unterdurchschnittliche Indexwerte, die sich im Vergleich zur Vorperiode verschlechtert haben. Der letzte Quadrant enthält die Anhänger. Sie haben einen überdurchschnittlichen Indexwert, haben sich jedoch zur Vorperiode verschlechtert.

Blasendiagramm (Bubble Chart)

Das Blasendiagramm (Abbildung 4.8d) ist ein Streudiagramm mit variabler Punktgröße. Es handelt sich um die grafische Entsprechung einer Kontingenztabelle. Im Blasendiagramm lassen sich insgesamt drei Merkmale gesamtheitlich betrachten. Zwei Merkmale werden in das Streudiagramm über Abszisse und Ordinate abgetragen und das dritte Merkmal über die Größe der Blase. Bei Bedarf kann diese Darstellung um ein weiteres Merkmal erweitert werden, welches über die Farbe der Blasen abgebildet wird. Bei der Nutzung von interaktiven Diagrammen, wie es z.B. auf Webseiten ohne Weiteres möglich ist, kann als fünfte Dimension die Zeit einbezogen werden. Diese wird über die Veränderung innerhalb eines laufenden Videos kenntlich gemacht.[20]

Diese Darstellungsform bietet sich für die Detailebene an. Beispielsweise können zwei Subdimensionen eines Index auf Abszisse und Ordinate abgetragen werden und der Indexwert selbst über die Größe der Blase. Damit lässt sich der Zusammenhang zwischen Subindex und Index visuell darstellen.

Choroplethenkarte

Wenn es um die Darstellung regionaler Unterschiede geht, ist eine attraktive Visualisierung über Karten möglich (Abbildung 4.8f), was man häufig z.B. nach politischen Wahlen sieht. Dabei wird der Anteil der Wähler über Farbabstufungen zwischen Regionen vergleichbar gemacht. Der Wert eines Merkmals ist jeweils einer Farbe zugeordnet und z.B. das jeweilige Bundesland entsprechend eingefärbt.

Diese Form der Darstellung eignet sich für die Ebene der Indexwerte. Diese werden zunächst z.B. je Bundesland berechnet, anschließend die Spannweite der Werte auf eine Farbskala abgetragen. Danach erhalten die einzelnen Bundesländer die ihrem Wert

[20] Ein Beispiel mit praktischer Umsetzung eines solchen „Motion Charts“ kann unter O.V. (2014) eingesehen werden.

entsprechende Farbe. Auf diese Weise lassen sich die Indexwerte übersichtlich und ansprechend vergleichbar machen.

Säulendiagramm

Das klassische Säulendiagramm (Abbildung 4.7p) nutzt horizontale oder vertikale Säulen, um diskrete, numerische Werte gegen Kategorien abzutragen. Dabei zeigt eine Achse die Kategorie, die andere Achse den entsprechenden Wert der Elemente.

In Diagramm B.3 ist ein praktisches Beispiel für ein Balkendiagramm (horizontales Säulendiagramm) zu sehen. Die numerischen Werte sind auf der horizontalen Achse abgetragen, die Populationen oder Kategorien auf der vertikalen Achse. Idealerweise wurde zusätzlich der Durchschnittswert in das Diagramm aufgenommen (vertikale Linie). Damit können die verschiedenen Populationen im Vergleich zum Durchschnitt beurteilt werden. Bei Bedarf kann ein solches Säulendiagramm um einzelne Indikatoren erweitert werden (gestapeltes und gruppiertes Säulendiagramm).

Gruppiertes Säulendiagramm

Gruppierte Säulendiagramme (Abbildung 4.8j) finden Verwendung, wenn Variablen Seite-an-Seite und gemeinsam gruppiert unter einer Kategorie abgebildet werden sollen. Wie beim einfachen Säulendiagramm steht die Länge einer Säule für den Wert der jeweiligen Gruppe und Kategorie. Damit lassen sich sowohl die Kategorienelemente als auch die Gruppenelemente miteinander vergleichen. Die Kategorien unterscheiden sich über die Abszisse, die Gruppen anhand verschiedener Farben.

Diese Visualisierungsform bietet sich für die Ebene der Indizes und die Detailebene an. Auf Ebene des Index lässt sich z.B. die zeitliche Veränderung über die Kategorien auf der Abszisse abbilden und im Vergleich zu verschiedenen Regionen oder Populationen setzen.

Auf der Detailebene können die verschiedenen Subindizes als Gruppe eingesetzt werden und z.B. ebenfalls im Zeitverlauf verglichen werden. Damit lassen sich kompensatorische Effekte leicht identifizieren, da leicht zu erkennen ist, wenn ein Subindex abgenommen, dafür ein anderer zugenommen hat.

Gestapeltes Säulendiagramm

Das gestapelte Säulendiagramm (Abbildung 4.8i) ist ideal für die Detailebene, d.h. die Ebene der Subindizes oder Indikatoren. Es unterscheidet sich vom gruppierten Säulendiagramm insofern, dass die Säulen nicht nebeneinander sondern übereinander angeordnet sind. Beim Säulendiagramm werden die Werte eines Segments auf das davor liegende gesetzt. Abbildung B.4 bietet ein praktisches Beispiel für die Nutzung dieses Diagrammtyps für Indexwerte. Im Diagramm wird z.B. das Segment *Diffusion of recent innovation* auf das Segment *Technology creation* gesetzt. Der Gesamtwert der Säule entspricht der Summe der einzelnen Segemente. Damit bietet sich diese Form der Darstellung beim Vergleich der Gesamtsummen der Segmente und Gruppen an.

In Abbildung B.4 steht jedes Segment für einen spezifischen Indikator. Die Übersichtlichkeit wird über eine individuelle Farbwahl erhöht. Die Summe der Indikatorsegmente ergibt den Indexwert. Damit kann der Beitrag jedes Indikators zum Index beurteilt und ins Verhältnis der Segmente anderer Länder gesetzt werden. Die absteigende Sortierung nach dem Indexwert steigert die Übersichtlichkeit und vereinfacht den Vergleich zwischen den Ländern.

Eine weitere Form der gestapelten Säulen ist die 100%-Darstellung. Die Gesamtsumme der Segmente wird mit 100% gleichgesetzt. Damit ist der prozentuale Beitrag jedes Segments zum Indexwert ablesbar und zwischen den Gruppen vergleichbar. Jedoch verschwinden dadurch die Absolutwerte. Ein Reihenfolge der Länder wäre somit z.B. nicht mehr möglich.

Radialsäulendiagramm

Radialsäulendiagramme (Abbildung 4.7o) ähneln den Säulendiagrammen. Die Abszisse wird zu einem Kreis gebogen, weshalb die einzelnen Variablen gleich skaliert sein sollten. Die größer wird über die Länge der Säulen abgebildet. Beginnend vom Mittelpunkt des Kreises werden die Säulen nach außen aufgebaut. Jedes Radialelement entspricht dabei einer gedrehten Säule. Diese Form der Darstellung bietet sich z.B. für die Darstellung der Detailebene an, d.h. für die Subindizes und Indikatoren. Alternativ können die verschiedenen Radialsegmente jedoch auch für den zeitlichen Vergleich eingesetzt werden. So könnte z.B. jedes Segment für den Indexwert eines Jahres stehen. Über die Erstellung

von Radialsäulendiagrammen je Gruppe können etwa auch verschiedene Länder miteinander verglichen werden. Die ausgefüllte Fläche des Kreises gibt dabei erste Hinweise auf die Zusammensetzung des Indexwertes.

Abbildung B.5 zeigt die praktische Verwendung des Diagrammtyps für die Visualisierung von Indexwerten auf der Detailebene. Dabei wurden die Indikatoren der Subindizes (äußerer Kreis) über sechs Farben zugeordnet. Die Bezeichnung der Indikatoren befindet sich außerhalb dieses Rings. Der Wert jedes einzelnen Indikators wird über die Größe des Radialsegments ausgehend von der Mitte des Kreises abgebildet.

Radialplot

Der Radialplot (auch Radialpolygone, Spider Plot, Radar Chart oder Netzdiagramm genannt) bietet sich an, wenn die abgetragenen Variablen alle in der gleichen Skalierung vorliegen. Radarplots (Abbildung 4.7n) sind kreisförmig angeordnet. Jedoch ist bei ihnen, anders als z.B. beim Kreis-/ Tortendiagramm, der Winkel um den Kreismittelpunkt konstant. Die Ausprägungen der Variablen werden auf Radien abgetragen. Ihre Länge entspricht dem Wert der Variablen. Die abgetragenen Punkte werden anschließend verbunden. Anders als beim Radialsäulendiagramm werden die Punkte direkt miteinander verbunden. Der Unterschied ähnelt dem zwischen Linien- und Säulendiagramm.

Bevölkerungspyramide

Bevölkerungspyramiden (Abbildung 4.8b) werden zur Darstellung von Verteilungen eingesetzt. Denkbar ist z.B. die Darstellung eines Indexwertes (z.B. Vertrauen) getrennt nach Alter und Geschlecht. Die Verteilung je Geschlecht wird dabei über eine Rücken-an-Rücken-Darstellung kenntlich gemacht. In der Mitte der Abszisse ist der Nullpunkt und nimmt von dort aus nach links und rechts in gleichen Verhältnissen zu. Ein Geschlecht wird nach links abgetragen, das andere nach rechts. Diese Anordnung hat den Nachteil, dass sich immer nur Zweier-Paare vergleichen lassen. Auf der Ordinate werden die Altersklassen abgetragen. Über diese Darstellung lassen sich übersichtlich Unterschiede und Gemeinsamkeiten in der Bevölkerungsstruktur bezüglich eines Metrikwertes erkennen.

Die Bezeichnung dieses Diagramms ist mittlerweile irreführend. Sie stammt aus Zeiten, in denen die Bevölkerungszahl, abgetragen nach Alter und Geschlecht, tatsächlich eine Pyramidenstruktur ergab. Im Vergleich gab es mehr jüngere als ältere Menschen, sodass

der untere (jüngere) Teil dieses Diagramms längere Balken nach links und rechts hatte als der obere (ältere) Teil.

Boxplot

Boxplots (Abbildung 4.8c) bieten sich für die Darstellung von Verteilungen an. Sie stellen bei beschränktem Platzangebot eine Alternative zu den Histogrammen dar.

Der Boxplot bildet einige Kennwerte einer Verteilung explizit ab. Der sogenannte Interquartilsabstand (Spannweite zwischen ersten und dritten Quartil) wird über einen Balken dargestellt. Innerhalb des Balkens befindet sich eine Markierung, die den Wert des Medians (zweites Quartil) abbildet. In der Regel wird diese Darstellung um jeweils eine Linie rechts und links neben dem Balken (sogenannte Whisker) erweitert. Die Definition dieser Linien kann verschieden ausfallen. Eine oft verwendete Variante ist, die Länge der Linien über die Punkte zu bestimmen, die maximal anderthalb Interquartilsabstände vom ersten und dritten Quartil entfernt sind.

Bubble Map

Das Bubble-Map-Diagramm (Abbildung 4.8e) findet Anwendung, wenn verschiedene Regionen bzgl. eines Wertes verglichen werden. Damit ist es funktional betrachtet mit der Choroplethenkarte gleichzusetzen. Der Unterschied besteht jedoch darin, dass beim Bubble Map der Wert des Merkmals über die Größe einer Blase dargestellt wird. Die Blase befindet sich im Diagramm auf der Fläche der jeweiligen Region, z.B. eines speziellen Bundeslandes. Bei Bedarf können über die Einfärbung der Region und/oder der Blase weitere Merkmale in das Diagramm integriert werden.

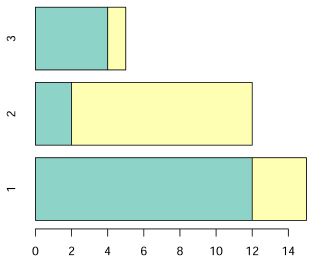
Bubble Maps haben gegenüber der Choroplethenkarte den Vorteil, dass sie mit einer reinen Strichdarstellung auskommen, was sich insbesondere beim Drucken bezahlt machen kann. Außerdem bleibt zu prüfen, inwieweit geringe Unterschiede im Merkmalswert besser über Größenunterschiede der Blase als über Farbunterschiede ersichtlich werden.

Flächendiagramm

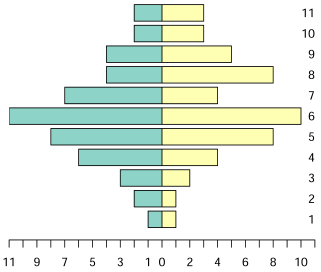
Flächendiagramme (Abbildung 4.8g) sind Liniendiagramme, bei denen die Flächen unterhalb der Linie gefüllt werden, z.B. über ein Muster oder eine Farbe. Dafür werden die Punkte in das Koordinatensystem eingetragen, miteinander verbunden und im letzten Schritt die Fläche darunter gefüllt.

Abbildung 4.8: Visualisierungstypen

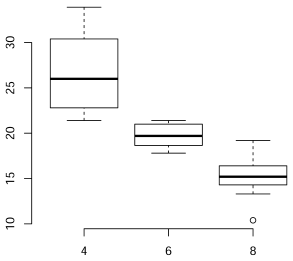
(a) Balkendiagramm



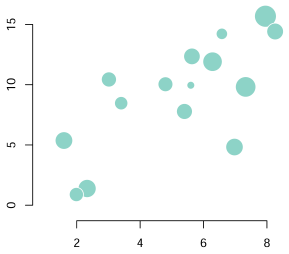
(b) Bevölkerungsp.



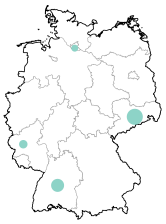
(c) Boxplot



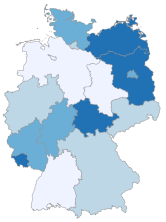
(d) Blasendiagramm



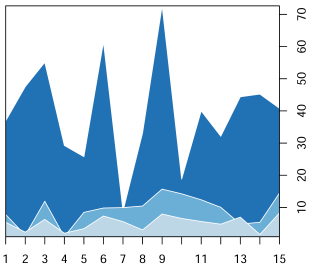
(e) Bubble Map



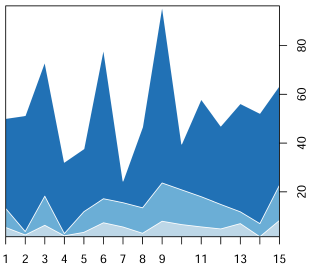
(f) Choroplethenkarte



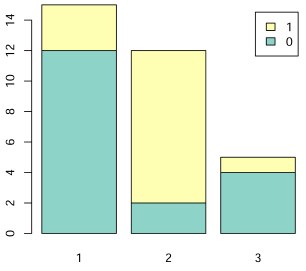
(g) Flächendiagramm



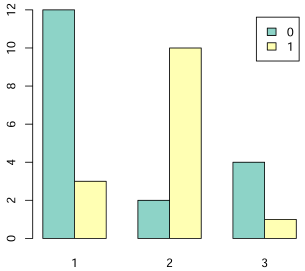
(h) Gestapelte Fläche



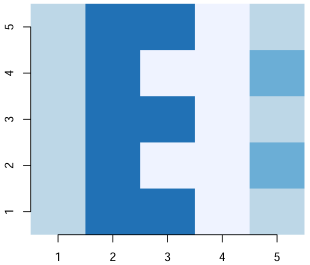
(i) Gestapelte Säule



(j) Gruppierte Säule



(k) Heatmap



(l) Histogramm

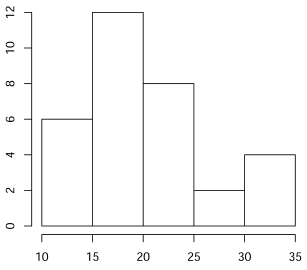
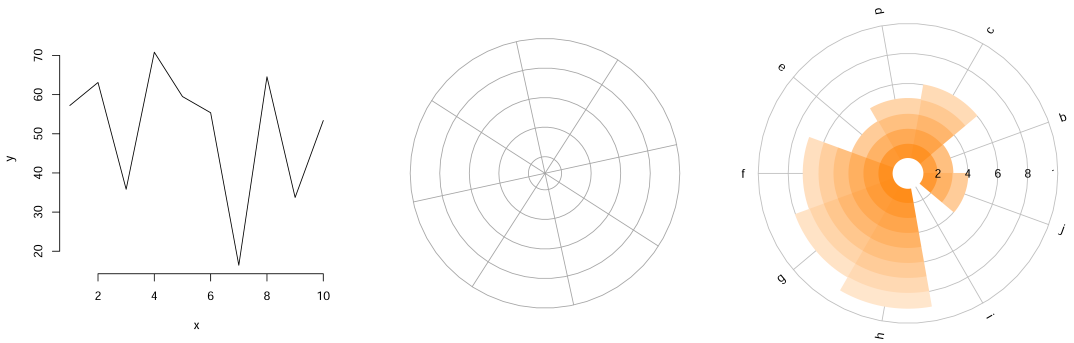
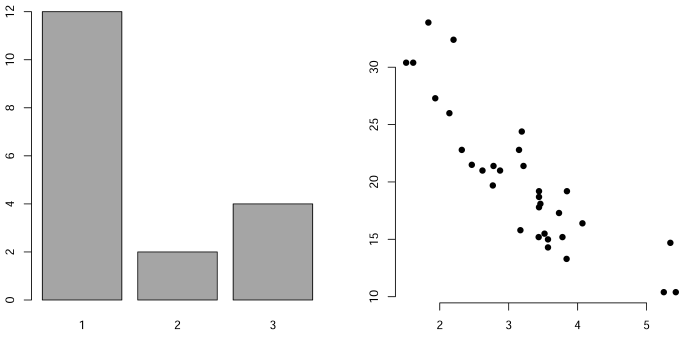


Abbildung 4.7: Visualisierungstypen (Fortsetzung)

(m) Liniendiagramm (n) Radialplot (o) Radialsäulendiagr.



(p) Säulendiagramm (q) Streudiagramm



Quelle: Eigene Darstellung

Wie Liniendiagramme finden Flächendiagramme Anwendung, wenn die Entwicklung eines bestimmtes Merkmal über ein Intervall beobachtet werden soll. Dabei lassen sich leicht Trends und Zusammenhänge erkennen.

Das Flächendiagramm kann für die Darstellung auf der Ebene des Index Anwendung finden. So kann ein Indexwert im Verlauf dargestellt und damit zeitliche Trends oder Entwicklungen einfach abgelesen werden.

Gestapeltes Flächendiagramm

Das gestapelte Flächendiagramm (Abbildung 4.8h) ist analog zum einfachen Flächendiagramm, bis auf die Eigenschaft, dass mehrere Gruppen visualisiert werden können. Dabei wird ein Gruppenelement auf das vorhergehende aufgesetzt. Der maximale Wert der obersten Reihe entspricht der Summe aller Einzelwerte.

Diese Darstellungsform lässt sich für die Detailebene nutzen. Dabei können z.B. einzelne Subindizes übereinander gestapelt werden und in Summe den absoluten Indexwert ausweisen. Damit lässt sich übersichtlich z.B. die Zusammensetzung des Index über mehrere Jahre hinweg analysieren.

Histogramm

Das Histogramm (Abbildung 4.8l) findet Verwendung, wenn die Verteilung der Häufigkeiten eines metrisch skalierten Merkmals angezeigt werden soll. Dafür wird das metrisch skalierte Merkmal in variable oder konstante Klassen aufgeteilt und die zugehörige Häufigkeit als Säule in das Diagramm eingezeichnet. Die Breite der Säulen steht dabei für die Breite des gewählten Intervalls oder Klasse. Der Höhe der Säule kann die absolute oder relative Häufigkeit abgelesen werden.

Diese Visualisierungsform eignet sich für die Ebene des Index, um z.B. zu prüfen, wie die einzelnen Indexwerte innerhalb einer Population verteilt sind. Zwar kann diese Auswertung auch für jeden Subindex durchgeführt werden, eine gesamtheitliche Betrachtung verschiedener Subindizes ist jedoch nur schwierig möglich.

Heatmap

Bei der Heatmap (Abbildung 4.8l) handelt es sich um eine zweidimensionale Matrix. Die einzelnen Zellen dieser Matrix werden auf Basis ihrer Werte eingefärbt. Die Reihenfolge der Spalten und Zeilen spielt dabei keine Rolle. Über die individuelle Einfärbung ist eine übersichtliche Visualisierung gegeben, mit deren Hilfe leicht Muster in den Kombinationen von Zeilen und Spalten identifiziert werden können. Dies kann u.U. erleichtert werden, wenn die Zeilen und Spalten entsprechend ihrer Werte sortiert werden.

Die Heatmap ermöglicht es, je zwei Merkmale paarweise zu vergleichen. Z.B. könnten alle Indikatoren eines Index verschiedenen Bundesländern gegenübergestellt werden. Eine rote Zelle könnte für einen hohen Indikatorwert stehen, eine blaue für einen niedrigen. Auf diese Weise könnten schnell die Treiber-Indikatoren eines Bundeslandes identifiziert werden.

4.4 Beurteilung der Qualität eines Index

Die Entwicklung eines Konzepts zur Beurteilung der Qualität von Indizes ist komplex, da die Qualität von Indizes von verschiedenen Aspekten abhängt. Sie wird einerseits über die zugrunde liegenden Daten und andererseits über die Prozesse, die zur Konstruktion des Index notwendig sind, beeinflusst. Dazu gehört u.a. die Zusammenstellung und Auswahl der Indikatoren, die Datenverarbeitung, die Analyse und Veröffentlichung der Ergebnisse. Damit muss die Qualität von Indizes in einem mehrdimensionalen Kontext bewertet werden.

Die Aspekte der Datenqualität und die der Prozessqualität sind kongruent. Die besten Verfahren können zu schlechten Ergebnissen führen, wenn die zugrundeliegenden Daten eine schlechte Qualität aufweisen. Demgegenüber kann trotz einer exzellenten Datenqualität ein schlechtes Ergebnis die Folge sein, wenn z.B. die falschen Indikatoren ausgewählt werden. Passen diese nicht zweifelsfrei in den theoretischen Rahmen des Indexkonstrukts, ist der Index schlecht oder gar nicht interpretierbar. (Nardo u. a., 2008: S. 44f.)

Zur Beurteilung der Qualität existieren viele Ansätze. Der Bereich des Datenqualitätsmanagements liefert z.B. eine Fülle an Kriterien zur Bewertung der Datenqualität. Darüber hinaus bietet das *Data Quality Framework* von *Eurostat* (Eurostat, 2011) und des *International Monetary Fund (IMF)* (International Monetary Fund, 2003) die Möglichkeit, neben der Datenqualität auch die Prozesse bei der Konstruktion des Index zu beurteilen. Tabelle 4.14 stellt einige Kriterien in diesem Zusammenhang zusammen.

Der *European Statistics Code of Practice* (Principles 11-15) von Eurostat betrachtet die Qualität aus Sicht des Nutzers (Eurostat, 2011). Eurostat konzentriert sich auf die folgenden sechs Qualitätsdimensionen für die Beurteilung der Basisdaten:

- **Die Relevanz** bezieht sich auf die aktuellen und potentiellen Bedürfnisse der Nutzer.
- **Die Genauigkeit** bezieht sich auf die Abweichung der Berechnung oder der Schätzung zum wahren oder „realen“ Wert.

Tabelle 4.14: Dimensionen und Kriterien des Data Quality Frameworks und Datenqualitätsmanagement

Die fünf Dimensionen des Data Quality Frameworks (International Monetary Fund, 2003)	Auswahl wichtiger Kriterien aus dem Datenqualitätsmanagement (vgl. Apel u. a. (2010: S. 19ff.))
• **Gewährleistung der Integrität** Welche Merkmale sichern die Objektivität bei der Erstellung von Statistiken und erhalten das Vertrauen der Nutzer? • **Methodologische Standards** Entsprechen die angewendeten Methoden internationalen Standards? • **Genauigkeit und Zuverlässigkeit** Sind die Quelldaten, die angewendeten statistischen Techniken etc. für das zu messende Phänomen angemessen? • **Gebrauchstauglichkeit** Werden die Bedürfnisse der Nutzer hinsichtlich der Aktualität der Statistiken, Frequenz, Konsistenz und Revisionszyklus beachtet? • **Zugang** Existiert ein einfacher Zugang zu den Statistiken und den Metadaten? Gibt es Unterstützung beim Zugang?	• **Korrektheit und Fehlerfreiheit** Entsprechen die Daten der Realität? • **Zuverlässigkeit und Nachvollziehbarkeit** Sind die Daten vertrauenswürdig? Ist die Entstehung der Daten nachvollziehbar? • **Vollständigkeit** Sind die Daten vollständig? Fehlen wichtige Entitäten? • **Genauigkeit** Liegen die Daten in der geforderten Genauigkeit vor? • **Aktualität, Zeitnähe und Zeitbezug** Entsprechen die Daten dem aktuellen Zustand der modellierten Welt? • **Redundanzfreiheit** Existieren Duplikate in den Daten? • **Relevanz** Deckt sich der Informationsgehalt der Daten mit dem Informationsbedarf? • **Einheitlichkeit** Ist die Repräsentationsstruktur der Daten einheitlich, d.h. werden sie fortlaufend gleich abgebildet? • **Verständlichkeit** Stimmen die Daten in ihrer Begrifflichkeit und Struktur mit den Vorstellungen der Nutzer überein?

Quelle: Apel u. a. (2010) und International Monetary Fund (2003), eigene Zusammenstellung

- **Die Aktualität und Zeitnähe** bezieht sich auf den zeitlichen Abstand zwischen der verfügbaren Information und dem zu beschreibenden Phänomen. Zeitnähe bezieht sich auf die Zeit zwischen Datenerhebung und dem Zeitpunkt der Veröffentlichung der Ergebnisse.

- **Die Verfügbarkeit und Klarheit** Die Verfügbarkeit bezieht sich auf den physischen Zugang, mit dem Nutzer auf die Statistiken zugreifen können: Vertriebskanäle, Bestellvorgänge, Lieferzeit, Preispolitik, Vertriebsbedingungen (Urheberrecht usw.), Verfügbarkeit von Mikro- oder Makrodaten, Medien (Papier, CD-ROM, Internet, etc.). Die Klarheit bezieht sich auf das Umfeld der statistischen Informationen: Metainformationen (Beschreibungen, Dokumentationen etc.), Grafiken, Karten und andere Abbildungen und Informationen zur Qualität der Statistiken (inkl. möglicher Grenzen bei der Weiterverwendung).

- **Die Vergleichbarkeit** bezieht sich auf die Quantifizierung möglicher Unterschiede bei regionalen, zeitlichen usw. Vergleichen, z.B. durch unterschiedliche Umfragetechniken oder statistische Konzepte.

- **Die Kohärenz** bezieht sich auf die Eignung der Daten, auf verschiedene Arten und in verschiedenen Fällen zuverlässig kombiniert zu werden.

Eurostat legt bei den o.g. Qualitätsdimensionen den Fokus auf den Output für den Nutzer. Damit soll sichergestellt werden, dass bei allen statistischen Produkten dieser Institution verschiedene Standards eingehalten werden und der Nutzer bei den Produkten immer mit der gleichen Qualität rechnen kann. Den Nutzer in die Qualitätsbetrachtung einzubeziehen, ist generell sinnvoll. Der Index entfaltet schließlich erst durch dessen Nutzung einen Wert. Werden die Ergebnisse des Index nicht genutzt, ist seine Existenz aus dieser Perspektive fragwürdig. Jedoch ist bei dieser Form der Betrachtung kritisch zu hinterfragen, ob sie den wissenschaftlichen Ansprüchen des Index entgegensteht.

Im Folgenden werden relevante Qualitätsdimensionen der Daten und Prozesse vorgestellt. Mit ihnen kann bereits im Prozess der Indexkonstruktion die Qualität überwacht werden.

Qualitätsdimensionen zur Beurteilung der Daten

Die Auswahl der Basisdaten beeinflusst direkt die Qualität des Index. Bei der Auswahl der Daten sollte laut Nardo u. a. (2008) daher besonderes Augenmerk auf folgenden Dimensionen liegen (S. 46-48):

Relevanz

Die Relevanz der Daten ist eine qualitative Beurteilung des Wertes, den diese Daten beitragen. Dieser ist dabei durch den Grad gekennzeichnet, mit dem die Statistiken den aktuellen und potentiellen Anforderungen der Nutzer entsprechen. Dafür muss er sowohl die entsprechenden Themen abdecken, als auch angemessene Konzepte nutzen. Die Relevanz in Bezug auf den Index muss an dessen Absicht beurteilt werden. Die Auswahl der Basisdaten sollte alle notwendigen Bereiche abdecken, damit das zu bestimmende Konzept ausgewogen über die Indikatoren numerisch abgebildet werden kann. Falls auf vorhandene Daten zurückgegriffen wird, ist es u.U. notwendig, das Konzept über Proxy-Variablen abzubilden. In diesem Fall ist der Bezug zur Zielgröße - wenn möglich - sicherzustellen.

Genauigkeit

Die Genauigkeit der Basisdaten gibt an, wie genau die Daten die gesuchte Zielgröße schätzen oder beschreiben können, d.h., wie weit die verfügbaren Werte vom wahren (unbekannten) Wert entfernt sind. Für die Bestimmung der Genauigkeit existiert eine ganze Reihe an Maßen. Generell werden sie unter dem Begriff *Fehler* ausgewiesen oder in Form von Konfidenzintervallen ausgedrückt.

Bei Schätzungen, die auf Zufallsstichproben basieren, geht der Fehler z.B. auf Stichprobendesign und Antwortausfall zurück. Bei abgeleiteten Schätzern, z.B. über Daten der statistischen Ämter, rührt der Fehler aus den jeweiligen Befragungen oder dem Zensus; bei Daten, die nicht exakt die Anforderungen erfüllen (z.B. zeitlich, inhaltlich etc.), hängt der Fehler von den Techniken ab, die zur Lösung des Problems eingesetzt werden. Außerdem kann der Fehler durch Saisonbereinigung (z.B. ARIMA) entstehen.

Ein weiterer Aspekt der Genauigkeit entsteht bei einer Revision. Erhalten einzelne Elemente aus der Indexkonstruktion über Erhebungswellen hinweg eine Revision, sind die Auswirkungen zu prüfen. Dabei können Änderungen entstehen durch:

1. Andere Daten.
2. Ersetzen von Prognosen durch Erhebungsdaten.
3. Andere Definitionen oder Methoden.
4. Anderer Referenzwert (z.B. bei Indizes, die über den Bezug zu einem Referenzjahr berechnet werden).

In Bezug auf Indizes ist die Genauigkeit der Basisdaten extrem wichtig. Sie wirkt sich unmittelbar auf die Glaubwürdigkeit der Daten aus. Bei einer geringen Glaubwürdigkeit wird auch das Vertrauen in den Index gering sein, weshalb es wichtig ist, die Objektivität der Daten zu gewährleisten. Es muss transparent dargestellt werden, dass die Daten professionell erhoben und geeignete statistische Verfahren angewendet wurden. Mit Transparenz kann verhindert werden, dass z.B. der Eindruck von Manipulation oder Umdatierung entsteht.

Aktualität

Die Aktualität von Datenprodukten wird über die Zeit, die zwischen ihrer Verfügbarkeit und dem zu beschreibenden Phänomen liegt, bestimmt. Auch die Zeitnähe steht in enger Beziehung dazu. Sie beschreibt den Zeitraum, der zwischen Datenerhebung und Veröffentlichung liegt und möglichst kurz sein sollte.

Bei der Indexkonstruktion ist die Aktualität besonders wichtig, um die Notwendigkeit der Schätzung von fehlenden Werten oder Revisionen vorhergehender Daten zu minimieren. Falls für den Index eigene Basisdaten erhoben werden, kann die Aktualität in der Regel sehr gut gesteuert werden. Teilweise ist jedoch der Kostendruck dafür verantwortlich, dass auf öffentliche Daten zurückgegriffen werden muss. Hier ist zwar oft die Genauigkeit gewährleistet, jedoch verschieben sich erfahrungsgemäß die Veröffentlichungstermine.

Verfügbarkeit

Diese Dimension umfasst den Zugang zu den Datenprodukten, der in erster Linie unkompliziert sein sollte. Außerdem sollte die Quelle der Originaldaten und deren Erreichbarkeit bekannt sein. Die Qualitätsdimension umfasst zusätzlich die Bedingungen, unter denen die Nutzer die Statistiken abrufen können (Vertriebskanäle, Copyright etc.).

Die Verschiedenheit der Nutzer führt ggf. zur Notwendigkeit einer breiten Palette an Verbreitungsformaten. Es ist u.U. auch sinnvoll, Metadaten selektiv darzustellen. Die Qualitätsdimension Verfügbarkeit umfasst somit auch, ob die Daten in geeigneter Form zugänglich gemacht werden. Die Datenprodukte sind über angemessene Medien zu verbreiten. Sie sollten einfach verfügbar sein, Metadaten sollten bereitgestellt werden und im Idealfall existieren Serviceleistungen, die mögliche Fragen zu den Datenprodukten beantworten können.

Grundsätzlich sollten Nutzer wissen, dass die Daten für sie zur Verfügung stehen. Außerdem sollte der Preis und Aufwand für die Daten im Verhältnis zum jeweiligen Nutzen stehen.

Für Indizes bedeutet die Verfügbarkeit der Basisdaten ggf. höhere Kosten. Sie kann jedoch die Glaubwürdigkeit des Index stark beeinflussen. Ein schlechter Zugang zu den Basisdaten macht es für Dritte schwer möglich, die Ergebnisse nachzuvollziehen.

Leider können u.U. Schwierigkeiten hinsichtlich der Kohärenz der Daten entstehen, wenn der Index z.B. über mehrere (z.B. öffentlich zugängliche) Datenquellen hinweg konstruiert wird. Daher sollte die Wahl der Datenquelle nicht ausschließlich auf Basis der Verfügbarkeit getroffen werden, sondern auch über den Einbezug anderer Qualitätsdimensionen.

Interpretierbarkeit

Bei der Interpretierbarkeit von Datenprodukten geht es um die Frage, ob Nutzer die Daten verstehen und angemessen nutzen und analysieren. Sie wird hauptsächlich durch das Konstruktionskonzept, die Nutzergruppen, die Variablen, die Begrifflichkeiten und etwaiger Limitationen (z.B. durch spezielle statistische Methoden) bestimmt. Ähnlich wie bei der Qualitätsdimension Verfügbarkeit ist zu überlegen, ob je nach Nutzergruppe eine Darstellungsform in unterschiedlicher Detailstufe gewählt wird. Unter Voraussetzung eines bestimmten Wissensstands können z.B. Metadaten zu den angewendeten Verfahren bei der Interpretation helfen. Werden diese jedoch von Laien betrachtet, können sie u.U. eher verwirren. Außerdem sollten die Metadaten nach Möglichkeit kohärent sein.

Bei der Konstruktion von Indizes werden Daten über eine breite Palette an Methoden bearbeitet. Die Art der Normalisierung, Konsolidierung, Aggregationsmethode etc. beeinflussen alle die Interpretation der Basisdaten. Verwendete Definitionen, Klassifikationen,

Annahmen usw., mit denen die Basisdaten erzeugt wurden, sind für einen Vergleich von Attributen oder verschiedenen Zeitpunkten notwendig. So sollten z.B. Veränderungen über verschiedene Wellen hinweg verfolgt und bewertet werden. Andernfalls können die Ergebnisse nicht miteinander verglichen werden. Die Verfügbarkeit von Metadaten ist damit ein wichtiges Element für die Qualität eines Index.

Kohärenz

Die Kohärenz von Datenprodukten gibt ein Maß für die logische Verbindung von Datenprodukten und ihre gegenseitige Konsistenz an. Das beinhaltet z.B. ihre Eignung, auf verschiedenen Wegen zuverlässig miteinander verbunden zu werden. Außerdem ist es sinnvoll, eine einheitliche Semantik zu verwenden: Ein Begriff darf nicht ohne Begründung für verschiedene Variablen oder Inhalte verwendet werden. Im Gegenzug sind unterschiedliche Begriffe nicht ohne Begründung für gleiche Variablen oder Inhalte zu nutzen. Ebenfalls sollte nicht grundlos von den Methoden abgewichen werden, wenn die Ergebnisse dadurch beeinflusst werden. Die Mindestanforderung für kohärente Datenprodukte ist, dass bei der Verbindung von Daten, die dasselbe Konzept operationalisieren und verschieden sind, die Unterschiede im Erhebungszeitraum, der Bewertung und der Erhebung identifiziert werden. Andernfalls ist eine Verbindung der Daten nicht möglich.

Bei der Betrachtung von Indizes ist die Kohärenz insbesondere beim zeitlichen oder räumlichen Vergleich entscheidend. Kohärenz über die Zeit wird erreicht, wenn keine Änderungen am Konzept, den Definitionen und den Methoden durchgeführt werden oder wenn alle Änderungen transparent nachvollzogen werden können. Bezüglich des räumlichen Vergleichs gelten die gleichen Parameter. Hierbei tritt jedoch in der Regel die Frage nach dem Untersuchungsdesign und der Repräsentativität in den Vordergrund.

Qualitätsdimensionen zur Beurteilung der Prozesse

Jeder Schritt der Indexkonstruktion ist wichtig und muss in Hinblick auf die Qualität verfolgt werden. So kann der theoretische Rahmen die Relevanz des Index beeinflussen. Die multivariaten Analysen sind wichtig, um die Reliabilität des Index zu erhöhen. Imputation, Normalisierung und Aggregation können die Genauigkeit des Index beeinflussen.

Mit Matrix 4.15 wird ein Bewertungsmodell eingeführt, mit dem die Qualität des Index bewertet werden kann. Dafür werden die verschiedenen Schritte der Indexkonstruktion

in Verbindung zu den oben genannten Qualitätsdimensionen der Basisdaten gebracht. Das Konzept basiert auf dem *OECD Quality Framework*.

Der theoretische Rahmen und die darin befindlichen Definitionen beeinflussen nicht nur die Relevanz des Index, sondern auch dessen Glaubwürdigkeit und Interpretation. Die Relevanz eines Index wird z.B. auf Basis von politischen, gesellschaftlichen oder analytischen Bedürfnissen bestimmt. Dafür benötigt der Rahmen jedoch auch ein solides theoretisches Fundament. Leider fehlt einigen Indizes dieses Fundament, weshalb die Idee von Indizes, insbesondere in Politik und Gesellschaft, nicht selten in Frage gestellt wird.

Das Imputieren fehlender Daten beeinflusst die Genauigkeit und die Glaubwürdigkeit. Falls die Imputation jedoch in zu großem Maße angewendet wird, besteht die Gefahr eines Qualitätsverlustes des gesamten Index.

Die Normalisierung beeinflusst sowohl die Genauigkeit, als auch die Kohärenz der finalen Ergebnisse. Eine unangemessene Normalisierungsmethode kann verzerrte und unzuverlässige Ergebnisse zur Folge haben. Außerdem hängt die Interpretierbarkeit des Index von der Normalisierungsmethode ab.

Auch die Qualität der Basisdaten wirkt sich stark auf die Genauigkeit und die Glaubwürdigkeit des Index aus. Ebenso die Aktualität wird maßgeblich durch die Wahl angemessener Daten beeinflusst.

Die Anwendung von multivariaten Analysen zur Identifikation der Datenstruktur kann sowohl Genauigkeit als auch die Interpretierbarkeit der Ergebnisse erhöhen. Diese Phase ist auch für die Identifikation von Redundanz und möglicher Datenlöcher wichtig.

Eines der Schlüsselelemente der Indexkonstruktion ist die Wahl des Gewichtungs- und Aggregationsmodells. Nahezu alle Qualitätsdimensionen werden von dieser Entscheidung beeinflusst - in besonderem Maße Genauigkeit, Kohärenz und Interpretierbarkeit. Dieser Schritt bedarf höchster Aufmerksamkeit, da sein Ergebnis die größte Angriffsfläche darstellen kann. Daher ist hier besonders auf die Vermeidung interner Widersprüche oder Fehler bei der Gewichtung und Aggregation zu achten.

Sensitivitäts- und Robustheitsanalysen sind wichtig, um das Risiko der Entstehung eines bedeutungslosen Index zu minimieren. Derartige Analysen können die Genauigkeit,

Glaubwürdigkeit und Interpretierbarkeit der finalen Ergebnisse erhöhen. Über Sensitivitätsanalysen kann z.B. geprüft werden, ob Unterschiede im Länderranking signifikant sind. So lässt sich das Risiko auf Fehlinterpretation oder falsche Nutzung minimieren.

Ein Vergleich mit anderen Indizes oder Maßzahlen zur Bestimmung des gleichen „Phänomens“ hilft dabei, die Relevanz des Index zu verdeutlichen und seine Fähigkeit der Informationsgenerierung zu klären. Relevanz und Interpretierbarkeit können über solch einen Vergleich erhöht werden. Außerdem trägt eine hohe Korrelation des Index zu den Referenzdaten zur Glaubwürdigkeit bei.

Die Darstellung der Ergebnisse hat Einfluss auf die Relevanz und die Interpretierbarkeit. Üblicherweise lesen Nutzer die methodischen Erläuterungen wegen der Komplexität des Index nicht. Das Verständnis der Ergebnisse wird daher stark über die Aussagen der aufbereiteten Tabellen oder Diagramme geprägt. Grundsätzlich sind Indexwerte als Ausgangspunkt von Analysen zu sehen, von denen aus tiefer in die Details gegangen wird. Daher kann dieser analytische Schritt sich auf die Relevanz und die Interpretierbarkeit auswirken. Falls der Index oder die Darstellung und Verbreitung der Werte keine detaillierte Analyse erlauben, kann die Glaubwürdigkeit in Frage gestellte werden. (Nardo u. a., 2008: S. 48f.)

Die Verbreitung und Publikation der Indexwerte hat Einfluss auf die Relevanz, Glaubwürdigkeit, Verfügbarkeit und Interpretierbarkeit. Leider wird diesem Schritt oftmals zu geringe Bedeutung geschenkt, indem z.B. nur der Gesamtzusammenhang veröffentlicht wird.

Tabelle 4.15: Matrix zur Beurteilung der Konstruktions-Qualität eines Index

	Qualitätsdimension						
Konstruktionsschritt	Relevanz	Genauigkeit	Glaubwürdigkeit	Aktualität	Verfügbarkeit	Interpretierbarkeit	Kohärenz
Theoretischer Rahmen	✓		✓			✓	
Datenerhebung / Variablenselektion		✓	✓	✓			
Behandlung fehlender Daten	✓	✓	✓	✓			
Prüfen der Datenstruktur		✓				✓	✓
Normalisierung		✓				✓	✓
Gewichtung und Aggregation	✓	✓	✓			✓	✓
Sensitivität und Robustheit		✓	✓			✓	
Vergleich mit anderen Variablen	✓		✓			✓	✓
Darstellung	✓					✓	
Drill-down	✓		✓			✓	
Verbreitung	✓		✓		✓	✓	

Quelle: Nardo u. a. (2008: S. 49), leicht modifiziert

Kapitel 5

Empirischer Teil: Inhaltsanalyse von Beiträgen mit Indexbezug

„Die 'Frankfurter Zeitung' hat sich mit ihren in großzügiger Weise und mit ungewöhnlicher Intensität ausgebauten 'Indexzahlen' in Deutschland an die Spitze [...] gestellt."

Paul Hermberg, 1920

Exakte, nachprüfbare Aussagen und präzise Vorhersagen sind der Erfolg moderner Wissenschaften. Insbesondere die Naturwissenschaften sind stets bestrebt, ihre Hypothesen mit immer genaueren Messmethoden zu prüfen und die Weiterentwicklung der Theorien voranzubringen. Zentrale Elemente mit einer weitreichenden Geschichte sind in diesem Zusammenhang das Zählen und Messen. Während diese Grundoperationen schon lange zum Handwerkszeug der Naturwissenschaften gehören, finden sie nach und nach auch in den Geisteswissenschaften Einzug.

Für die Medienwissenschaften kann die quantitative Betrachtung von redaktionellen Beiträgen Mehrwert bezüglich der Prüfung und Generierung von Hypothesen sowie eine beträchtliche Reduktion des personellen Aufwands bedeuten. Insbesondere vor dem Hintergrund der Digitalisierung bieten quantitative Methoden zusammen mit maschineller Verarbeitung die Möglichkeit einer steigenden Anzahl von Inhalten gerecht zu werden. Hilfestellung bieten Methoden anderer Disziplinen, wie z.B. dem quantitativen Zweig der

Linguistik, der auf eine noch junge, ca. 50-jährige Geschichte zurückblickt. (Köhler, 2005: S. 2ff.)

5.1 Vorgehen

Für die Analyse der medialen Funktion von Indizes innerhalb der Verbreitungsmedien werden in diesem Kapitel eine Reihe von Kriterien erarbeitet, mit denen die Beiträge mit Indexbezug einer einschlägigen deutschen Tageszeitung bewertet werden (Abschnitt 5.5).

Bereits zu Beginn des Dissertationsvorhabens wurde deutlich, dass längere Zeitreihen für aussagekräftige Ergebnisse notwendig sind. Einige Indizes werden in den Verbreitungsmedien nur wenige Male innerhalb eines Jahres aufgegriffen. Die manuelle Selektion und Durcharbeitung von Beiträgen erscheint vor dem Hintergrund eines mehrjährigen Zeitfensters nicht umsetzbar. In Folge dessen wird für den empirischen Teil der Arbeit eine Methode entwickelt, mit der die Kriterien der Inhaltsanalyse so weit wie möglich automatisiert erhoben werden könne.

Eine Automatisierung besitzt den Vorteil, dass gegenüber der herkömmlichen Inhaltsanalyse keine Untermenge der Beiträge gebildet werden muss, sondern die Grundgesamtheit an Beiträgen ausgewertet werden kann. Gleichzeitig entsteht die Chance einer kontinuierlichen Überwachung. Sobald die Kriterien der Inhaltsanalyse automatisiert wurden, können sie auf jegliche Datenbasis angewendet werden. Die Kriterien können ohne wiederholten Aufwand auf neue und andere Indizes oder auf später veröffentlichte Beiträge angewendet werden, um z.B. Zeitreihen aktuell zu halten.

Wie in allen anderen empirischen Wissenschaften wird die Inhaltsanalyse dadurch erschwert, dass niemals vollständige Informationen zum Untersuchungsgegenstand existieren. Der Untersuchungsgegenstand kann entweder prinzipiell unendlich sein oder er ist aus praktischen Gründen nicht vollständig beschreibbar. Daher wird immer nur ein Ausschnitt einbezogen. Dieser besteht in der vorliegenden Arbeit aus einer einschlägigen Tageszeitung sowie regelbasiert einbezogener Indizes.

Die Schwierigkeit von Stichproben-Verzerrungen (Selektions-Bias) rückt durch die Berücksichtigung aller Beiträge in den Hintergrund. Die Herausforderung liegt vielmehr

in der präzisen Definition der Untersuchungskriterien und einer adäquaten Aggregation der Daten. Nur über diese Aggregation lassen sich relevante Informationen ableiten.

Wie zu sehen sein wird, ist eine automatisierte Erhebung nicht bei allen Kriterien möglich oder sinnvoll. Aus diesem Grund wird für einen Teil der Kriterien eine manuelle Inhaltsanalyse durchgeführt. Somit soll die Aussagekraft der automatisiert erhaltenen Ergebnisse möglichst umfassend ergänzt werden.

Um den Aufwand bei der manuellen Analyse nicht überhandnehmen zu lassen und dennoch die statistischen Vorzüge einer Vollerhebung der Beiträge zu genießen, werden für die manuelle Prüfung Stichproben gezogen. Die Anzahl an Beiträgen wird relativ festgelegt. Es erfolgt die Ziehung einer einprozentigen Stichprobe mit mindestens zehn Beiträgen je Index.

Sicherlich ist auch für die automatisiert zu erhebenden Kriterien eine Stichprobenziehung denkbar. Da jedoch absolute Zahlen im Vordergrund stehen und die benötigten Rechnerkapazitäten in Anbetracht der aufkommenden Datenmenge überschaubar sind, finden sich keine Vorzüge einer Stichprobenziehung gegenüber dem Einbeziehen aller Beiträge.

5.1.1 Explorative Datenanalyse

Die Festlegung der Kriterien für die Analyse der Beitragsinhalte orientiert sich nicht nur an existierenden Forschungsfragen und Hypothesen. Die Datenerhebung und Definition der Kriterien soll im Gegenteil eine explorative Datenanalyse ermöglichen, mit deren Hilfe u.a. Hypothesen generiert werden können:

> "In der explorativen Datenanalyse gehen wir nun einen Schritt weiter und versuchen, mit Hilfe von geeigneten Darstellungen und Berechnungen die Daten nach Mustern oder Zusammenhängen zu untersuchen. Daher auch der Begriff 'explorativ' - wir forschen (explorieren) in den Daten nach interessanten Informationen, die man bei der einfachen Betrachtung in der deskriptiven Analyse nicht auf den ersten Blick sehen kann." (Schäfer, 2010)

Dieses Vorgehen bietet sich insbesondere an, wenn nur wenig Vorwissen über die Zusammenhänge der Daten zur Verfügung steht:

> „I hope that I have shown that exploratory data analysis is actively incisive rather than passively descriptive, with real emphasis on the discovery of the unexpected.“ Tukey (1977: S. lxii)

Laut Jambu (1992) habe die Datenanalyse dabei im Kern die Zielstellung, Hypothesen zu generieren. Dies werde mit deskriptiven Verfahren und induktiven Testmethoden unter der Leitfrage: *Was ist an einer Verteilung eines Merkmals bemerkenswert oder ungewöhnlich?* erreicht. Im Anschluss werden die statistischen Befunde interpretiert und in eine wissenschaftlichen Hypothese überführt.

Die Ziele der explorativen Datenanalyse werden nach Schröder (1997) wie folgt zusammengefasst:

- Systematische oder versuchsweise Reduzierung der Datenmenge, um Strukturen in den Daten aufzudecken und einfache überschaubare Zusammenhänge auszuarbeiten.
- Herausarbeiten latenter Faktoren.
- Aufdecken von Regelmäßigkeiten oder Gemeinsamkeiten in der Verteilung eines Merkmals oder mehrerer Merkmale.
- Erstellung von Hypothesen zu möglichen Datenstrukturen und Zusammenhängen.
- Testen der Daten auf statistische Eigenschaften und Formulieren passender statistischer Werkzeuge und Techniken (z.B. Annahme einer Normalverteilung).

Für das Auffinden von Mustern und Zusammenhängen werden in der vorliegenden Arbeit besondere Formen von Grafiken verwendet (vgl. Abschnitt 5.7.1). Ebenfalls werden grundlegende Methoden der Korrelationsbestimmung verwendet.

5.1.2 Umsetzung der Automatisierung

Für die technische Umsetzung der im Folgenden dargestellten Methode zur Automatisierung der Inhaltsanalyse findet die populäre Open Source-Statistik-Umgebung R Verwendung. Über Pakete ist diese nahezu beliebig erweiterbar und erfreut sich neben dem akademischen Bereich mittlerweile auch in anderen Bereichen zunehmender Beliebtheit.

Die vorliegende Fragestellung profitiert insbesondere von den R-Paketen zur Abfrage von Webdatenbanken über SPARQL (Paketname: SPARQL) und den Paketen für „Web scraping“, d.h. für das Abrufen und Verarbeiten von Webseiteninformationen (Paketname: xml2, rvest).

Die Entwicklung und Umsetzung der Methode mit R wurde in den Online-Dienst Github eingebettet. Dieser ermöglicht die verschiedenen Versionen des Entwicklungsprojekts zu verwalten. Außerdem ergibt sich der Vorteil, dass die entwickelte Anwendung frei für die Öffentlichkeit zur Verfügung steht und leicht in R installierbar ist. Die Anwendung kann des Weiteren nach Veröffentlichung der Dissertation unter höheren Versionsnummern von anderen Nutzern weiterentwickelt oder durch das Eintragen von Verbesserungsvorschläge in ein angeschlossenes Ticketsystem optimiert werden.

Das für die Dissertation entwickelte Paket analyseMedia kann direkt auf Github eingesehen oder über folgende Anweisung in R installiert werden[1]:

```
1 library(devtools)
2 install_github("digitaleMedien/analyseMedia")
```

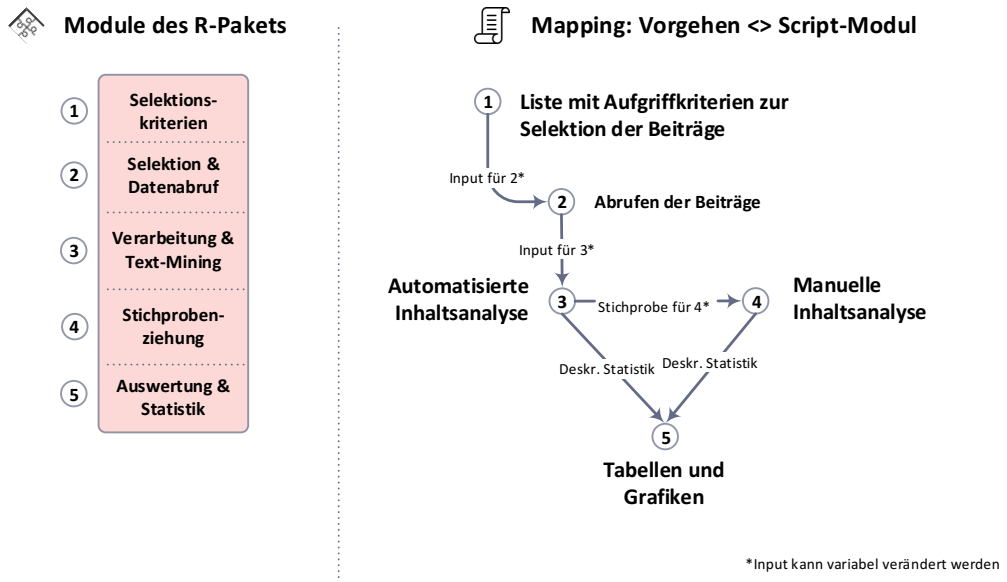
Das Paket analyseMedia enthält verschiedene Funktionen, mit denen digitale Medien automatisiert verarbeitet werden können. Sie sind modular gestaltet, d.h. sie können je nach Fragestellung aufeinander aufbauend oder einzeln verwendet werden.

Die Qualität des Pakets wurde mittels R CMD check geprüft. Die verschiedenen Funktionen des Pakets werden in einer eigenständigen Dokumentation beschrieben. Diese kann ebenfalls auf Github als PDF-Dokument abgerufen werden. Ein Auszug aus der Dokumentation ist in Anhang E, Abbildung E.1 einsehbar.

Das im Folgenden beschriebene Vorgehen zur medialen Inhaltsanalyse von Beiträgen mit Indexbezug lässt sich auf die einzelnen Module des R-Pakets mappen. Abbildung 5.1 zeigt, welches Modul für welchen Schritt der Inhaltsanalyse Anwendung findet.

[1] Das Repository analyseMedia ist unter `https://github.com/digitaleMedien/analyseMedia` erreichbar (besucht am 15. Nov. 2016).

Abbildung 5.1: Module des R-Pakets mit Mapping zum Vorgehen



Quelle: Eigene Darstellung

5.2 Untersuchungsinstrument

Wie im Umgang mit quantitativen Inhaltsanalysen üblich wird auch in diesem Fall ein Codebuch als Untersuchungsinstrument verwendet. Nach Raupp und Vogelgesang (2009) beinhaltet ein Codebuch sämtliche Aufgriffskriterien[2] und Codierregeln, die der Wissenschaftler für die Beantwortung seiner Forschungsfragen definiert hat (S. 151). Die Aufgriffskriterien beziehen sich dabei auf die Identifikation der Untersuchungseinheit. In der vorliegenden Arbeit sind das Beiträge mit Bezug zu einem bestimmten Index. Die Codierregeln hingegen beziehen sich auf die Eigenschaften dieser Untersuchungseinheiten. Diese können mit Hilfe der in Kapitel 2.2.4 vorgestellten Skalen operationalisiert werden.

Im Folgenden werden verschiedene Kriterien zur Beschreibung der Eigenschaften einer Untersuchungseinheit hergeleitet. Um dem Vorzug der Automatisierung mehr Beachtung schenken zu können, werden die Kriterien getrennt nach automatisiert und manuell zu erhebenden betrachtet.

[2]Der Begriff *Aufgriffkriterium* wird aus Raupp und Vogelgesang (2009) übernommen. Er bezieht sich auf die Kriterien, die dazu führen, dass eine Einheit aufgegriffen wird.

Bei der Entwicklung eines Codebuchs oder eines Kriterienkatalogs ist der Ausgangspunkt ein theoretischer Begriff, der über das Codebuch eine numerische Abbildung finden soll. Wie in Kapitel 2 vorgestellt wurde, sind für die Bestimmung eines theoretischen Begriffs mehrere Schritte notwendig. Im ersten Schritt muss der Begriff einer Beobachtung zugänglich gemacht werden (Quantifizierung), im zweiten Schritt muss er über geeignete Verfahren (vgl. vertiefend Abschnitt 2.2.4) in den numerischen Raum überführt werden.

Die numerische Abbildung theoretischer Begriffe ist komplex, da sie meist keinen direkten Bezug zum „realweltlichen Phänomen“ besitzt. Eine Operationalisierung sei immer dann gelungen, wenn sie durch Präzision und Empiriefähigkeit ausgezeichnet ist (Mayntz, 1978). Es gilt daher, die Qualität der Kriterien anhand dieser Qualitätsdimensionen zu beurteilen.

Folgende Einteilung soll das Codebuch erhalten:

1. **Variablenname:** Er kann aus einem oder mehreren Wörtern bestehen und beschreibt das Kriterium, z.B. *„Ressort“*.
2. **Variablenkürzel:** Für eine übersichtlichere und kompaktere Dokumentation der Ergebnisse wird der Variablenname abgekürzt. Im Fall von *„Ressort“* z.B. durch *„R“*.
3. **Code der Variablenausprägung:** Er wird zusammen mit seiner Ausprägung in Wortform (z.B. *„1– Politik“*) oder bei einer quantitativen Inhaltsanalyse vorrangig in Form von Zahlen festgehalten. Wobei im letzten Fall auch die Dokumentation in Worten möglich ist.

5.2.1 Variablen und Variablenausprägung

Wie bereits in Abschnitt 2.2.4 ausführlich diskutiert wurde, existiert für die Abbildung eines empirischen Relationalsystems eine ganze Reihe von Skalen. Zur Erinnerung werden die Skalen, die im Folgenden Anwendung finden, in Tabelle 5.1 zusammengefasst.

Eine dichotom oder binär skalierte Variable hat lediglich zwei Ausprägungen, die sich gegenseitig ausschließen, z.B. Ja/Nein. Es handelt sich um disjunkte Elemente. Eine multinomial skalierte Variable hat mehr als zwei Ausprägungen, wobei kein Rückschluss auf

Tabelle 5.1: Variablentypen der Inhaltsanalyse

	Beispiel	
Variablentyp	Kürzel	**Variable**/Ausprägung
Binär skalierte Variable	**T**	**Indexnennung auf Titelseite**
	1	Ja
	2	Nein
Multinomial skalierte Variable	**R**	**Ressort**
	1	Politik
	2	Wirtschaft
	...	...
Metrisch skalierte Variable	**WB**	**Wortanzahl im Beitrag**

eine Reihenfolge zugelassen wird. Es handelt sich im Gegensatz zu einer ordinal skalierten Variable lediglich um eine Zuordnung von Kategorien, die zwar eine Verschiedenartigkeit, jedoch keine Rangfolge zum Ausdruck bringen (z.B. das Ressort eines Beitrags). Beide Variablentypen, dichotome und multinomiale, liegen damit auf dem untersten Skalenniveau. Das nächsthöhere Skalenniveau der Ordinalskala findet an dieser Stelle keine Anwendung, dafür jedoch die Kardinalskala. Die zugehörigen metrisch skalierten Variablen erlauben eine quantitative Charakterisierung der Merkmalseigenschaften und diverse statistische Methoden (z.B. Wortanzahl der Beiträge).

Abgesehen von den metrischen Variablen stellt sich bei den dichotomen und multinomialen Variablen die Frage nach der Adäquatheit der festgelegten Ausprägungen und Kategorien:

> „Es ist unmöglich, feste Regeln oder Patentrezepte für die Kategorienbildung zu nennen. Jedoch können unabhängig von einer konkreten Fragestellung und bestimmtem Datenmaterial einige Merkmale genannt werden, die unbedingt beachtet werden müssen. Zunächst einmal muß das Kategorienschema erschöpfend sein, was aber nicht heißt, daß die Möglichkeit bestehen muß, sämtliche Elemente eines Textes einer der Kategorien zuzuordnen, sondern die Forderung bezieht sich nur auf den in bezug auf die Fragestellung relevanten Inhalt eines Textes. Dieser muß mit dem Kategorienschema abgedeckt werden, um nicht auf die Erfassung relevanter Informationen zu verzichten. Eine weitere Voraussetzung dafür, daß ein Kategorienmuster zur Abbilung eines Textes geeignet ist, besteht in der Eindeutigkeit der Kategorien.“ (Lisch, 1978: S. 70f.)

Bei der Festlegung der Kategorien muss berücksichtigt werden, dass folgende Qualitätsdimensionen Beachtung finden:

1. **Angemessenheit des Skalenniveaus einer Variablen:** Es ist ein Skalenniveau zu wählen, bei dem die Variable den Gegenstand bei der Datenauswertung angemessen beschreiben kann. Soll z.B. die durchschnittliche Wortanzahl je Monat dargestellt werden, ergibt es kaum Sinn, die zugrunde liegende Variable Wortanzahl über eine Klassifizierung wie 0 50, 50 300 etc. zu bestimmen. Der Mittelwert liefert hier ungenaue oder u.U. sogar falsche Ergebnisse. Stattdessen ist daher die tatsächliche Anzahl an Wörtern zu verwenden.

2. **Vollständigkeit:** Ausprägungen sind so zu wählen, dass die Variable den Gegenstand erschöpfend erfassen kann. So sollte z.B. nicht der Fall eintreten, dass ein Beitrag nach einem Kriterium bewertet werden soll, jedoch alle vorab definierten Kategorien nicht zutreffend sind. In diesem Fall würde mindestens eine Kategorie fehlen. Entweder sind alle möglichen Kategorien vorab zu definieren oder eine Sammelkategorie wie z.B. „Restliche“ anzulegen.

3. **Exklusivität:** Es sind die Ausprägungen so zu wählen, dass sie trennscharf und exklusiv für eine Variable sind. Das heißt, Ausprägungen dürfen sich nicht überschneiden. Zu jedem Gegenstand darf nur eine einzige Ausprägung existieren.

5.2.2 Beurteilung der Güte der Variablen

Die Güte der Kriterien kann über die Qualitätsdimensionen Objektivität, Reliabilität und Validität beurteilt werden. Hinter diesen Begriffen steht der bereits erwähnte Anspruch, dass eine „Messung“ unabhängig vom Untersucher wiederholt werden kann und zu denselben Ergebnissen führt.

So wäre ein Untersuchungsinstrument wenig objektiv, wenn eine Person bei der Anwendung des Codebuchs auf einen spezifischen Beitrag zu anderen Ergebnisse käme als eine andere Person. Es gilt daher, Objektivität zu gewährleisten. Eine Maßnahme in diesem Zusammenhang ist die Standardisierung des Kategoriensystems (Raupp und Vogelgesang, 2009: S. 168). Der Standardisierungsgrad eines Kategoriensystems oder eines Codebuchs ist hoch, wenn die darin enthaltenen Variablen vorab definierte Ausprägungen

besitzen, zwischen denen der Codierer wählen kann. Wird hingegen eine Variable in Form eines Freitextes vom Codierer beantwortet, ist Objektivität schwer zu gewährleisten. So kann sich z.B. die Ausführlichkeit oder Genauigkeit in der Begriffswahl je nach Codierer unterscheiden.

Das Untersuchungsinstrument ist reliabel, wenn es wiederholt zu denselben Ergebnissen führt, d.h. die Untersuchung reproduzierbar ist. Bei der vorliegenden Arbeit ergibt sich in dieser Hinsicht ein wesentlicher Unterschied zu komplett manuell durchgeführten Inhaltsanalysen. Die automatisierten Kriterien werden komplett transparent, nachvollziehbar und wiederholbar sein, da sie ausschließlich regelbasiert sind. Anders verhält es sich bei den manuell zu erhebenden Kriterien. Eine niedrige Reliabilität der Kriterien kann nach Neuendorf (2002) durch folgende Fehler beeinflusst werden:

- Präzision: Eine mangelhafte Präzision bei der Definition der Codierregeln, z.B. durch unvollständige oder fehlende Erläuterungen des Codes.
- Schulung: Die Codierer müssen für die Codierung des Untersuchungsmaterials geschult werden. Die Reliabilität wird gefährdet, wenn keine ausreichende Schulung der Codierer vorausging.
- Beanspruchung: Durch überbeanspruchte Codierer wird ebenfalls eine niedrige Reliabilität riskiert. Überbeanspruchung kann z.B. durch sehr viele Variablen je Untersuchungseinheit oder durch eine überdurchschnittlich hohe Zahl an zu prüfenden Untersuchungseinheiten auftreten.

Grundsätzlich kann der Grad der Reliabilität über den Reliablitätskoeffizienten gemessen werden. Dieser basiert auf dem Prinzip einer Messwiederholung. Die Reliabilität wird dabei über den Grad der Übereinstimmung bei der mehrmaligen Untersuchung desselben Untersuchungseinheit bestimmt. Die mehrmalige Untersuchung kann dabei auf zwei Arten durchgeführt werden: 1. die Reliabilität eines Codierers steht im Vordergrund und 2. die Reliabilität zwischen den Codierern steht im Vordergrund:

1. Intrarater-Reliabilität: Hierbei geht es um die Reliabilität eines Codierers. Dieser untersucht einen kleinen Teil des Untersuchungsmaterials zweimal. Danach wird überprüft, ob die Genauigkeit des Codierens bei der ersten Untersuchung mit der der zweiten übereinstimmt.

2. Interrater-Reliabilität (Urteilerübereinstimmung): Bei mehreren Codierern kann zusätzlich die Reliabilität zwischen den Codierern berechnet werden. Dafür codieren verschiedene Codierer (zeitgleich) dasselbe Untersuchungsmaterial. Die Reliabilität wird über den Grad der Übereinstimmung der verschiedenen Codierungsergebnisse bestimmt.

Neuendorf (2002: S. 142) ist der Ansicht, dass Analysen, bei denen der Entwickler des Codebuchs auch der Codierer des Untersuchungsmaterials ist, nicht als Inhaltsanalyse, sondern als Expertenanalyse bezeichnet werden sollten. Raupp und Vogelgesang (2009: S. 171) widerspricht dieser Auffassung unter der Bedingung, dass der Grad der Intra-Coderreliabilität dokumentiert wird.

Für einen Reliabilitätstest werden übereinstimmende und nicht übereinstimmende Codierentscheidungen aus den Codierergebnissen abgeleitet. Auf dieser Datenbasis wird der Reliabilitäskoeffizient berechnet. In der hiesigen Literatur findet sich eine ganze Reihe an Reliabilitätskoeffizienten. An dieser Stelle wird auf eine zusammenfassende Diskussion in Kolb (2004), Krippendorff (2004, 2013) und Lauf (2001) verwiesen.

Die oben genannten Qualitätsdimensionen Objektivität und Reliabilität sind erst aussagekräftig, wenn das tatsächlich angestrebte Objekt „gemessen“ wird. Es geht damit um den Gültigkeitsanspruch der „Messung“. Daher wird bei der Validitätsprüfung sichergestellt, dass eine angemessene Operationalisierungsentscheidung getroffen wurde. Die oben dargestellten Prinzipien der Variablenkonstruktion (Skalenniveau, Vollständigkeit und Exklusivität) sind dabei entscheidend. Wird z.B. überdurchschnittlich oft die Ausprägung *Sonstiges* gewählt, ist das ein Zeichen für ein unvollständiges Set an Ausprägungen und eine ungenügende Abbildung des zu operationalisierenden Gegenstands (Rössler, 2010: S. 194f.). Eine weitere Möglichkeit, die Validität zu prüfen, liegt im Vergleich mit anderen (überschneidenden) Studienergebnissen oder Inhaltsanalysen (Raupp und Vogelgesang, 2009: S. 173).

Abschließend bleibt in der Qualitätsbetrachtung das Zusammenspiel von Objektivität, Validität und Reliabilität zu diskutieren:

- Die Objektivität ist eine Voraussetzung für die Validität. Wenn das Untersuchungsinstrument nur in Abhängigkeit zu dem Codierer funktioniert, ist keine valide Messung möglich.

- Außerdem ist die Reliabilität eine Voraussetzung für die Validität. Wenn die Codierer wenig reliabel sind, d.h. bei Wiederholung andere Ergebnisse zu erwarten sind, kann nicht davon ausgegangen werden, dass die empirische Messung mit dem Messziel des theoretischen Modells übereinstimmt. Umgekehrt kann auch nicht angenommen werden, dass eine hohe Reliabilität eine hohe Validität zur Folge hat. Die Aussage ist lediglich, dass der Codierer reliabel also reproduzierbar arbeitet, verhindert jedoch z.B. nicht, dass etwas anderes gemessen und kodiert wird, als theoretisch festgelegt.

Da in der vorliegenden Arbeit die Automatisierung im Vordergrund steht, stellen sich Fragen der Reliabilität in dieser Weise weniger. Die Ergebnisse der Kriterien sind transparent über den R-Code nachvollziehbar und komplett reproduzierbar. Reliabilitätstest erscheinen vor diesem Hintergrund nicht sinnvoll.

Für die wenigen manuell zu erhebenden Kriterien wird aus pragmatischen Gründen ebenfalls auf Reliabilitätstests verzichtet. Der Aufwand einer mehrmaligen Analyse der Beiträge wird als nicht verhältnismäßig erachtet. Die möglichen Ungenauigkeiten für die wenigen manuellen Kriterien werden akzeptiert.

5.3 Kriterien für die Selektion der Beiträge

Kriterien, die dazu führen, dass ein Beitrag aufgegriffen wird, werden gemäß Raupp und Vogelgesang (2009) als Aufgriffkriterien bezeichnet. In der vorliegenden Arbeit wird für die Selektion der Untersuchungseinheiten lediglich ein einziges Aufgriffkriterium angewendet: Die Nennung eines Index.

5.3.1 Beitrag mit Indexbezug

Die Grundgesamtheit aller Untersuchungseinheiten wird durch Beiträge mit Bezug zu einem Index gestellt. Der Indexbezug wird als gegeben angesehen, wenn der Index namentlich im Beitrag erwähnt wird. Problematisch sind dabei abweichende Schreibweisen, z.B. durch Rechtschreibfehler, Synonyme oder Abkürzungen.

Über die Anbindung einer Webdatenbank, in der zu diversen Begriffen und Themen u.a. Falschschreibungen, Synonyme oder Abkürzungen hinterlegt werden, wird versucht dieser Problematik entgegenzuwirken. Außerdem definiert das Kategoriensystem der Webdatenbank die Grundgesamtheit, indem es z.B. die Namen existierender Indizes unter dem Konzept *Index* eingruppiert. In Abschnitt 5.6 wird dieses Verfahren näher vorgestellt.

5.4 Kriterien für die Bewertung der Beiträge

5.4.1 Entwicklung der automatisch zu erhebenden Kriterien

Automatisch zu erhebende Kriterien können aus den Beiträgen entweder direkt übernommen oder regelbasiert abgeleitet werden. So kann das Veröffentlichungsdatum eines Beitrags nach der Identifikation direkt in die Ergebnisliste zum Codebuch kopiert werden. Für andere Kriterien wie z.B. Satzlänge muss der Wert über Regeln (z.B. „Zähle alle Wörter“), Funktionen oder Aggregationen aus dem Beitrag abgeleitet werden.

5.4.1.1 Direkte Kriterien

Bei der Veröffentlichung der Beiträge werden bereits wertvolle Informationen bereitgestellt, die jeweils als Kriterium für die spätere Analyse der Indizes aus den Beiträgen übernommen werden können.

5.4.1.1.1 Titel des Beitrags Der Titel lässt sich direkt übernehmen. Mit ihm lässt sich der Beitrag im Nachhinein transparenter recherchieren und auffinden. Er wird in der vorliegenden Arbeit jedoch nur in Sonderfällen Berücksichtigung finden, z.B. wenn der Inhalt des Titels eine entscheidende Zusatzinformation über die mediale Funktion von Indizes bereitstellt. Ein anderes Vorgehen wäre unübersichtlich, da z.B. allein der ifo-Index mit über 4000 Beiträgen eine nicht fassbare Zahl von Titeln mit sich bringen würde. Mit dem noch vorzustellenden Kriterium *Anzahl Wörter in der Überschrift* (Abschnitt 5.4.1.2.1.1) wird ein aggregierbares Kriterium präsentiert, welches z.B. über Häufigkeitsdiagramme oder Streudiagramme die Beiträge zu einem Index ganzheitlich und übersichtlich zu betrachten vermag.

5.4.1.1.2 Datum Über die Metadaten der Beiträge lässt sich i.d.R auch das Datum als Kriterium ausgelesen. Es wird insbesondere bei der quantitativen Auswertung der Beiträge eine entscheidende Rolle spielen. So werden z.B. Wochen-, Monats- oder Jahresbetrachtungen in folgender Form möglich: *Wie viele Beiträge werden je Monat zu einem bestimmten Index veröffentlicht?* Außerdem kann die Entwicklung der Beitragshäufigkeit über eine Zeitreihe hinweg betrachtet werden. Interpretationen hinsichtlich Relevanz oder Potential liegen nahe.

Außerdem ist das Datum entscheidend für die Bestimmung der Aktualität eines Beitrags. In Verbindung mit dem Datum der Indexveröffentlichung kann ein weiteres Kriterium gebildet werden.

5.4.1.1.3 Ressort Ein relevantes Beitragsattribut ist das Ressort. Tageszeitungen sind i.d.R. nach Ressorts organisiert. Das Ressort ist dabei ein Teil der Vollredaktion und bearbeitet ein bestimmtes Themengebiet oder eine bestimmte Rubrik. In der vorliegenden Arbeit wird das ausgewiesene Ressort daher als Proxy für das Themengebiet des Indexes genutzt und im späteren Teil der Arbeit mit dem durch die Konstruktion definierten Thema abgeglichen. In der vorliegenden Arbeit stehen die Beiträge der Frankfurter Allgemeinen Zeitung (FAZ) und Frankfurter Allgemeinen Sonntagszeitung (FAS) im Zentrum. Die Redaktion dieser Zeitung ist detailliert organisiert. Sie beinhaltet die folgenden elf Ressorts:

1. Politik
2. Wirtschaft
3. Finanzen
4. Sport
5. Feuilleton
6. Technik und Motor
7. Natur und Wissenschaft
8. Reiseblatt
9. Immobilien

Tabelle 5.2: Tabelle für die Zuordnung der Ressorts von FAZ und FAS

FAZ	FAS
Politik	Politik
Wirtschaft	Wirtschaft
Finanzen	Geld und Mehr
Sport	Sport
Feuilleton	Feuilleton
Technik und Motor	Technik und Motor
Natur und Wissenschaft	Wissenschaft
Reiseblatt	Reise
Immobilien	Immobilien
Beruf und Chance	Beruf und Chance
Rhein-Main Zeitung	Rhei-Main
Sonstiges	Gesellschaft

10. Beruf und Chance

11. Rhein-Main Zeitung

Die vorliegende Arbeit übernimmt für das Kriterium Ressort die elf Ressorts der FAZ. Die Ressorts von FAZ und FAS werden über die Tabelle 5.2 konsolidiert.

5.4.1.1.4 Autor Ein weiteres leicht automatisiert zu überführendes Kriterium ist der Autor oder Urheber des Beitrags. Die Tageszeitungen weisen in der Regel zu jedem Beitrag den Autor aus. Der Autor kann ein zeitungsinterner und zeitungsexterner Urheber, eine Nachrichtenagentur oder ein Gastautor sein.

Fragen im Zusammenhang mit dem Autor sind, ob er einen signifikant höheren Einfluss auf die Nennung eines Index innerhalb eines Beitrags hat als das Ressort. Ist die Nennung eines Index bei einem bestimmten Autor wahrscheinlicher als bei anderen? Aufgrund der thematischen Spezialisierung der Journalisten ist jedoch damit zu rechnen, dass eine derartige Hypothesenprüfung zu verzerrten Ergebnissen führt. Ist z.B. ein Journalist im Ressort Wirtschaft ausschließlich auf DAX-Meldungen spezialisiert, so wird er vermutlich einen signifikanten Einfluss auf die Wahrscheinlichkeit der Indexnennung haben. Jedoch ist dies nur eine Scheinkorrelation, da er sich thematisch mit diesem Bereich beschäftigen muss.

Für die Relevanz eines Index kann diese Kennzahl zumindest als Proxy in Betracht gezogen werden. Hypothese ist dabei, je mehr verschiedene Autoren einen Index zitieren,

desto höher ist seine Relevanz. Diese Hypothese wird freilich von der Mitarbeiterfluktuation und der absoluten Anzahl an Beiträgen mit diesem Index beeinflusst. Dennoch muss ein Index einen gewissen Bekanntheitsgrad besitzen, um zitiert zu werden. Ein Quotient wie Anzahl der Autoren je Beitrag mit Indexbezug kann die absolute Anzahl an Beiträgen relativieren.

Auf eine Gruppierung der Autoren wird hinsichtlich einer leichteren Automatisierung verzichtet. Folgende Gruppierung bleibt daher der Weiterentwicklung des R-Pakets vorbehalten:

1. Zeitungsinterner Urheber (Name, Redaktion oder Kürzel)
2. Nachrichtenagentur(en)
3. Gastautor
4. Zeitungsinterner und zeitungsexterner Urheber (z.B. dpa usw.)
5. Anderes Medium
6. Nicht feststellbar/keine Quelle

5.4.1.2 Indirekte Kriterien

Indirekte Kriterien werden über Regeln, Funktionen oder Aggregationen aus den Beiträgen abgeleitet. Die wichtigste Aggregationsform ist an dieser Stelle die *Anzahl*, die wichtigste Funktion der *Quotient* und die wichtigste Regel die "Wenn-Dann-Beziehung".

5.4.1.2.1 Wortanzahl In der Linguistik gibt es eine reichhaltige Diskussion zu dem „vagen" Begriff *Wort*. Sie wird unter dem Sammelbegriff Morphologie zusammengefasst. Als die Lehre von der Struktur der Wörter (vgl. Bauer (2003: S. 3ff.)) befasst sie sich mit der Form und der Bildung von Wörtern. Es geht um die sprachlichen Einheiten (Wortformen, Wortformentoken, Wortformentypes, Lexeme, Lemmata) und um ihre Kategorisierung (Inhaltswörter vs. Funktionswörter, offene vs. geschlossene Wortklassen, Wortarten) (Prün und Steiner, 2005: S. 227). Die erwähnten Begriffe wurden von der quantitativen Linguistik jedoch noch nicht hinreichend klar operationalisiert (ebd.: S. 227). Die Brauchbarkeit des vagen Begriffs *Wort* für die Sprachwissenschaft

ist äußerst umstritten. Er ist lexikalisch mehrdeutig und hängt sowohl vom Untersuchungsgesichtspunkt als auch von der interessierenden sprachlichen Ebene ab. Daher wird das *Wort* von den oben genannten Einheiten abgegrenzt. Für die Untersuchung der inneren Struktur werden Morpheme (Bestandteile von Wortgruppen) beschrieben und klassifiziert. Doch ist auch deren Operationalisierung strittig:

> „Es gibt zahlreiche Methoden zur Segmentierung des Textes in Morphe(me) [...] und die Zahl automatischer Segmentierungsverfahren wächst von Jahr zu Jahr. Sie liefern jedoch nie 'hundertprozentige' Resultate, was sie keineswegs disqualifiziert, sondern eher auf die Unschärfe der Morphemgrenzen hinweist."

Rothe (1988) schlägt eine simple und praktikable Lösung zur Operationalisierung von Morphen im Deutschen vor:

> „In order to distinguish a component from affixes, prefixes or further morphological elements, we took as criterion its autonomous existence in the German lexicon, that is, its existence as a lexeme that can also be used in isolation". (S. 125)

Das Verfahren läuft demnach auf einen Vergleich mit einem Wortlexikon hinaus, was für eine Automatisierung keine größeren Schwierigkeiten bedeuten sollte. Jedoch ist auch bei diesem Vorgehen mit Fehlern zu rechnen (vgl. Prün und Steiner (2005: S. 227)).

In der vorliegenden Arbeit wird die Häufigkeit der Wörter (im vagen Sinne) als Proxy für die Textlänge eingesetzt. Neben der Häufigkeit werden keine weiteren quantitativen Kennzahlen[3] erhoben, weshalb eine deutlich simplere Bestimmungsregel, als in Köhler, Altmann und Piotrowski (2005) diskutiert wird, angemessen erscheint. Das *Wort* wird für eine leicht umsetzbare Operationalisierung in der vorliegenden Arbeit definiert als durch ein Leerzeichen begrenzter Textausschnitt.

5.4.1.2.1.1 Anzahl Wörter in der Überschrift Die Überschrift ist über die Metadaten der Beiträge i.d.R. eindeutig identifiziert. Die Anzahl der Wörter in einer

[3] Wie z.B. Länge, Distribution, Polysemie, Polytextie, derivate Produktivität, kompositionale Produktivität, Wortarten, Alter, Herkunft.

Überschrift wird über das Trennungszeichen Leerzeichen bestimmt. Als Wort wird alles gezählt, was mindestens einen Buchstaben hat und entweder von einem Leerzeichen am Anfang und Ende begrenzt ist oder wenn das Wort am Anfang oder Ende der Überschrift steht. Bei dieser Kennzahl ist zu prüfen, ob die Länge der Überschrift in Zusammenhang mit dem Index gebracht werden kann.

5.4.1.2.1.2 Anzahl Wörter des Beitrages Die Anzahl der Wörter des Beitrages wird auf die gleiche Weise bestimmt wie die *Anzahl der Wörter im Titel* (5.4.1.2.1.1). Die Wörter des Titels werden an dieser Stelle nicht in die Berechnung einbezogen. Die Anzahl der Wörter des Beitrages wird in der vorliegenden Arbeit als Proxy für die Relevanz des Themas benutzt. Das Kriterium ist jedoch immer in Bezug auf die Anzahl an Beiträgen eines Zeitraums zu betrachten. Ein langer Beitrag spricht nicht für eine höhere Relevanz als viele kurze Beiträge. Abhilfe kann eine abgeleitete Kennzahl bieten, die die Summe an Wörtern aller Beiträge in einem Zeitraum umfasst.

Die Anzahl der Wörter eines Beitrages wird außerdem für diverse abgeleitete Quotienten eine Rolle spielen (z.B. bei der Länge eines Satzes: Abschnitt 5.4.1.2.4).

Für eine leichte Vergleichbarkeit der Indizes hinsichtlich ihrer durchschnittlichen Länge werden drei Kategorien gebildet. Als Bezug haben sie alle die Verteilung der durchschnittlichen Beitragslänge von allen Indizes:

- Kurzer Beitrag: Unter diese Kategorie fallen alle Indizes, bei denen die durchschnittliche Beitragslänge zu den unteren 25% gehört (Mittelwert $\leq$ 25%-Quantil).
- Mittlerer Beitrag: Zu dieser Kategorie gehören alle Indizes, deren durchschnittliche Beitragslänge größer ist als bei den unteren 25% und kleiner als bei den oberen 25% (25%-Quantil $<$ Mittelwert $<$ 75%-Quantil).
- Langer Beitrag: Unter diese Kategorie fallen alle Indizes, bei denen die durchschnittliche Beitragslänge zu den oberen 25% gehört (Mittelwert $\geq$ 75%-Quantil).

Auf Basis der empirischen Daten ist das 25%-Quantil definiert über 542,7 Wörter und das 75%-Quantil über 773,3 Wörter.

5.4.1.2.2 Anzahl Absätze Die Anzahl Absätze eines Beitrags ist eine wesentliche Zusatzinformation zu der Anzahl der Wörter, da über die Absätze eine Form der Beitragsgliederung erfolgt. Dieses Kriterium wird automatisch erhoben und erlaubt eine Vielzahl abgeleiteter Größen. Eine entscheidende Information ist, in welchem Absatz der Index erstmals genannt wird oder als relative Kennzahl betrachtet, nach wie viel Prozent aller Absätze die erste Nennung stattfand. Die beiden Kennzahlen können insbesondere bei längeren Beiträgen unterschiedliche Informationen bereitstellen.

5.4.1.2.3 Anzahl der Indexnennungen im Beitrag Bei dieser Kennzahl geht es um die Frage, wie oft der Index im Beitrag genannt wird. Dafür wird der Beitrag nach dem entsprechenden Index durchsucht. Es geht dabei um die explizite Nennung des Indexnamen. Indirekte Bezüge wie *„Er ist in den letzten 12 Monaten nicht gestiegen“* werden nicht berücksichtigt. Ebenfalls werden Falschschreibungen oder Abkürzungen nicht oder nur bedingt einbezogen. Im Abschnitt 5.7 werden nähere Informationen zu den Suchbegriffen je Index gegeben.

Folglich werden in der vorliegenden Arbeit sämtliche indirekte Nennungen ignoriert. Andernfalls würde die automatisierte Erhebung dieses Kriteriums unnötig kompliziert und fehleranfällig. Es bleibt weiterführenden Untersuchungen vorbehalten, den Fehler dieser Vereinfachung zu quantifizieren. Möglich wäre das z.B. über eine Stichprobe aus der Gesamtheit aller Beiträge mit Indexbezug. Diese Beiträge werden manuell nach der Anzahl der Indexnennungen analysiert und den Ergebnissen der automatischen Bestimmung gegenübergestellt. Der Effekt wird an dieser Stelle als zu vernachlässigen und der Zeitaufwand für die Untersuchung als zu hoch angesehen, gemessen an dem Mehrwert der Information.

5.4.1.2.4 Satzlänge Erste sprachstatistische Untersuchungen zur Satzlänge gehen auf Sherman 1988 zurück. Er untersuchte englische Prosaautoren und stellte eine allgemeine Tendenz zur Verkürzung der Sätze fest. Außerdem untersuchte er die Komplexität von Sätzen und die Variation der Satzlängen in verschiedenen Werken eines Autors. Die folgende Forschung behandelt u.a. die Satzlänge als Charakteristikum für den Stil oder die Lesbarkeit und Textverständlichkeit. (Sherman, 1888)

Zur Bestimmung eines *Satzes* oder der *Satzlänge* existieren verschiedene Varianten. In der quantitativen Linguistik hat sich eine pragmatische Definition für den *Satz* durchgesetzt:

> "[Ein] 'Satz' wird bei schriftsprachlichen Texten gemäß orthografischen Kriterien als die Folge von Wörtern bestimmt, die durch satzabschließende Interpunktion <.>, <!>, <?> markiert wird". (Best, 2005: S. 300)

Probleme, die aufgrund dieser Definition entstehen können, lassen sich über zusätzliche Kriterien ausräumen. Es ergibt sich z.B. die Frage, ob es sich bei der direkten Rede *Herr Müller sagte: „Heute versuche ich pünktlich zu sein"* um einen oder zwei Sätze handelt. Best (ebd.) halten beide Varianten für zulässig, solange es nicht um absolute Werte geht (S. 300). Weitere Probleme können durch Interjektionen oder Parenthesen entstehen, wenn z.B. Ausrufezeichen oder Fragezeichen innerhalb von Sätzen eingesetzt werden. Doch kann auch dies über Regeln gelöst werden.

Auch für die Bestimmung der Satzlänge bieten sich zahlreiche Varianten an. Die Länge einer Einheit kann z.B. nach der Zahl ihrer Konstituenten bestimmt werden. Jedoch beinhaltet die komplexe Einheit *Satz* mehrere direkte und indirekte Konstituenten. Eine andere Variante definiert die Satzlänge über die Zahl der Wörter (bezogen auf den deutschen Sprachraum). Auch denkbar ist die Bestimmung über *clauses* oder Teilsätze. Ein *clause* erfordert dabei das Vorkommen eines finiten Verbs. Verblose Sätze erfordern zusätzliche Regeln. (ebd.: 300) Hierarchisch betrachtet kann die Länge eines Satzes im Wesentlichen über folgende Einheiten bestimmt werden:

1. Laute/ Phoneme/Buchstaben
2. Silben oder Morphe
3. Wörter
4. rhythmische Einheiten oder Satzglieder/Phrasen
5. *clauses* bzw. Teilsätzen

Für die vorliegende Arbeit wird die Satzlänge pragmatisch über die Anzahl der Wörter gemessen. Die Definition eines Satzes erschöpft sich dabei an der satzabschließenden

Interpunktion <.>, <!>, <?>. Auf weitere Regeln wird verzichtet, da die oben dargestellten Probleme als zu vernachlässigen erachtet werden. Insbesondere weil der Aufwand für die Erstellung der Regeln nicht im Verhältnis zum Nutzen gesehen wird.

Die Satzlänge wird über die Anzahl der Wörter bestimmt und als Durchschnittswert je Beitrag ausgegeben. Anders gesagt berechnet er sich über die Anzahl der Wörter geteilt durch die Anzahl der Sätze. Es ergibt sich damit die durchschnittliche Anzahl Wörter je Satz. Die Kennzahl wird als Proxy für den Stil des Beitrags genutzt. Es wird zu prüfen sein, ob signifikante Unterschiede bei der Satzlänge existieren, z.B. je Aktualität, Ressort, Index u.v.m.

5.4.1.2.5 Nennung des Index im Titel Hierbei handelt es sich um ein Ja/Nein-Kriterium. Die Nennung des Index im Titel wird als Proxy für die Relevanz des Index für den Beitrag angesehen. Wird der Index bereits im Titel genannt, wird angenommen, dass er eine wesentliche Relevanz für den Inhalt des Beitrags besitzt. Analog zu dem Kriterium *Anzahl Indexnennung im Beitrag* können sich hierbei durch die Automatisierung Probleme mit Abkürzungen, Falschschreibungen oder nicht berücksichtigter Synonyme ergeben. Dieser Umstand wird aufgrund der Vorteile einer Automatisierung[4] in Kauf genommen.

5.4.1.2.6 Nennung des Index im ersten Absatz Ähnlich wie die Nennung des Index im Titel wird dieses Kriterium als Proxy für die Relevanz des Index für den Beitrag angenommen. Falls der Index bereits im ersten Absatz genannt wird (dazu gehört auch der Teaser oder eine Zwischenüberschrift), wird er als wesentlich für den Inhalt des Beitrags angesehen. Für die automatisierte Bestimmung dieser Kennzahl spielt die Länge des Absatzes keine Rolle. Eine sehr kurze Zwischenüberschrift wird damit einem langen ersten Absatz gleichgesetzt. Relevant ist nicht der Inhalt, sondern lediglich das Absatzzeichen. Es bleibt Folgeuntersuchungen vorbehalten, ob Einschränkungen wie *„Durchsuche den ersten Absatz mit mindestens X Wörtern“* zu aussagekräftigeren Ergebnissen führen.

[4] Die Vorteile sind u.a. Entlastung von Codierern, Verarbeitung größerer Datenbestände, hohe Reliabilität.

5.4.1.2.7 Nennung des Index im Beitrag Dies ist ein Ja/Nein-Kriterium. Es geht darum aufzuzeigen, ob der Index wörtlich im Beitrag genannt wird. Damit ist es ein weiteres Anzeichen für die Relevanz des Index für den Beitrag. Grundsätzlich kann jedoch davon ausgegangen werden, dass dieses Kriterium beim Großteil der Beiträge erfüllt sein wird. Schließlich basieren die Beiträge auf einer Suche nach den Indizes. Falls der Index nicht im Beitrag vorkommt, ist es denkbar, dass er ausschließlich in der Überschrift genannt wurde oder, dass die angewendete Selektion (z.B. der Suchmaske) ihn aus bestimmten Gründen in der Ergebnisliste angezeigt hat. Diese Gründe sind in jedem Fall zu recherchieren.

5.4.1.2.8 Aktualität Dieses Kriterium ist die Basis für das Kriterium *Anlass* (Kapitel 5.4.1.2.9) und wird über die Zeitspanne zwischen Indexveröffentlichung und Beitragsdatum berechnet. Die Obergrenze der möglichen Tage wird damit im Wesentlichen vom Intervall der Indexveröffentlichung gesteuert.

Die Berechnung dieses Merkmals ermöglicht eine differenziertere Analyse der Aktualität eines Beitrags, da die Anzahl der Tage zwischen Indexveröffentlichung und Beitragsdatum in Rohform betrachtet werden kann. Jedoch entwickelt diese Maßzahl erst durch das Definieren von Grenzwerten (siehe Kapitel 5.4.1.2.9) oder über eine aggregierte Betrachtung eine Information. Zwei Grafiken werden in diesem Zusammenhang wesentlich sein: Ein simples Balkendiagramm mit der relativen Anzahl der veröffentlichten Beiträge, abgetragen auf die Aktualität des Beitrags in Tagen (Tage zwischen Indexveröffentlichung und Beitragsdatum). Über die Verteilung können wertvolle Informationen abgeleitet werden. Folgende Fragestellungen sind in diesem Zusammenhang denkbar:

1. Wann wurden seit Indexveröffentlichung die meisten Beiträge veröffentlicht?
2. Folgt die Beitragsveröffentlichung einer Verteilung? Dabei könnte eine Gleichverteilung für eine Unabhängigkeit vom Ereignis der Veröffentlichung stehen, eine Exponentialverteilung für eine Abhängigkeit, z.B. indem kurz nach der Indexveröffentlichung relativ viele Beiträge publiziert werden und die Häufigkeit mit jedem weiteren Tag verteilungsgemäß abnimmt.
3. Außerdem kann ein Anstieg der relativen Häufigkeit zum Ende der Zeitreihe für eine Ankündigung der Indexveröffentlichung sprechen. Sollte solch ein Anstieg im

Abbildung 5.2: Beispiel für ein Liniendiagramm mit kumulierten Häufigkeiten

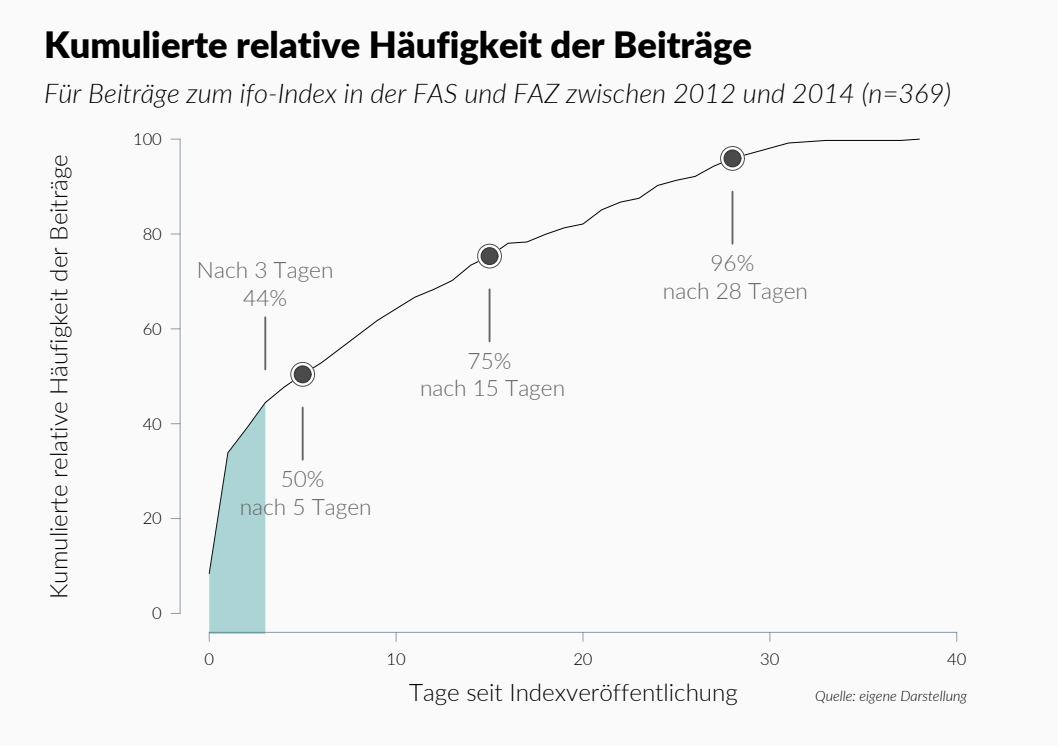


Diagramm erkennbar sein, sind die entsprechenden Beiträge manuell auf die genauen Gründe zu prüfen.

Die zweite sinnvolle Darstellung ist das Liniendiagramm mit den kumulierten relativen Häufigkeiten (Abbildung 5.2). Es werden die im Balkendiagramm je Tag dargestellten relativen Häufigkeiten kumuliert und z.B. in Form einer Linie dargestellt. Endpunkt der Linie sind immer 100%, d.h. die Gesamtheit aller einbezogenen Beiträge. Dem Diagramm kann u.a. entnommen werden, wie viel Prozent der Beiträge nach vier Tagen seit Indexveröffentlichung publiziert wurden (vgl. Kapitel 5.4.1.2.9). Je höher dieser Anteil ist, desto eher kann davon ausgegangen werden, dass Beiträge zu diesem Index in Folge der Indexveröffentlichung publiziert wurden.

Grundsätzlich wird die Vermutung aufgestellt, dass insbesondere im Ressort Wirtschaft ein großer Anteil von Beiträgen innerhalb der ersten Tage nach Indexveröffentlichung publiziert wird. Danach sollte der Zuwachs der kumulierten relativen Häufigkeit kontinuierlich abnehmen, d.h. die Kurve wird flacher. Je linearer der Kurvenverlauf, desto weniger sollte der Beitrag mit der Veröffentlichung des Index zusammenhängen. Je höher der Anstieg am Nullpunkt des Diagramms, desto mehr ist davon auszugehen, dass die

Aktualität eine große Relevanz hat. Folgende Fragestellungen sind im Zusammenhang mit dem Liniendiagramm denkbar:

1. In welchem Zeitraum ist der Anstieg des Graphen am steilsten, d.h. in welchem Zeitraum wurde der verhältnismäßig größte Teil an Beiträgen veröffentlicht?
2. Wie viel Prozent der Beiträge wurden z.B. in den ersten vier Tagen veröffentlicht?

5.4.1.2.9 Anlass Bei dem Kriterium Anlass geht es im Wesentlichen darum, ob das Zitieren des Index auf eine Veröffentlichung des Index zurückgeht. Als Proxy dient die Zeitspanne zwischen der Veröffentlichung des Index und dem Datum des Beitrags. Im ersten Schritt werden zu einem bestimmten Index die Veröffentlichungstermine recherchiert und gespeichert. Beim ifo-Index schwanken diese z.B. monatlich zwischen dem 15. und 26. Tag. Im zweiten Schritt wird zu jedem Beitragsdatum der unmittelbar davor liegende Veröffentlichungstermin zugeordnet. Die Zeitspanne zwischen diesem Veröffentlichungstermin und dem Beitragsdatum in Tagen stellt die Grundlage für das Kriterium Anlass dar. Ein Beitrag geht genau dann auf einen Veröffentlichungstermin zurück, wenn die Zeitspanne zwischen diesen beiden Tagen maximal vier Tage beträgt. Alle anderen Beiträge werden per Definition aus anderen Gründen publiziert.

Einschränkungen bei dieser Berechnungslogik entstehen bei Vorankündigungen. Beispielsweise kann der Fall eintreffen, dass unmittelbar vor der Veröffentlichung des Index eine Terminmeldung herausgegeben wird. Dies ist oftmals bei Kennzahlen zur Börse der Fall. Die Anzahl der Ankündigungen ist an dieser Stelle zu vernachlässigen. Es wird darauf verzichtet, die Berechnung der Zeitspanne um einen Tag zu verschieben, d.h. Beiträge mit einem Datum einen Tag vor der Veröffentlichung zu dieser Veröffentlichung zu zählen. Vielmehr ist über die grafische Darstellung der kumulierten relativen Häufigkeit je Index zu prüfen, wie viele Beiträge unmittelbar vor einer Indexveröffentlichung publiziert wurden.

Für das Kriterium Anlass dient die Zeitspanne von Indexveröffentlichung und Beitragsveröffentlichung als Basis. Die Artikelveröffentlichung wird als Folge der Indexveröffentlichung angesehen, wenn die Zeitspanne kleiner als vier Tage ist. Als Nebenbedingung gilt, dass die Beitragsveröffentlichung nicht vor der Indexveröffentlichung liegen darf, d.h. die Zeitspanne keine negativen Werte annehmen kann. Es ergibt sich damit die dichotome Skala:

1. Aktuell, wenn 0 Tage $\leq$ Zeitspanne $<$ 4 Tage
2. Nicht aktuell, wenn Zeitspanne $\geq$ 4 Tage

Der zusätzlicher Skalenpunkt $<$ 24h wird als zu gering erachtet, der Skalenpunkt $<$ 7 Tage als zu groß, um von Aktualität sprechen zu können. Es wird damit bewusst von der Zeiteinteilung tagesaktuell, wochenaktuell, weniger als wochenaktuell Abstand genommen.

5.4.1.2.10 Auf Titelseite Dieses Kriterium fragt, ob der Beitrag auf der Titelseite veröffentlicht wurde. Es handelt sich um ein Ja/Nein-Kriterium. Es wird als Proxy für die Relevanz des Index benutzt. Ein grober Einblick in die Daten lässt vermuten, dass ein Großteil der Indizes nicht auf der Titelseite genannt wird.

Im FAZ-Archiv existiert zu jedem Beitrag der FAZ die Information zur Seitenzahl in der Form *„Frankfurter Allgemeine Zeitung, 18.03.2014, Nr. 65, S. 23“*. Die Seitenzahl lässt sich automatisiert aus diesem Text-String entnehmen.

5.4.1.2.11 Illustration vorhanden Über dieses Kriterium wird festgestellt, ob innerhalb des Beitrages eine Grafik und/oder ein Foto vorhanden ist. Es handelt sich um ein Ja/Nein-Kriterium. Basierend auf der vorhandenen Datengrundlage ist dies lediglich ein halb-automatisierbares Kriterium. Die Daten aus dem FAZ-Archiv erfordern einen zu hohen Automatisierungsaufwand, da neben dem reinen Textbeitrag die Darstellung der Seite als PDF-Dokument untersucht werden müsste. Ebenfalls ist es nicht aussagekräftig, den Beitragstext auf Schlüsselwörter wie z.B. Diagramm, Grafik etc. zu durchsuchen. Zwar könnte davon ausgegangen werden, dass bei einem Schlüsselwort wie *„siehe Diagramm“* tatsächlich eine Grafik im Beitrag vorhanden ist, jedoch kann damit nur der Teil erfasst werden, in dem die Grafik explizit genannt wird, was z.B. bei rein illustrierendem Bildmaterial (Foto) nicht zwingend gegeben ist. Es ist damit zu rechnen, dass die Ergebnisse insbesondere durch die nicht identifizierten Diagramm-/Grafik-Beiträge verzerrt sind. Eine manuelle Prüfung des Kriteriums ist in Zusammenhang zur Genauigkeit zu bevorzugen.

5.4.1.2.12 Anzahl Zahlen Die Anzahl der Zahlen im Beitrag zeigt, ob ein Indexwert genannt wird. Grundsätzlich kann ein Index innerhalb eines Beitrages verschieden

behandelt werden. Neben der reinen inhaltlichen Nennung dienen auch Vergleiche wie größer/kleiner als Aussage. Auch die Erwähnung eines spezifischen Indexwertes ist denkbar. Grundsätzlich ist es sinnvoll, dies anhand eines Kriteriums festzumachen. Doch stellt die automatisierte Feststellung eine Herausforderung dar. Reicht die bloße Suche nach einer Zahl im Beitrag aus? Davon ist nicht auszugehen. Schließlich würden dann auch Jahreszahlen, Seitenzahlen oder das Alter von Personen zu einem positiven Ergebnis führen.

Eine zweite Annäherung könnte durch die Nähe einer Zahl und namentlicher Indexnennung hergestellt werden. Jedoch entstehen dann Probleme, wenn der Indexwert erst im Folgesatz oder noch später genannt wird und der Indexname durch Pronomen ersetzt wird, z.B. in der Form *„er stieg um 3 %“*. Sicherlich könnte durch ein kompliziertes Netz aus Bedingungen ein Großteil der Eventualitäten und Ungenauigkeiten abgefangen werden, jedoch werden in den meisten Fällen mit einer bestimmten Regel teilweise auch korrekte Treffer ausgeschlossen werden. Aus diesem Grund ist eine simple und überschaubare Regel vorzuziehen. Einen Ansatz bieten die Indexwerte selbst. Diese werden von den Herausgebern zur Verfügung gestellt und können als Filterkriterium eingesetzt werden, beispielsweise in Form eines Wertebereichs. Beim Durchsuchen des Beitrags nach Indexwerten werden über diesen Filter lediglich Zahlen berücksichtigt, die innerhalb des definierten Wertebereichs liegen. Zwischen 1991 und 2015 lag der ifo-Index z.B. immer zwischen den Werten 83,5 und 114,2. Alle Beiträge mit Bezug zum ifo-Index und einer Zahl zwischen 83,5 und 114,2 würden nach dieser Regel zur Nennung des Indexwertes gezählt werden. Problematisch sind dabei Indizes mit einer größeren Spannweite. Der DAX schließt z.B. in der historischen Zeitreihe Indexwerte ein, die potenziell als Jahreszahl angesehen werden können und umgedreht (z.B. 2003: ca. 2000). Wegen des erhöhten Aufwands einer Automatisierung wird das Kriterium in der vorliegenden Arbeit nicht erhoben, jedoch für die Weiterentwicklung des R-Pakets eingeplant.

5.4.1.3 Zusammenfassung der automatisierten Kriterien

Aus der vorangegangenen Diskussion ergeben sich die automatisch zu erhebenden Kriterien. Von den diskutierten fasst Tabelle 5.3 nur jene zusammen, die in der vorliegenden Arbeit Verwendung finden.

Tabelle 5.3: Automatisch zu erhebende Kriterien für die Inhaltsanalyse

Erhebungsart	Kriterium	Ausprägung
Direkt	Titel	Text
	Datum	Datum
	Ressort	Politik, Wirtschaft, Finanzen, Sport, Feuilleton, Technik und Motor, Natur und Wissenschaft, Reiseblatt, Immobilien, Beruf und Chance, Rhein-Main Zeitung
	Autor	Text
Indirekt	Wörter/ Titel	Zahl
	Wörter/ Beitrag	Zahl
	Anzahl Absätze	Zahl
	Satzlänge	Zahl
	Anzahl Indexnennung/ Beitrag	Ja/ Nein
	Indexnennung/ Titel	Ja/ Nein
	Indexnennung/ 1. Absatz	Ja/ Nein
	Indexnennung/ Beitrag	Ja/ Nein
	Anlass	Ja/ Nein
	Aktualität	Zahl
	Titelseite	Ja/ Nein

5.4.2 Entwicklung der nicht automatisch zu erhebenden Kriterien

Nicht alle Kriterien können automatisiert erhoben werden. Entweder weil sie grundsätzlich zu komplex sind, um in Form von Regeln, Filtern oder Funktionen zusammengefasst zu werden oder weil der Aufwand dafür unverhältnismäßig groß ist.

Um den manuellen Aufwand möglichst gering zu halten, wird für die Bewertung der Kriterien eine Stichprobenziehung angewendet (vgl. Abschnitt 5.7). Mit überschaubaren Auswirkungen auf die Aussagekraft der Ergebnisse wird damit der manuelle Aufwand auf das Minimum reduziert.

5.4.2.1 Indirekte Kriterien

5.4.2.1.1 Genre Journalistische Beiträge unterteilen sich in verschiedene Genres. Die Wahl einer bestimmten Textform kann eine Aussage zum Inhalt liefern. Es wird erwartet, dass die Klassifikation der Beiträge nach dem Genre eine Korrelation zu den Indizes hat. Eine Hypothese in diesem Zusammenhang ist, dass die Beiträge zu einem bestimmten Index keine signifikanten Unterschiede bezüglich des Genres haben. Beim

Vergleich der Beiträge unterschiedlicher Indizes hingegen wird ein signifikanter Unterschied der Genres erwartet. Im Abschnitt 5.7 werden diese Hypothesen auf statistische Signifikanz getestet.

Die Genres der Beiträge werden über die folgende Festlegung klassifiziert. Sie basiert im Wesentlichen auf dem Lehrbuch *Medienresonanzanalyse* von Juliana Raupp und Jens Vogelgesang. Dieses Werk behandelt inhaltsanalytische Verfahren, die unter dem Begriff *Medianresonanzanalyse* zusammengefasst werden. Die theoretischen Erkenntnisse dienen dort als Grundlage für die Entwicklung eines medienresonanzanalytischen Untersuchungsinstruments. (Raupp und Vogelgesang, 2009: S. 151ff.)

1. (Kurz-) Meldung, Kurznachricht, Notiz: Eine Meldung, Kurznachricht oder Notiz ist die kürzeste Textform. In der hierarchischen Abfolge steht das Wichtigste am Anfang. Vorgänge werden kurz und sachlich dargestellt. Das allgemein interessierende Ereignis wird auf das Minimum beschränkt. In der Regel handelt es sich dabei um die Darstellung eines bereits stattgefundenen Ereignisses, jedoch ist auch eine Ankündigung für zukünftige Ereignisse denkbar - z.B. für die bevorstehende Veröffentlichung neuer Werte zum ifo-Index.

2. Nachricht: Bei der Nachricht handelt es sich um die kompakte Darstellung eines allgemein interessierenden Ereignisses. Alle für das Thema relevanten W-Fragen werden beantwortet.[5] In der hierarchischen Abfolge steht das Wichtigste am Anfang. Danach folgen Elemente wie Einzelheiten, Quelle und Hintergrund. Jedoch kann ihre Reihenfolge variieren. Sie richtet sich nach der Zweckmäßigkeit.[6]

3. (Nachrichten-) Bericht: Der (Nachrichten-) Bericht ist eine Nachricht in ausführlicher Fassung, was in der Länge deutlich wird. Außerdem informiert er umfassend über Ereignisse und vermittelt Details, Zusammenhänge, Vorgeschichte und Hintergründe, die über den Nachrichtenkern hinausgehen. Der Stil ist zwar sachlich-objektiv, kann ferner jedoch Urteile, Einschätzungen und Stellungnahmen von Experten beinhalten.

4. Veranstaltungshinweise mit kurzem redaktionellen Text.

[5] Dazu gehören die Fragen: Wer? Wann? Wo? Wie? Warum? Woher?
[6] Vgl. Bausteine-Modell und Prinzip der umgekehrten Pyramide in Schwiesau (2016: S. 33ff.).

5. Interview: Interviews werden häufig in der dialogischen Struktur, d.h. explizit in der Frage- und Antwortform dargestellt. Sie basieren auf einem realen Gespräch zwischen Journalist(en) und Interviewpartner.

6. Reportage: Bei der Reportage handelt es sich um einen Erlebnisbericht. Sie ist tatsachenbetont, jedoch persönlich gefärbt. Man spricht von einer lebendigen journalistischen Darstellungsform, die den Leser an den Ort des Geschehens führt.

7. Kommentar, Glosse, Kritik oder Leitartikel: Bei Kommentar, Glosse, Kritik oder dem Leitartikel handelt es sich um meinungsäußernde Darstellungsformen. Der Kommentar besteht aus der Meinung eines einzelnen Verfassers zu einem Sachverhalt, über den bereits an anderer Stelle berichtet wurde. Bei der Glosse handelt es sich um eine kurze meinungsäußernde Darstellungsform, oft mit einer sehr spitzen Argumentation und einem heiteren und ironischen Unterton. Die Kritik dagegen bezieht sich auf eine Publikation, eine Institution oder eine Veranstaltung. Der Leitartikel befasst sich weitestgehend mit Entwicklungen und Trends und weniger mit tagesaktuellen Themen. Der Leitartikel gibt die Meinung des Verfassers oder z.T. auch die einer ganzen Redaktion wieder.

8. Dokumentation: Wird ein Dokument oder ein Vorfall 1:1 wiedergegeben, handelt es sich um eine Dokumentation.

9. Leserbrief: Bei dem Leserbrief handelt es sich um eine schriftliche Meinungsäußerung oder Information eines Lesers zu einem bestimmten Thema. In der Regel bezieht sich der Verfasser auf einen vorangegangenen Beitrag, indem er zustimmt, widerspricht, richtigstellt oder ergänzt.

10. Eigenständige Grafik oder Tabelle: Auch eigenständige Grafiken und Tabellen sind in Form von Infografiken oder Diagrammen denkbar. Sie werden unabhängig von den o.g. Genres präsentiert. Fotoreportagen, Bilder oder Fotos spielen bei Indizes keine Rolle und werden an dieser Stelle vernachlässigt.

11. Sonstiges: Diese Kategorie beinhaltet alle Beitragsarten, die den oberen nicht zugeordnet werden können.

5.4.2.1.2 Konflikt/Kontroverse/Problem Bei diesem Merkmal geht es um die Thematisierung von Konflikten, Kontroversen, Problemen oder Schaden. Es handelt sich

um eine dichotome Variable mit den Ausprägungen *Ja* und *Nein*, d.h. kommt vor und kommt nicht vor. Bei positiver Ausprägung wird von einer negativen Akzentuierung ausgegangen.

5.4.2.1.3 Überraschung Mit dem Kriterium Überraschung wird der Beitrag dahingehend charakterisiert, ob sein Autor ihr keinen, einen geringen oder einen großen Überraschungsgrad zugeschrieben hat. Überraschend ist dabei ein Ereignis, das nicht ankündbar ist oder welches im Widerspruch zu bestehenden Erwartungen steht. Keine Überraschung bedeutet in diesem Zusammenhang, dass das Ereignis ankündbar ist. Dazu gehören z.B. Veröffentlichungstermine, Eröffnungen oder Jubiläen. Der Mittelpunkt der multinomialen Skala ist zugleich der Normalfall: die geringe Überraschung. Es handelt sich dabei um ein Ereignis, welches passiert und nicht ankündbar ist. Jedoch ist aus dem Beitrag nicht ersichtlich, dass spezifische Erwartungen bestanden, die im Widerspruch zu dem Ereignis stehen. Der Endpunkt der Skala ist die große Überraschung. Ein Beitrag, der diesem Skalenpunkt zugeordnet wird, enthält ein Ereignis, welches den Erwartungen widerspricht. Die Ankündbarkeit des Ereignisses ist irrelevant, jedoch muss der Beitrag Elemente enthalten, mit denen die enttäuschte Erwartung kenntlich wird (z.B. der ifo-Index stieg *„überraschend“/„unerwartet“* stark).

5.4.2.1.4 Zentralität Das Ziel dieses Kriteriums ist es, die inhaltliche Verwendung des Index zu charakterisieren. Bezogen auf den Konstruktionsprozess (vgl. Abschnitt 4.3.1) besitzt jeder Index eine Definition des theoretisch zu messenden Phänomens. Bewegt sich die Argumentation des Beitrags innerhalb dieser Begriffsdefinition, wird das dichotome Kriterium mit der Ausprägung *Ja* besetzt, andernfalls mit *Nein*.

Außerdem wird dieses Kriterium von den statistischen Methoden innerhalb der Indexkonstruktion beeinflusst. Die statistischen Methoden beeinflussen wesentlich die Verwendung der Indexwerte. Bei einem zu geringen Skalenniveau (ordinale Skala) ist z.B. eine Mittelwertbildung nur bedingt geeignet. Wird im Beitrag gegen statistische Regeln widersprochen, wird der Index außerhalb seines Definitionsbereichs benutzt und das Kriterium wird mit *Nein* besetzt. Ein *Nein* bei der theoretischen oder statistischen Definition ist ausreichend, um das Kriterium generell mit *Nein* zu bewerten.

Da die Trennschärfe des theoretischen Rahmens mit dessen Begriffsdefinition oftmals kritisiert werden kann und nicht immer transparent zur Verfügung steht, wird an dieser Stelle auf dieses Kriterium verzichtet. Zusätzliche Schwierigkeiten kann die klare thematische Eingrenzung der Verwendung des Index im Beitrag mit sich bringen. Die Objektivität wird stark von der Transparenz und Ausführlichkeit der Indexkonstruktion beeinflusst, die Reliabilität vom Wissensstand und dem Überblick des Codierers. Die Validität dieses Kriteriums kann damit nur schwer gewährleistet werden.

5.4.2.1.5 Quellenangabe Die explizite Angabe der originalen Quelle über die Veröffentlichung des Index muss manuell erhoben werden. Die Regel ist, dass dieses dichotome Ja/Nein-Kriterium nicht zutrifft. Beiträge, wie z.B. zum *Happy Planet Index*, in denen die URL zu dem aktuellen Happy Planet Index Report genannt wird, sind bemerkenswert und ggf. signifikant über eine Teilmenge der Indizes (vgl. Abschnitt 5.7.3). Daher wird dieses Kriterium stichprobenartig erhoben.

5.4.2.1.6 Vergleich Hierbei wird geprüft, ob der Indexwert eine vergleichende Funktion erfüllt. Dieses multinomiale Kriterium hat die folgenden Ausprägungen:

- Kein Vergleich: Diese Ausprägung ist zutreffend, wenn kein Indexwert genannt oder lediglich der absolute Wert des Index ohne Vergleiche genannt wird.
- Zeitlich: Diese Ausprägung ist zutreffend, wenn ein Indexwert mit vorhergehenden Werten in Bezug gesetzt wird (z.B. *ifo-Index steigt deutlich* oder *der ifo-Index steigt im Vergleich zur Vorperiode um 2%*).
- Regional, national: Diese Ausprägung ist zutreffend, wenn ein Indexwert in Beziehung zu dem Wert einer anderen Region gesetzt wird (z.B. *Vanuatu ist das glücklichste Land, weil es im Ländervergleich des Happy Planet Index auf Platz 1 liegt.*).
- Sozial, gesellschaftlich: Diese Ausprägung ist zutreffend, wenn ein Indexwert in Beziehung zu dem Wert einer anderen Population gesetzt wird (z.B. der Body Mass Index von Jugendlichen in Beziehung zu dem von Erwachsenen).
- Sachlich: Der Indexwert wird in Beziehung zu anderen Größen oder Indizes gesetzt (z.B. der ifo-Index steigt und der DAX zieht nach).

5.4.2.1.7 Transparenz Es wird untersucht, ob der Index vom Leser zweifelsfrei identifiziert werden kann. Die bloße Bezeichnung Vertrauensindex ist z.B. intransparent, weil eine ganze Reihe von *„Vertrauensindizes“* in den Medien existiert (GPRA Vertrauensindex, CCI Consumer Confidence Index etc.). Daher ist es je nach Index notwendig, zusätzliche Informationen, wie z.B. den Herausgeber des Index zu nennen. Dieses multinomiale Kriterium hat die folgenden Ausprägungen:

- Keine: Die Informationen des Beitrags sind unzureichend, um weitere Informationen zu dem Index (über andere Quellen) abrufen zu können.
- Gering: Der Beitrag nennt mindestens den eindeutigen Namen des Index oder den Namen der Institution, die den Index herausgegeben hat.
- Moderat: Kriterien von *gering* + Nennung von mindestens einem Attribut des Index (erste Veröffentlichung, Items, Befragungstyp, Zyklus etc).
- Hoch: Kriterien von *moderat* + mehr als ein Index-Attribut

5.4.2.1.8 Argumentation Über dieses Kriterium wird analysiert, ob der Autor den Index für die Argumentation benutzt und ob er zentrales Element des Beitrags ist. Diese dichotome Variable hat die Ausprägung Ja/Nein oder trifft zu und trifft nicht zu.

5.4.2.2 Zusammenfassung der manuell zu erhebenden Kriterien

Tabelle 5.4 fasst von den oben diskutierten Kriterien nur jene zusammen, die in der vorliegenden Arbeit Verwendung finden.

5.4.3 Zusammenfassung der Bewertungskriterien

Die o.g. Bewertungskriterien werden individuell für jeden Index und jeden Beitrag bestimmt und in eine Datenbank gespeichert. Sie werden in der vorliegenden Arbeit jedoch nur in aggregierter und teilweise transformierter Form präsentiert. Die Daten werden damit strukturiert und überschaubar gemacht.

Tabelle 5.4: Manuell zu erhebende Kriterien für die Inhaltsanalyse

Kriterium	Ausprägung
Genre	Meldung/ Kurznachricht, Nachricht, Bericht, Veranstaltungshinweise, Interview, Reportage, Kommentar/ Glosse/ Kritik/ Leitartikel, Dokumentation, Leserbrief, Grafik/ Tabelle, Sonstiges
Konflikt/Kontroverse/Problem	Ja/ Nein
Überraschung	Liste
Quellenangabe	Ja/ Nein
Vergleich	Liste
Transparenz	Liste
Argumentation	Ja/ Nein

Für die automatisch erhobenen Kriterien Titelseite, Wörter/Beitrag, Anzahl/Absätze, Wörter/Titel, Anzahl Indexnennung/Beitrag, Satzlänge und Aktualität werden das Minimum und Maximum, der Median und der Mittelwert und der mittlere Fehler bestimmt. Grundlage sind die absoluten Zahlen. Für die automatisch erhobenen Kriterien Indexnennung/Titel, Indexnennung/1. Absatz und Indexnennung/Beitrag wird jeweils die relative Häufigkeiten in Prozent als Grundlage für die bereits genannten Aggregationsmethoden benutzt.

Für das automatisch erhobene Kriterium Ressort wird die absolute und relative Häufigkeit an Beiträgen ausgegeben.

Für die automatisch erhobenen Kriterien Datum und Aktualität wird jeweils ein Diagramm für die Darstellung der Häufigkeitsverteilung ausgegeben. Aggregiert wird beim Veröffentlichungsdatum nach Monat oder Jahr und bei der Aktualität nach Tagen.

5.5 Wahl des Untersuchungsmediums

In Deutschland erscheinen derzeit mehrere hundert Tageszeitungen mit unterschiedlichen Titeln (Wilke, 2009: S. 459). Zu den Tageszeitungen zählen dabei Publikationen, die im Wesentlichen an jedem Werktag erscheinen, mindestens jedoch zweimal pro Woche. Außerdem enthalten sie einen aktuellen politischen Teil mit abgegrenzter Nachrichtenvermittlung. Nahzu alle Zeitungen erscheinen dabei als Morgenzeitungen. (ebd.: S. 460ff.)

In der vorliegenden Arbeit wird die *Frankfurter Allgemeinen Zeitung* (FAZ) sowie die *Frankfurter Allgemeinen Sonntagszeitung* (FAS) als Medium für die automatisierte und manuelle Inhaltsanalyse benutzt.

Die FAZ gehört zu einer der sieben überregionalen Tageszeitungen. Das bedeutet, dass ihre Auflage zum größten Teil im gesamten Bundesgebiet verbreitet ist. Neben *Die Welt* wird sie als Qualitätszeitung bezeichnet. Die verkaufte Auflage der FAZ[7] vom 4. Quartal 2010 bis zum 3. Quartal 2015 liegt zwischen 363.620 und 264.628. Die der FAS liegt im gleichen Zeitraum zwischen 361.133 und 276.030.[8]

Die FAZ gehört zur Verlagsgruppe *Frankfurter Allgemeine Zeitung/Frankfurter Societäts-Druckerei GmbH*. Sie hat einen Marktanteil von 3,0% (Wilke, 2009: S. 468). Politisch betrachtet kann die FAZ als gemäßigt rechts eingestuft werden (ebd.: S. 472).

Neben ihrer Funktion als einschlägige Tageszeitung qualifiziert die FAZ für die Inhaltsanalyse ihre historische Vorreiterrolle in der Verwendung von Indizes. Im *Weltwirtschaftlichen Archiv* bemerkte Paul Hermberg bereits 1920 in seinem Aufsatz "Die Indexzahlen der Frankfurter Zeitung" die Pionierarbeit der Zeitung:

> „Die 'Frankfurter Zeitung' hat sich mit ihren in großzügiger Weise und mit ungewöhnlicher Intensität ausgebauten 'Indexzahlen' in Deutschland an die Spitze dieser Bewegung gestellt. Sie will selbstverständlich in erster Linie der Praxis dienen und bringt die Resultate ihrer Berechnungen mit einer staunenswerten, vor allem in Deutschland angenehm überraschenden Geschwindigkeit." (S. 537)

Zu jener Zeit ging es noch hauptsächlich um Indizes von Preisen, Löhnen, Lebenserhaltung, Valuta oder Börsenkursen. Hermberg fasst den Wert und die Funktion der Indizes innerhalb der Kommunikation folgendermaßen zusammen:

> "Naturgemäß darf man diese Zahlen nicht wie eine wissenschaftliche Arbeit beurteilen. Die Art der Berechnung wird komplizierte Feinheiten vermeiden müssen, nicht immer allen Anforderungen welche die Wissenschaft an die Methode stellt, gerecht werden können. Auch kleine Ungenauigkeiten

[7] Montag bis Samstag.

[8] Zahlen wurden bereitgestellt über die *Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern e. V. (IVW)* unter `http://www.ivw.eu` (besucht am 17. Apr. 2015).

> wird man in den Kauf nehmen. Immerhin kann es nicht gerade als erfreulich begrüßt werden, daß die 'Frankfurter Zeitung' ihre Berechnungen auf eine Grundlage stellt, die im gleichen Augenblick in England als unzulänglich aufgegeben wird [...]. Im ganzen muß man aber doch den Versuch, den die Zeitung mit Kühnheit und Energie unternommen hat, durchaus freudig begrüßen. Auch das Ausland hat den neuen deutschen Indexzahlen weitgehende Beachtung geschenkt." (S. 537)

In diesem Zusammenhang ist auch der FAZ-Index zu nennen, der seit 1961 ermittelt wird und bis zur Einführung des *Deutschen Aktienindex* (DAX) 1988 der wichtigste Index auf dem deutschen Aktienmarkt war. Mittlerweile verteilt die FAZ eine ganze Reihe von Indizes. Das Konzept sowie detaillierte Konstruktionsinformationen sind transparent und nachvollziehbar im *Leitfaden zur F.A.Z. Europa Indexfamilie* zusammengefasst (aktuelle Fassung: Version 1.2 vom 22. November 2011). Die FAZ nutzt somit nicht nur Indizes innerhalb der Kommunikation, sondern sie erstellt und verbreitet diese mit einer langen Tradition auch im eigenen Haus.

Bei der Entwicklung des R-Paketes zur Analyse von digitalen Medien (vgl. Abschnitt 5.1.2) wurde darauf geachtet, dass das Untersuchungsmedium variabel und damit unabhängig vom hier gewählten Medium der FAZ ist. Somit kann das R-Paket prinzipiell für jedes digitale Medium verwendet werden. Da die Plattformen für digitale Inhalte in der Regel jedoch unterschiedlich strukturiert sind, und sich darüber hinaus auch oft die zugrunde liegende Metasprache der digitalen Inhalte unterscheidet, müssen für jedes Medium bestimmte Plattform-Parameter hinterlegt werden. In der aktuellen Version des Pakets wurden die Parameter für drei verschiedene Plattformen hinterlegt.

Obwohl das entwickelte Paket somit die automatisierte Analyse von drei verschiedenen Nachrichten-Webseiten unterstützen würde, wurden für die empirische Inhaltsanalyse ausschließlich Beiträge der FAZ und FAS verwendet. Das Einbeziehen weiterer Medien hätte den Rahmen der vorliegenden Arbeit gesprengt. Die empirische Untersuchung ist damit aussagekräftig für die FAZ und FAS, jedoch nur unter Einschränkungen auf die Gesamtheit der überregionalen Tageszeitungen zu verallgemeinern.

5.6 Selektion der Beiträge

Sollen über einen längeren Zeitraum Printmedien bezüglich eines Gegenstandbereiches untersucht werden, ist mit einem erhöhten zeitlichen und kognitiven Aufwand zu rechnen. Vor diesem Hintergrund liegt es nahe, die Chance zu nutzen und digitalisierte Printmedien maschinell zu verarbeiten. Der Personalaufwand sinkt dadurch und Ergebnisse sind in deutlich kürzeren Zeiträumen lieferbar. Jedoch besteht die Herausforderung, Textinhalte maschinenlesbar zu machen und sicherzustellen, dass die maschinellen Ergebnisse die gleiche Qualität erreichen, wie manuelle. Die Vorteile einer maschinellen Selektion lassen sich wie folgt zusammenfassen:

1. **Aufwand:** geringerer personeller Aufwand.
2. **Umfang:** Möglichkeit der Auswertung kompletter Zeitreihen (z.B. alle Ausgaben zwischen 1950-2015).
3. **Dauer:** sofortige Bereitstellung von Ergebnissen.
4. **Fehler:** Fehlselektionen sind ausschließlich auf die entwickelten Regeln zurückzuführen. Im Vergleich zur manuellen Selektion besteht z.B. keine Gefahr von konzentrationsbedingten Fehlern.
5. **Monitoring:** Möglichkeit der kontinuierlichen Überwachung ganzer Themengebiete.

In der vorliegenden Arbeit geht es ausschließlich um die Selektion von Beiträgen mit Indexbezug. Indizes sind, wie in Kapitel 4 ausführlich dargestellt wurde, theoretische Konzepte. In der Regel hat der Namen des Index einen Bezug zu diesem theoretischen Konstrukt. Für die Selektion von Beiträgen mit Indexbezug genügt zunächst der trivialste Ansatz, die Beitragsinhalte nach dem Namen der Indizes zu durchsuchen. Dieses Vorgehen läuft auf eine Schlüsselwortsuche hinaus. Sollen z.B. Beiträge mit Bezug zum *DAX* selektiert werden, müssen alle verfügbaren Beiträge hinsichtlich des Schlüsselwortes „DAX“ durchsucht werden.

An dieser Stelle offenbaren sich zwei Herausforderungen. Die erste besteht in der Identifikation des Index. Das eingangs erwähnte Beispiel der Klimaerwärmung legt nahe, dass eine im Printmedium genannte Maßzahl vermutlich nicht immer zweifelsfrei über

die im Beitrag selbst gelieferten Informationen als Index identifiziert werden kann, und somit z.B. als naturwissenschaftliche Messung erscheint. Es müssen externe, gesicherte Informationen zur Selektion der Beiträge verwendet werden. Im einfachsten Fall wäre dies eine Liste der Namen von mehrdimensionalen Konstrukten.

Das Beispiel des „DAX“ bringt unter der Nutzung solch einer Liste jedoch eine zweite Herausforderung zum Vorschein: In der Realität existieren zu dem Index verschiedene Schreibweisen oder Abkürzungen. Bei der Betrachtung ist es notwendig, auch solche Beiträge zu selektieren, die mit abweichenden Bezeichnungen, wie *Deutscher Aktienindex* oder *DAX30* arbeiten. Um die Gesamtheit aller relevanten Beiträge zu filtern, muss daher mit einer Keyword-Liste gearbeitet werden, die im Idealfall sämtliche abweichende Schreibweisen und Falschschreibungen enthält.

Zusammenfassend gilt es für eine maschinelle Selektion von Beiträgen folgende Herausforderungen zu bewältigen:

1. Erstellung einer Liste mit existierenden Indizes.
2. Erweiterung der Liste um alle existierenden Schreibweisen.

Der Qualitätsunterschied von maschineller und manueller Selektion ist schwer abschätzbar. Insbesondere bei komplexeren Entscheidungen ist zu befürchten, dass die dargestellte maschinelle Methode über Schlüsselwörter Beiträge vernachlässigt, die eigentlich selektiert werden müssen. Beispielhaft könnte dafür eine Formulierung wie *Index für die 30 größten Aktiengesellschaften* sein. Diese komplexe Regel ist nur mit größerem Aufwand maschinell umsetzbar. Die Frage ist in diesem Zusammenhang, bei wie vielen Beiträgen der Index inhaltlich zwar vorkommt, aber nicht namentlich genannt wird.

In der vorliegenden Arbeit wird von einer ausreichenden Qualität der maschinell selektierten Beiträge ausgegangen und in Anbetracht des beträchtlichen Mehraufwands auf eine Quantifizierung des Qualitätsunterschieds verzichtet. Dies soll Folgeuntersuchungen vorbehalten bleiben. Eine Methode zur Quantifizierung des Qualitätsunterschieds sei an dieser Stelle kurz eingeführt.

Um die Abweichung zwischen manueller und maschineller Selektion zu quantifizieren, kann eine Stichprobe hilfreich sein. Dafür werden aus einem längeren Zeitraum zufällig

eine Reihe von Ausgaben selektiert.[9] Im vorliegenden Fall handelt es sich dabei um Ausgaben der *Frankfurter Allgemeinen Zeitung*, die sowohl maschinell als auch manuell untersucht werden. Die Selektionsergebnisse werden danach gegenübergestellt und eine Fehlerquote berechnet. In solch einem Test muss auch mit gegenteiligen Ergebnissen gerechnet werden. So sind durch Ungenauigkeit, Konzentrationsprobleme oder Überarbeitung auch bei der manuellen Selektion Fehler denkbar.

Das dargestellte Vorgehen zur Quantifizierung des Unterschieds kann auch für einen iterativen Prozess zur Verbesserung der maschinellen Selektion genutzt werden. Sobald der Unterschied einen gewissen Grenzwert überschreitet, werden die Selektionsregeln überarbeitet und der Unterschied erneut quantifiziert. Diese Anpassungsschritte werden so lange durchgeführt, bis eine vorab definierte Fehlerschwelle erreicht wurde.

5.6.1 Erstellung der Schlüsselwortliste

Wie in Abschnitt 3.2 gezeigt wurde, ist die Verbreitung von Indizes seit Anfang des 19. Jhd. stark angestiegen und es gibt keine Anzeichen, dass sich diese Entwicklung ändern könnte. Um auch in Zukunft auf neue entstehende Indizes reagieren zu können, ist eine Schlüsselwortliste dynamisch zu erstellen. Das bedeutet, dass sie bei Neuveröffentlichungen erweitert wird.

In der vorliegenden Arbeit wird die Schlüsselwortliste über einen Zugriff auf die Online-Datenbank *DBpedia* erstellt. Die Inhalte von *DBpedia* basieren auf den Inhalten einer freien Enzyklopädie und werden fortwährend erweitert, womit auch zukünftig erscheinende Indizes abgedeckt sein sollten. Vor jeder Ausführung der maschinellen Beitragsselektion wird die Liste der Schlüsselwörter mit *DBpedia* abgeglichen, womit jederzeit die größtmögliche Aktualität sichergestellt wird.

Bei dem dargestellten Vorgehen ist zu bemerken, dass die resultierende Liste an Indizes aktuell nicht vollständig ist. Das bereits genannte Beispiel der Klimaerwärmung ist beispielsweise nicht enthalten. Vielmehr konzentriert sich die Liste auf mehrdimensionale Konstrukte, die meist die Bezeichnung „Index“ im Namen tragen. Zum aktuellen Zeitpunkt stellt sie dennoch die umfangreichste Sammlung dar und wird deshalb als Basis

[9] Die notwendige Anzahl an Ausgaben richtet sich nach der geforderten Messgenauigkeit. Für weitere Informationen zu Stichproben sei auf Albers (2009: S. 79ff.) verwiesen.

für die Selektion von Beiträgen verwendet. Das entwickelte R-Paket ist mit seinem modularen Aufbau in dieser Hinsicht flexibel. Für Folgeuntersuchungen können alternative Listen angebunden werden und die Unterschiede der Ergebnisse untersucht werden.

5.6.1.1 Abfrage der DBpedia

Als Grundlage für die Suche und Identifikation existierender Indizes werden Informationen auf Basis der *Wikipedia* benutzt. Dem Verfasser ist dabei die weitreichende Diskussion über die Qualität der Inhalte dieser Plattform bewusst. Es wird an dieser Stelle ausschließlich auf das Kategoriensystem, d.h. auf die Beziehungen zwischen Entitäten sowie den Subjekt-Titeln zurückgegriffen, nicht jedoch auf die Inhalte. Aus pragmatischen Gründen ist die Abfrage der Daten daher nachvollziehbar: Da die Plattform eine freie Enzyklopädie darstellt, kann davon ausgegangen werden, dass Neuerscheinungen oder Neuentwicklungen rasch Beachtung finden.

Die Informationen der *Wikipedia* sind jedoch maschinell schwer zu verarbeiten. Eine strukturierte Extraktion von *Wikipedia*-Daten ist äußerst anspruchsvoll. Aus diesem Grund wurde als Endpunkt für den Zugriff auf die *Wikipedia*-Daten das Projekt *DBpedia* genutzt. Dieses Gemeinschaftsprojekt wird von der *Universität Leipzig*, der *Universität Mannheim* des *Hasso-Plattner-Instituts* und von *OpenLink Software* bearbeitet. Ziel ist es, strukturierte Daten aus der *Wikipedia* zu extrahieren und als solche zur Verfügung zu stellen. Grundsätzlich existieren in *Wikipedia* sowohl unstrukturierte Daten in Form von Fließtext als auch strukturierte Daten, z.B. in Form von Infoboxen, Tabellen, Kategorien, geographische Koordinaten und Weblinks. Das Projekt extrahiert die Daten aus *Wikipedia* und speichert sie als *Resource Description Framework* (RDF) ab. Mit RDF lässt sich eine formale Beschreibung der Semantik festlegen sowie Beziehungen von Subjekt, Prädikat und Objekt definieren. Die Beziehungen sind beispielhaft zum Eintrag *Max Planck* in Abbildung 5.3 dargestellt.

Die extrahierten Daten werden zusätzlich mit weiteren freien Datensammlungen[10] über RDF verbunden.

[10]Dazu gehören z.B. Freebase, Open Cyc, UMBEL, GeoNames, MusicBrainz, CIA World Factbook, das Linked Open Data-Projekt der New York Times, Digital Bibliography & Library Project, Project Gutenberg, Jamendo, Eurostat oder die US-Volkszählung.

Ein ähnliches Projekt *Yago* wird vom *Max Planck Institut* entwickelt. *Yago* ist eine semantische Datenbank, die nicht nur auf Wissen aus der *Wikipedia* basiert, sondern zusätzlich auf *WordNet*[11] und *GeoNames*[12]. 2012 beinhaltet die Datenbank mehr als 10 Millionen Entitäten (z.B. Personen etc.) mit insgesamt über 120 Millionen Fakten. Die Beziehungen wurden nach Angaben von *Yago* manuell geprüft und als 95% korrekt bestätigt.

In *Yago* wird es ermöglicht, die Taxonomie von *WordNet* mit dem Kategorie-System von *Wikipedia* zu verbinden. Da *Wordnet* jedoch ausschließlich für die englische Sprache existiert, kann der Zugang *Yago* an dieser Stelle nicht weiterverfolgt werden. Eine beispielhafte Abfrage für den Typ *Index* über *Yago* findet sich in Anhang C.

Ein Teil des internationalen *DBpedia*-Projektes ist *DBpedia Deutschland*. Genau wie im Gesamtprojekt geht es um die Bereitstellung strukturierter Daten – jedoch mit dem Fokus auf deutsche Benutzer. Ziel ist die Integration der deutschen Informationen der *Wikipedia* in die globale *DBpedia*, in der nationale Inhalte miteinander verknüpft sind. Da die Schlüsselwörter als Basis für die Beitragsselektion der deutschen *Frankfurter Allgemeinen Zeitung* dienen sollen, wird an dieser Stelle die deutsche Version von *DBpedia* abgefragt. Dafür findet der Online-Abfrage-Editor unter der URL `http://de.dbpedia.org/sparql` Verwendung. Um die gewünschten Informationen zu erhalten, müssen die Entitäten der Datenbank gefiltert werden. Dafür werden nur Ressourcen selektiert, deren zugeordnetes Konzept oder die Subjektbezeichnung den Term „index“ enthält (vgl. Syntax 5.1: Zeile 13, 14). Der Term darf dabei nicht unmittelbar von Buchstaben gefolgt sein (z.B. „Po**index**ter“). Neben der Seiten-URL werden, falls vorhanden, Informationen zur Länge des *Wikipedia*-Beitrags (Syntax 5.1: Zeile 7), zum Namen (Syntax 5.1: Zeile 11), zum Konzept (Syntax 5.1: Zeile 5,) und zum Typ (Syntax 5.1: Zeile 9) abgefragt. Zusätzlich werden noch Teile des Fließtextes (Einleitung, Kommentar) des Wikipedia-Beitrags (Syntax 5.1: Zeile 20, 18) extrahiert.

Syntax 5.1: SPARQL-Code zum Abrufen aller Einträge zum Konzept oder Label „Index“

```
1 PREFIX dcterms: <http://purl.org/dc/terms/>
2 SELECT DISTINCT ?Seite ?wikiPageLength ?subject_label ?concept ?type
3     ?comment ?abstract
```

[11] WordNet ist ein lexikalisch-semantisches Netz der englischen Sprache. Es wird am Cognitive Science Laboratory der Princeton University entwickelt.

[12] GeoNames ist eine geografische Datenbank. Sie enthält z.B. geografische Namen, topografische Objekte, Koordinaten und Daten über die Bevölkerungsgröße.

```
4     WHERE
5       { ?Seite dcterms:subject ?concept .
6     OPTIONAL
7       {?Seite <http://dbpedia.org/ontology/wikiPageLength> ?wikiPageLength}
8     OPTIONAL
9       {?Seite rdf:type ?type}
10    OPTIONAL
11      {?Seite rdfs:label ?subject_label}
12    FILTER
13      (regex(STR(?concept), 'index') || (regex(?subject_label,
14      '.Index[^a-z]','i'))).
15    OPTIONAL
16      {?Seite <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> ?type}
17    OPTIONAL
18      {?Seite <http://www.w3.org/2000/01/rdf-schema#comment> ?comment}
19    OPTIONAL
20      {?Seite <http://dbpedia.org/ontology/abstract> ?abstract}
21  }
22  LIMIT 10000
```

Die gesamte Abfrage kann Syntax 5.1 entnommen werden. Ein Auszug der Ergebnisse der Abfrage befindet sich im Anhang in der Tabelle A.1.

Aus den Namen (*subject_label*) der Einträge, die dem Konzept „Index“ zugeordnet sind oder den Term „index“ beinhalten, wird eine Liste von Schlüsselwörtern erstellt, mit denen relevante Beiträge ausgewählt werden. Zusätzlich können die verschiedenen Schreibweisen über die *Property* „is dbpedia-owl:wikiPageRedirects of“ zugespielt werden. Dabei handelt es sich um vorab definierte Weiterleitungen (Redirects), mit denen der Seitenbesucher auf die Hauptseite des Artikels weitergeleitet wird. Für den Hauptartikel „DAX“ werden folgende Weiterleitungen und damit Schreibweisen identifiziert:

1. ĜDAXI
2. DAX30
3. L-DAX
4. L/E-Dax
5. DAX 30
6. ShortDAX
7. DAX-Index

8. Deutscher Aktienindex

Bei der Erweiterung der Such-Liste des *DAX* um die o.g. Weiterleitungen werden verschiedene Entscheidungen vorweggenommen. Z.B. muss die Entscheidung, ob *DAX* und *L-DAX* zusammengefasst betrachtet werden sollen, hinsichtlich der Relevanz der Tageszeit für die Auswertung getroffen werden. Denn der *L-DAX* folgt den gleichen Konstruktionsregeln wie der DAX, nur werden bei ihm lediglich die Kurswerte zwischen 17:30 Uhr und 20 Uhr einbezogen (*L* steht dabei für englisch „late“, d.h. „spät“).

Für internationale Untersuchungen kann die Such-Liste zusätzlich um weitere Sprachen ergänzt werden. Dafür ist das *Property* „owl:sameAs“ hilfreich. Es stellt u.a. die Beziehung zu den Einträgen der *Wikipedia* anderer Länder her.

Für spätere Auswertungsschritte kann es nützlich sein, zusätzlich die übergeordneten Kategorien der Einträge abzufragen. Dafür wird die *Property*“dcterms:subject“ genutzt. Bei Anwendung auf den *DAX* ergeben sich die übergeordneten Kategorien „Aktienindex“ und „Deutsche Börse AG“. Über die übergeordneten Kategorien können damit z.B. Aktienindizes von sozialwissenschaftlichen Kennzahlen unterschieden werden.

Die komplett automatisierte Umsetzung der oben dargestellten Methode mit R kann über die Funktion *getConceptData* des R-Pakets *analyseMedia* nachvollzogen werden (vgl. auch Abschnitt 5.1.2). Mit dieser Funktion wird automatisch die *DBpedia* zum Konzept „Index“ abgefragt. Optional lässt sich über die Funktion der SPARQL-Endpunkt, das Konzept und das Subjekt anpassen. In Anhang D.1 kann die Funktion direkt eingesehen werden.

Vor der Verwendung einer so erstellten Schlüsselwort-Liste ist eine manuelle Prüfung der Ergebnisse vorzunehmen. Teilweise werden z.B. Bezeichnungen für einen Index zurückgegeben, die nicht eindeutig sind und deren Anwendung zu unkorrekten Ergebnissen führen würde.

5.6.2 Auswahl zu analysierender Indizes

In der vorliegenden Arbeit wird aufgrund des umfangreichen manuellen Teils nur eine Submenge der möglichen Indizes untersucht. Die Selektionskriterien für die zu analysierenden Indizes werden in Abbildung 5.4 schematisch dargestellt. Im ersten Schritt wird

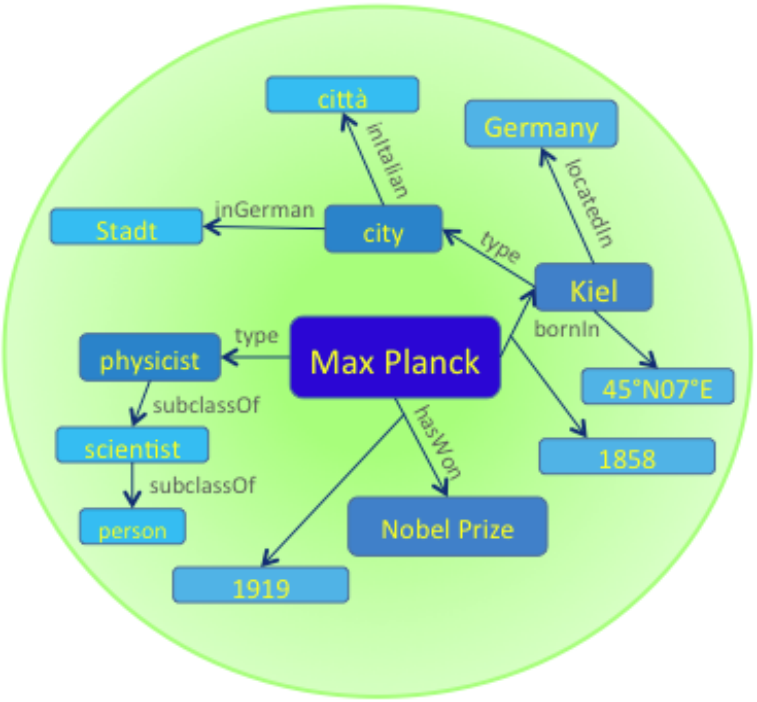
mit Hilfe der Abfrage 5.1 eine Liste mit in der Wikipedia gelisteten Indizes erstellt (siehe Tabelle A.1). Im folgenden Schritt werden alle Indizes ausgeschlossen, zu denen in der FAZ/FAS kein Beitrag zwischen 2010 und 2014 erschienen ist. Für die Inhaltsanalyse werden aus dieser Liste 15% stichprobenartig ausgewählt. Der DAX wird als täglich zitierter Index wegen seines großen Gewichts außerhalb der Stichprobenziehung in die Analyse aufgenommen. Damit werden im empirischen Teil der vorliegenden Arbeit die Beiträge folgender Indizes analysiert:

1. Better Life Index
2. Body-Mass-Index
3. DAX
4. GfK-Konsumklimaindex
5. Happy Planet Index
6. Human Development Index
7. ifo-Geschäftsklimaindex
8. Korruptionswahrnehmungsindex
9. University of Michigan Consumer Sentiment Index

5.6.3 Aufgreifen der Beiträge

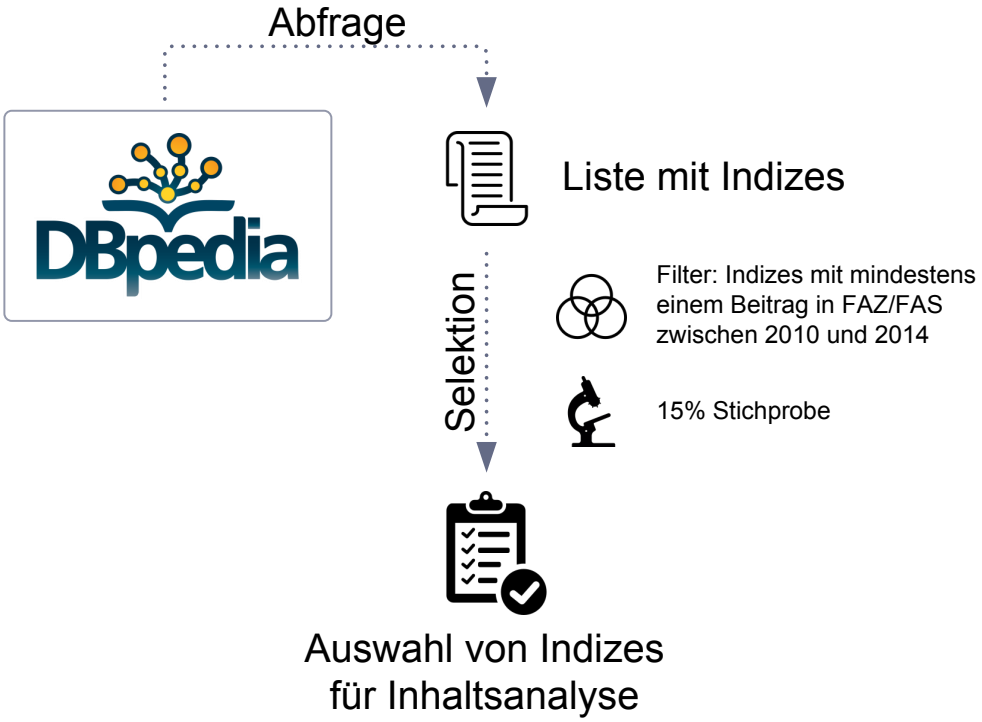
Nachdem eine Liste von Suchbegriffen erstellt wurde, können mit ihrer Hilfe Beiträge mit Indexbezug selektiert werden. Grundsätzlich kommen an dieser Stelle alle digitalen Medien mit Textbeiträgen in Frage. Idealerweise verfügt die Plattform, auf der die Beiträge zur Verfügung gestellt werden, bereits über eine integrierte Suchfunktion. Die Schlüsselwörter werden dann Index für Index in die entsprechende Suchmaske übertragen und die Ergebnisse abgelegt. Dieser Schritt muss nicht manuell durchgeführt werden. Bei der Nutzung von R kann im *R-Code* die entsprechende Such-Seite hinterlegt werden, sodass R die Abfrage der Ergebnisse übernimmt. Vorteilhaft ist dabei, dass die zurückgegebenen Seiten sofort auf relevante Informationen durchsucht und in Form

Abbildung 5.3: Yago-Graph



Quelle: Max Planck Institute for Informatics (2014)

Abbildung 5.4: Schema für die Auswahl der zu analysierenden Indizes



einer strukturierten Tabelle abgelegt werden können. Jedoch ist vor solch einem automatisierten *Webcrawling*[13] zu prüfen, ob der Herausgeber der Webseite oder Plattform dieses Vorgehen gestattet oder ob die Rechtsgrundlage das Vorgehen zulässt.

Bei der Durchführung des o.g. Verfahrens werden die Beiträge in der Regel als *HTML-Dokumente* abgespeichert. Diese enthalten meist eine auf der XML-Spezifikation basierende Metasprache, die für die strukturierte Zusammenstellung der relevanten Daten genutzt werden kann. Die Überschrift eines Artikels kann z.B. in folgendem Beispiel über *class="headline"* eindeutig identifiziert werden, was die automatische Weiterverarbeitung möglich macht:

Syntax 5.2: Beispiel für ein XML DOM - Node

```
1  <span class="'headline"'>
2    Stimmung an Weltbörsen kippt
3  </span>
```

Neben dem Fließtext werden auf diese Weise in den digitalen Beiträgen meist Attribute wie Datum, Titel, Teaser und Autor über die Metadaten ausgewiesen.

In der vorliegenden Arbeit werden Beiträge der *Frankfurter Allgemeinen (Sonntags-) Zeitung* betrachtet. Die gesamte Zeitreihe ist digital über das FAZ-Archiv abrufbar. Ein spezieller Zugang zum Archiv über das Bibliotheksportal verbietet jedoch explizit automatisierte Abfragen:

> „Das F.A.Z.-Bibliotheksportal ist ein Produkt des Frankfurter Allgemeine Archiv, powered by GENIOS. Nutzungshinweise: Artikel dürfen nur für den eigenen Gebrauch abgerufen werden. Eine Vervielfältigung der Artikel, die Weitergabe an Dritte und eine unlizenzierte Veröffentlichung ist untersagt. Der Download großer Artikelmengen, automatisierte Abrufe und missbräuchliche Nutzung ziehen eine Sperrung des Zugangs nach sich." (Frankfurter Allgemeine Archiv, 2015)

Daher muss der Abruf der Suchergebnisse manuell erfolgen. Andernfalls würde dieses Vorgehen von einem Rechner übernommen werden, der dem Archiv den Suchbegriff übermittelt und alle zurückgegebene Ergebnisse deskriptiv verarbeitet.

[13] Im vorgestellten Beispiel umfasst das Webcrawling das Sammeln aller durch die Suchmaschine zurückgegebenen Beitrags-URLs, die danach einzeln besucht und für die Analyse vorgehalten werden.

Da aus der automatisierten Inhaltsanalyse lediglich deskriptive Informationen hervorgehen, müssen keine Beitragsinhalte abgespeichert werden. Die deskriptiven Informationen sind über die Ausgabe (inkl. Datum, Titel und Seitenzahl) der einbezogenen Beiträge eindeutig reproduzierbar.

Die Digitalisierung von Inhalten bietet einen beträchtlichen Vorteil in Bezug auf quantitative Analysen. In diesem Zusammenhang erscheint es generell nützlich, insbesondere der Wissenschaft einen uneingeschränkten Zugang zu ermöglichen. Schließlich ist sie ein Treiber für Entwicklung.

5.6.3.1 Die Suchmaske des FAZ-Archivs

Neben einer Stichwortsuche für den gesamten Text, Titel und Serientitel bietet das FAZ-Archiv die Möglichkeit, nach Zeitraum, Ressort, Rubrik, Land und Branche zu filtern. Außerdem kann per Freitext nach Personen und Firmen gefiltert werden. Für die Selektion von Beiträgen für die vorliegende Arbeit sind lediglich die Stichwortsuche für den gesamten Text in Verbindung mit dem Zeitraum relevant. Letztere hat eine große Bedeutung für die Reproduzierbarkeit der Ergebnisse.

Die Stichwortsuche ermöglicht auch die Verwendung von Operatoren. So sind die Vergleichs-Operatoren „Und“ und „Oder“ anwendbar, sowie der Verknüpfungs-Operator „Nicht“. Außerdem lassen sich zusammengehörige Begriffe über Anführungszeichen definieren. Sollen Anfang oder Ende eines Begriffs offen bleiben, kann mit Platzhaltern gearbeitet werden. Dieser wird über ein Sternchen kenntlich gemacht (Bsp. *Benzin findet Superbenzin, Feuerzeugbenzin etc.).

Für das oben genannte Beispiel DAX ergibt sich folgender Such-String für das FAZ-Archiv:

```
1  DAX OR GDAXI OR DAX30 OR "‘L-DAX"’ OR "‘L/E-Dax"’
2  OR ShortDAX OR "‘DAX-Index"’ OR "‘Deutscher Aktienindex"’
```

Die in Abschnitt 5.6.1.1 genannte Weiterleitung „DAX 30“ kann ausgelassen werden, da dieser Begriff bereits über „DAX“ abgedeckt wird.

5.6.3.2 Die Ergebnisse des FAZ-Archivs

Die über die Suche zurückgegebenen Ergebnisse sind strukturiert. Neben dem Fließtext werden folgende Artikelattribute ausgegeben:

- Datum
- Quelle
- Ressort
- Titel
- Autor
- Teaser
- Anzahl Wörter
- Beitragstext

Das Attribut *Quelle* ist im Rahmen der vorliegenden Untersuchung für die weitere Einschränkungen notwendig. Neben der *Frankfurter Allgemeinen Zeitung* deckt das FAZ-Archiv noch andere Medien ab (FAZ.NET, Rhein Main Zeitung etc.). In der vorliegenden Arbeit werden jedoch nur die *Frankfurter Allgemeine Zeitung* und *Frankfurter Allgemeine Sonntagszeitung* betrachtet. Diese können über die Informationen des Attributs *Quelle* gefiltert werden. Jedoch ist eine weitere Filterung auf das Attribut *Ressort* notwendig, um die *Rhein Main Zeitung* von der *Frankfurter Allgemeinen Zeitung* zu differenzieren.

Das R-Paket ist, wie in Abschnitt 5.1.2 bereits erwähnt, modular aufgebaut. Das bedeutet, dass auch manuell digitalisierte oder manuell abgerufene Beiträge automatisiert in deskriptive Informationen überführt werden können. Dieser Vorteil wird für die deskriptive Beschreibung der Beiträge von FAZ und FAS genutzt.

5.7 Quantitative Inhaltsanalyse

5.7.1 Schema für die quantitative Inhaltsanalyse

Die Zusammenstellung der Ergebnisse aus der quantitativen Inhaltsanalyse erfolgt für jeden Index nach einem festgelegten Schema (vgl. Abbildung 5.5). Dabei erfolgt eine Trennung der Analyse in einen konstruktionsanalytischen Teil und einen inhaltsanalytischen Teil. Erstgenannter orientiert sich an den in Kapitel 4 erarbeiteten Schritten der Indexkonstruktion und an dem Konzept zur Beurteilung der Qualität eines Index aus Abschnitt 4.4. Der inhaltsanalytische Teil beinhaltet die Analyse der im vorangegangenen Teilkapitel 5.4 erarbeiteten Kriterien für die Bewertung von Beiträgen.

Die Analyse der Konstruktion wird im Wesentlichen von der Transparenz des Index beeinflusst, die vom Herausgeber des Index bestimmt wird. Einige Institute stellen z.B. die mathematische Konstruktion ihres Index nicht öffentlich zur Verfügung, andere hingegen machen selbst die zugrunde liegenden Individualdaten zugänglich, wodurch sie eine Reproduzierbarkeit der Ergebnisse ermöglichen. Diese Abweichungen in der Transparenz der einzelnen Indizes und die Abhängigkeit von den veröffentlichten Daten hat zur Folge, dass die jeweiligen Analysen im Umfang stark voneinander abweichen können. Der Grundaufbau folgt jedoch stets folgendem Schema – vorausgesetzt die entsprechenden Informationen stehen zur Verfügung:

1. Diskussion des theoretischen Rahmens (vgl. Abschnitt 4.3.1).
2. Diskussion der mathematisch/statistischen Grundeigenschaften (vgl. Abschnitt 4.3.2 bis 4.3.5).
3. Diskussion der verwendeten Darstellungsform (vgl. Abschnitt 4.3.7).
4. Beurteilung der Qualität des Index (vgl. Abschnitt 4.4).

Für die darauffolgende quantitative Inhaltsanalyse werden wie bereits angemerkt, die im Teilkapitel 5.4 erarbeiteten Kriterien als Basis benutzt. Sie werden zu verschiedenen Tabellen und Diagrammen verdichtet.

Je Index werden jeweils die folgenden drei Tabellen erstellt:

Tabelle 5.5: Beschreibung der numerische Kennzahlen

Kennzahl	Aussage
Titelseite	Anzahl der Beiträge, die auf der Titelseite veröffentlicht wurden.
Wörter/Beitrag	Durchschnittliche Anzahl an Wörtern je Beitrag.
Anzahl Absätze	Durchschnittliche Anzahl an Absätzen je Beitrag.
Wörter/Titel	Durchschnittliche Anzahl an Wörtern je Titel.
Anzahl Indexnennung/Beitrag	Durchschnittliche Anzahl an Nennungen des Index je Beitrag.
Indexnennung/Titel (Prozent)	Prozentualer Anteil der Titel mit Nennung des Index.
Indexnennung/1. Absatz (Prozent)	Prozentualer Anteil der Beiträge mit Nennung des Index im ersten Absatz.
Indexnennung/Beitrag (Prozent)	Nach wieviel Prozent des Beitrags wird der Index im Durchschnitt das erste Mal genannt?
Satzlänge	Durchschnittlich Satzlänge über alle Beiträge.
Aktualität	Durchschnittliche Anzahl Tage zwischen Beitrag und Pressemitteilung.

1. **Numerische Kennzahlen:** Diese Tabelle enthält alle automatisch erhobenen Kennzahlen (vgl. Abschnitt 5.4.1). Je Kennzahl werden Minimum, Median, Maximum, Mittelwert und Standardfehler ausgegeben, soweit dies inhaltlich für sinnvoll erachtet wird. Für eine bessere Verständlichkeit wird in Tabelle 5.5 der Inhalt der numerischen Kriterien kurz zusammengefasst. Da es sich bei dieser Tabelle um eine Aggregation aller Beiträge handelt, wurden neben dem Mittelwert auch Median, Minimum und Maximum ausgegeben. Dadurch soll die Beurteilung der zugrunde liegenden Verteilung erleichtert werden.

2. **Ressort-Statistik:** Diese Tabelle zeigt die absolute und die relative Anzahl an Beiträgen aufgeschlüsselt nach dem Ressort der Zeitung.

3. **Ergebnisse der manuellen Inhaltsanalyse:** Diese Tabelle enthält alle Ergebnisse der manuellen Inhaltsanalyse.

Neben den drei Tabellen dienen zwei standardisierte Diagramme der besseren Interpretierbarkeit der Ergebnisse:

1. Diagramm „Aktualität der FAZ und FAS Beiträge“: Als Zeitreihe beinhaltet dieses Diagramm die kumulierte Häufigkeit an Beiträgen über die Anzahl der Tage seit der Pressemitteilung oder Veröffentlichung des Index. Der eingefärbte Bereich zwischen 0 und 3 Tagen bezieht sich auf die Definition in Abschnitt 5.4.1.2.9. Alle Beiträge aus diesem Bereich werden unmittelbar der Pressemitteilung oder Veröffentlichung des Index zugeordnet. Der zugehörige Prozentwert wird direkt im

Diagramm ausgegeben. Außerdem werden drei weitere Werte (falls empirisch vorhanden) ausgegeben: das 50%-, 75%- und 95%- Quantil. Mit ihrer Hilfe kann die Aussage getroffen werden, nach wie vielen Tagen 50%, 75% und 95% der Beiträge veröffentlicht wurden. Das Diagramm ist insbesondere zur Bewertung der Relevanz eines Index zu nutzen.

2. Diagramm „Anzahl Artikel je Jahr in FAS und FAZ“: Dieses Diagramm zeigt im oberen Teil die Anzahl an Beiträgen pro Jahr. Die Länge der Zeitreihe wird dabei über die Lebensdauer des Index und über die Datenqualität bestimmt. Als letzter Datenpunkt wurde der 31.12.2014 festgelegt. Alle nach diesem Zeitraum veröffentlichten Beiträge finden keine Berücksichtigung. Im unteren Teil des Diagramms wird die mittlere Anzahl an Wörtern pro Beitrag und Jahr dargestellt. Die X-Achse der beiden Teildiagramme verläuft deckungsgleich. Damit kann für jedes Jahr auf einen Blick die Anzahl der Beiträge und die zugehörige durchschnittliche Anzahl an Wörtern pro Beitrag entnommen werden.

Da durch die Erstellung der drei Tabellen und zwei Diagramme eine erhebliche Informations- und Datenmenge entsteht, behält sich der Autor bei der Kommentierung und Interpretation an dieser Stelle ein gewisses Maß an Selektion vor. Für ihn und den Argumentationsstrang nicht relevant erscheinende Werte kann er somit ignorieren oder nicht vertiefend behandeln. Damit soll der Lesefluss gewahrt und eine zielführende Argumentation gesichert werden. In Abschnitt 5.7.11 werden hingegen alle Werte einbezogen, um einen umfangreichen Vergleich der betrachteten Indizes zu gewährleisten.

Nach der Konstruktionsanalyse des Index und der Inhaltsanalyse der Beiträge soll ein abschließender Abschnitt die erarbeiteten Ergebnisse zusammenführen und hinsichtlich der Funktion innerhalb der Kommunikation bewerten.

5.7.2 Human Development Index

Der *Human Development Index* (HDI) wird seit 1990 von den Vereinten Nationen berechnet und veröffentlicht. Er dient als Indikator für den Wohlstand der Länder und erscheint jährlich im Rahmen des Human Development Report des Entwicklungsprogramms der Vereinten Nationen.

Ausgangspunkt und Motivation des Index ist die eingeschränkte Aussagekraft des BIP. Das theoretische Konzept des Index basiert auf dem Anspruch, den Entwicklungsstand eines Landes messbar zu machen. Dieser sei wichtig, da er maßgeblich die Möglichkeiten der Menschen vergrößere. Dabei wird der Begriff Entwicklung betont nicht nur auf materiellen Wohlstand bezogen:

> „People are the real wealth of a nation. The basic objective of development is to create an enabling environment for people to enjoy long, healthy and creative lives. This may appear to be a simple truth. But it is often forgotten in the immediate concern with the accumulation of commodities and financial wealth." (Ross-Larson und Hanlon, 1990: S.9)

Das theoretische Konzept *menschliche Entwicklung* besteht aus drei Dimensionen. Die Konstruktion des Index wurde in der Vergangenheit mehreren Revision unterzogen, wodurch die Zeitreihe nur bedingt vergleichbar ist. Die aktuelle Indexkonstruktion basiert auf Festlegungen von 2010 und beinhaltet die folgenden drei Dimensionen (vgl. Abbildung 5.6): "Langes und gesundes Leben", „Wissen" und „Lebensstandard". Sie werden über die Subindizes Lebenserwartung (LEI), Bildung (BI) und Einkommen (EI) quantifiziert.

Der Subindex *Lebenserwartung* wird über den Indikator *Lebenserwartung zum Zeitpunkt der Geburt* bestimmt (vgl. Abbildung 5.6). Er wird für den Index normiert.

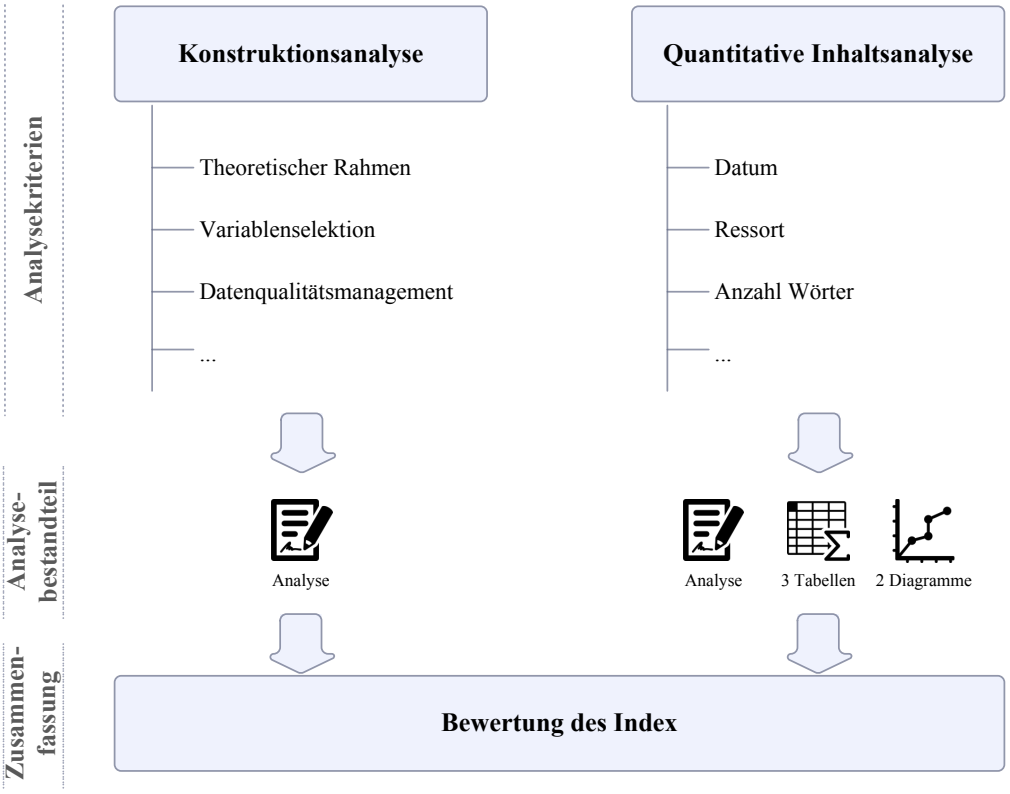
$$LEI = \frac{LE - 20}{83.6 - 20} \tag{5.1}$$

Bei der Normierung handelt es sich um eine leichte Abwandlung der in Abschnitt 4.3.5.1 beschriebenen Min-/Max-Normalisierung. Die maximale Lebenserwartung wird dabei von Japan mit 83,6 Jahren bestimmt. Die minimale Lebenserwartung wird per Definition auf 20 Jahre festgelegt. Der Subindex schwankt in Folge der Normierung zwischen 0 und 1.

Der Subindex *Bildung* (BI) setzt sich aus den Indizes *durchschnittliche Schulbesuchsdauer* (DSDI) und der *voraussichtlichen Schulbesuchsdauer* (VSDI) zusammen.[14] Beide

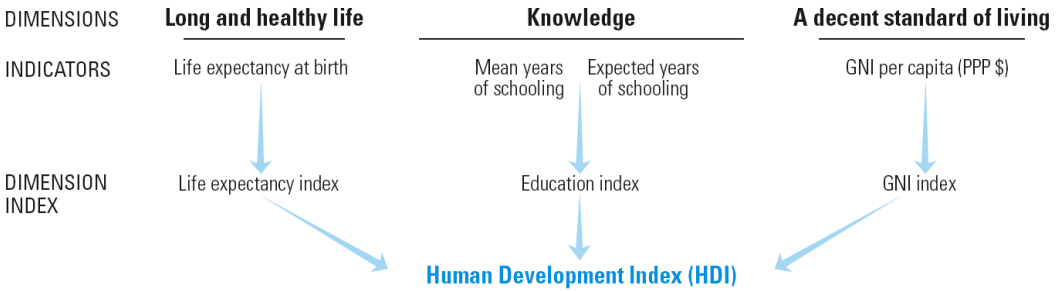
[14] Für den HDI Report 2007 wurde der Subindex *Bildung* über die Indikatoren Alphabetisierungsrate und Bruttoeinschulungsquote berechnet (vgl. Watkins (2007: S. 355)).

Abbildung 5.5: Schema für die Analyse eines Index



Quelle: Eigene Darstellung

Abbildung 5.6: Konstruktion des Human Development Index



Quelle: Khalid (2013: S.1)

gehen normiert in den Subindex BI ein. Wie in Formel 5.1 handelt es sich bei der Normierung um eine modifizierte Min-/Max-Normalisierung.

$$DSDI = \frac{DSD - 0}{13.3 - 0} \tag{5.2}$$

Der Index *Durchschnittliche Schulbesuchsdauer* wird über die maximal beobachtete Schulbesuchsdauer in den USA mit 13,3 Jahren normiert. Das Minimum liegt bei 0 Jahren.

$$VSDI = \frac{VSD - 0}{18.0 - 0} \tag{5.3}$$

Der Index *Voraussichtliche Schulbesuchsdauer* wird über das definitorische Maximum von 18 Jahren und einem Minimum von 0 Jahren normiert.

Der Bildungsindex besteht aus dem geometrischen Mittel von DSDI und VSDI und wird über das beobachete Maximum von Neuseeland 2010 (0,971) normiert. Der Bildungsindex schwankt nach der Normierung zwischen 0 und 1.

$$BI = \frac{\sqrt{DSDI * VSDI}}{0.971} \tag{5.4}$$

Die Dimension *Lebensstandard* wird über den Einkommensindex konstruiert. Dieser basiert auf dem Bruttonationaleinkommen pro Kopf (BNE_{pk}) und geht normiert in den Index ein.

$$EI = \frac{\ln(BNE_{pk}) - \ln(100))}{\ln(87478) - \ln(100)} \tag{5.5}$$

Es handelt sich in Formel 5.5 wieder um eine Min-/Max-Normalisierung. Dabei wird das Maximum über den Wert von Katar 2014 gebildet (87.478 US$) und das Minimum definitorisch auf 100 US$ gesetzt.

Der *Human Development Index* wird über das geometrische Mittel aus den drei Subindizes LEI, BI und EI gebildet. Die drei Komponenten besitzen alle den gleichen Wertebereich [0,1].

$$HDI = \sqrt[3]{LEI * BI * EI} \tag{5.6}$$

Das geometrische Mittel wird beim HDI dem arithmetischen vorgezogen, da es bei den Aussagen des HDI um Wachstumsraten geht. Eine schlechte Leistung in einer der Dimensionen soll nicht linear durch eine hohe Leistung einer anderen Dimension kompensiert

werden können (wie es z.B. beim arithmetischen Mittel möglich wäre). Das geometrische Mittel reduziert die Substituierbarkeit zwischen den Dimensionen (vgl. Abschnitt 4.3.5.2). Damit hat z.B. ein Prozent Rückgang bei dem Index der Lebenserwartung den gleichen Einfluss auf den HDI wie ein Prozent Rückgang des Einkommensindex. (UNDP, 2015)

Vor der Revision des HDI 2010 wurde statt des geometrischen Mittels mit dem arithmetischen Mittel gearbeitet. Die Revision 2010 ist als Eliminierung der Nachteile des arithmetischen Mittels zu betrachten (Watkins, 2007: S. 356). Jedoch ist aufgrund der starken inhaltlichen Anpassung 2010 (Substituierbarkeit der Dimensionen) ein Vergleich mit früheren Werten nur unter Einschränkung möglich.

Mit Bezug zu Tabelle 4.15 auf Seite 156 ergeben sich hinsichtlich der Qualität des Human Development Index folgende Ergebnisse:

Der theoretische Rahmen des Index wird beschrieben sowie die Relevanz des Index hergeleitet (vgl. Ross-Larson und Hanlon (1990: S. 9ff.)). Die Verbindung zwischen den beobachtbaren Indikatoren und dem theoretischen Konstrukt wird ebenfalls verständlich dargestellt (vgl. ebd.: S. 13ff.). Durch mehrere Revisionen ergibt sich jedoch der Eindruck, dass das Konstrukt nicht stabil ist. Mehrere Indikatoren wurden ausgetauscht sowie die Funktion zur Verbindung von beobachtbaren Indikatoren und Konzept verändert. Das lässt vermuten, dass auf Ebene der „indirekten Messung“ Unklarheit besteht sowie auf der Ebene des Konzepts, indem die Beziehung zwischen den Indikatoren und dem latenten Konstrukt festgelegt wird.

Die Qualität des Index bzgl. der Selektion von Daten ist grundsätzlich positiv zu werten. Bei den Indikatoren handelt es sich um *hard data*, d.h. quantitative Daten (vgl. Abschnitt 4.3.2). Lebenserwartung, Schuljahre und Bruttonationaleinkommen pro Kopf werden in der Regel von den statistischen Ämtern erhoben und können als gesicherte Informationen angesehen werden. Außerdem ist davon auszugehen, dass diese Daten auch in Zukunft zur Verfügung stehen.

Die Gewichtung und Aggregation der Daten ist transparent und nachvollziehbar. Grundsätzlich stehen die Individualdaten zur Verfügung[15], womit die Indexwerte problemlos reproduziert werden können.

[15] Für den europäischen Raum z.B. über eurostat oder als Anhang in den HDI Berichten.

Die Verbreitung der Daten findet im Wesentlichen über den *Human Development Report* statt. Die Webseite des Programms beinhaltet die bisherigen Publikationen, jedoch keine expliziten Tabellen oder Diagramme.[16]

Der HDI Report verwendet verschiedene in Abschnitt 4.3.7 genannte Darstellungsformen.[17] Die Visualisierung der Werte ist dabei ansprechend, leicht verständlich und sachgerecht. Jedoch ist das Fehlen einer interaktiven Präsentation innerhalb der Webseite als negativ (im Vergleich zu anderen Indizes) zu werten.

5.7.2.1 Inhaltsanalyse zum HDI

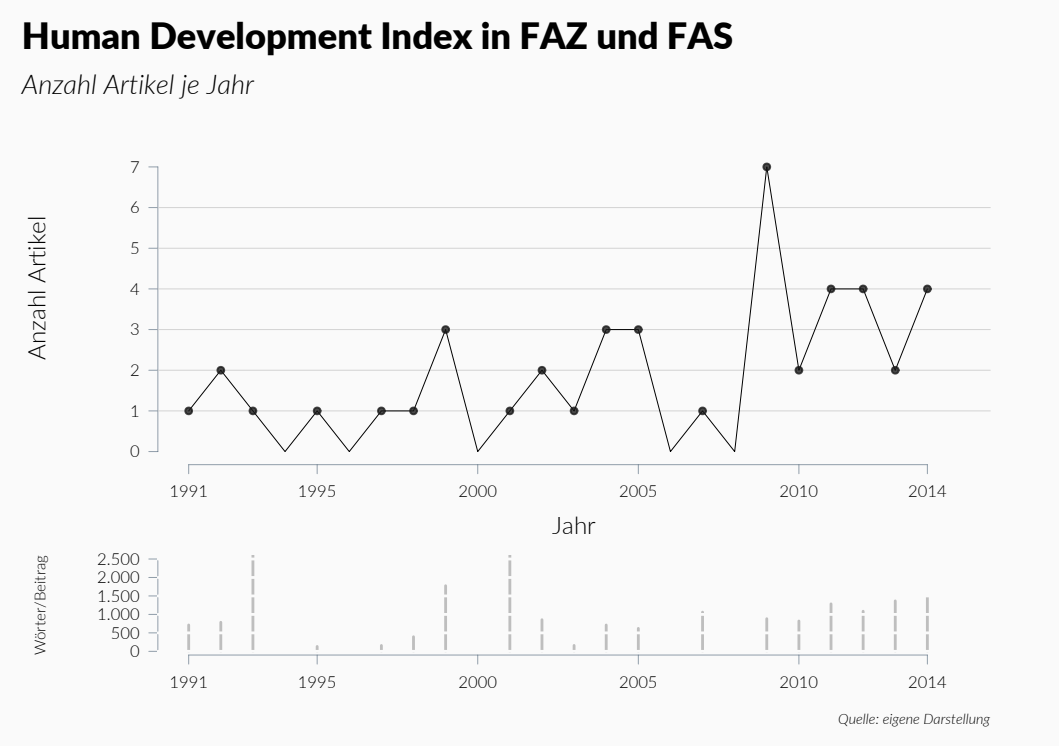
Der Human Development Index wird seit 1990 jährlich berechnet und in Form von globalen und regionalen Berichten veröffentlicht. Die globalen Human Development Reports wurden zwischen 1990 und 2014 in der Regel jährlich veröffentlicht. Die Jahre 2007 und 2008 wurden zu einem Bericht zusammengefasst. 2012 ist kein Bericht erschienen. Die nationalen Berichte werden uneinheitlich veröffentlicht und werden daher nicht in die Analyse einbezogen.

Die Veröffentlichungstermine der globalen HDI Reports ab 2010 wurden den Pressemitteilung der UN entnommen.[18] Die Termine werden für die Analyse der Aktualität der FAZ und FAS Beiträge verwendet.

Der HDI sieht sich, ähnlich wie der *Better Life Index* (vgl. Abschnitt 5.7.4) und *Happy Planet Index* (vgl. Abschnitt 5.7.3), als Alternative zum BIP. Anders als der *Better Life Index* (BLI) und *Happy Planet Index* (HPI), die Größtenteils unter dem *Ressort* Wirtschaft erscheinen, streut der HDI über eine Vielzahl von Ressorts. Der größte Teil entfällt dabei auf das Ressort Politik (43,2%). Danach folgen die Ressorts *Wirtschaft* und *Geld und Mehr* mit zusammen 29,5%. Die Verteilung auf die restlichen Ressorts ist Tabelle 5.6 zu entnehmen. Die verhältnismäßig starke Streuung über eine Vielzahl von Ressorts ist als Indiz für eine allgemeine Relevanz des Index zu deuten. Wegen der vermehrten Beiträge im Politik-Teil und dem politischen Engagement des Herausgebers *United Nations* wird dem Index ein politischer Bezug unterstellt. Als Grund für die Dominanz des Index im Vergleich zu HPI und BLI kommt die längere Historie des Index

[16] Vgl. `http://hdr.undp.org/en` (besucht am 12. Okt. 2014).
[17] Z.B. Tabelle, Säulendiagramm, Balkendiagram, Streudiagramm, Liniendiagrammm, Kreisdiagramm, Flächendiagramm, Blasendiagramm, Choroplethenkarte, Bevölkerungspyramide.
[18] Vgl. `http://hdr.undp.org/en/global-reports` (besucht am 12. Okt. 2014).

Abbildung 5.7: Anzahl Beiträge zum Human Development Index



sowie die Reputation und Relevanz der *United Nations* in Betracht. Die Qualität der Indexkonstruktion vermag den Unterschied nicht zu erklären.

Abbildung 5.8 spricht ebenfalls für eine allgemeine Relevanz des Index. Zwar besitzt der Index keine Aktualität, da der Großteil der Beiträge von FAZ und FAS erst deutlich nach der Veröffentlichung des HDI Reports erscheint (frühestens nach 21 Tagen). Jedoch ist auch ableitbar, dass der HDI zwischen den Veröffentlichungen regelmäßig aufgegriffen wird (tendenziell linearer Anstieg der Kurve). Daraus lässt sich die Vermutung anstellen, dass eher inhaltliche Aspekte der Beiträge für die Nennung des Index verantwortlich sind als z.B Pressemitteilungen oder neue Erkenntnisse. Bei der manuellen Analyse des Inhalts wird dieser Vermutung weiter nachgegangen.

Abbildung 5.8: Aktualität der Beiträge zum Happy Planet Index

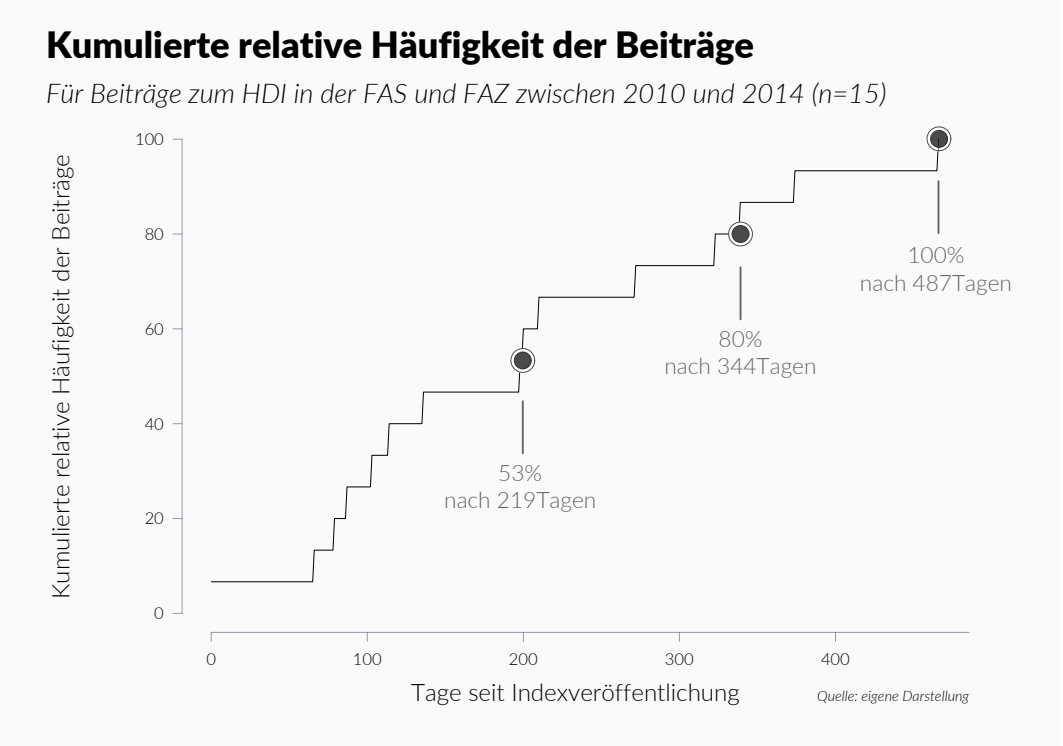


Tabelle 5.6: Ressort-Statistik zum Human Development Index

Ressort	rel. Anzahl	Anzahl
Politik	43.2	19
Wirtschaft	22.7	10
Feuilleton	9.1	4
Geld Und Mehr	6.8	3
Natur Und Wissenschaft	4.5	2
Reise	4.5	2
Beruf Und Chance	2.3	1
Gesellschaft	2.3	1
Titelseite Sonntagszeitung	2.3	1
Wissenschaft	2.3	1
Ohne Angabe	0	0
Gesamt	100	44

Im Durchschnitt wurde der HDI zwischen 1991 und 2014 jährlich in 1,8 Beiträgen genannt. Die durchschnittliche Länge der Beiträge fällt mit 1057 Wörtern in die Kategorie lang. (vgl. Abbildung 5.7).

Für die Relevanz des Index spricht, dass er seit seiner Veröffentlichung mehrmals auf der Titelseite genannt wurde (11,4% der Beiträge). Jedoch scheint seine Stellung innerhalb des Beitrags eher nicht zentral zu sein, da er nicht im Titel genannt wird und in der Regel auch nicht im ersten Absatz (97,7%).

Tabelle 5.7: Numerische Kennzahlen zum Human Development Index

Kennzahl	Min	Median	Max	Mittelwert	m. F.
Titelseite	0	0	100	11.4	32.1
Wörter/ Beitrag	130	826.5	3751	1057.2	811.7
Anzahl Absätze	1	8.5	31	11.7	9.2
Wörter/ Titel	1	4	11	4.5	2.3
Anzahl Indexnennung/ Beitrag	1	1	2	1.1	0.3
Indexnennung/ Titel (Prozent)	-	-	-	0	-
Indexnennung/ 1. Absatz (Prozent)	-	-	-	2.3	-
Indexnennung/ Beitrag (Prozent)	-	-	-	100	-
Satzlänge	12.8	17.2	22.3	17.3	2.6
Aktualität	21	219	487	218.8	132.8

In der manuellen Inhaltsanalyse wurden inhaltliche Überschneidungen zu den Beiträgen von HPI und BLI festgestellt. Dabei steht im Wesentlichen die Diskussion um alternative Kennzahlen zum BIP im Vordergrund sowie die Frage, welchen Stellenwert z.B. Nachhaltigkeit, Umwelt, Gesundheit oder Zufriedenheit bei der Entwicklung einer Gesellschaft haben.

Im Gegensatz zum BLI und HPI wird der HDI nicht überwiegend zur inhaltlichen Diskussion von Entwicklung und Wohlstand verwendet. In 90% der Beiträge werden seine Werte eingesetzt, um (a) Nationen (50%), (b) die Entwicklung über einen Zeitraum (20%) oder (c) beides zusammen (20%) miteinander in Beziehung zu setzen. Der zeitliche Vergleich hebt ihn dabei deutlich vom BLI und dem HPI ab, die ausschließlich für den nationalen Vergleich eingesetzt werden.

Die Transparenz des Index ist mit 60% überwiegend gering. Eventuell wird der Index als Allgemeinwissen vorausgesetzt und daher nicht ausführlich beschrieben. Immerhin wird der HDI z.T. im Lehrplan der gymnasialen Oberstufe behandelt.[19] In den restlichen 40% der Beiträge ist die Transparenz hingegen hoch. Der Index wird ausführlich vorgestellt – teilweise mit seinen Komponenten.

> „Kernstück dieses Jahresberichts [Human Development Report] ist der Human Development Index, ein Mittelwert aus Lebenserwartung, Alphabetisierungsgrad und Kaufkraft pro Kopf der Bevölkerung eines Landes.“ (Brosch, 31. Dezember 1991: S. N3)

[19] Z.B. innerhalb des Geographie- und Erdkundeunterrichts der Liebfrauenschule Bonn oder Sedanstraße Wuppertal.

Andere Beiträge setzen das Wissen zum Konzept des Index voraus und bringen ihn in Verbindung mit dem Begriff *Armut*:

> „Bangladesch zählt zu den ärmsten Ländern der Welt. Im Human Development Index der Vereinten Nationen nimmt das Land den 129. von 169 Plätzen ein." (Klette, 2. Februar 2011: S. 48)

Der Armutsbegriff ist in diesem Zusammenhang deutlich weiter als üblich aufzufassen. Denn der HDI bezieht eben nicht nur materielle Werte ein, sondern auch Lebenserwartung und Bildung. Ähnlich vage kann folgender Auszug aufgefasst werden:

> „[...] dagegen steigt der Human Development Index (Index menschlicher Entwicklung) der Vereinten Nationen für Deutschland kontinuierlich weiter, weil die Lebenserwartung stetig wächst, aber die Umwelt nicht berücksichtigt wird." (Bernau, 20. September 2009: S. 45)

Hier entsteht für den Leser ohne Hintergrundwissen der Eindruck, der HDI würde nur über den Lebensstandard definiert.

Der Großteil der Beiträge zum HDI wurde im Genre Bericht verfasst (60%). Dieses Ergebnis deckt sich mit der durchschnittlichen Länge der Beiträge. Die Anzahl der Wörter kann damit als Proxy für das Genre gesehen werden. Keiner der Beiträge enthält eine negative Akzentuierung. Außerdem beinhalten die Beiträge keine (40%) oder nur eine geringe Überraschung (60%). Jedoch wird der Index in allen Beiträgen argumentativ eingesetzt. Keiner der Beiträge enthält eine Quellenangabe, was der Regel entspricht.

5.7.2.2 Der HDI im Kontext der Kommunikation

Für die Quantifizierung von Wohlstand bedarf es eines theoretischen Konzepts. Weder die fundamentale noch die abgeleitete Messung eignen sich zur Quantifizierung, da Wohlstand nicht direkt beobachtbar und als mehrdimensional anzunehmen ist. Wohlstand ist einer jener qualitativen Begriffe, die in der Evolution der „Messung" zur indirekten oder indikatorbasierten Messung überleiten. Wegen seines mehrdimensionalen Charakters ist er prädestiniert für eine Indexbildung. Jedoch bestehen dabei große Unterschiede zu naturwissenschaftlichen Messungen.

Tabelle 5.8: Ergebnisse der manuellen Inhaltsanalyse zum HDI

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
06.08.2009	Ein Kindersegen?	FAZ, 06.08.2009, Nr. 180, S. 27	(Kurz) Meldung, Kurznachricht, Notiz	Nein	Gering	Nein	National	Gering	Ja
08.06.2012	Mehr Monarchie wagen!	FAZ, 08.06.2012, Nr. 131, S. 35	Sonstiges	Nein	Nein	Nein	National	Gering	Ja
14.10.2011	Das Wesen des Wachstums	FAZ, 14.10.2011, Nr. 239, S. 12	Bericht	Nein	Nein	Nein	National, Zeitlich	Gering	Ja
28.09.2009	Das BIP als fragwürdige Größe	FAZ, 28.09.2009, Nr. 225, S. 12	Bericht	Nein	Nein	Nein	Zeitlich	Hoch	Ja
09.07.2003	Nur Platz 18 für Deutschland	FAZ, 09.07.2003, Nr. 156, S. 4	(Kurz) Meldung, Kurznachricht, Notiz	Nein	Gering	Nein	National, Zeitlich	Hoch	Ja
25.04.2004	Kleiner Taschenführer für forsche Ost-Investoren	FAS, 25.04.2004, Nr. 17, S. 68	Bericht	Nein	Gering	Nein	National	Hoch	Ja
28.12.1999	Das afrikanische Emirat muss sparen	FAZ, 28.12.1999, Nr. 302, S. 4	Bericht	Nein	Gering	Nein	National	Gering	Ja
20.09.2009	Was die Völker der Welt glücklich macht	FAS, 20.09.2009, Nr. 38, S. 45	Bericht	Nein	Gering	Nein	Zeitlich	Gering	Ja
31.12.1991	Theoriehunger	FAZ, 31.12.1991, S. N3	Bericht	Nein	Gering	Nein	Nein	Hoch	Ja
12.02.2011	Niedrige Löhne als Lockstoff	FAZ, 12.02.2011, Nr. 36, S. 48	Nachricht	Nein	Nein	Nein	National	Gering	Ja

Der Human Development Index blickt auf eine 25-jährige Historie zurück und stellt eine der frühen Alternativen zum BIP dar. Die Operationalisierung des theoretischen Konzepts *Wohlstand* scheint dabei die wesentliche Herausforderung zu sein. Sie wurde in der Vergangenheit über die Anpassung der Konstruktionsregeln und der Zusammensetzung der Indikatoren mehrfach modifiziert. Hinsichtlich der Definition des theoretischen Konzepts kommt somit die Frage auf, ob dieses – über die Unterschiedlichkeit der Operationalisierung in verschiedenen Wellen – jemals adäquat numerisch wiedergegeben wurde. Außerdem ist die Definition von *Wohlstand* kulturell bedingt und damit womöglich je Nation über andere Indikatoren zu operationalisieren. Ob der HDI damit als interpretationsfreie Beschreibung angesehen werden kann, bleibt zweifelhaft. Zudem scheint der Begriff „Wohlstand“ zu umfangreich, um ihn mit lediglich drei Indikatoren quantifizieren zu können. Der über den Index operationalisierte Begriff ist stark vereinfacht.

Die Relevanz des HDI ergibt sich insbesondere aus der langen Historie, dem Bekanntheitsgrad und der guten Reputation der *United Nations*. Dennoch ist die Präsenz des Index in FAZ und FAS mit durchschnittlich rund zwei Beiträgen pro Jahr gering. Als mögliche Begründung kommt die geringe Aktualität in Betracht, da die Datenerhebung i.d.R. weit vor der Indexveröffentlichung liegt. Hinzu kommt, dass der Index keinen Prognosecharakter aufweist, was die Kommunikation im Wesentlichen auf retrospektive Betrachtungen beschränkt.

Der HDI wird jährlich berechnet. Damit können kurzfristige Veränderungen nicht im Index abgebildet werden. Darüber hinaus basiert der HDI auf verhältnismäßig stabilen Indikatoren. Lebenserwartung und Schulbesuchsdauer werden sich nur langsam verändern – insbesondere in hochentwickelten Nationen. Überraschungen werden daher eher selten zu erwarten sein und eher bei schwach entwickelten Nationen auftauchen. Für eine Zeitung bedeutet das, dass Nachrichten prinzipiell nur über die Veröffentlichung des Index generiert werden können, was sich über die Länge der Beiträge und der Wahl des Genres bestätigt.

Die breite Streuung der Beiträge über verschiedene Ressorts zeigt den allgemeinen Bekanntheitsgrad des Index. Dieser Faktor kann – neben den generellen Eigenschaften von Zahlen (vgl. Abschnitt 3.3) – die Akzeptanz für Kommunikation erhöhen.

Der Index scheint kein Domänen-Wissen darzustellen, sondern wird in verschiedenen

Bereichen aufgegriffen. Dennoch bleibt das Potential des Index im Wesentlichen ungenutzt. Die wenigen Nennungen des Index dienen hauptsächlich dem Vergleich von Nationen oder Zeitpunkten. Längere Zeitreihen wären durchaus denkbar, werden aber nicht genutzt. Damit wäre es z.B. möglich, Entwicklungstrends zu argumentieren. Auch werden die Indexwerte selten in Beziehung zu Ereignissen, Maßnahmen oder anderen Kennzahlen gesetzt. Ursache ist sicherlich, dass die Werte und Aussagen hauptsächlich auf den HDI Berichten oder anderen Quellen basieren – nicht auf den Rohdaten.

Die mathematische Konstruktion des Index wurde gemäß dem theoretischen Konzept gewählt. Da sich z.B. die einzelnen Dimensionen von Wohlstand nicht kompensieren dürfen, wurde das geometrische Mittel zum Zusammenfassen der Indikatoren gewählt. Außerdem wurde über die Min-/Max-Normalisierung sichergestellt, dass der Index die Ränder bedient. Das bedeutet, dass die Nationen über den kompletten Wertebereich des Index streuen. Damit lassen sich die verschiedenen Länder leichter differenzieren. Bezüglich der Ergebnisse des Index bleibt jedoch zu bezweifeln, dass die Genauigkeit und Präzision, mit der die Länder untereinander abgestuft werden – bis auf die dritte Nachkommastelle – angebracht ist. Zwar werden die Ergebnisse der Öffentlichkeit i.d.R. hauptsächlich in Form eines Länderrankings präsentiert, dieses basiert auf sehr genauen metrischen Werten, deren Unterschied in der „Realität“ ggf. gar nicht existiert oder wahrnehmbar ist. Das beinhaltet auch die Frage, ob der statistische Fehler solch ein Ranking überhaupt zulässt (*Ist der Wohlstand zwischen zwei bestimmten Ländern überhaupt signifikant verschieden?*).

Lange vor der Kritik von Wirschafts-Nobelpreisträger Joseph Stiglitz am BIP und dem Abschlussbericht der *Stiglitz-Sen-Fitoussi-Kommission* 2009 zur Messung von Wohlstand nimmt der HDI eine Vorreiterrolle bei der Diskussion um Alternativen zum BIP ein. Die Relevanz und Funktion des Index basiert damit nicht ausschließlich auf der „Messung“ des komplexen Begriffs, sondern wird auch vom theoretischen Konzept dahinter generiert. Es dient als Grundlage für die Diskussion über Voraussetzungen für den Wohlstand einer Gesellschaft.

Der HDI wird in den Beiträgen der FAZ und FAS in zwei Funktionen verwendet. Einerseits zur Diskussion des *Wohlstandsbegriffs* ohne einen expliziten Bezug zu Indexwerten und anderseits zur Kommunikation komplexer Zusammenhänge, wie z.B. der Vergleich des Wohlstands zweier Länder. Erst an dieser Stelle können Parallelen zu den, in Kapitel

3.3 dargelegten, Eigenschaften von Zahlen als symbolisch generalisiertes Kommunikationsmedium gesucht werden. Wie aus der manuellen Inhaltsanalyse hervorgeht, werden bei dem HDI i.d.R. Zahlen nur in Form einer Rangliste genannt. Der Bezug zu einer naturwissenschaftlichen Messung lässt sich weder durch die Konstruktion, noch durch Indexwerte erkennen. Vielmehr müsste dieser implizit dem präsentierten Länder-Ranking zugeschrieben werden. Aussagen wie z.B.: „Unter den Top zehn der Staaten mit dem höchsten Lebensstandard befinden sich laut Human Development Index der Vereinten Nationen allein sieben Monarchien." (Allmeling und Weber, 8. Juni 2012: S. 35) besitzen nur einen indirekten Bezug zu Zahlen. Dieser Aussage Objektivität zuzugestehen, erfordert dreifaches Vertrauen: (1) in die Zuverlässigkeit der Messverfahren, über die die Rohdaten gewonnen wurden, (2) in die Zuverlässigkeit der Operationalisierung des theoretischen Konzepts und (3) in die Zuverlässigkeit der verwendeten statistischen Verfahren sowie deren korrekte Anwendung. Förderlich in diesem Zusammenhang ist die Reputation der *United Nations*. Sie tritt als Garant für die Anwendung anerkannter Methoden ein und erhöht die Glaubwürdigkeit in die Ergebnisse.

Rein auf die wissenschaftliche Kommunikation bezogen, wirkt sich negativ auf die Glaubwürdigkeit des Index aus, dass das Konstruktionsmodell in der Vergangenheit mehrfach Revisionen unterzogen wurde. Dies lässt zweifeln, was eigentlich quantifiziert werden soll und was in früheren Wellen quantifiziert wurde. Revisionen senken die Zurechnung von Objektivität, indem sie signalisieren, dass bestimmte Konstruktionsentscheidungen nicht eindeutig motivierbar sind.

Hinzu kommt der vage Charakter der Begriffe *Wohlstand* oder *Lebensstandard*. Die Begriffe selbst bergen die Gefahr von Missverständnissen, da ihr Verständnis individuell ist. Sicherlich vereinfacht in diesem Zusammenhang ein Index die Kommunikation dieses komplexen Begriffs, indem er ihn sachlogisch definiert. Jedoch bleibt dann der Einwand bestehen, ob mit dem Index nicht ein völlig neuer Begriff konstituiert wird.

5.7.3 Happy Planet Index

Der *Happy Planet Index* (HPI) wird von der New Economics Foundation (Stiftung für neue Volkswirtschaftslehre), kurz NEF, veröffentlicht. Er kommt dem Bestreben nach, die Zufriedenheit und ökologische Effizienz zu messen. Hintergrund des Index ist die eingeschränkte Aussagekraft vom BIP. Damit erhebt er den gleichen Anspruch wie z.B.

der Human Development Index (vgl. Abschnitt 5.7.2) oder der Better Life Index (vgl. Abschnitt 5.7.4). Im Gegensatz zu den beiden genannten Indizes wird beim HPI der Faktor *Nachhaltigkeit* einbezogen. Der Erfinder des Index Herman Daly[20] begründet die Relevanz zusätzlich über die generelle Funktion von Indizes, Komplexität zu reduzieren (Grossarth, 5. Mai 2012).

Das Konzept des Index beinhaltet drei Indikatoren: (1) Lebenserwartung, (2) Zufriedenheit und (3) ökologischer Fußabdruck. Ziel des Index ist die durchschnittliche Anzahl glücklicher Lebensjahre einer Gesellschaft je verbrauchter Ressource (Einheit) des ökologischen Fußabdrucks wiederzugeben. Dafür werden die Indikatoren Lebenserwartung und Zufriedenheit ins Verhältnis zum ökologischen Fußabdruck gesetzt. Die Bestimmung der Indikatoren wird an dieser Stelle für den HPI 2012 zusammengestellt. (Abdallah u. a., 2012: S. 19f.)

1. **Lebenserwartung:** Dieser Indikator entspricht der durchschnittlichen Anzahl an Jahren, die einem Neugeborenen auf Basis der Sterbetafel des jeweiligen Landes vorausgesagt werden kann.

2. **Zufriedenheit:** Dieser Indikator basiert auf der Frage zur „Lebensleiter“[21] der Gallup World Befragung. In den vorhergehenden zwei HPI-Berichten wurde der Indikator jedoch anders bestimmt.[22]

3. **Ökologischer Fußabdruck:** Dieser Indikator wurde vom *Global Footprint Network* übernommen, außerdem für 9 von 151 Länder über ein lineares Regressionsmodell geschätzt. Das Regressionsmodell beinhaltete die Variablen: CO2 Emission,

[20] War früher Entwicklungsökonom bei der Weltbank und lehrt aktuell an der Universität Maryland.

[21] „Please imagine a ladder with steps numbered from zero at the bottom to 10 at the top. Suppose we say that the top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time, assuming that the higher the step the better you feel about your life, and the lower the step the worse you feel about it? Which step comes closest to the way you feel?“

[22] Für den ersten HPI-Bericht wurden mehrerer Datenquellen zur Bestimmung des Indikators *Zufriedenheit* zusammengeführt. Im zweiten HPI-Bericht wurde die Frage zur *Lebenszufriedenheit* der Gallip World Poll und der World Value Study verwendet. Laut NEF seien jedoch die Indexwerte über die verschiedenen Zeiträume durch die Anwendung statistischer Modellierungstechniken vergleichbar. Nähere Informationen zu den Techniken werden nicht gegeben. (Abdallah u. a., 2012: S. 19) Zwischen dem zweiten und dritten HPI-Bericht ist die Frage zur *Lebenszufriedenheit* der Gallup World Poll weggefallen. Damit stand für den dritten HPI-Bericht nur die Frage zur *Lebensleiter* zur Verfügung. Laut NEF ist jedoch die Korrelation zwischen den Fragen zur *Lebensleiter* und *Lebenszufriedenheit* hoch korreliert. Damit tragen beide die gleiche Information, womit es möglich sei, auf eine der Fragen zu verzichten. Somit konnte auf alternative Datenquellen verzichtet werden. Für den HPI ist die Aufnahme neuer Datenquellen schwierig, da die Daten für eine Vielzahl von Ländern zur Verfügung stehen müssen. Andernfalls verliert der HPI an Abdeckung, da der Index nicht für Länder berechnet werden kann, zu denen Daten fehlen. Außerdem entstehen bei mehreren Datenquellen u.U. Verzerrungseffekte. (ebd.: S. 19f.)

> Bruttoinlandsprodukt pro Kopf, Industrialisierung, Urbanisierung, Bevölkerungsdichte und die geografische Breite.[23] (Abdallah u. a., 2012: S. 20)

Vor der Zusammenführung werden die Indikatoren normiert. Das ist notwendig, da sie in unterschiedlichen Einheiten vorliegen. Die Berechnung des Index ist recht simpel und lässt sich inhaltlich über die *durchschnittliche Anzahl glücklicher Jahre je ökologischem Fußabdruck* zusammenfassen.

Vor der Berechnung des HPI steht damit die Bildung des Subindex „Glückliche Lebensjahre“ (GLJ).

$$GLJ = \frac{(Zufriedenheit + \alpha) * durchschnittliche\ Lebenserwartung}{10 + \alpha}, \quad \text{mit } \alpha = 2.93 \tag{5.7}$$

In Formel 5.7 wird zu dem Wert der *Zufriedenheit* eine Konstante α addiert, damit *Zufriedenheit* und *durchschnittliche Lebenserwartung* denselben Variationskoeffizienten besitzen. Andernfalls würde eine der Variablen u.U. ein höheres Gewicht auf den finalen Indexwert besitzen. Beide Indikatoren sollen jedoch mit gleichem Gewicht eingehen. Die skalierte *Zufriedenheit* wird mit der *durchschnittlichen Lebenserwartung* multipliziert und durch $10 + \alpha$ dividiert, um den GLJ-Score zwischen 0 und der *durchschnittlichen Lebenserwartung* jedes Landes zu skalieren (der maximal erreichbare Wert bei *Zufriedenheit* ist 10). Im zweiten Schritt wird der HPI berechnet.

$$HPI = \delta * \frac{GLJ - \gamma}{Ökologischer\ Fußabdruck + \beta}, \quad \text{mit } \delta = 7.77,\ \gamma = 5.67,\ \beta = 4.38 \tag{5.8}$$

Vom Subindex „Glückliche Lebensjahre“ (GLJ) wird die Konstante γ abgezogen. Die Konstante γ wird dabei derart bestimmt, dass ein Zufriedenheitswert von 0 und eine durchschnittliche Lebenserwartung von 25 Jahren einen HPI-Wert von 0 ergibt. Zusätzlich wird zu dem *ökologischen Fußabdruck* die Konstante β addiert, sodass sein Variationskoeffizient dem des skalierten GLJ entspricht. Der HPI errechnet sich schließlich über den Quotienten aus skaliertem GLJ und skaliertem *ökologischen Fußabdruck* und einer Multiplikation dieses Quotienten mit der Konstante δ. Mit dieser Konstante

[23] Da für zwei Länder nicht ausreichend Daten verfügbar waren, wurde eine zweites Modell geschätzt. In diesem Modell fehlt die Variable *Industrialisierung*. Stattdessen wurden Dummyvariablen für den mittleren Osten und die geografische Breite eingefügt.

wird festgelegt, dass ein Land mit einer Zufriedenheit von 10, einer durchschnittlichen Lebenserwartung von 85 Jahren und einem ökologischen Fußabdruck von 1,78 g/ha pro Kopf einen HPI von 100 erhält.

Zusammenfassend können Formel 5.7 und 5.8 auch folgendermaßen geschrieben werden.

$$HPI = \Phi * \frac{((Zufriedenheit + \alpha) * durchschnittlicheLebenserwartung) - \pi}{\ddot{O}kologischer\ Fußabdruck + \beta}, \quad \text{mit } \alpha = 2.93,\ \beta = 4.38,\ \pi = 73.35,\ \gamma = 0.6 \tag{5.9}$$

In älteren HPI-Berichten wurde die *Zufriedenheit* vor dem Multiplizieren mit der durchschnittlichen Lebenserwartung nicht skaliert.

Mit Bezug zu Tabelle 4.15 auf Seite 156 ergeben sich hinsichtlich der Qualität des Happy Planet Index folgende Ergebnisse:

Der theoretische Rahmen des Index wird beschrieben sowie die Relevanz des Index hergeleitet (vgl. Abdallah u. a. (2006: S. 6ff.)). Die Verbindung der beobachtbaren Indikatoren und dem theoretischen Konstrukt werden ebenfalls verständlich dargestellt (vgl. ebd.: S. 10ff.), wenngleich Begrifflichkeiten wie Zufriedenheit und Glück vage sind.[24] Die Vagheit der Begriffe zeigt sich auch darin, dass die Messung der Indikatoren über die Zeit verändert wurde – ein Faktor, der sich direkt auf die Glaubwürdigkeit des Index auswirkt.

Die Notwendigkeit, die Definition anzupassen ist der Abhängigkeit des HPI von externen Datenquellen geschuldet. Für die Berechnung des HPI führt die NEF keine eigene Datenerhebung durch, sondern greift auf frei zugängliche Datenquellen zurück.[25] Durch den Zugriff auf externe Datenressourcen hat die NEF keine direkte Kontrolle über die Datenerhebung. Dies führte dazu, dass der Indikator *Zufriedenheit* über die verschiedenen Wellen hinweg umdefiniert werden musste und im aktuellen Bericht von 2012 nur noch über eine Frage bestimmt wird.[26] Bei genauerer Betrachtung der Messung des Indikators *Zufriedenheit* entsteht der Eindruck, dass der Zusammenhang zwischen Konzept und Indikator eher über die zur Verfügung stehenden Daten bestimmt wird als über eine

[24] Jan Grossarth unterstreicht in seinem Zeitungsartikel *Die Berechnung des Glücks* die Vagheit der Begriffe Glück und Zufriedenheit z.B. mit ihren verschiedenen Nuancen wie Triumph, Erfüllung, Eroberung, Sehnsucht, Genuss, Bestätigung, Liebe. (Grossarth, 5. Mai 2012)

[25] Für den HPI 2006 wurden für den Indikator *Lebenserwartung* Daten des UN *Human Development Report 2005* verwendet, für den Indikator *Ökologischer Fußabdruck* Daten vom Global Footprint Network und für den Indikator *Zufriedenheit* Daten der World Database of Happiness.

[26] Siehe Fußnote 22.

Theorie und damit der Wissenschaftlichkeit des Index entgegenwirkt. Die Zusammenstellung der beobachtbaren Elemente ist zwar transparent, jedoch äußerst komplex und theoretisch schwer nachvollziehbar. Insgesamt werden vier Datenquellen mit diversen Umfragen einbezogen. Damit lässt sich der HPI für eine Vielzahl von Ländern berechnen. Um die verschiedenen Fragen vergleichbar zu machen, wurden Transformationen und statistische Modelle angewendet. Des Weiteren wurden für Länder ohne Daten auf Basis von statistischen Modellen Werte imputiert (ausführlich in Abdallah u. a., 2006: S. 48ff.). Im aktuellen Bericht von 2012 werden die Unterschiede zusammengefasst:

> „The data used to measure average levels of well-being in this HPI report differ from the previous two HPI reports. In the first HPI report, data on well-being were gathered from a wide range of disparate sources, and modelling techniques were used to estimate values for countries where no well-being data were available. In the HPI 2.0 report, data on well-being were obtained from responses to the satisfaction with life questions in the Gallup World Poll and World Values Survey, and statistical modelling techniques were applied to take into account differences between the two surveys to ensure that the well-being data used to construct the final index were comparable.“ (Abdallah u. a., 2012: S. 19)

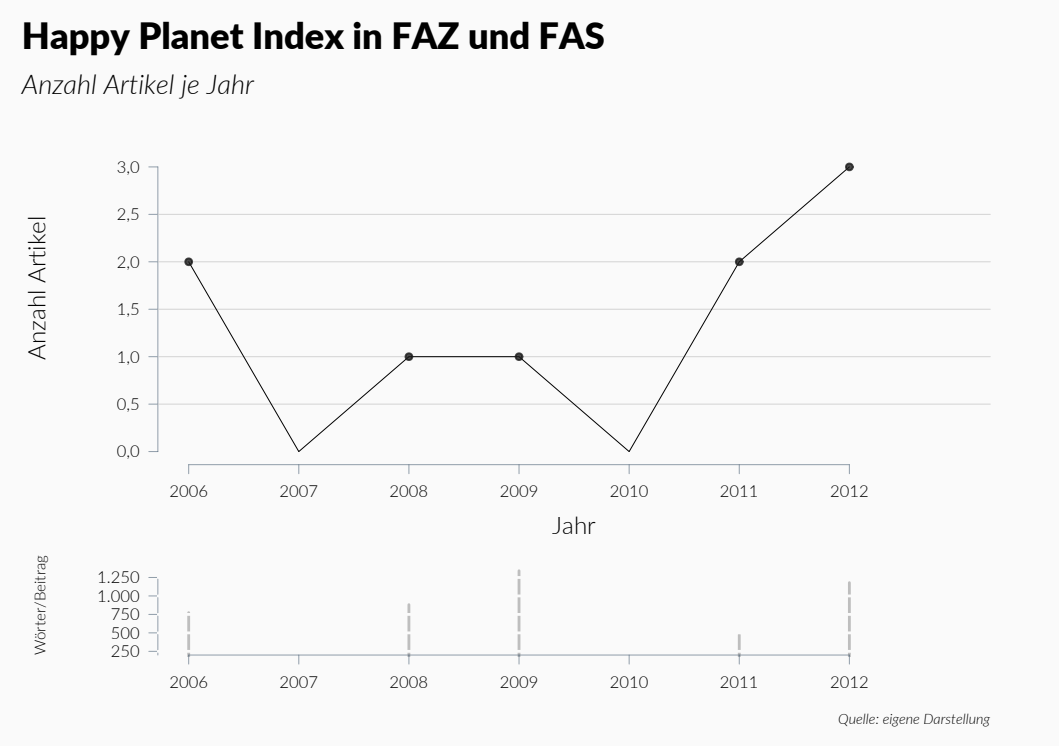
Die Qualität des Index bzgl. der Selektion von Daten ist damit eher als gering zu bewerten (vgl. Abschnitt 4.3.2). Die Transparenz der Daten ist zwar grundsätzlich hoch, da sämtliche Quellen genannt werden, jedoch fehlen Informationen zu den diversen angewendeten statistischen Methoden, um Rechnungen nachvollziehen zu können.

Die Gewichtung und Aggregation der (vorbereiteten) Daten ist transparent und nachvollziehbar. Dennoch ist die Verwendung diverser Konstanten für den Laien irreführend.

Die Verbreitung der Daten findet über die Webseite `http://www.happyplanetindex.org` und über den *Happy Planet Report* statt. Über einen Datenexport können die HPI Daten zu sämtlichen Ländern und Indikatoren heruntergeladen werden. Damit ist auf der letzten Stufe die Berechnung des Index aus den Indikatoren nachvollziehbar. Der Bericht enthält verschiedene in Abschnitt 4.3.7 genannte Darstellungsformen.[27]

[27] Z.B. eingefärbte Tabellen, Choroplethenkarte, Liniendiagrammm, Streudiagramm.

Abbildung 5.9: Anzahl Beiträge zum Happy Planet Index



5.7.3.1 Inhaltsanalyse zum HPI

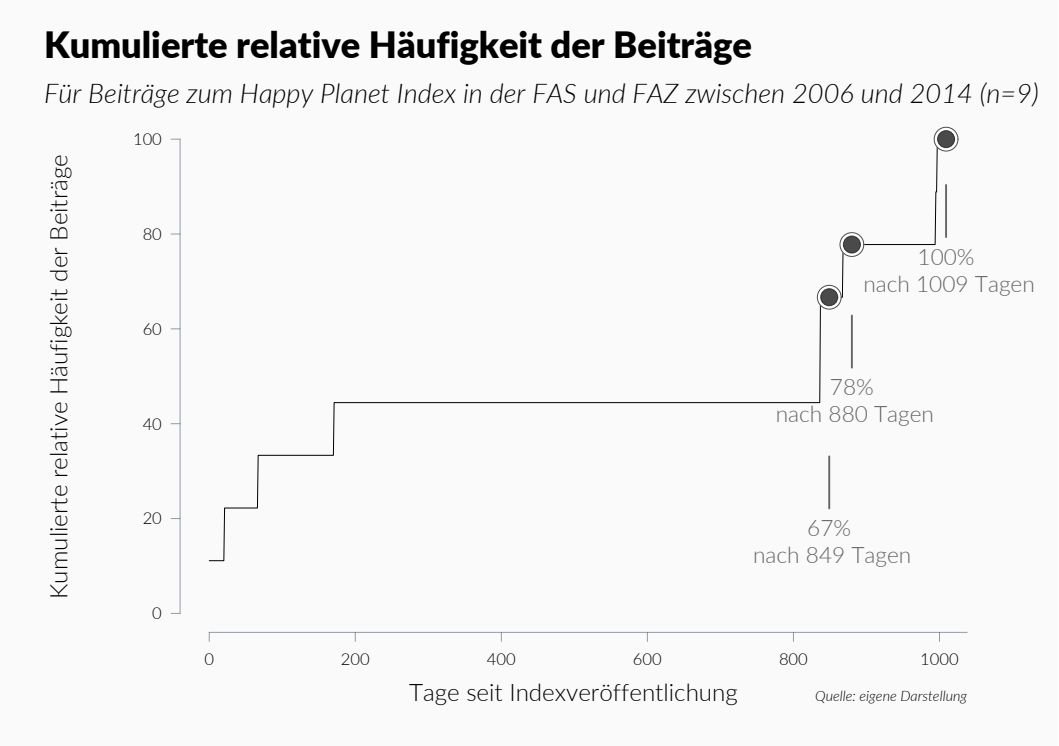
Der Happy Planet Report wird seit Juli 2006 im 3-Jahres-Rhythmus veröffentlicht. Die exakten Veröffentlichungstermine wurden der Webseite mit den Publikationen des NEF entnommen.[28] Da andernorts keine belastbare Zusammenstellung der Veröffentlichungstermine existiert, wurden die Pressemitteilungen aller Monate zwischen 2006 und 2014 einzeln durchgearbeitet und der jeweilige Veröffentlichungstermin entnommen. Alle drei Veröffentlichungen waren zwischen 13.06 und 11.07.

Der Großteil der Beiträge in FAZ und FAS wurde unter dem Ressort *Wirtschaft* und *Geld und mehr* veröffentlicht. Inhaltlich beziehen sich die meisten Beiträge (56%) auf die Glücksökonomie und die Frage, ob das BIP als Kennzahl noch zeitgemäß ist und ob das Glück z.B. als Maß für den Nutzen verwendet werden kann, den die Menschen aus dem Konsum ziehen. Allgemein gesprochen beziehen sich die Beiträge damit im Wesentlichen auf ökonomische Theorien und Erkenntnisse.

Der Happy Planet Index wurde seit seiner Veröffentlichung bis 2014 insgesamt neunmal in der FAZ und FAS aufgegriffen. Mit einer durchschnittlichen Länge von 918 Wörtern

[28] Siehe `http://www.neweconomics.org/search/publications` (besucht am 1. Nov. 2015).

Abbildung 5.10: Aktualität der Beiträge zum Happy Planet Index



handelt es sich um lange Beiträge (vgl. Definition in Abschnitt 5.4.1.2.1). Der längste Beitrag verfügt über 2205 Wörter. Jedoch wird der Index selbst meist erst relativ spät, nämlich nach 64% des Beitrags genannt. Der Name wird nicht im Titel genannt und nur bei 11% im ersten Absatz. Diese Werte sprechen eher für eine geringe Relevanz des Index für den Beitrag.

Die Beiträge zum Index gehen nicht auf die Veröffentlichung des Happy Planet Reports zurück (vgl. Abbildung 5.9 und Definition in Abschnitt 5.4.1.2.9). Für eine geringe Aktualität der Ergebnisse spricht auch das Alter der verwendeten Daten. Für den Bericht 2012 wurden für die meisten Länder Daten verwendet, die zwischen 2010 und 2011 erhoben wurden. Für einige wenige Länder wurden auch Daten verwendet, die nur bis zum Jahr 2007 erhoben wurden (Abdallah u. a., 2012: S. 19). Das Alter der Daten ist in diesem Fall größer als der Zeitraum bis zur Veröffentlichung eines neuen Happy Planet Reports.

Die geringe Aktualität zeigt sich auch über die manuelle Inhaltsanalyse (vgl. Tabelle 5.11). Keiner der Beiträge verweist auf ein aktuelles Ereignis oder besitzt einen aktuellen Bezug.

Tabelle 5.9: Ressort-Statistik zum Happy Planet Index

Ressort	rel. Anzahl	Anzahl
Wirtschaft	55.6	5
Geld Und Mehr	33.3	3
Reise	11.1	1
Ohne Angabe	0	0
Gesamt	100	9

Tabelle 5.10: Numerische Kennzahlen zum Happy Planet Index

Kennzahl	Min	Median	Max	Mittelwert	m. F.
Titelseite	0	0	0	0	0
Wörter/ Beitrag	182	768	2205	917.9	630.1
Anzahl Absätze	2	11	30	12.8	10.5
Wörter/ Titel	2	4	7	4.4	1.4
Anzahl Indexnennung/ Beitrag	1	1	3	1.6	0.9
Indexnennung/ Titel (Prozent)	-	-	-	0	-
Indexnennung/ 1. Absatz (Prozent)	-	-	-	11.1	-
Indexnennung/ Beitrag (Prozent)	-	-	-	63.7	-
Satzlänge	10.6	14.5	23.2	15.3	3.6
Aktualität	12	849	1009	544.6	450

Über die Hälfte (56%) der Beiträge wurde als Bericht verfasst. Mit durchschnittlich 1178 Wörtern sind es auch die längsten Beiträge. Die restlichen Genres sind mit jeweils 11% *(Kurz) Meldung, Kurznachricht, Notiz*; *Kommentar, Glosse, Kritik oder Leitartikel*; *Nachricht* und *Reportage*. Letztere wurde im Ressort *Reise* veröffentlicht. Der Indexwert wurde eingesetzt, um die Zufriedenheit der karibischen Staaten auf unterhaltsame Weise zu untermauern. Der Index wurde damit, wie in 89% aller Beiträge, für einen nationalen Vergleich eingesetzt. Die Transparenz des Index war überwiegend moderat bis hoch (55%). Das bedeutet, dass zusätzlich zum Namen des Index oder Herausgebers Informationen zur Konstruktion oder den enthaltenen Indikatoren gegeben wurden. In allen restlichen Beiträge wurde zumindest der Name und/oder der Herausgeber genannt.

Keiner der Beiträge enthält einen Konflikt oder ist negativ akzentuiert. Der Großteil der Beiträge enthält eine geringe Überraschung (78%), was der Normalfall ist. Sie bezieht sich weitestgehend auf das nicht intuitiv erwartbare Länderranking.

Anders als z.B. beim *Better Life Index* werden beim HPI keine Quellen genannt. Der Index wird in 56% der Beiträge argumentativ eingesetzt.

Tabelle 5.11: Ergebnisse der manuellen Inhaltsanalyse zum HPI

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
30.10.2011	Für Nachhaltige	FAS, 30.10.2011, Nr. 43, S. 48	(Kurz) Meldung, Kurznachricht, Notiz	Nein	Gering	Nein	National	Hoch	Ja
07.04.2012	BIP, Wachstum und Glück	FAZ, 07.04.2012, Nr. 83, S. 11	Bericht	Nein	Hoch	Nein	National	Moderat	Nein
13.08.2006	Jede ist eine Trauminsel	FAS, 13.08.2006, Nr. 32, S. V3	Reportage	Nein	Gering	Nein	National	Moderat	Ja
05.04.2012	Die Berechnung des Glücks	FAZ, 05.04.2012, Nr. 82, S. 12	Bericht	Nein	Gering	Nein	National	Hoch	Nein
23.07.2006	NEULICH IN MEINEM CAFE	FAS, 23.07.2006, Nr. 29, S. 29	Kommentar, Glosse, Kritik oder Leitartikel	Nein	Gering	Nein	National	Gering	Ja
20.09.2009	Was die Völker der Welt glücklich macht	FAS, 20.09.2009, Nr. 38, S. 45	Bericht	Nein	Gering	Nein	Kein	Moderat	Ja
30.10.2011	Das Glück für jeden Typ	FAS, 30.10.2011, Nr. 43, S. 48	Bericht	Nein	Gering	Nein	National	Gering	Nein
13.12.2012	Das andere Leben des Hartmut Ostrowski	FAZ, 13.12.2012, Nr. 291, S. 16	Nachricht	Nein	Keine	Nein	National	Gering	Nein
07.12.2008	Wir steigern das Bruttoglücksprodukt!	FAS, 07.12.2008, Nr. 49, S. 36	Bericht	Nein	Gering	Nein	National	Gering	Ja

5.7.3.2 Der HPI im Kontext der Kommunikation

Über den Happy Planet Index wird das theoretische Konzept quantifiziert, mit welcher Effizienz eine Nation glückliche Lebensjahre generiert. Diese sind bezogen auf den ökologischen Fußabdruck. Operationalisiert wird damit der qualitative Begriff *Glückliche Lebensjahre* sowie der komplexe quantitative Begriff *ökologischer Fußabdruck*. Beide werden zueinander ins Verhältnis gesetzt.

Die Messung dieses theoretischen Konzepts ist weder fundamental noch abgeleitet realisierbar. In Folge dessen ist es prädestiniert für eine Indexbildung. Sowohl der Teilaspekt *Glückliche Lebensjahre* als auch *ökologischer Fußabdruck* sind komplexe mehrdimensionale Begriffe und erfordern eine Operationalisierung über ein indikatorbasiertes System. Der HPI ist damit als Trend der aktuellen Welt zu betrachten, komplexe Sachverhalte über ein numerisches System scheinbar fassbar zu machen und Aussagen wie denen von Heinrich von Kleist zu trotzen:

> „Das Glück kann nicht, wie ein mathematischer Lehrsatz, bewiesen werden, es muss empfunden werden, wenn es da sein soll."

Jedoch ist die Aussagekraft der quantitativen Ergebnisse nicht mit der naturwissenschaftlicher Messungen vergleichbar.

Als Ausgangspunkt für das theoretische Konzept des HPI kann die Diskussion als Alternative zum BIP gesehen werden. Im Mittelpunkt steht die Frage, ob materieller Wohlstand ausschlaggebend ist und ob das BIP noch zeitgemäß ist. Sie wird kontrovers in Politik und Wissenschaft diskutiert und konnte bisher nicht abschließend geklärt werden. In diesem Zusammenhang scheint der HPI eher ein vorläufiges Nebenprodukt der neueren Glücksökonomie zu sein als eine abschließend definierte Maßzahl.

Konzeptionell stellt sich bei dem HPI insbesondere die Frage nach der Unabhängigkeit von *Zufriedenheit*, *Lebenserwartung* und *ökologischem Fußabdruck*. Es ist davon auszugehen, dass Faktoren wie *Umweltverschmutzung* auf alle drei Indikatoren einwirken. So ist damit zu rechnen, dass bei einer großen *Umweltverschmutzung* (Smogbelastung in Peking Dezember 2015) das Wohlbefinden durch die negativen Rahmenbedingungen sinkt, die Lebenserwartung durch die Feinstaubbelastung längerfristig abfällt und der

ökologische Fußabdruck größer wird. Bezogen auf Formel 5.9 wirkt der Faktor *Umweltverschmutzung* damit auf Zähler und Nenner zugleich. Die Indikatoren sind theoretisch nicht ausreichend distinkt definiert. Außerdem lässt die Formulierung des theoretischen Rahmens die Abhängigkeit der drei Indikatoren voneinander nicht vermuten. Vielmehr wird der Zähler über die zusammenfassende Bezeichnung „Glückliche Lebensjahre“ irreführend auf ausschließlich positive Elemente bezogen. Die z.T. kommunizierte negative Korrelation[29] von *Glücklichen Lebensjahren* und *ökologischer Fußabdruck* wird bei Extremwerten eher nicht zutreffend sein (als Beispiel kann die USA genannt werden, bei der einem großen ökologischen Fußabdruck eine hohe Lebenserwartung und Zufriedenheit gegenübersteht). Der Indikator *ökologische Fußabdruck* hat im Index ein deutliches Übergewicht im Vergleich zu den anderen Indikatoren.

Das Übergewicht des *ökologische Fußabdrucks* kann in der Kommunikation irreführend sein. So kann der Index in der FAS u.a. als Maßzahl für das Glück der Länder falsch begriffen werden:

> „Weswegen es gar nicht verwundert, daß jetzt bei einer Studie der New Economics Foundation über die glücklichsten Länder der Welt fünf karibische Staaten unter die ersten zwanzig kamen [...]“. (Rüther, 13. August 2006: S. V3)

Die spezifische Definition von *Glück* hat in diesem Fall das Potential, zu einer Störung in der Kommunikation zu führen. Denn das *Glück der Länder* impliziert an dieser Stelle auch den ökologischen Fußabdruck. In einem anderen Beitrag wird *Glück* in Verbindung zur Bevölkerung gebracht. Diese Aussage entspricht ebenfalls nicht dem theoretischen Rahmen des Index:

> „Laut dem Happy Planet Index leben dort die glücklichsten Menschen der Welt.“ (Ritter, 13. Dezember 2012: S. 16)

In anderen Beiträgen der FAZ wird der Rahmen des Index deutlicher beschrieben:

> „Und für alle Länder der Welt berechnet in London die New Economics Foundation (Stiftung für neue Volkswirtschaft) den Happy Planet Index, der

[29] Mehr von einem Element führt zu weniger bei dem anderen und umgedreht.

> zeigen soll, in welchen Ländern die Menschen mit sparsamem Ressourceneinsatz lange und glücklich leben.“ (Bernau, 20. September 2009: S. 45)

> „Ziemlich kuriose Ergebnisse bringt der Happy Planet Index der New Economics Foundation, der den ökologischen Fußabdruck stark hervorhebt.“ (Plickert, 7. Juli 2012: S. 11)

Obwohl die beiden Auszüge am Ende den theoretischen Rahmen des Index genauer wiedergeben als die ersten beiden, ist auch ihre Definition sehr verschieden. Offensichtlich führt die Komplexität des theoretischen Konzepts des HPI in Verbindung zum Namen des Index zu Widersprüchen, die sich auch in den Beiträgen niederschlagen.

Die Operationalisierung des HPI-Konzepts hat insbesondere vor dem Hintergrund des Klimawandels seine Berechtigung. Jedoch scheint die Art der Umsetzung eine Kommunikation nicht zu begünstigen. Bisher wurde der Index in der FAS und FAZ lediglich für die Diskussion genutzt, wie Wohlstand zu messen sei oder wie Alternativen zum BIP aussehen könnten. Damit stehen nicht die Werte des HPI im Mittelpunkt, sondern seine theoretischen Konzepte Nachhaltigkeit und Wohlbefinden sowie deren Zusammenspiel zu definieren. In keinem der FAZ- und FAS-Beiträge werden spezifische Werte genannt. Auszugsweise wird in einigen Beiträgen maximal auf eine Rangliste von Ländern eingegangen.

Der Bezug zu einer naturwissenschaftlichen Messung lässt sich weder durch die Konstruktion, noch durch Indexwerte erkennen. Vielmehr müsste dieser implizit dem präsentierten Länder-Ranking zugeschrieben werden. Die Zurechnung von Objektivität zum kommunizierten Länderranking erfordert dreifaches Vertrauen: (1) in die Zuverlässigkeit der Messverfahren, über die die Rohdaten gewonnen wurden, (2) in die Zuverlässigkeit der Operationalisierung des theoretischen Konzepts und (3) in die Zuverlässigkeit der verwendeten statistischen Verfahren sowie deren korrekte Anwendung. Insbesondere vor dem Hintergrund der kritischen Kommentierung der HPI-Werte durch die Autoren bleibt zweifelhaft, ob dieses Vertrauen vom Leser aufgebracht werden kann. Zumal bei tieferem Verständnis in die Konstruktionsmethoden des Index Zweifel an der Objektivität zu erwarten sind (nicht kontrollierbare Datengrundlage, Veränderung des Konstruktionsmodells usw.). Außerdem verringert die Komplexität der Selektion durch das Indexkonzept

die Transparenz, da schwer nachvollzogen werden kann, was und wie mit den Indikatoren eigentlich „gemessen“ wird, und wie diese mit dem Konzept zusammenhängen.

Neben den beiden Funktionen (a) Diskussionsgrundlage für den *Wohlstandsbegriff* und (b) Kommunikation mehrdimensionaler Zusammenhänge lässt sich beim HPI thematisch eine weitere Funktion vermuten. So war z.B. die Enquetekommission des Bundestages auf der Suche nach einer allumfassenden Maßzahl für den menschlichen Fortschritt, u.a. wegen der größeren PR-Chancen (Bernau, 30. Oktober 2011: S. 48).

Generell hat der HPI wegen seiner Nähe zum *Better Life Index* und *Human Development Index* große Konkurrenz, zumal beide mit qualitativ hochwertigeren Daten arbeiten.

5.7.4 OECD Better Life Index

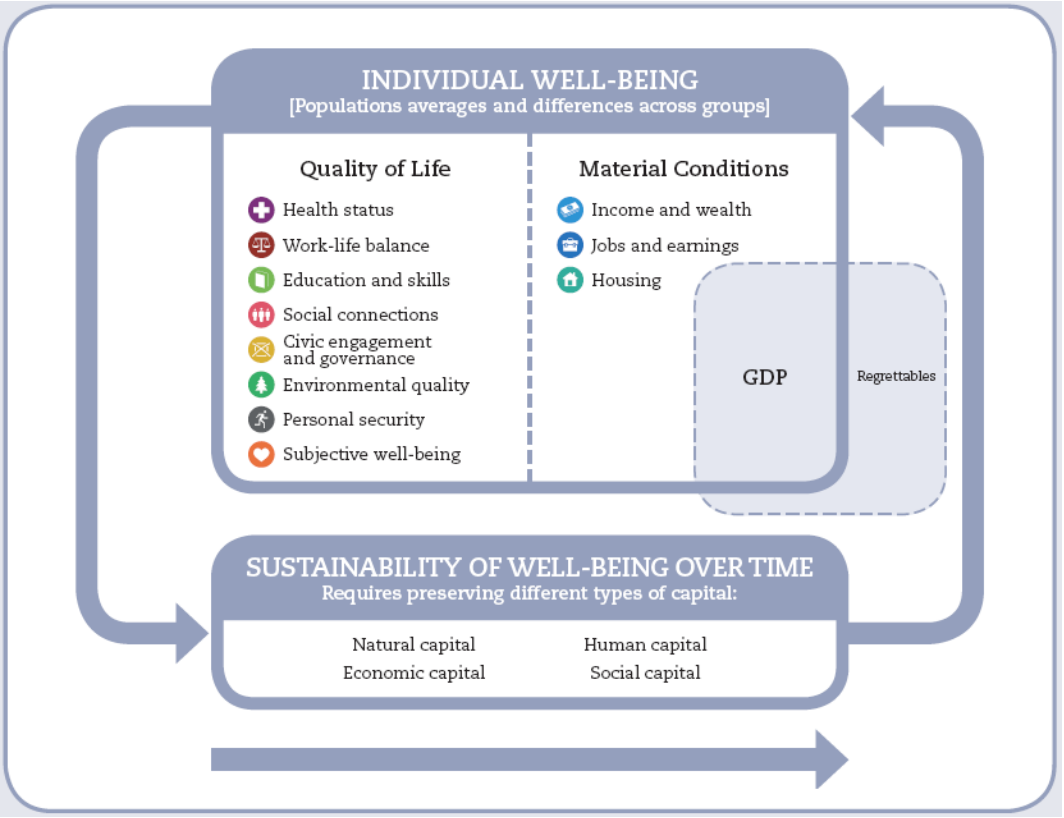
Der *Better Life Index* (BLI) wird seit Mai 2011 von der *Organisation für wirtschaftliche Zusammenarbeit* (OECD) jährlich erstellt und veröffentlicht (OECD, 2013b: S. 1). Sein Ziel ist es, den Wohlstand und sozialen Fortschritt zu messen. Nach eigenen Angaben ist er politisch motiviert (OECD, 2011: S. 15). Ausgangspunkt für die Konstruktion des Index stellt die eingeschränkte Aussagekraft des Bruttoinlandprodukts (BIP) dar. Diesem fehlen laut OECD „viele Aspekte, die das Leben der Menschen prägen wie etwa Sicherheit, Freizeit, Einkommensverteilung und eine saubere Umwelt“ (OECD, 2013a: S. 1). Hinter der Konstruktion des Index stehen laut OECD zehn Jahre Pionierarbeit.[30]

Die Ergebnisse der Forschungsarbeit zum Index werden über die im Mai 2011 gegründete *Better Life Initiative* präsentiert. Konzeptionell setzt sich der Index aus zwei Dimensionen zusammen: *Materielle Bedingungen* und *Lebensqualität* (vgl. Abbildung 5.11). In Anlehnung an die Empfehlungen der *Stiglitz-Sen-Fitoussi-Kommission* hat die *Better Life Initiative* elf Aspekte identifiziert, die für die Messung der zwei Dimensionen und damit für das Wohl der Menschen ausschlaggebend sind (ebd.: S. 1). Die elf Aspekte wiederum setzen sich aus ein bis vier Einzelindikatoren zusammen. Insgesamt handelt es sich um 24 Einzelindikatoren, die für die Bestimmung der Aspekte folgendermaßen verteilt sind (OECD, 2014):

1. Wohnverhältnisse

[30] An dieser Stelle bleibt jedoch zu erwähnen, dass auch der Human Development Index einen ähnlichen Ansatz enthält.

Abbildung 5.11: Konzept zur Messung des Wohlbefindens



Quelle: OECD (2013b: S. 4)

- Wohnungen ohne Grundausstattung (Anteil der Wohnungen ohne eigene Toilette mit Wasserspülung innerhalb der Wohnung)
- Wohnungsbauausgaben (in Prozent des adjustierten Haushaltsnettoeinkommen)
- Zimmer pro Person

2. Einkommen

   - Adjustiertes Haushaltsnettoeinkommen
   - Haushaltsnettofinanzvermögen

3. Beschäftigung

   - Beschäftigungsquote
   - Arbeitsplatzsicherheit (Wahrscheinlichkeit des Arbeitsplatzverlustes)
   - Langzeitarbeitslosenquote
   - Jahresdurchschnittslohn je Vollzeitäquivalent eines Beschäftigten

4. Gemeinsinn

- Qualität des eigenen Netzwerks (Anteil der positiven Zustimmungen auf die Frage *„If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?“*)

5. Bildung
   - Bildungsniveau
   - Fähigkeiten (PISA)
   - Ausbildungsjahre
6. Umwelt
   - Luftverschmutzung
   - Wasserqualität
7. Politischer Rahmen
   - Rechtsberatung (Gewichteter Durchschnitt der Ja/Nein-Antworten auf verschiedene Fragen über die Existenz von Rechtsberatungen durch Bürger, mit denen die Öffentlichkeit Maßnahmen und Aktionen der Regierung beeinflussen kann.)
   - Wahlbeteiligung
8. Gesundheit
   - Lebenserwartung
   - Gesundheitszustand (Anteil der Personen, die mindestens gut auf die Frage *„How is your health in general?“* geantwortet haben. Antwortmöglichkeiten: *„It is very good/ good/ fair/ bad/ very bad“*.)
9. Lebenszufriedenheit
   - Lebenszufriedenheit (basiert auf der gewichteten Summe verschiedener Fragen zum eigenen Leben im Ganzen)
10. Sicherheit
    - Angriffsrate (Anteil der Personen, die auf die Frage *„Within the past 12 months: have you been assaulted or mugged?“* angaben, angegriffen oder ausgeraubt worden zu sein.)
    - Mordrate
11. Vereinbarkeit von Berufs- und Privatleben

- Beschäftigte mit sehr langer Arbeitszeit (Anteil der Personen mit mehr als 50 Arbeitsstunden pro Woche).
- Zeit für Freizeit und persönliche Angelegenheiten (Stunden pro Tag)

Die o.g. Aspekte und Einzelindikatoren sind aktuell nicht endgültig definiert. Es wird betont, dass die Konstruktion und Zusammensetzung des Index nach wie vor im Prozess und damit vorläufig ist. Dabei wird explizit um das Mitwirken (Empfehlungen, Kritik, Wünsche etc.) jedes Einzelnen gebeten (OECD, 2013a: S. 2), um den Index in Zukunft mit weiteren aussagekräftigen Indikatoren anreichern zu können.

Die aktuelle Auswahl der elf Teilbereiche basiert auf der Beteiligung der OECD an der Debatte zur Messung des „Wohlergehens“:

> „Die elf Teilbereiche des Better Life Index (BLI) entsprechen den Aspekten, die die OECD im Rahmen dieser Debatte als unerlässlich für das Wohlergehen der Menschen identifiziert hat, sowohl in Bezug auf ihre materiellen Lebensbedingungen (Wohnverhältnisse, Einkommen, Beschäftigung) als auch auf ihre Lebensqualität insgesamt (Gemeinsinn, Bildung, Umwelt, politischer Rahmen, Gesundheit, Lebenszufriedenheit, Sicherheit und Vereinbarkeit von Berufs- und Privatleben).“ (OECD, 2015c)

In Zukunft sollen insbesondere Indikatoren einbezogen werden, die es ermöglichen, Dimensionen wie *Nachhaltigkeit* und *Lebensqualität* abzubilden.

Für die Berechnung des *Better Life Index* werden im ersten Schritt die verschiedenen Indikatoren normiert. Das ist notwendig, da sie teilweise in verschiedenen Einheiten vorliegen (z.B. Dollar, Jahre etc.). Hierfür wird die in Abschnitt 4.3.5.1 beschriebene Min-/Max-Normalisierung angewendet. Je nachdem ob der Indikator einen positiven oder negativen Faktor abbildet, jedoch in leicht abgewandelter Form:

$$I_{normiert,positiv} = 1 - \frac{Indikatorwert - Mindestwert}{Höchstwert - Mindestwert} \tag{5.10}$$

$$I_{normiert,negativ} = \frac{Indikatorwert - Mindestwert}{Höchstwert - Mindestwert} \tag{5.11}$$

Nach der Normierung werden die einzelnen Indikatoren eines Aspekts über den arithmetischen Durchschnitt zusammengefasst:

$$Aspekt = \frac{\sum_{i=1}^{n} Indikator_i}{n} \tag{5.12}$$

Die Verdichtung der elf Aspekte zum BLI ist nicht fest definiert. Im ersten *How's Life?*-Bericht wird der Index unter Anwendung von drei verschiedenen Gewichtungsmodellen vorgestellt (vgl. OECD (2011: S. 26)). Bei der ersten Variante gehen alle Aspekte gleichgewichtet in den Index ein (jeder Aspekt zu 1/11).[31] Bei der zweiten Variante werden die Aspekte indirekt über die Dimensionen gewichtet (1/6 für die Aspekte unter der Dimension *Materielle Bedingungen* und jeweils 1/16 unter *Lebensqualität*). Die dritte Variante basiert auf der Erkenntnis, dass verschiedene Länder auch verschiedene Präferenzen besitzen und die Gewichte der einzelnen Aspekte individuell pro Land erfolgen sollten (ebd.: S. 26). Um dem zu entsprechen, wurde unter dem Namen *Your Better Life Index* eine individuelle Gewichtungsmethode eingeführt. Dafür wurde eine Webseite eingerichtet (OECD, 2015c), auf der jeder Besucher jeden Aspekt individuell über eine Lickert-Skala mit fünf Ausprägungen (von *nicht wichtig* bis *sehr wichtig*) festlegt. Auf Basis der Bewertung wird danach der Index individuell berechnet. Dieses Vorgehen entspricht der in Abschnitt 4.3.5.3.8 beschriebenen Gewichtungsmethode.

Bis November 2015 wurden auf der Webseite von mehr als 80.000 Nutzern aus 180 Ländern individuelle Gewichte eingegeben. Zusätzlich sind Informationen zu Land, Geschlecht und Alter zu den Nutzern verfügbar. Mit diesen Daten versucht die *Better Life Initiative* neben der Berechnung des Index den Zusammenhang zwischen dem Entwicklungsstand (wird über die o.g. 24 Indikatoren bestimmt) und der Wichtigkeit der einzelnen Aspekte zu bestimmen. (OECD, 2015b)

Trotz der umfassenden Beschreibung des theoretische Rahmens zur Messung von Wohlbefinden fehlt eine tiefer gehende Diskussion der individuellen Gewichtung. Die unterschiedlichen Präferenzen verschiedener Länder sollten erklärt und im theoretischen Konzept verankert werden, damit der Zusammenhang zwischen Indikatoren und Konzept eindeutig expliziert ist. Aktuell befindet sich das Konzept nach Angabe der OECD noch in einer explorativen Phase (OECD, 2011: S. 14ff.). Für die Kommunikation des Index

[31] Oftmals wird diese Gewichtung für die Kritik zum BLI verwendet. In diesem Fall entsteht z.B. eine hohe Korrelation zum BIP und damit Redundanz.

ist dieser Umstand jedoch nachteilig, da er die Wahrscheinlichkeit für Missverständnisse erhöht.

Mit Bezug auf Tabelle 4.15 auf Seite 156 ergeben sich hinsichtlich der Qualität des Better Life Index folgende Ergebnisse: Der theoretische Rahmen des Index wird umfassend beschrieben und hergeleitet. Dennoch ergeben sich Einschränkungen in der Glaubwürdigkeit, da das Konzept nicht endgültig ist. Vielmehr wird der Prozesscharakter betont und die Allgemeinheit aufgefordert, neue Indikatoren vorzuschlagen und existierende kritisch zu diskutieren.

Bei der Selektion der Daten wurde auf eine hohe Relevanz für Politik und auf eine gute Interpretierbarkeit geachtet. Ersteres hat vermutlich einen Konflikt zwischen Interesse und Wissenschaftlichkeit zur Folge. Des Weiteren wurde eine hohe Datenqualität sichergestellt. Fehlende Daten wurden über Standardmethoden imputiert. (Boarini, 2011)

Die Daten zum Index werden umfassend unter `http://stats.oecd.org` zur Verfügung gestellt. Es können die Werte der Indikatoren nach Land und Geschlecht abgefragt werden. Zudem werden in der interaktiven Maske verschiedene Pivotierungs- und Visualisierungsmöglichkeiten angeboten. Zu jedem Indikator und jeder Kovariablen (z.B. Land, Alter) werden Informationen zu den Individualdaten und ihrer Herkunft bereitgestellt. Damit wird ein Höchstmaß an Transparenz ermöglicht, was die Glaubwürdigkeit der Indexwerte positiv beeinflusst.

Die Gewichtung und Aggregation der Daten ist ebenfalls transparent. Diese wird z.B. kurz und verständlich im FAQ zum Index erklärt.[32] Positiv in diesem Zusammenhang ist, dass der Besucher auf die Gewichtung des Index selbst Einfluss hat und damit spielerisch die Sensitivität des Index nachvollziehen kann.

Die statistische Qualität des Index wurde seitens der OECD über verschiedene Tests sichergestellt. Die Ergebnisse zeigen, dass der Index bei verschiedenen Annahmen robust ist. (ebd.)

Für die Darstellung der Indexwerte wurde die Webseite `www.oecdbetterlifeindex.org` eingerichtet. Sie entspricht modernsten interaktiven Standards und ist in diversen Sprachen verfügbar. Optisch ansprechend, übersichtlich und einfach verständlich werden die

[32] Siehe `http://www.oecdbetterlifeindex.org/de/about/better-life-initiative` (besucht am 30. Sep. 2015).

Indexwerte verschiedener Länder gegenübergestellt. Außerdem kann individuell die Gewichtung für jeden Teilaspekt des Index definiert und abgespeichert werden. Neben der Webseite existieren die *How's Life?*-Berichte. Sie werden in der Regel jährlich veröffentlicht und enthalten diverse Darstellungsformen aus Abschnitt 4.3.7.[33]

5.7.4.1 Inhaltsanalyse zum BLI

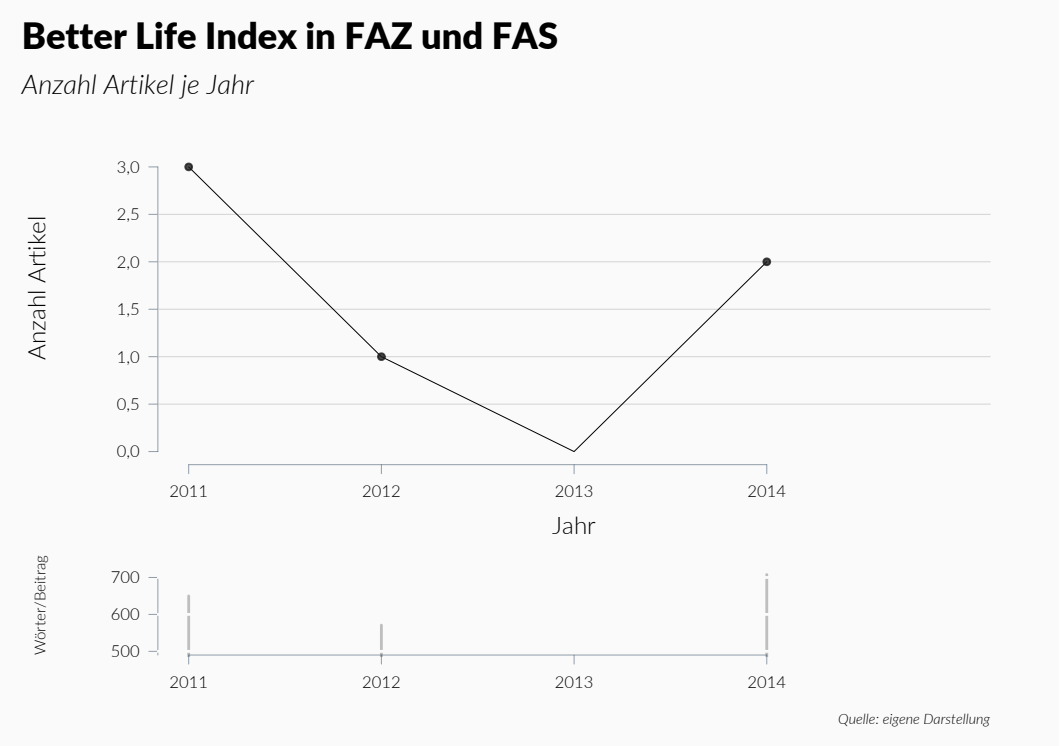
Der BLI wird jährlich erhoben. Jedoch fehlt es an konkreten Pressemitteilungen, weshalb der jeweilige Veröffentlichungstermin nur schwer nachvollziehbar ist. Für die Festlegung des Veröffentlichungsdatums kommt grundsätzlich die Hauptpublikation der *Better Life Initiative* in Betracht – *How's Life?*. Einerseits wird in diesen Berichten explizit auf die Ergebnisse des Index eingegangen, andererseits existiert ein eindeutiger Bezug zu diesem Bericht in einem der manuell geprüften FAZ-Beiträge (vgl. Bernau (11. Mai 2014: S. 20)).

Zwischen 2011 und 2015 wurde der Index viermal in Form des *How's Life?*-Berichts veröffentlicht (jedes Jahr, bis auf 2012). Problematisch in diesem Zusammenhang ist jedoch, dass die Angaben zum Datum der Publikation des Berichts in der *OECD Library*[34] unplausibel sind. So wird der Bericht *How's Life in Your Region?* für 2014 auf den 06.10.2014 datiert. Jedoch hat ein Beitrag der FAZ bereits am 11.05.2014 auf diesen Bericht (inkl. URL) Bezug genommen. Aus diesem Grund sind die Daten kritisch zu verwenden. Im angesprochenen Fall wurde statt des Datums der *OECD Library* das Datum der Erstellung der PDF-Datei des Berichts benutzt. Es liegt nur wenige Tage vor dem FAZ-Beitrag.

Seit seiner Veröffentlichung im Mai 2011 wurde der Index insgesamt in sechs Beiträgen der FAZ und FAS genannt (vgl. Abbildung 5.12). Bereits 2009 erschien eine Vorstellung der Entwicklung einer Maßzahl für Wohlergehen in der FAZ, damals noch ohne explizite Nennung des Indexes (vgl. Bernau (20. September 2009: S. 45)). Danach gab es im Jahr der Veröffentlichung drei Nennungen. In zwei der Beiträge kam es zu ausführlichen Erläuterungen des Index. Der andere Beitrag ist ein Rätsel der FAS mit 27 Fragen. Die letzte Frage betrifft den BLI:

[33] Z.B. eingefärbte Tabelle, Choroplethenkarte, Säulen-, Streu- oder Radialsäulendiagramm etc..
[34] Siehe `http://www.oecd-ilibrary.org` (besucht am 30. Sep. 2015).

Abbildung 5.12: Anzahl Beiträge zum Better Life Index



„Auf die Kritik am BIP als Wohlstandsindikator reagieren internationale Organisationen mit neuen Richtgrößen. So hat die OECD jüngst einen Better-Life-Index eingeführt. In welchem Industrieland leben demnach die glücklichsten Menschen?“ (o. V, 9. Oktober 2011: S. 20)

Es ist davon auszugehen, dass der *Better Life Index* im ersten Jahr über die FAZ- und FAS-Beiträge an Bekanntheit gewonnen hat. Insbesondere das Rätsel sollte dazu geführt haben, dass die Index-Webseite zusätzliche Besuche erhalten hat. Denn in den vorausgehenden FAZ- und FAS-Beiträgen wurde die Antwort für die o.g. Frage nicht genannt und konnte demnach nur mit individueller Recherche beantwortet werden.

Obwohl auch 2013 ein *How's Life?* Bericht publiziert wurde, gab es in diesem Jahr keine Beiträge mit Bezug zum Index. Erst drei Monate nach Veröffentlichung des Berichts findet sich ein Beitrag in der FAZ (Januar 2014).

Die Beiträge zum BLI fallen mit 656 Wörtern in die Kategorie mittlere Länge (vgl. Tabelle 5.13). Der Indexname selbst befindet sich in den aufgegriffenen Beiträgen weder im Titel des Beitrags noch im ersten Absatz. Dies kann als Zeichen für eine geringe Relevanz für den Beitrag oder einen geringen Bekanntheitsgrad des Index gewertet werden. Die

Abbildung 5.13: Aktualität der Beiträge zum Better Life Index

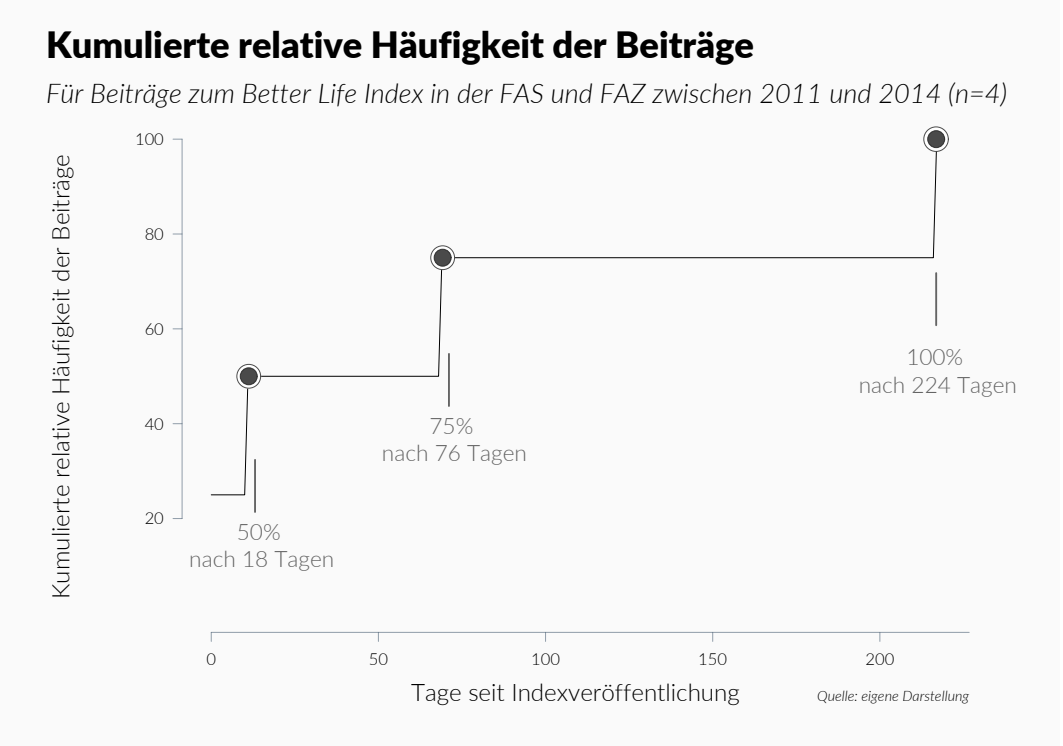


Tabelle 5.12: Ressort-Statistik zum Better Life Index

Ressort	rel. Anzahl	Anzahl
Wirtschaft	83.3	5
Geld Und Mehr	16.7	1
Ohne Angabe	0	0
Gesamt	100	6

Inhaltsanalyse stützt Letzteres, da der Index in den Beiträgen in der Regel beschrieben und eingeführt wird.

Abbildung 5.13 ist wie oben beschrieben unter Vorbehalt zu analysieren. Aus der Abbildung geht hervor, dass die Beiträge nicht in Verbindung zu den o.g. Veröffentlichungen der OECD stehen. Es bleibt jedoch offen, ob andere PR-Maßnahmen seitens der OECD dafür ursächlich sind.

Alle Beiträge zum Index werden unter dem Ressort *Wirtschaft* oder *Geld und Mehr* veröffentlicht. Hier wäre auch das Ressort *Politik* zu erwarten gewesen, da der Index, wie oben bereits beschrieben, politisch motiviert ist. Der wirtschaftliche Bezug ergibt sich aus der Nähe des Index zum BIP.

Tabelle 5.13: Numerische Kennzahlen zum Better Life Index

Kennzahl	Min	Median	Max	Mittelwert	m. F.
Titelseite	0	0	0	0	0
Wörter/ Beitrag	273	578.5	1091	656.2	306.2
Anzahl Absätze	4	7	142	30.2	54.9
Wörter/ Titel	2	4.5	6	4.2	1.5
Anzahl Indexnennung/ Beitrag	1	1	3	1.3	0.8
Indexnennung/ Titel (Prozent)	-	-	-	0	-
Indexnennung/ 1. Absatz (Prozent)	-	-	-	0	-
Indexnennung/ Beitrag (Prozent)	-	-	-	53	-
Satzlänge	5.9	15.2	18.4	14.2	4.3
Aktualität	7	47	224	81.2	99.9

Über die manuelle Inhaltsanalyse zeigt sich, dass nahezu alle Beiträge tiefgehende Informationen zum Index liefern. 67% der Beiträge haben eine hohe Transparenz, 17% eine moderate. Das bedeutet, dass neben dem Namen des Index und dem herausgebenden Institut noch weiterführende Informationen geliefert werden – meist sogar über die einzelnen Komponenten des Index.

Im Zentrum der Beiträge steht der Index selbst und erst an zweiter Stelle seine Werte. Dennoch werden in 83% der Beiträge Nationen über die Indexwerte miteinander in Beziehung gesetzt. Bemerkenswert ist dabei, dass meistens nicht die Werte von Index und Indikatoren in den Vordergrund gestellt werden, sondern die über die Webseite individuell festgelegte Gewichtung für den Gesamtindex. Dabei wird nur einem kritischen Leser bewusst werden, dass es sich bei der Gewichtung und der Messung der Indikatoren um zwei völlig verschiedene Umfragen handelt. Die Gewichte werden über die Webseite des *Better Life Index* erhoben, die Indikatoren z.B. über Umfragen. Letztere sind im Gegensatz zu der Befragung auf der Webseite repräsentativ. Es entstehen damit drei Herausforderungen: (1) Repräsentativität, (2) unterschiedliche Stichproben und (3) Ungleichzeitigkeit. Letztere bezieht sich auf den Umstand, dass die Gewichte auf der Webseite kontinuierlich erhoben werden, die Werte der Indikatoren jedoch nur maximal jährlich.

Der Großteil der Beiträge (50%) ist dem Genre *Bericht* zuzuordnen, 33% erscheinen als *Kommentar, Glosse, Kritik oder Leitartikel*. Im Wesentlichen werden Details zum Index kommuniziert.

Im Vergleich zu anderen Indizes ist auffällig, dass in der Hälfte der Beiträge eine genaue Quellenangabe in Form der URL gegeben wird. Auch dies ist dahingehend zu werten,

Tabelle 5.14: Ergebnisse der manuellen Inhaltsanalyse zum BLI

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
30.10.2011	Für Individualisten	FAS, 30.10.2011, Nr. 43, S. 48	Kommentar, Glosse, Kritik oder Leitartike	Nein	Gering	Ja	National	Hoch	Ja
09.10.2011	Was wissen Sie über Wirtschaft?	FAS, 09.10.2011, Nr. 40, S. 28	Anderes	Nein	Nein	Nein	National	Gering	Nein
10.07.2011	Glück statt BIP	FAS, 10.07.2011, Nr. 27, S. 39	Kommentar, Glosse, Kritik oder Leitartike	Ja	Gering	Nein	Kein	Hoch	Ja
20.01.2014	Wie gut geht es den Deutschen?	FAZ, 20.01.2014, Nr. 16, S. 20	Bericht	Nein	Gering	Ja	National	Hoch	Ja
11.05.2014	Das Geld der anderen	FAS, 11.05.2014, Nr. 19, S. 20	Bericht	Nein	Gering	Ja	National	Hoch	Ja
23.05.2012	Männer sind reicher, Frauen zufriedener	FAZ, 23.05.2012, Nr. 119, S. 10	Bericht	Ja	Gering	Nein	National	Moderat	Ja

dass der Index aktuell noch eine eher untergeordnete Relevanz in der Öffentlichkeit hat und die Bekanntheit eingeschränkt ist.

Der Großteil der Beiträge (83%) beinhaltet eine geringe Überraschung. Diese ergibt sich weitestgehend aus dem Länderranking des Index oder im Bezug zum BIP. Genauso groß ist der Anteil an Beiträgen, bei denen der Index für die Argumentation benutzt wird. Dies liegt daran, dass der Index in der Regel der Kern des Beitrags selbst ist.

5.7.4.2 Der BLI im Kontext der Kommunikation

Der *Better Life Index* der OECD wird seit 2011 jährlich berechnet und publiziert. Die Quantifizierung des theoretischen Konzepts *Wohlstand* ist dabei aber nicht neu. Bereits seit 1990 wird dieses Konzept auch über den *Human Development Index* operationalisiert (vgl. Abschnitt 5.7.2). Auch der *Happy Planet Index* versucht seit 2006 ein ähnliches Konzept zu standardisieren (vgl. 5.7.3). Mit den drei Ansätzen wird deutlich, welchen Schwierigkeiten indikatorbasierte Messsysteme gegenüberstehen: Aus einer Vielzahl möglicher Indikatoren sollen all jene ausgewählt werden, mit deren Hilfe das zu operationalisierende Phänomen mit dem geringsten Fehler bestimmt werden kann. Dieses Optimierungsproblem fördert Alternativkonzepte, da es selten zweifelsfrei gelöst werden kann.

Die Beiträge von FAS und FAZ thematisieren in der Regel nicht die Werte des BLI. Stattdessen steht das Konzept des Index mit seinen Eigenschaften im Vordergrund. Da inhaltliche Aspekte des Index betrachtet werden, kann er in den Beiträgen nicht mit den Eigenschaften eines symbolisch generalisierten Kommunikationsmediums in Verbindung gebracht werden (vgl. Abschnitt 3.3). Seine Funktion ist es vielmehr, eine Diskussionsgrundlage für die Messung von Wohlstand zu bieten. Dafür wird in den Beiträgen sein theoretischer Rahmen thematisiert.

In Folge des starken Fokus der Beiträge auf das Konzept des Index, bietet er nur einen eingeschränkten Neuigkeitswert. Außerdem wird der Index jährlich erhoben und kann daher nicht auf kurzfristige Ereignisse reagieren. Das ist insbesondere vor dem Hintergrund relevant, dass er ausschließlich im Bereich Wirtschaft genannt wurde – ein Bereich, der u.a. durch die Börse von großer Volatiltät geprägt ist. Es sollte nur schwer möglich sein, kurzfristige Ereignisse in Verbindung zu den Indexwerten zu bringen.

Von seiner inhaltlichen Ausrichtung wäre der Index eher im Bereich Politik und Gesellschaft zu erwarten. Um so überraschender ist es, dass er in Beiträgen der zugehörigen Ressorts nicht aufgegriffen wird. Vermutlich könnten dort seine Werte zielgerichteter eingesetzt werden.

Generell ist die Relevanz des Index gering. Bis 2014 wurde er lediglich sechsmal in der FAS und FAZ aufgegriffen. Zukünftig wird die Herausforderung des Index sein, ein endgültiges und nachvollziehbares theoretisches Konzept zu entwickeln. Derzeit ist die Verbindung zwischen Indikatoren und theoretischem Konzept nicht abschließend definiert. Es ist damit ungeklärt, was der Index kommunizieren soll. Im Zuge dessen wurde der Index in den Beiträgen bisher ausschließlich inhaltlich behandelt.

5.7.5 GfK-Konsumklima-Index (KKI)

Der *Konsumklimaindex* der Nürnberger Gesellschaft für Konsumforschung (GfK) wird als Indikator für die Konsumneigungen der Privathaushalte angesehen und ist somit ein Wegweiser für die konjunkturelle Entwicklung Deutschlands. Durch die Regression der Teilkomponenten auf die tatsächlichen Konsumausgaben der Privathaushalte soll er prognosefähig sein.

Der Konsumklimaindex gilt als vielbeachteter Indikator für die konjunkturelle Entwicklung in Deutschland. Die Werte des Index werden zur Bildung des EU-Verbrauchervertrauens genutzt. Gewinnbringend ist das monatliche Erscheinen des Index, das eine größere Aktualität gewährleistet als z.B. Quartalszahlen. Für die Berechnung der Indexzahl werden im Auftrag der EU-Kommission monatlich 2000 repräsentativ ausgewählte Personen ab 14 Jahren zur gesamtwirtschaftlichen Situation in Deutschland und zur Lage der Haushalte befragt. Dabei sind insbesondere die Erwartungen gegenüber den kommenden 12 Monaten entscheidend. Folgende Mehrthemen- und Omnibusumfragen werden „face-to-face“ gestellt:[35]

1. Wie wird sich Ihrer Ansicht nach die finanzielle Lage Ihres Haushaltes in den kommenden zwölf Monaten entwickeln? Sie wird ...

    a. sich wesentlich verbessen

[35] Es werden nur Fragen einbezogen, die für die Berechnung des Index benötigt werden.

b. sich etwas verbessern

c. in etwa gleich bleiben

d. sich etwas verschlechtern

e. sich wesentlich verschlechtern

f. weiß nicht

2. Glauben Sie, dass es in Anbetracht der allgemeinen Wirtschaftslage derzeit günstig ist, größere Anschaffungen (Möbel, elektrische/elektronische Geräte usw.) zu tätigen?

   a. ja, jetzt ist der Augenblick günstig

   b. der Augenblick ist weder besonders günstig noch besonders ungünstig

   c. nein, der Augenblick ist nicht günstig

   d. weiß nicht

3. Glauben Sie, dass es in Anbetracht der allgemeinen Wirtschaftslage derzeit ratsam ist, zu sparen?

   a. ja, auf alle Fälle

   b. wahrscheinlich ja

   c. eher nicht

   d. auf keinen Fall

   e. weiß nicht

Für die Bildung des Konsumklima-Index werden die drei Einzelindikatoren *Einkommenserwartung* sowie *Anschaffungs-* und *Sparneigung* gebildet. Grundlage sind dabei die Salden von positiven und negativen Antworten. Bei der Bildung des Konjunkturindikators wird z.B. der Anteil der Personen mit Antwort *sich etwas verschlechtern* oder *sich wesentlich verschlechtern* von dem Anteil der Personen mit der Antwort *sich wesentlich verbessen* oder *sich etwas verbessern* abgezogen. Es werden also die negativen Antworten von den positiven Antworten der jeweiligen Frage abgezogen. Minimum und Maximum dieser Salden werden durch +100 und -100 bestimmt, +100 wenn alle Respondenten positiv antworten, -100 wenn alle negativ Antworten.

Antwortausprägungen der Einzelindikatoren werden auf die Skala positiv – neutral – negativ transformiert. Das bedeutet, dass bei Frage 1 die Ausprägungen *sich wesentlich verbessen* und *sich etwas verbessern* zu verbessern (positiv) zusammengefasst wird, *in etwa gleich bleiben* zu gleich bleiben (neutral) und *sich etwas verschlechtern* und *sich wesentlich verschlechtern* zu verschlechtern (negativ).

Nach diesem Schritt werden die Salden über die Zeitreihe standardisiert, sodass der langfristige Durchschnitt eines Indikators 0 beträgt. Die standardisierten Werte werden dann auf den Wertebereich +100 bis -100 Punkte mit 0 Punkten als langfristiger Durchschnitt transformiert.

In der Realität haben die Indikatoren seit 1980 i.d.R. Werte zwischen +60 und -60 Punkten angenommen. Positive Werte des Index stehen im langfristigen Vergleich als überdurchschnittlich, negative Werte entsprechend umgedreht.

Der Konsumklima-Index wird über ein Regressionsmodell bestimmt. Dabei ist die abhängige Variable die Veränderungsrate der privaten Konsumausgaben und die erklärenden Variablen die o.g. Indikatoren *Einkommenserwartung*, *Anschaffungs-* und *Sparneigung*. Mit welcher Skalierung diese in das Modell gehen und auf welcher Basis die abhängige Variable im Detail beruht, ist nicht recherchierbar.

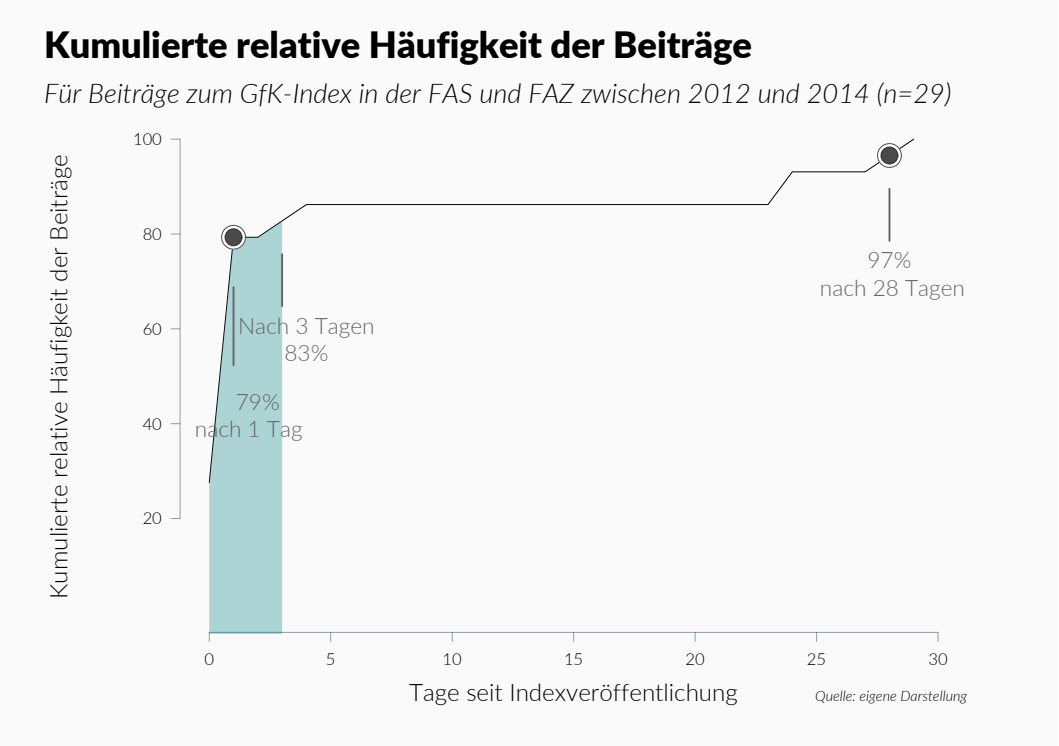
Alles in allem wird mit dem Regressionsmodell die Veränderung des realen privaten Verbrauchs prognostiziert. Die Veränderung bezieht sich dabei immer auf das Vorjahr. Hat der Index z.B. einen Wert von 4,9, zeigt er an, dass der Konsum um 0,49 (4,9/10) höher liegt als ein Jahr zuvor. Der Fakt, dass der Indexwert durch 10 geteilt werden muss, um die Wachstumsquote zu erhalten, liegt am Design der Indexkonstruktion. (Bürkl, 2009, 2013)

5.7.5.1 Inhaltsanalyse zum GfK-Index

Wie bereits angemerkt, wird der GfK-Index einmal pro Monat veröffentlicht. Die exakten Veröffentlichungstermine wurden der Webseite der GfK entnommen.[36] Da andernorts keine belastbare Zusammenstellung der Veröffentlichungstermine existiert, wurden die Pressemitteilungen aller Monate zwischen 2012 und 2014 einzeln durchgearbeitet und der

[36] Siehe `http://www.gfk.com/de/news-und-events/presse/pressemitteilungen/Seiten/default.aspx` (besucht am 10. Jul. 2015).

Abbildung 5.14: Aktualität der Beiträge zum Gfk-Index

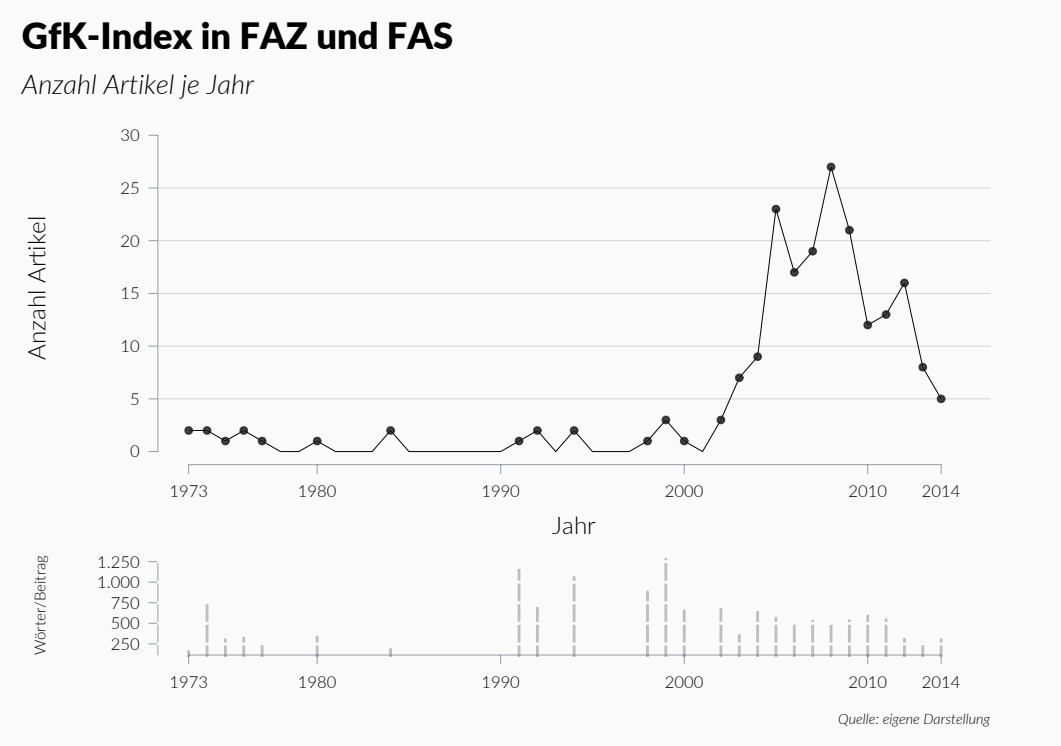


jeweilige Veröffentlichungstermin übernommen. Die Analyse wurde auf diesen Zeitraum beschränkt, da er als aussagekräftig genug bewertet wird und gleichzeitig den Aufwand im Rahmen hält. Bis auf die Dezembermonate, in denen die Veröffentlichungen jeweils zwischen dem 19. und 21. Monatstag stattfanden, wurden die Pressemitteilungen jeweils zwischen dem 24. und 31. eines Monats veröffentlicht.

Aus Abbildung 5.14 wird ersichtlich, dass 79% der FAZ- und FAS-Beiträge bereits maximal einen Tag nach der Pressemitteilung der GfK veröffentlicht wurden. Erst zum Ende des Veröffentlichungszeitraums zwischen Tag 23 und 30 findet ein weiterer größerer Anstieg von ca. 13% statt. Dieser ist u.U. auf Vorankündigungen zurückzuführen. Da der Großteil der Beiträge bereits nach einem Tag des Erscheinens veröffentlicht wurde, ist dem GfK-Index eine hohe Aktualität und Relevanz zuzusprechen. Außerdem werden die Beiträge laut Definition in Abschnitt 5.4.1.2.9 zum überwiegenden Teil (83%) auf die Veröffentlichung durch die GfK zurückgeführt.

Neben den Veröffentlichungen von Beiträgen in unmittelbarer zeitlicher Nähe zu den Pressemitteilungen der GfK gab es im mittleren Bereich zwischen Tag 4 und 23 keine weiteren Beiträge. In dieser Beobachtung ist ebenfalls ein Argument für den starken Aktualitätscharakter des Index zu sehen.

Abbildung 5.15: Anzahl Beiträge zum Gfk-Index



In seiner über vierzigjährigen Geschichte hat der Gfk-Index seit 2001 eine steigende Relevanz, bezogen auf die Nennung in FAZ und FAS (vgl. Abbildung 5.15). Seit 2005 wurde der Index pro Jahr im Schnitt über 12 Mal zitiert. Die höchste Anzahl an Beiträgen mit Indexbezug wurde zu Beginn der Finanzkrise 2008 mit 27 Beiträgen erreicht.

Mit 510 Wörtern im Durchschnitt sind die Artikel zum GfK-Index verhältnismäßig kurz (vgl. Abschnitt 5.4.1.2.1). Im Zeitverlauf (vgl. Abbildung 5.15) wird ersichtlich, dass die Artikellänge seit 2000 zwischen einer Anzahl von 300 bis 750 Wörtern streut. Die durchschnittliche Länge der Beiträge spiegelt sich auch in dem manuell erhobenen Kriterium *Genre* wider. Insgesamt wurden 70% der Beiträge in der Genreform *Meldung/Kurznachricht* geschrieben.

In 34,3% der Beiträge wird der Name des Index bereits im Teaser oder im ersten Absatz genannt. Das spricht für die Relevanz des Index. Nur äußerst selten (0,5%) taucht er dagegen schon in der Überschrift auf. Im Durchschnitt erscheint der Name des Index nach 56,5% des Beitrags (bezogen auf die Anzahl der Absätze des Beitrags).

Nahezu alle betrachteten Beiträge (99,5%) sind dem Ressort *Wirtschaft* und *Geld und*

Tabelle 5.15: Numerische Kennzahlen zum GfK-Index

Kennzahl	Min	Median	Max	Mittelwert	m. F.
Titelseite	0	0	0	0	0
Wörter/ Beitrag	50	499	1449	509.6	361
Anzahl Absätze	1	6	141	32.2	42.3
Wörter/ Titel	1	4	9	3.9	1.4
Anzahl Indexnennung/ Beitrag	1	1	7	1.1	0.6
Indexnennung/ Titel (Prozent)	-	-	-	0.5	-
Indexnennung/ 1. Absatz (Prozent)	-	-	-	34.3	-
Indexnennung/ Beitrag (Prozent)	-	-	-	56.5	-
Satzlänge	3.2	12.4	27.1	12	4.9
Aktualität	0	1	29	4.4	9

Mehr zugeordnet. Das ist nachvollziehbar, da der Index als Wegweiser für die konjunkturelle Entwicklung Deutschlands angesehen wird und damit einen starken Bezug zur Wirtschaft aufweist. Auch börsenorientierte Plattformen wie *www.handelsblatt.com*, *www.finanzen.net* oder *www.wallstreet-online.de* widmen sich in regelmäßigen Beiträgen dem GfK-Index oder stellen die Zeitreihe der Indexwerte in Chartform zur Verfügung.

Durch die manuelle Inhaltsanalyse (vgl. Tabelle 5.17) zeigte sich, dass 50% der Beiträge unter dem Titel *Termine der Woche* und *Wochenplaner* erscheinen. Es handelt sich dabei jeweils um eine Kurzmitteilung, in der darauf hingewiesen wird, dass die Zahlen zum GfK-Index am betreffenden Tag veröffentlicht werden. Dies unterstreicht ebenfalls den hohen Aktualitäts- sowie den Seriencharakter des Index.

In den Kurzmitteilungen wird nur sehr selten Bezug auf historische Indexwerte genommen (20%), was ebenfalls den Fokus auf die reine Terminankündigung zeigt. Gerade in der Gegenüberstellung mit den übrigen Beiträgen, die jeweils historische Indexwerte aufweisen, wird dies deutlicher.

Die Transparenz des Index ist überwiegend gering (80%) bis moderat (20%). Das bedeutet, dass der Index immer namentlich genannt wird, teilweise in Verbindung mit dem herausgebenden Institut. In seltenen Fällen wurde zusätzlich die zugrundeliegende Frage genannt (20%).

Die Hälfte der Beiträge zum GfK-Index enthält keine Überraschung. Eine große Überraschung wird in 20% der Beiträge festgestellt. 30% der Beiträge enthalten eine geringe Überraschung. Die Überraschung bezieht sich dabei meist auf einen Unterschied zwischen Realität und Erwartung.

Tabelle 5.16: Ressort-Statistik zum GfK-Index

Ressort	rel. Anzahl	Anzahl
Wirtschaft	98.5	198
Geld Und Mehr	1	2
Medien	0.5	1
Ohne Angabe	0	0
Gesamt	100	201

Der GfK-Index als konjunktureller Frühindikator ist eng verknüpft mit typischen ökonomischen Größen wie Produktion, Beschäftigung, Zinssatz oder Preisen. In diesem Zusammenhang können auch die 40% der Beiträge mit einem Konflikt oder einem Problem gesehen werden. Inhaltlich beziehen sich die ökonomischen Probleme, die in Zusammenhang mit den Indexwerten der Stichprobenbeiträge gebracht werden, auf Ereignisse wie beispielsweise die Finanzkrise ab 2007 oder der Zweite Golfkrieg 1990. Zum überwiegenden Teil (70%) leitet der Index oder Indexwert jedoch nicht die Argumentation des Beitrags.

Abgesehen von den Beiträgen mit dem Titel *Termine der Woche* wird auffällig, dass alle übrigen Titel der Stichprobe einen Negativismus beinhalten. Diese Beobachtung sollte in einer Folgeuntersuchung über eine Sentimentanalyse[37] generalisiert werden. Zusätzlich könnten diese Ergebnisse mit den Indexwerten und mit dem Delta von Periode und Vorperiode korreliert werden. So könnte die Frage beantwortet werden, ab welchem Indexwert oder welcher Veränderung zur Vorperiode eine negative Haltung im Beitrag eingenommen wird.

5.7.5.2 Der GfK-Index im Kontext der Kommunikation

Der GfK-Index operationalisiert das theoretische Konzept *Konsumneigung*. Es handelt sich dabei um ein qualitatives Merkmal, was weder fundamental noch abgeleitet bestimmt werden kann. Es wird mehrdimensional über verschiedene Sentimente definiert und ist daher prädestiniert für die Indexbildung.

Der GfK-Index ist als Beispiel für die steigende Anforderung an Messapparaturen zu betrachten. Komplexe Zusammenhänge innerhalb der Wirtschaft sollen kontrollierbar,

[37] Das Vorgehen ist hier, für jeden Beitrag die Grundhaltung in „positiv“ und „negativ“ zu trennen. Dies kann über existierende Sentimentwörterbücher und die Statistiksoftware R mit einem überschaubaren Aufwand durchgeführt werden.

überschaubar und kommunizierbar gemacht werden. Der Index spielt dabei insbesondere im Zusammenspiel mit anderen wirtschaftsbezogenen Indizes eine große Rolle.

Er blickt bereits auf eine jahrzehntelange Geschichte zurück und scheint als vielbeachteter Index in den Medien allgemein und in FAZ und FAS im Speziellen etabliert zu sein. So reagieren FAZ und FAS regelmäßig in kürzester Zeit auf die Veröffentlichung des Index, indem sie diese terminlich ankündigen.

Neben der Nennung der Indexwerte in den Beiträgen wird inhaltlich nicht weiter auf den Index eingegangen. Die Verbreitung und Geschichte des Index macht es anscheinend überflüssig, ihn in den Beiträgen näher zu beschreiben oder zu diskutieren. Der Index genießt großes Vertrauen in der Öffentlichkeit. Er wird oft mit dem Prädikat „vielbeachtet“ ausgezeichnet. Außerdem reagieren die Finanzmärkte sensibel auf die Veröffentlichung.

Dennoch ist seine Transparenz eingeschränkt. Auf der Webseite der GfK fehlt eine detaillierte Konstruktionsbeschreibung. Diese kann mit größeren Einschränkungen dem PDF-Dokument *Methode der Erhebung und Berechnung des GfK-Konsumklimas* entnommen werden (Bürkl, 2009). Der Aufbau und Fehler des Regressionsmodell oder ein Zugang zu den zugrunde liegenden Daten kann hier jedoch nicht nachvollzogen werden. Die Validierung des theoretischen Konzepts sollte sich ebenfalls als schwierig erweisen, da Erwartungen gegenüber der Zukunft operationalisiert werden. Diese müssen nicht stabil sein, weswegen eine nachträgliche Validierung des Konzepts tendenziell fehlerbehaftet ist.[38]

Die Veröffentlichung des GfK-Index wird verhältnismäßig oft angekündigt. In diesem Zusammenhang spielt die Nennung von Werten i.d.R. keine Rolle. Die Eigenschaften von Zahlen gemäß Abschnitt 3.3 kommen somit nicht zum tragen. Bei der eher seltenen Nennung von Indexwerten ist für die Zurechnung von Objektivität dreifaches Vertrauen erforderlich: (1) in die Zuverlässigkeit der Messverfahren, über die die Rohdaten gewonnen wurden, (2) in die Zuverlässigkeit der Operationalisierung des theoretischen Konzepts und (3) in die Zuverlässigkeit der verwendeten statistischen Verfahren sowie deren korrekte Anwendung. Der GfK-Index hat in diesem Zusammenhang den Vorteil, dass er prognosefähig ist. Somit können seine Aussagen retrospektiv mit „der“ Realität

[38] Hinzu kommt, dass gesamtwirtschaftliche Entscheidungen sich früher oder später auf die Konsumenten auswirken. Da jedoch für die gesamtwirtschaftlichen Entscheidungen auch der GfK-Index eine Rolle spielen sollte, ist hier ein bedingter Zirkelschluss zu vermuten.

abgeglichen werden. Sobald der Unterschied zwischen Prognose und Realität klein genug ist, sollten die Verfahren zur Herstellung des Index in den Hintergrund rücken. Seine Glaubwürdigkeit wird vielmehr von der Fähigkeit beeinflusst, zukünftige Entwicklungen treffsicher vorherzusagen. Dies konnte er in der Vergangenheit öfter unter Beweis stellen, womit das Prädikat „vielbeachtet“ als Zeichen für seinen starken kommunikativen Wert angesehen werden kann.

Dies erscheint aus wissenschaftlicher Sicht überraschend. Neigungen sind von einer Vielzahl von Faktoren abhängig, und ihre Beschreibung über drei Indikatoren erscheint stark vereinfacht. Von einer Messung im naturwissenschaftlichen Sinne kann ist nicht im Geringsten gesprochen werden. Dennoch suggeriert der Index eine hohe Genauigkeit über die metrischen Werte, obwohl die Ausgangsvariablen ordinal skalliert sind. Bei dem GfK-Index scheinen die Konstruktionsüberlegungen jedoch eher nebensächlich zu sein, da seine Qualität stattdessen über die Genauigkeit der Prognose „gemessen“ wird.

5.7.6 ifo-Index

Der *ifo-Geschäftsklimaindex* vom ifo Institut – Leibniz-Institut für Wirtschaftsforschung an der Universität München e.V. (ifo) gilt als Frühindikator für die konjunkturelle Entwicklung in Deutschland (Krystek und Müller-Stewens, 1993: S. 17f.). Der Indexwert kann, laut einem ifo-Diskussionsbeitrag, für die Prognose der konjunkturellen Entwicklung oder aufkommenden Inflation genutzt werden (Kunkel, 2003).

Der Index wird monatlich erstellt und seit 1972 regelmäßig veröffentlicht (Finanz-Lexikon, 2015). Für die Berechnung werden pro Monat ca. 7000 Unternehmen des verarbeitenden Gewerbes, des Bauhauptgewerbes, des Großhandels und des Einzelhandels befragt. Inhalt der Befragung ist u.a. die Beurteilung der gegenwärtigen Geschäftslage und die Erwartungen gegenüber den nächsten sechs Monaten. Die Unternehmen haben die Wahl zwischen jeweils drei Antwortmöglichkeiten. Die Geschäftslage kann mit den Antworten *gut*, *befriedigend* und *schlecht* beurteilt werden, die zukünftige Erwartung mit *günstiger*, *gleich bleibend* und *ungünstiger*. Die Antworten der Unternehmen werden über die Bedeutung der Branchen gewichtet und aggregiert. (Ifo Institut, 2015)

Tabelle 5.17: Ergebnisse der manuellen Inhaltsanalyse zum GfK-Index

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
29.03.2005	Die F.A.Z.-Termine der Woche	FAZ, 29.03.2005, Nr. 72, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
28.04.2009	Talfahrt mit gebremstem Tempo	FAZ, 28.04.2009, Nr. 98, S. 12	Nachricht	Ja	Groß	Nein	Zeitlich	Gering	Ja
23.06.2007	GfK-Konsumklima-Index	FAZ, 23.06.2007, Nr. 143, S. 19	Meldung/ Kurznachricht	Nein	Gering	Nein	Zeitlich	Gering	Nein
25.01.2004	Wochenplaner	FAS, 25.01.2004, Nr. 4, S. 34	Meldung/ Kurznachricht	Ja	Gering	Nein	Zeitlich	Gering	Nein
25.09.2006	Termine der Woche	FAZ, 25.09.2006, Nr. 223, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
27.09.2014	Konsumenten verlieren Optimismus	FAZ, 27.09.2014, Nr. 225, S. 18	Nachricht	Ja	Gering	Nein	Zeitlich	Moderat	Ja
31.12.1991	Die Deutschen bleiben auch für das neue Jahr zuversichtlich	FAZ, 31.12.1991, S. 12	Bericht	Ja	Groß	Nein	Zeitlich	Moderat	Nein
29.08.2012	Verbraucher-stimmung kühlt ab	FAZ, 29.08.2012, Nr. 201, S. 12	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich	Gering	Ja
18.12.2009	TERMINE DER WOCHE	FAZ, 18.12.2009, Nr. 294, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
27.10.2008	TERMINE DER WOCHE	FAZ, 27.10.2008, Nr. 251, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein

Ein Teilschritt bei der Konstruktion des ifo-Geschäftsklimaindex ist die Bildung von Salden. Der Saldo der gegenwärtigen Geschäftslage wird über die Differenz der Prozentanteile der Antworten *gut* und *schlecht* errechnet. Die Antwort *befriedigend* wird als neutrales Merkmal ignoriert.

$$Lage = p_{gut} - p_{schlecht} \tag{5.13}$$

Der Saldo der Erwartungen wird über die Differenz der Antworten *günstiger* und *ungünstiger* berechnet. Die Antwort *gleich bleibend* wird als neutrales Merkmal ignoriert.

$$Erwartung = p_{günstiger} - p_{ungünstiger} \tag{5.14}$$

Mit diesen Werten wird das Geschäftsklima über das geometrische Mittel bestimmt. Das geometrische Mittel hat dabei gegenüber dem arithmetischen den Vorteil, dass sich die Dimensionen nicht gegenseitig kompensieren können.

$$Geschäftsklima = \sqrt{(Lage + 200)(Erwartungen + 200)} - 200 \tag{5.15}$$

Aus Formel 5.15 folgt, dass das Geschäftsklima zwischen -100 und +100 liegen kann.[39] Dieser Umstand wird u.a. durch die Verwendung der Konstante 200 in Formel 5.15 erreicht.

Für die Berechnung der Indexwerte des *Geschäftsklimas* sowie der Teilkomponenten *Geschäftslage* und *Erwartungen* wird eine Normierung bzgl. eines Referenzjahres durchgeführt (vgl. weiterführend Abschnitt 4.3.5.1). Dazu werden die Salden des betrachteten Monats wie bereits in Formel 5.15 um 200 erhöht und auf den Durchschnitt eines Basisjahres normiert (aktuell 2005) (Ifo Institut, 2015).

$$Index = \frac{Saldo\ im\ Berichtsmonat + 200}{Durchschnittlicher\ Saldo\ im\ Basisjahr + 200} * 100 \tag{5.16}$$

Ein Indexwert von 100 entspricht damit dem Durchschnittswert von 2005.

Die Gewichtung der Branchen sowie die Normierung auf ein Basisjahr machen Zeitreihen des Index anfällig hinsichtlich Veränderungen. In einer Bekanntmachung von 2011, in der das ifo-Institut mitteilt, den Index künftig vom Basisjahr 2000 auf 2005 sowie die

[39] *Lage* und *Erwartungen* besitzen den Wertebereich -100 bis +100.

Klassifikation der Wirtschaftszweige von WZ 2003 auf WZ 2008[40] umzustellen, heißt es (Ifo Institut, 2011):

> „Der ifo Geschäftsklimaindex wird künftig auf Basis einer neuen Klassifikation und Gewichtung berechnet. [...] Dadurch werden sich die Ergebnisse deutlich verändern."

Für die Vergleichbarkeit von Werten mehrerer Zeitperioden führt das zu Schwierigkeiten, da Werte mit unterschiedlicher Basis und Gewichtung nicht absolut miteinander vergleichbar sind.

In Presseartikeln wird der ifo-Index oft als ein vielbeachteter, anerkannter und aussgekräftiger Index beschrieben. Die Vorteile des Index gegenüber der amtlichen Statistik sind die höhere Frequenz und Aktualität der Veröffentlichungen. Dennoch existiert eine umfangreiche Diskussion über die Aussagekraft des Index. Grundsätzlich ist festzustellen, dass das ifo-Institut die Konstruktion des Index nicht komplett offenlegt. Es sind z.B. keine genauen Informationen zur Gewichtung der einzelnen Branchen recherchierbar. Außerdem stellte der Senat der Leibniz-Gemeinschaft 2006 in der *Stellungnahme zum ifo Institut für Wirtschaftsforschung an der Universität München (ifo)* im Allgemeinen fest:

> „Hinsichtlich der Konjunkturanalysen und der Branchenforschung hat die Bewertungsgruppe festgestellt, dass sich die Qualität der zugrunde liegenden ökonometrischen Modelle, statistischen Ansätze und Umfragemethoden stärker an dem heutigen Stand der Forschung orientieren sollte. Darüber hinaus wird die langjährige und einzigartige Sammlung von ökonomischen Daten hinsichtlich der Zugriffs- und Auswertungsmöglichkeiten nicht vollständig ausgeschöpft." [S. 3]

Bezugnehmend auf die Tabelle 4.15 auf Seite 156 ergeben sich hinsichtlich der Qualität des ifo-Index folgende Ergebnisse: Der theoretische Rahmen des Index wird nicht (öffentlich) hergeleitet. Es existiert keine klare Definition des theoretischen Konzepts

[40] Für weiterführende Informationen zu der Klassifikation der Wirtschaftszweige des Statistischen Bundesamtes wird auf Statistisches Bundesamt (2008) verwiesen.

Geschäftsklima und keine theoretische Begründung des Zusammenhangs der beobachtbaren Indikatoren *Geschäftslage* und *Erwartungen* zu diesem Konstrukt. Dieser Umstand kann die Glaubwürdigkeit und Interpretierbarkeit des Index beeinträchtigen.

Hinsichtlich der Datenerhebung ist nicht ersichtlich, wie viele der ca. 7000 befragten Unternehmen tatsächlich geantwortet haben. Es ist ebenfalls nicht nachvollziehbar, auf welche Weise die Stichprobe gebildet wird. Um repräsentative Werte für ganz Deutschland zu erhalten, ist hier u.a. eine Schichtung nach Region und Unternehmensgröße zu erwarten. Neben der Rücklaufquote ist ebenfalls unbekannt, wie mit fehlenden Werten umgegangen wird. Die fehlende Transparenz beeinflusst dabei die Glaubwürdigkeit der Daten sowie die Zurechnung von Objektivität.

Über die Gewichtung und Aggregation der Daten wird Auskunft gegeben. Damit ist die grundsätzliche Berechnungslogik des Index transparent und nachvollziehbar. Jedoch wird die Anwendung der verwendeten Formel nicht begründet. So bleibt unklar, warum z.B. ausgerechnet einfache Saldowerte und das geometrische Mittel Verwendung finden.

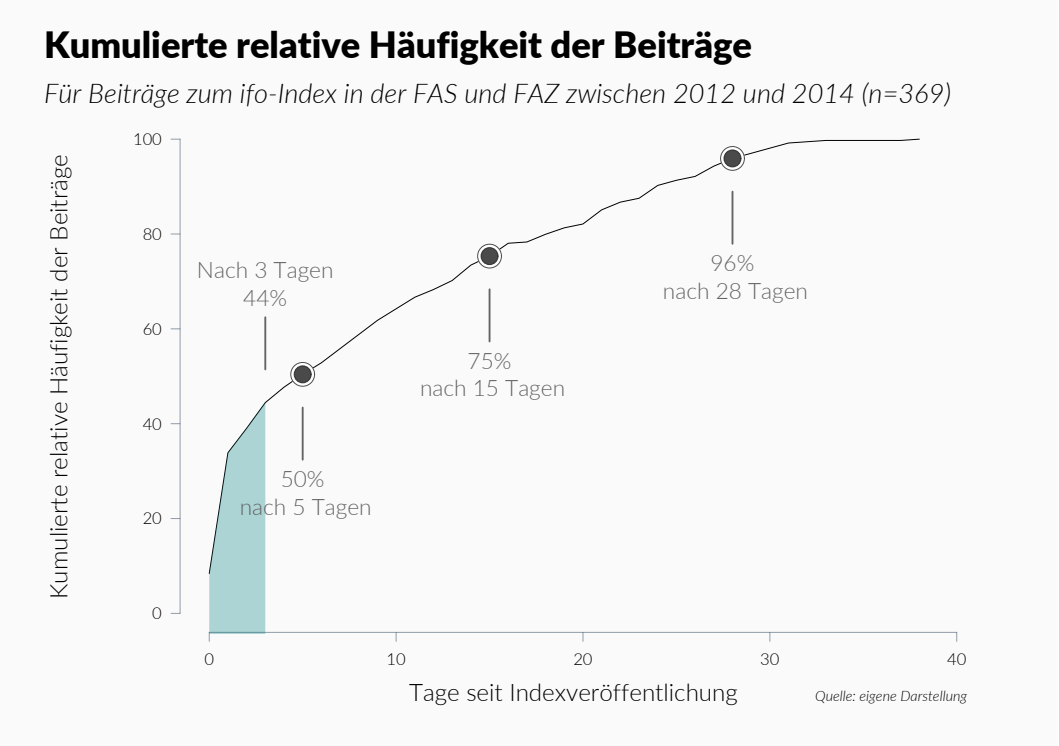
Die Darstellung und Verbreitung der Indexwerte wird von der Webseite des ifo Instituts über Excel-Tabellen und Zeitreihendiagramme umgesetzt. In der Tabelle sind Teilkomponenten und Geschäftsklima in Salden- und Indexform enthalten. Die Indexwerte können somit aus den Teilkomponenten „nachgerechnet“ und validiert werden.

5.7.6.1 Inhaltsanalyse zum ifo-Index

Der ifo-Geschäftsklimaindex wird wie bereits angemerkt monatlich veröffentlicht. Die exakten Veröffentlichungstermine wurden der Webseite des ifo-Instituts entnommen.[41] Da andernorts keine belastbare Zusammenstellung der Veröffentlichungstermine existiert, wurden die Pressemitteilungen aller Monate zwischen 2012 und 2014 einzeln durchgearbeitet und der jeweilige Veröffentlichungstermin in einer Datenbank hinterlegt. Die Analyse beschränkt sich auf diesen Zeitraum, da er als aussagekräftig bewertet wird und gleichzeitig den Aufwand im Rahmen hält. Bis auf die Dezembermonate, in denen die Veröffentlichungen jeweils zwischen dem 18. und 19. Monatstag stattfanden, werden die Pressemitteilungen jeweils zwischen dem 20. und 27. eines Monats, meistens jedoch am 24. veröffentlicht.

[41] Siehe `http://www.cesifo-group.de/de/ifoHome/facts/Survey-Results/Business-Climate/Geschaeftsklima-Archiv.html` (besucht am 6. Nov. 2015).

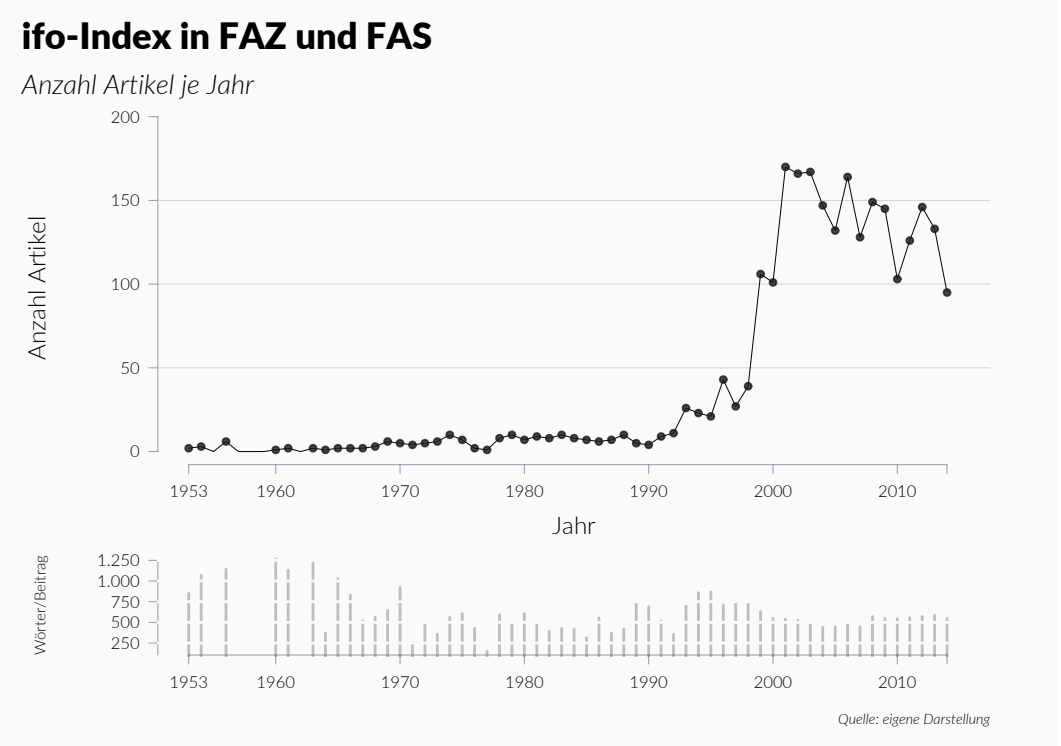
Abbildung 5.16: Aktualität der Beiträge zum ifo-Index



Über den Anstieg der Kurve in Abbildung 5.16 wird ersichtlich, dass die meisten Beiträge innerhalb eines Tages nach der Pressemitteilung des ifo-Instituts veröffentlicht werden (34%). Bis zum dritten Tag erscheinen weiterhin verhältnismäßig viele Beiträge. Insgesamt werden 44% der FAZ- und FAS-Beiträge maximal drei Tage nach der Pressemitteilung publiziert. Anschließend bleibt die Häufigkeit der Beiträge mit Nennung des ifo-Index zwischen dem 3. und 31. Tag nach der Pressemitteilung nahezu konstant. Das Abflachen der Kurve nach dem 31. Tag kann inhaltlich darauf zurückgeführt werden, dass die relevanten Informationen von der nächsten Pressemitteilung überlagert werden, da der Veröffentlichungszeitraum der Meldungen selten mehr als 31 Tage beträgt. Das ist insbesondere zwischen Dezember und Januar der Fall, da im Dezember, wie oben beschrieben wurde, sehr früh veröffentlicht wird.

Zusammenfassend lässt sich aus Abbildung 5.16 ableiten, dass der ifo-Index eine hohe Aktualität besitzt. 44% der Beiträge werden laut Definition in Abschnitt 5.4.1.2.9 auf die Veröffentlichung durch das ifo-Institut zurückgeführt. Anders als beispielsweise der Gfk-Index behält der Index jedoch über den gesamten Monatszeitraum eine gleichbleibende Relevanz. Diese Aussage ergibt sich aus dem nahezu linearen Kurvenverlauf zwischen Tag 3 und 31. Durchschnittlich wurden in dem betrachteten Zeitraum über 10 Beiträge

Abbildung 5.17: Anzahl Beiträge zum ifo-Index



pro Monat mit Bezug zum ifo-Index veröffentlicht (*n=369*).

In seiner über vierzigjährigen Geschichte hat der ifo-Index seit 1991 eine steigende Relevanz erhalten, bezogen auf die Nennung in FAZ und FAS (vgl. Abbildung 5.16). Bis zu diesem Zeitpunkt erschienen im Durchschnitt ca. sieben Beiträge pro Jahr mit Indexbezug. Seit 1999 hält die Beitragshäufigkeit ein hohes Niveau mit im Schnitt über 136 Beiträgen pro Jahr (elf pro Monat). Mit 543 Wörtern im Durchschnitt fallen die Artikel zum ifo-Index verhältnismäßig kurz aus. Der Zeitverlauf aus Abbildung 5.17 lässt einen gewissen Zusammenhang zwischen der Anzahl der Beiträge und der durchschnittlichen Artikellänge ab 1991 vermuten. Ein Test auf Korrelation der beiden Metriken zeigt einen signifikanten moderaten negativen Zusammenhang.[42] Daraus ist zu schließen, dass die Zunahme durch verhältnismäßig kurze Beiträge verursacht wurde. In 31,5% der Beiträge wird der Name des Index bereits im Teaser oder im ersten Absatz genannt. In 4,4% der Beiträge hingegen bereits in der Überschrift. Im Durchschnitt erscheint der Name des Index nach 54,6% des Beitrags (bezogen auf die Anzahl der Absätze des Beitrags). Auffällig ist die verhältnismäßig hohe Anzahl an Nennungen im Titel. Dies ist als deutliches Zeichen für die Relevanz und Bekanntheit des Index zu werten.

[42] Ab 1991 wurden alle Jahre einbezogen. Die davorliegenden Jahre wurden aufgrund geringer Fallzahlen ausgeschlossen. Die Korrelation nach Pearson beträgt -0.5 auf einem Signifikanzniveau von 0.01.

Der Großteil der Beiträge (97,7%) ist dem Ressort *Wirtschaft* und *Geld und Mehr/ Anlage* zugeordnet. Das ist insofern nachvollziehbar, da der Index als Frühindikator für die konjunkturelle Entwicklung Deutschlands einen starken Wirtschaftsbezug aufweist. Auch börsenorientierte Plattformen wie *www.handelsblatt.com*, *www.finanzen.net* oder *www.wallstreet-online.de* widmen sich in regelmäßigen Beiträgen dem ifo-Index.

Über die manuelle Artikelanalyse lässt sich feststellen, dass insgesamt 60% der Beiträge als *Nachricht* oder *Meldung oder Kurznachricht* erscheinen (jeweils 30%). Dieser Umstand passt zu der o.g. durchschnittlichen Länge der Beiträge. Im Vergleich z.B. zum Gfk-Index tauchen jedoch auch umfangreichere Genres auf. So können 37% der Beiträge dem *Bericht* und 4% der *Dokumentation* zugeordnet werden.

Wie der Gfk-Index erscheint auch der ifo-Index in dem regelmäßigen FAZ-Beitrag *Termine der Woche*. Hier finden sich hauptsächlich die Veröffentlichungstermine von wichtigen Kennzahlen und Ereignissen für die bevorstehende Woche. Anders als beim Gfk-Index, bei dem 50% der Beiträge unter diesen Titel fallen, sind es beim ifo-Index aufgrund der deutlich häufigeren Zitierung in anderen Beiträgen lediglich 13%.

In der manuellen Inhaltsanalyse wurden außerdem 17% der Beiträge dem regelmäßigen FAZ-Beitrag mit dem Untertitel *Der monatliche Konjunkturbericht der Frankfurter Allgemeinen Zeitung* zugeordnet. Dieser monatliche Beitrag berichtet umfassend über die aktuelle Konjunktur und vermittelt u.a. Details, Zusammenhänge und Hintergründe, die über den Nachrichtenkern hinausgehen. Bei Anwendung des Untertitels[43] für eine Freitextsuche über alle gesammelten Beiträge (2548) ergeben sich in Summe 187 Beiträge dieser Reihe, in denen der ifo-Index auftaucht. Die durchschnittliche Länge der Beiträge umfasst 5980 Wörter.[44] Das Genre *Bericht* kann damit auch längerfristig über die Anzahl der Wörter approximiert werden. Aus Abbildung 5.18 lässt sich nachvollziehen, dass der ifo-Index zeitweise fester Bestandteil des Konjunkturberichts ist (z.B. 1994, 1995), in anderen Jahre jedoch nicht (z.B. 2003). An dieser Stelle kann die Vermutung aufgestellt werden, dass der Index nur dann in dieser Beitragsreihe aufgegriffen wird, wenn er einen sachlichen Beitrag zur eigentlichen Nachricht leisten kann. Diese Vermutung basiert auf der Grundlage, dass die Spannweite zwischen keiner Nennung innerhalb eines Jahres und zwölf Nennungen extrem ist und systematische Gründe haben sollte.

[43] Wobei FAZ, F.A.Z. und Frankfurter Allgemeine Zeitung Synonym verwendet wird.

[44] Prozentual entspricht das 7,3% der Beiträge und damit ca. 9,6% weniger als über die Stichprobe identifiziert wurden. Es bleibt an dieser Stelle offen, ob der Unterschied durch die geringe Größe der Stichprobe verursacht wird.

Abbildung 5.18: Der ifo-Index im monatlichen Konjunkturbericht der FAZ

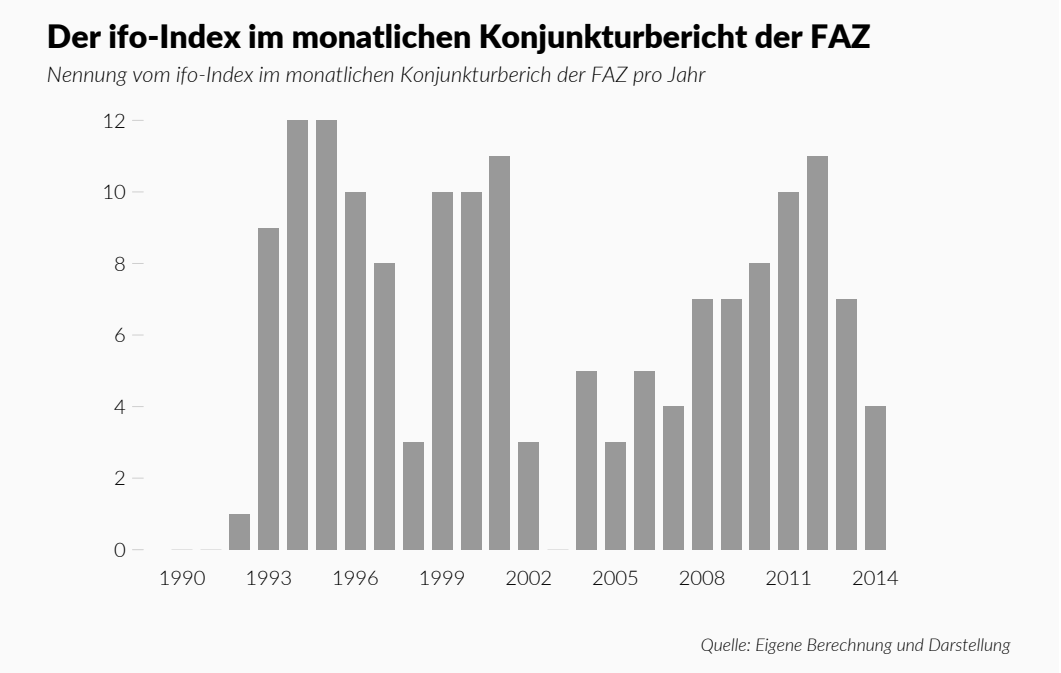


Tabelle 5.18: Numerische Kennzahlen zum ifo-Index

Kennzahl	Min	Median	Max	Mittelwert	m. F.
Titelseite	0	0	100	2.9	16.8
Wörter/ Beitrag	36	489.5	2961	542.7	346.8
Anzahl Absätze	1	6	148	16.9	31.3
Wörter/ Titel	1	4	12	4.4	1.7
Anzahl Indexnennung/ Beitrag	1	1	12	1.3	0.8
Indexnennung/ Titel (Prozent)	-	-	-	4.4	-
Indexnennung/ 1. Absatz (Prozent)	-	-	-	31.5	-
Indexnennung/ Beitrag (Prozent)	-	-	-	54.6	-
Satzlänge	2.7	14.4	37.2	14.4	4.8
Aktualität	0	5	38	9.1	9.4

In etwas mehr als der Hälfte der Beiträge werden die Indexwerte im zeitlichen Vergleich betrachtet (53%). Es zeigt sich dabei eine Korrelation zum Genre des Beitrags.[45] Dieses Ergebnis ist über die manuelle Artikelanalyse nachvollziehbar, da innerhalb eines Berichts der Indexwert viel umfassender behandelt wird als beispielsweise in einer Kurznachricht, in der eine bevorstehende Veröffentlichung angekündigt wird.

Die Transparenz des Index ist überwiegend gering (80%) bis moderat (17%). Das bedeutet, dass der Index in der Regel namentlich genannt wird, teilweise in Verbindung mit dem herausgebenden Institut. In seltenen Fällen (Kategorie: *moderat*) wurden darüber

[45] Chi-Quadrat-Test nach Pearson signifikant auf 5%-Niveau.

Tabelle 5.19: Ressort-Statistik zum ifo-Index

Ressort	rel. Anzahl	Anzahl
Wirtschaft	94.5	2407
Geld Und Mehr	2.3	58
Geld Und Anlage	0.9	24
Politik	0.8	21
Immobilienmarkt	0.3	8
Verlagsbeilage	0.3	8
Redaktionsbeilage	0.2	5
Ohne Angabe	0.2	5
Feuilleton	0.1	3
Das Wetter	0	1
Der Euro In Not	0	1
Gesamt	100	2548

hinaus Informationen zu den Respondenten oder den zugrunde liegenden Fragen gegeben. In 3% der Beiträge wurde kein eindeutiger Bezug zum Index ermöglicht. Es ist zu vermuten, dass die eher geringen Zusatzinformationen zum Index ihren Ursprung in der Bekanntheit des ifo-Index haben, der oft als *vielbeachtet* und *weitverbreitet* zitiert wird. In Wirtschaft und Politik wird er als Allgemeinwissen vorausgesetzt.

30% der Beiträge enthalten keine Überraschung. Eine große Überraschung wird nur in 3% der Beiträge festgestellt. 67% der Beiträge enthalten eine geringe Überraschung. Diese bezieht sich dabei meist auf einen Unterschied zwischen Realität und Erwartung. Eine beispielhafte Formulierung ist in diesem Zusammenhang *„Der unerwartet deutliche Rückgang des Indexes [...]“* (RE., 22. März 2001: S. 42).

Der ifo-Index als Frühindikator für die konjunkturelle Entwicklung in Deutschland ist eng verknüpft mit typischen ökonomischen Größen wie Produktion, Beschäftigung, Zinssatz oder Preisen. In diesem Zusammenhang erklären sich auch die 27% der Beiträge mit einem Konflikt oder einem Problem. Inhaltlich beziehen sich die ökonomischen Schwierigkeiten, die in Zusammenhang mit den Indexwerten der Stichprobenbeiträge gebracht werden, auf Ereignisse wie beispielsweise die Finanzkrise ab 2007.[46] Zum überwiegenden Teil (63%) fügt sich der Index in die Argumentation des Beitrags ein und wird verwendet, um diese zu stützen.

[46]Siehe beispielhaft PLICKERT (23. Dezember 2008: S. 10).

Tabelle 5.20: Ergebnisse der manuellen Inhaltsanalyse zum ifo-Index (I)

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
02.07.2006	Konjunkturdaten beflügeln den Dax	FAS, 02.07.2006, Nr. 26, S. 50	Nachricht	Nein	Nein	Nein	Nein	Gering	Ja
23.08.2000	Der Aufschwung hat an Kraft gewonnen	FAZ, 23.08.2000, Nr. 195, S. 1	Bericht	Nein	Gering	Nein	Nein	Keine	Ja
26.01.2007	Yen stabilisiert sich	FAZ, 26.01.2007, Nr. 22, S. 26	Meldung/ Kurznachricht	Nein	Gering	Nein	Nein	Gering	Ja
04.11.1999	Vorboten des Aufschwungs	FAZ, 04.11.1999, Nr. 257, S. 17	Bericht	Nein	Gering	Nein	Zeitlich	Gering	Ja
26.07.2012	Euro-Krise überschattet deutsche Konjunktur	FAZ, 26.07.2012, Nr. 172, S. 1	Bericht	Ja	Gering	Nein	Zeitlich	Moderat	Ja
20.06.2009	Wechsel im S-Dax	FAZ, 20.06.2009, Nr. 140, S. 17	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
21.08.2009	TERMINE DER WOCHE	FAZ, 21.08.2009, Nr. 193, S. 19	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
25.01.2013	Ifo-Index höher erwartet	FAZ, 25.01.2013, Nr. 21, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
23.12.2008	Deutsche Wirtschaft steht vor einem langem Winter	FAZ, 23.12.2008, Nr. 300, S. 10	Bericht	Ja	Gering	Nein	Zeitlich	Moderat	Ja
31.12.2004	Eine kurze Erholung	FAZ, 31.12.2004, Nr. 306, S. S7	Dokumentation	Nein	Gering	Nein	Zeitlich	Gering	Ja

Tabelle 5.21: Ergebnisse der manuellen Inhaltsanalyse zum ifo-Index (II)

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
18.03.2012	Kursprognosen mit Facebook	FAS, 18.03.2012, Nr. 11, S. 34	Bericht	Nein	Gering	Nein	Nein	Gering	Nein
03.04.2002	Aufschwungzeichen und Risikosorgen	FAZ, 03.04.2002, Nr. 77, S. 15	Bericht	Ja	Gering	Nein	Zeitlich	Moderat	Ja
06.11.2009	TERMINE DER WOCHE	FAZ, 06.11.2009, Nr. 258, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
28.02.2004	Rauf mit der Dividende	FAZ, 28.02.2004, Nr. 50, S. 23	Bericht	Ja	Gering	Nein	Zeitlich	Gering	Ja
28.07.2008	Wenig Licht am Horizont	FAZ, 28.07.2008, Nr. 174, S. 26	Bericht	Ja	Gering	Nein	Zeitlich	Gering	Ja
21.02.2007	Eine Stimmungsdelle	FAZ, 21.02.2007, Nr. 44, S. 12	Nachricht	Nein	Gering	Nein	Zeitlich	Gering	Ja
21.03.2005	Die F.A.Z.-Termine der Woche - Wirtschaft	FAZ, 21.03.2005, Nr. 67, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
21.02.1992	Ifo: Furcht vor Rezession ist unbegründet	FAZ, 21.02.1992, S. 16	Nachricht	Nein	Gering	Nein	Zeitlich	Gering	Ja
25.02.1998	Ifo: Die Stimmung ist gespalten	FAZ, 25.02.1998, Nr. 47, S. 15	Meldung/ Kurznachricht	Ja	Gering	Nein	Nein	Gering	Ja
03.11.1995	Abkühlungstendenzen in der Chemieindustrie	FAZ, 03.11.1995, Nr. 256, S. 17	Nachricht	Ja	Gering	Nein	Zeitlich	Gering	Ja

Tabelle 5.22: Ergebnisse der manuellen Inhaltsanalyse zum ifo-Index (III)

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
12.01.2012	Europa bleibt an den Märkten das wichtigste Thema	FAZ, 12.01.2012, Nr. 10, S. 19	Bericht	Nein	Gering	Nein	Nein	Gering	Ja
24.11.1989	Zuwanderung stärkt privaten Verbrauch	FAZ, 24.11.1989, S. 17	Nachricht	Nein	Gering	Nein	Zeitlich	Gering	Ja
26.05.2004	Renten mit leichten Kursgewinne	FAZ, 26.05.2004, Nr. 121, S. 30	Nachricht	Nein	Nein	Nein	Zeitlich	Moderat	Ja
23.06.2001	Daimler-Chrysler im Aufwind	FAZ, 23.06.2001, Nr. 143, S. 28	Nachricht	Nein	Gering	Nein	Zeitlich	Gering	Nein
01.06.2012	Europas Wirtschaft droht ein Schwächeanfall	FAZ, 01.06.2012, Nr. 126, S. 16	Bericht	Ja	Gering	Nein	Zeitlich	Gering	Ja
22.03.2001	Renten präsentieren sich fester	FAZ, 22.03.2001, Nr. 69, S. 42	Nachricht	Nein	Groß	Nein	Zeitlich	Gering	Ja
23.11.2004	Weitere Kursgewinne am Rentenmarkt	FAZ, 23.11.2004, Nr. 274, S. 28	Nachricht	Nein	Nein	Nein	Nein	Gering	Nein
17.03.1997	Eine Flucht in die D-Mark ist nicht zu beobachten	FAZ, 17.03.1997, Nr. 64, S. 23	Bericht	Nein	Gering	Nein	Nein	Gering	Nein
25.09.2012	Ifo-Index enttäuscht die Börsianer abermals	FAZ, 25.09.2012, Nr. 224, S. 24	Meldung/ Kurznachricht	Nein	Gering	Nein	Zeitlich	Moderat	Nein
20.03.2009	TERMINE DER WOCHE	FAZ, 20.03.2009, Nr. 67, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein

5.7.6.2 Der ifo-Index im Kontext der Kommunikation

Mit dem ifo-Index wird das theoretische Konzept *Geschäftsklima* operationalisiert. Es setzt sich aus den qualitativen Dimensionen *Erwartungen* und *Lage* zusammen und wird über Indikatoren der Beobachtungswelt zugänglich gemacht.

Das Konzept des ifo-Index ähnelt dem des GfK-Index. Beide besitzen als Dimensionen die Erwartung gegenüber der Zukunft und die aktuelle Situation. Jedoch sind die Indikatoren unterschiedlich und auch die Konstruktionsregeln weichen stark voneinander ab. Hinsichtlich der Zielgruppe zielt der GfK-Index auf die Konsumenten, wohingegen der ifo-Index die Unternehmen befragt. Wegen ihrer unterschiedlichen Ausrichtung haben beide Konzepte eine Berechtigung. Sie besetzen jeweils Teilaspekte der Einschätzung der konjunkturellen Entwicklung Deutschlands, womit sie als Beispiel für den gestiegenen Anspruch an Messapparaturen betrachtet werden können (vgl. Kapitel 3).

Der ifo-Geschäftsklimaindex ist schon einige Jahrzehnte in den Medien zu finden und scheint allgemein und auch speziell in den untersuchten Zeitungen vielbeachtet und etabliert zu sein. Die Relevanz des ifo-Index übersteigt dabei die des GfK-Index in FAZ und FAS bei Weitem.

FAZ und FAS reagieren regelmäßig in kürzester Zeit auf die Veröffentlichung des Index und kündigen diese sogar im Vorfeld an. Der überwiegende Teil der Beiträge wird dabei im Bereich Wirtschaft veröffentlicht. Dort entfaltet der Index seine Stärke darin, Komplexität zu reduzieren, indem die Vielschichtigkeit der konjunkturellen Entwicklung in einer Maßzahl verdichtet wird. Die Werte des ifo-Index werden dabei in einen historischen Kontext und in Zusammenhang zur aktuellen politischen, gesellschaftlichen oder wirtschaftlichen Lage gesetzt.

Mit der Nennung von Zahlenwerten stellt sich bei dem Index die Frage, welche Eigenschaften diese mit auf Messung basierten Zahlen teilen. Trotz der Nähe zu Wissenschaft und Forschung ist die Transparenz des Index eingeschränkt, was die Glaubwürdigkeit der Zahlen aus wissenschaftlicher Sicht beeinträchtigt. Anders als viel umstrittenere Indizes, wie z.B. dem *Better Life Index* der OECD (vgl. Abschnitt 5.7.4), fehlt es beim ifo-Index an einer umfassenden und transparenten Dokumentation zur Indexkonstruktion. Fragen zum theoretischen Konzept sowie zu den Hintergründen der Berechnungslogik und der Wahl der Indikatoren bleiben ungeklärt. Außerdem stehen die Daten nicht zur

Verfügung oder sind den Teilnehmern der Umfrage vorbehalten, was sicherlich z.T. auf Gründe des Datenschutzes zurückzuführen ist. Im Zusammenspiel zur intransparenten Dokumentation ergibt sich jedoch bei genauerem Hinsehen der Eindruck einer „Black-box“. Die Zurechnung von Objektivität ist damit entscheidend von dem Vertrauen (1) in die Zuverlässigkeit der Messverfahren, über die die Rohdaten gewonnen wurden, (2) in die Zuverlässigkeit der Operationalisierung des theoretischen Konzepts und (3) in die Zuverlässigkeit der verwendeten statistischen Verfahren sowie deren korrekte Anwendung abhängig. Förderlich ist dabei der Prognosecharakter des Index. So kann die Qualität der Operationalisierung retrospektiv über den Vergleich der Prognose mit „der“ Realität beurteilt werden. Jedoch wird, wissenschaftlich betrachtet, auch dieser Vergleich auf eine Konzeptualisierung hinauslaufen, da ‘Realität“ in diesem Zusammenhang erst definiert werden muss. Dafür kommen mehr oder weniger verbundene Konzepte, wie z.B. das Wirtschaftswachstum, in Betracht.

Förderlich für die Glaubwürdigkeit und das Vertrauen in den Index sind neben korrelierenden Konzepten (Wirtschaftswachstum, Arbeitslosigkeit, Löhne, etc.) auch die in Abschnitt 3.3 angesprochenen vertrauensbildenden Maßnahmen. So sollte das Vertrauen in die Objektivität des Index über die Reputation der Leibniz Gemeinschaft, der Nähe zur Universität München und die mediale Präsenz des ifo-Instituts selbst sowie in der Vergangenheit personell durch Hans-Werner Sinn positiv beeinflusst werden. In diesem Zusammenhang ist insbesondere auch das Auditing der Leibniz Gemeinschaft zu nennen.

5.7.7 Body-Mass-Index

Der *Body-Mass-Index* (BMI)[47] ist eine Maßzahl zur Bewertung des Verhältnisses von Körpergewicht und Körpergröße. Er wurde 1832 von dem belgischen Mathematiker, Astronom und Statistiker Adolphe Quetelet entwickelt. Damals jedoch unter der Bezeichnung Quetelet-Index. Erst 1972 fand er durch einen Artikel von Ancel Keys unter dem Namen Body-Mass-Index (BMI) Verbreitung. (Eknoyan, 2008)

Motivation bei der Definition des Index waren die bahnbrechenden Querschnittsstudien von Quetelet. Sie führten ihn zu dem Schluss, dass – bis auf die Wachstumsschübe nach der Geburt und vor der Pubertät – sich das Gewicht zum Quadrat der Körpergröße

[47] Auch bekannt unter Körpermasseindex, Körpermassenzahl oder Quetelet-Kaup-Index.

erhöht. Aufgrund seiner Anwendung der vergleichenden Statistik auf soziale Bedingungen und moralische Fragen gilt Quetelet als Begründer der Sozialwissenschaften. (Eknoyan, 2008)

Nach dem zweiten Weltkrieg erhielt der Index erneute Aufmerksamkeit durch einen Artikel von Keys. Zu dieser Zeit erhielt das Verhältnis von Körpergewicht und Sterblichkeit größere Bedeutung, insbesondere durch Herzerkrankungen und Diabetes. Das Problem des relativen Körpergewichts wurde von der epidemiologischen und klinischen Forschung aufgegriffen und der Quetelet-Index innerhalb vergleichender Studien von Keys validiert. Er empfahl, den Index für den statistischen Vergleich von Populationen zu verwenden, in Abgrenzung zur Beurteilung von individuellem Übergewicht (Keys u. a., 1972). Danach gewann er weitere Beachtung durch den Einsatz bei US-amerikanischen Lebensversicherungen und der BMI Klassifikation der Weltgesundheitsorganisation (WHO) (Eknoyan, 2008; WHO, 2015). Außerdem wird in einigen Bundesländern für die Verbeamtung der BMI-Wert einbezogen. (Mennen, 2013)

Der BMI besteht aus zwei Dimensionen: dem Körpergewicht und der Körpergröße. Beide werden zueinander ins Verhältnis gesetzt.

$$BMI = \frac{m}{l^2} \tag{5.17}$$

Das Körpergewicht in *Kilogramm* wird ins Verhältnis zum Quadrat der Körpergröße in *Meter* gesetzt. Grundsätzlich ist der Index unabhängig von den Einheiten bestimmbar und zwischen verschiedenen Populationen vergleichbar. Über Empfehlungen zum individuellen BMI-Wert oder Klassifikationstabellen hat sich der Standard kg/m^2 durchgesetzt.

Der BMI kann verschiedentlich eingesetzt werden. Neben dem statistischen Vergleich von Populationen wird er im Wesentlichen als Indikator für die Körpermasse oder den Ernährungsstatus verwendet.[48]

Die Beurteilung der Qualität des Index in Bezug auf Tabelle 4.15 auf Seite 156 wird ausschließlich anhand der Daten und Konstruktion durchgeführt, da kein konkreter Herausgeber im Hintergrund existiert.

[48] Die WHO unterscheidet zwischen den Ausprägungen: Underweight, Normal weight, Pre-obesity, Obesity class I, Obesity class II und Obesity class III. (WHO, 2015)

Der theoretische Rahmen des Index wurde von KEYS u. a. (1972) und in verschiedenen Aufsätzen, z.B. in der epidemiologischen und klinischen Forschung diskutiert und hergeleitet. Die Relevanz des Index ergibt sich aus der eingeschränkten Aussagekraft des Körpergewichts für z.B. die Identifikation von Adipositas und der Prävention gesundheitlicher Probleme, da das Körpergewicht abhängig ist von der Körpergröße. Dessen ungeachtet wird auch der BMI vielfach kritisiert, da die Körpermasse und der Ernährungsstatus von weiteren Faktoren wie Geschlecht, Alter, Knochenbau und Sportlichkeit maßgeblich beeinflusst wird. Alternative anthropometrische Größen zum Ernährungsstatus sind die Folge.[49] Es ist daher nur von einer ungenügenden Operationalisierung auszugehen, da zu wenige beeinflussende Faktoren im theoretischen Konzept enthalten sind. Diese werden z.B. über eine Trennung der Klassifikationstabellen nach Alter und Geschlecht konzeptextern verarbeitet.

Die Qualität des Index bzgl. der Selektion von Daten ist grundsätzlich positiv zu werten. Bei den Indikatoren handelt es sich um *hard data* und quantitative Daten (vgl. Abschnitt 4.3.2). Gewicht und Körpergröße werden in der Regel über genormte Messgeräte bestimmt und können als gesicherte Informationen angesehen werden.

Die Gewichtung und Aggregation der Daten ist einfach, transparent und nachvollziehbar. Die simple Handhabung ist sicherlich ein Faktor, der zur umfangreichen Verbreitung des Index beigetragen hat. Er kann von Laien selbst berechnet und benutzt werden und passt daher zum eingangs erwähnten Selbstvermessungstrend.

Da der Index seinen Ursprung in Forschung und Wissenschaft hat, wurde er wissenschaftlich ausgiebig diskutiert und kritisiert. Detaillierte Informationen zum Konzept und der Aussagekraft sind leicht zugänglich.

5.7.7.1 Inhaltsanalyse Body-Mass-Index

Der Body-Mass-Index ist seit 1972 unter dieser Bezeichnung bekannt. Für die vorliegende Inhaltsanalyse wurden jedoch auch die alternativen Bezeichnungen einbezogen.[50] Da der Index nicht von einem konkreten Herausgeber veröffentlicht wird, erfolgt keine Untersuchung der Aktualität.[51]

[49] Z.B. Area Mass Index, Ponderal-Index oder Broca-Index.

[50] Suchstring: *Body-Mass-Index OR „Body Mass Index“ OR Körpermasseindex OR Körpermassenzahl OR Quetelet-Kaup-Index OR „Quetelet Kaup Index“.*

[51] Grundsätzlich sind jedoch diverse Studienergebnisse als Ursache für einen Beitrag denkbar.

Abbildung 5.19: Anzahl Beiträge zum Body-Mass-Index

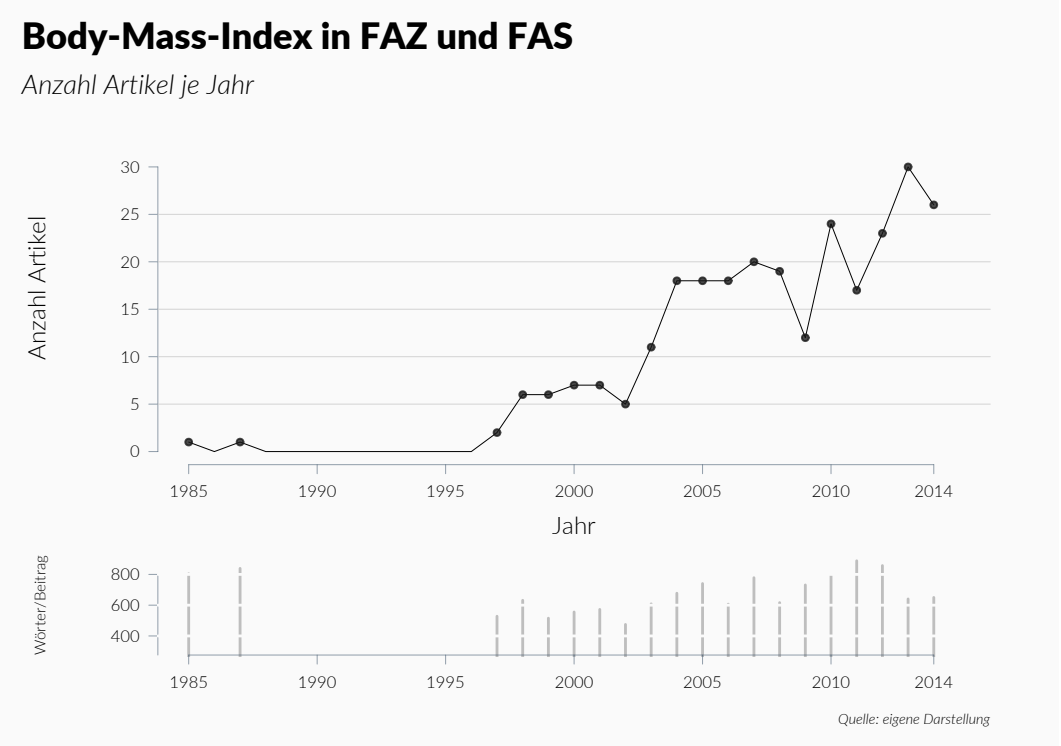


Tabelle 5.23: Ressort-Statistik zum Body-Mass-Index

Ressort	rel. Anzahl	Anzahl
Politik	15.5	42
Natur Und Wissenschaft	15.1	41
Sport	12.9	35
Wirtschaft	9.2	25
Feuilleton	7.4	20
Wissenschaft	7.4	20
Ohne Angabe	3.7	10
Frankfurt	2.6	7
Gesellschaft	2.6	7
Jugend Schreibt	2.2	6
Technik Und Motor	2.2	6
Gesamt	100	271

In der FAZ und FAS kam es 1985 erstmalig zur Nennung des BMI. Aus Abbildung 5.19 ist eine steigende Anzahl der Beiträge seit 1997 erkennbar. Im Durchschnitt wurde der Index seit 1985 jährlich in 9,0 Beiträgen genannt (ohne die Jahre vor 1996 sogar in 14,8 Beiträgen). Die Beiträge verfügen im Mittel über 698 Wörter. Damit liegen sie im Mittelfeld – verglichen mit anderen Indizes (vgl. Definition in Abschnitt 5.4.1.2.1). Der Trend in Abbildung 5.19 lässt vermuten, dass die Beiträge seit 1996 eher umfangreicher werden. Jedoch nur unter Ausblendung der Jahre 2013 und 2014. Das Vorliegen einer Systematik hinter dem Anstieg wird in der manuellen Analyse geprüft.

Tabelle 5.24: Numerische Kennzahlen zum Body-Mass-Index

Kennzahl	Min	Median	Max	Mittelwert	m. F.
Titelseite	0	0	100	8.5	27.9
Wörter/ Beitrag	50	634	2872	697.8	485.4
Anzahl Absätze	1	7	44	8.5	7.2
Wörter/ Titel	1	4	10	4.5	1.9
Anzahl Indexnennung/ Beitrag	1	1	6	1.2	0.6
Indexnennung/ Titel (Prozent)	-	-	-	0.7	-
Indexnennung/ 1. Absatz (Prozent)	-	-	-	27.7	-
Indexnennung/ Beitrag (Prozent)	-	-	-	56	-
Satzlänge	6.7	15.9	27.7	16.2	3.2
Aktualität					NaN

Über die größtenteils gleichmäßige Streuung der Beiträge über eine Vielzahl von Ressorts kann dem Index eine allgemeine Relevanz beigemessen werden (vgl. Tabelle 5.23). Der größte Anteil der Beiträge erscheint unter den Ressorts Politik (15,5%), Natur und Wissenschaft (15,1%), Sport (12,9%) und Wirtschaft (9,2%).

Die Relevanz des Index spiegelt sich auch in dem Anteil der Beiträge auf der Titelseite wieder. Nahezu jeder elfte Beitrag (8,5%) zum Index wird auf der Titelseite veröffentlicht. Der Name des Index erscheint nur sehr selten im Titel des Beitrags (0,7%), jedoch verhältnismäßig oft bereits im ersten Absatz (27,7%). Da in der Hierarchie eines Beitrags in der Regel das Wichtigste zuerst genannt wird, lässt sich die Vermutung aufstellen, dass der Index eine verhältnismäßig hohe Relevanz im Beitrag hat.

Über die manuelle Inhaltsanalyse zeigt sich, dass der größte Teil der Beiträge als Bericht (38,5%) und Nachricht (30,8%) veröffentlicht wurde.[52] Die Nachrichten basieren dabei größtenteils auf der Kommunikation aktualisierter Studienergebnisse, die in den Berichten ausführlicher dargestellt werden.

Neben dem Inhalt wird auch über die Beitragstitel deutlich, dass der überwiegende Teil thematisch auf Gesundheit und Ernährung ausgerichtet ist (vgl. Tabelle 5.25).[53] Die Funktion des Index ist verschieden. Er wird zum sozialen und gesellschaftlichen Vergleich (38,5%), zum sachlichen (23,1%) und zum Vergleich von Nationen (7,7%) eingesetzt.[54] In 46% der Beiträge wird er nicht vergleichend eingesetzt.

[52] Die restlichen Beiträge verteilen sich auf *Kommentar, Glosse, Kritik oder Leitartikel* (15,4%), *Leserbrief* (7,7%) und *Sonstige* (7,7%).

[53] Diese Behauptung resultiert u.a. aus den Begriffen: „Burger“, „Zusatzstoffe“, „Schokolade“, „Obst“, „Diabetes“, „Übergewicht“, „kochen“, „laufen“.

[54] Einige Beiträge verwenden den Index für mehrere Vergleichsarten. Daher summieren sich die einzelnen Prozentangaben nicht zu 100.

Der soziale und gesellschaftliche Vergleich geht im Wesentlichen aus der Definition von Übergewicht hervor. Der Begriff *Übergewicht* ist ohne Bezug vage. Mit messbaren Indikatoren und definierten Grenzwerten lässt er sich jedoch eindeutig definieren:

> „Besonders Übergewichtige sind gefährdet. Medizinisch gesehen übergewichtig ist man bei einem Body-Mass-Index (BMI, Gewicht in kg/Größe in cm2) [sic] von 25, von Fettleibigkeit oder Adipositas spricht man bei einem BMI über 30.“ (Rödder, 19. Januar 2003: S. 10)

Der sachliche Vergleich basiert weitestgehend auf Studienergebnissen. So wird der BMI z.B. ins Verhältnis zum Diabetes-Risiko gebracht:

> „Wenn der BMI um einen Punkt steigt, erhöht sich das Diabetes-Risiko etwa um das 1,6fache, sagt Joachim Spranger, einer der Autoren der Studie.“ (ebd.: S. 10)

In einem anderen Beitrag wird seine Verbindung zum Risiko, bei einem Autounfall ums Leben zu kommen, aufgezeigt:

> „Das Todesrisiko von Menschen mit einem BMI zwischen 35 und 39 ist ebenfalls mehr als doppelt so hoch wie bei Menschen mit einem BMI um zwanzig.“ (Rhk., 5. April 2002: S. 10)

Der Bezug der Maßzahl geht damit nicht auf den Autor des Beitrags zurück, sondern auf den Herausgeber der Studie.

Teilweise wurden Aussagen zum BMI auf ein Minimum reduziert. Bezogen auf den nationalen Vergleich z.B. auf die Aussage: *„dass deutsche Männer in Europa den Body-Mass-Index anführen“* (Körte, 25. Februar 2007: S. 29). Die dahinterstehende Aussage, dass Deutschland im Schnitt die meisten Übergewichtigen Männer hat, wird über die Nennung des Index numerisch belegt und scheinbar objektiver.

Im Großteil der Beiträge wird der Index argumentativ eingesetzt (69%). In den restlichen erfüllt er keine spezifische Funktion. So z.B. in einem Kommentar im Ressort *Reise*.

> „Möglicherweise verstehe ich das als Europäerin nicht, aber ich dürfte ja auch niemals ein Singapore Girl werden. Der Vorschrift (natürlich) nach dürfen das nur Asiatinnen. Angeblich gibt es das Gewicht betreffend keine Regeln. Roselyn Chua schrieb mir auf meine Frage hin, was passiere, wenn ein Singapore Girl plötzlich ganz viel zunehme, dass man bei Singapore Airlines weniger auf das Gewicht denn auf den Body Mass Index achte.“ (Baum, 2. November 2014: S. V3)

Die Transparenz des Index ist überwiegend gering (77%). Das bedeutet, dass neben dem Namen des Index keine weitere Information zum Index gegeben wird. Es wird in den Beiträgen also davon ausgegangen, dass der Leser über Hintergrundwissen verfügt. Da die Berechnung des Index und die Klassifikationstabellen u.a. im Tafelwerk der Sekundarstufen I und II enthalten ist sowie im Internet zahlreiche BMI-Rechner frei zugänglich sind, sollte der Index als Allgemeinwissen angesehen werden können (vgl. Erbrecht u. a. (1999: S. 141)).

Die restlichen Beiträge besitzen bezüglich des Index eine moderate (7,7%) oder hohe (15,4%) Transparenz. Beiträge mit einer hohen Transparenz nennen neben dem Namen des Index auch seine Dimensionen und weitere Informationen, z.B. zur Konstruktion des Index. Trotz des großen Bekanntheitsgrades des BMI wird etwa in einem Beitrag im Ressort Politik der Index umfangreich beschrieben.

> „Diesmal zogen die Forscher nicht das tatsächliche Gewicht, sondern den sogenannten Body mass index (BMI) zu Rate. Der BMI läßt sich einfach ermitteln: Körpergewicht in Kilogramm dividiert durch das Quadrat der Körpergröße in Metern. Ein BMI bis 25 gilt als Normalgewicht, ein BMI über 30 zeigt Übergewicht an. Der BMI ist von der Körpergröße vollkommen unabhängig: Ein 1,50 Meter großer Mensch mit 88 Kilogramm Körpergewicht hat den gleichen BMI von 39 wie ein 1,80 Meter großer Mensch mit 126 Kilogramm Gewicht.“ (Rhk., 5. April 2002: S. 10)

Missverständlich ist in diesem Auszug jedoch die Formulierung zur Unabhängigkeit. Diese ist bereits konzeptimmanent nicht gegeben. Schließlich ist der BMI eine Funktion von Körpergröße und Körpergewicht. Der Autor scheint vielmehr klar machen zu wollen, dass das Gewicht eines Menschen in Relation zu seiner Größe behandelt werden muss.

Das bedeutet, dass große Menschen natürlicherweise ein größeres Gewicht haben als kleinere.

Keiner der Beiträge ist negativ akzentuiert. Sie enthalten keine Konflikte. Darüber hinaus besitzt der Großteil keine Überraschung (69%) oder nur eine geringe (30,7%).

5.7.7.2 Der Body-Mass-Index im Kontext der Kommunikation

Die Quantifizierung des Body-Mass-Index (BMI) basiert auf Indikatoren, die über eine fundamentale Messung bestimmt werden. Das bedeutet, sie besitzen eine quantitative Struktur und sind als naturwissenschaftliche Messung zu betrachten. Die Messergebnisse sind unter gleichen Bedingungen reproduzierbar – unabhängig von Person[55], Raum und Zeit. (vgl. Kapitel 2)

In den Beiträgen der FAZ und FAS zeigt sich, dass der Index in zweierlei Hinsicht Verwendung findet. Neben den reinen Indexwerten, mit denen z.B. Nationen, soziale Schichten oder Zeitpunkte gegenüberstellt werden können, entwickelt der BMI in der Kommunikation zusätzlich eine Relevanz über die numerische Eingrenzung von unscharfen Begriffen wie z.B. Übergewicht oder Adipositas. Eine Klassifikation (z.B. von der WHO) ordnet dafür BMI-Werte definitorisch einzelnen Gruppen zu (z.B. Normalgewicht, Übergewicht etc.). In dieser Hinsicht scheint der BMI nicht zu versuchen, einen theoretischen Begriff zu operationalisieren sondern diesen über Werte zu definieren. Es geht also nicht darum, zu messen, was übergewichtig ist, sondern darum, dies zu definieren – z.B. über Grenzen. In den Beiträgen wurde diese Funktion i.d.R. innerhalb der breiten Themen Ernährung und Gesundheit verwendet. Hier bietet der Index eine konzeptionelle Basis für die Bewertung der Körpermasse, genau genommen das Verhältnis von Körpergewicht und Körpergröße. Diese Funktion zeigt sich in einer Vielzahl von Ressorts.

Neben dem Indexkonzept zeigte sich insbesondere in Beiträgen zu wissenschaftlichen Themen eine Relevanz der Indexwerte. Dabei ging es um den Zusammenhang der BMI-Werte zu anderen Kennzahlen und inwiefern der BMI diese erklären kann. Außerdem schafft der Index eine Basis, die es erlaubt, das subjektive Augenmaß des körperlichen Erscheinungsbildes auf eine scheinbar objektive Weise zu kommunizieren. Sender und

[55] Gemeint ist die Person, die die Messung durchführt.

Tabelle 5.25: Ergebnisse der manuellen Inhaltsanalyse zum BMI (I)

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
25.02.2007	Der mndige Burger	FAS, 25.02.2007, Nr. 8, S. 29	Kommentar, Glosse, Kritik oder Leitartikel	Nein	Nein	Nein	National	Gering	Ja
03.09.2008	Zusatzstoffe im Essen frdern hheres Gewicht	FAZ, 03.09.2008, Nr. 206, S. N2	Nachricht	Nein	Gering	Nein	Sozial	Gering	Ja
19.01.2003	Neue Waffe gegen Volkskrankheit Diabetes	FAS, 19.01.2003, Nr. 3, S. 10	Nachricht	Nein	Gering	Nein	Gesellschaftlich, Sachlich	Hoch	Ja
02.04.2007	Ambulant gegen bergewicht	FAZ, 02.04.2007, Nr. 78, S. 50	Nachricht	Nein	Nein	Nein	Sozial	Moderat	Ja
01.07.2007	Chilischote statt Schokolade	FAS, 01.07.2007, Nr. 26, S. R5	Bericht	Nein	Nein	Nein	Nein	Gering	Nein
11.12.2007	Deutschland nimmt Ma	FAZ, 11.12.2007, Nr. 288, S. 11	Bericht	Nein	Nein	Nein	Nein	Gering	Nein
26.04.2007	Viel Obst schadet nicht	FAZ, 26.04.2007, Nr. 97, S. 9	Bericht	Nein	Gering	Nein	Sachlich	Gering	Ja
30.11.1999	Wie luft's denn so?	FAZ, 30.11.1999, Nr. 279, S. T2	Sonstiges	Nein	Nein	Nein	Nein	Gering	Ja
05.04.2002	bergewicht erhht Risiko bei Unfllen	FAZ, 05.04.2002, Nr. 79, S. 10	Nachricht	Nein	Gering	Nein	Sozial, Sachlich	Hoch	Ja

Tabelle 5.26: Ergebnisse der manuellen Inhaltsanalyse zum BMI (II)

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
06.12.2014	Hollywood kocht ber	Frankfurter Allgemeine Magazin, 06.12.2014, Nr. 17, S. M96	Bericht	Nein	Nein	Nein	Nein	Gering	Nein
15.08.2010	Alkoholverbot I	FAS, 15.08.2010, Nr. 32, S. 28	Leserbrief	Nein	Nein	Nein	Nein	Gering	Ja
31.12.2004	Deutsche Springer landen in der zweiten Klasse	FAZ, 31.12.2004, Nr. 306, S. 34	Bericht	Nein	Nein	Nein	Gesellschaftlich	Gering	Ja
02.11.2014	Kebaya und Liebe	FAS, 02.11.2014, Nr. 44, S. V3	Kommentar, Glosse, Kritik oder Leitartikel	Nein	Nein	Nein	Nein	Gering	Nein

Empfänger haben mithilfe des Index den gleichen Ausgangspunkt, sodass er in seiner Eigenschaft als Kommunikationsmedium Störungen durch subjektive Missverständnisse vermeidet.

Die Definition eines gesunden Verhältnisses von Körpergewicht und Körpergröße geht auf das 18. Jhd. zurück, als erstmals Korpulenz nicht mehr automatisch mit Wohlstand gleichgesetzt wurde. Von da an entwickelte sich die Wichtigkeit des Index gleichzeitig mit jener des Körpergewichts im Allgemeinen und hinsichtlich der Verbindung zu Krankheitserscheinungen im Speziellen.

Die Zurechnung von Objektivität bei seinen Aussagen ist in zweierlei Hinsicht zu betrachten. Bei der Verwendung des Index für die Eingrenzung von unscharfen Begriffen (z.B. übergewichtig) wird die Glaubwürdigkeit sicherlich stark vom individuellen Körperempfinden abhängen. Steht dieses im Widerspruch zu den Aussagen, sinkt die Wahrscheinlichkeit für die Annahme einer Kommunikation. Anders ist es bei der wissenschaftlichen Kommunikation. Da der Index auf fundamentalen Messungen basiert, lassen sich die Eigenschaften von auf Messung basierender Zahlen aus Abschnitt 3.3 auf ihn übertragen, solange die zugrunde liegenden Messungen den Voraussetzungen entsprechen (Reproduzierbarkeit etc.). Die Glaubwürdigkeit des Index als Zusammenfassung der fundamentalen Messergebnisse wird im Wesentlichen davon abhängen, ob er neue und belastbare Informationen generieren kann. In diesem Rahmen stellt sich die Frage, ob der Index gewisse Phänomene (z.B. Krankheiten) exakter erklären kann als andere Maßzahlen.

Trotz umfangreicher Kritik an der Aussagekraft des Index und diversen Bemühungen einer Weiterentwicklung, nimmt die Relevanz bezogen auf die Anzahl der Beiträge in FAZ und FAS im Mittel kontinuierlich zu. Vorteilhaft in diesem Zusammenhang ist sicherlich das Verhältnis zwischen Komplexität des Konzepts und Einfachheit der Konstruktion sowie Bestimmung.[56]

[56] Die Aussage bezieht sich auf den Großteil der Gesellschaft. Für spezielle Sondergruppen (z.B. Menschen mit fehlenden Extremitäten oder Kleinkindern) existieren spezielle Klassifikationen und Ausnahmelösungen, die die Komplexität des Konzepts erhöhen.

5.7.8 Michigan Consumer Sentiment Index

Der *Konsumklimaindex der Universität Michigan* (MCSI[57]) wird als Indikator für die Konsumneigungen der Privathaushalte in den USA angesehen. Seine Werte werden u.a. für die Prognose der konjunkturellen Entwicklung oder einer aufkommenden Inflation genutzt. Insofern kann er auch als Frühindikator für die wirtschaftliche Entwicklung in den USA angesehen werden.

Das Konzept des MCSI ähnelt dem des GfK-Index (vgl. Abschnitt 5.7.5), der die Konsumneigungen der deutschen Bevölkerung misst. Dennoch existieren bei der Operationalisierung des Konzepts große Unterschiede. Auch die mediale Aufmerksamkeit unterscheidet sich. Der MCSI entwickelt über die Verbindung der internationalen Finanzmärkte und der dominanten Rolle der USA eine große Relevanz für die deutschen Märkte. So reagieren die Finanzmärkte sensibel auf unerwartete Veränderungen der Indexzahl.

Das theoretische Konzept des MCSI basiert auf fünf Indikatoren und wurde in den späten 1940er von Professor George Katona an der Universität Michigan entwickelt. Er ist der älteste Konsumklimaindex (Odekon, 2015: S. 152). Thematisch gibt es große Ähnlichkeit zu den Indikatoren des Gfk-Index, jedoch beinhaltet der MCSI zwei Dimensionen mehr. Für die Bildung der Indikatoren des MCSI werden repräsentativ ausgewählte Konsumenten zu den folgenden fünf Themenbereichen befragt (vgl. UMich (2014)):

1. Beurteilung der aktuellen finanziellen Situation des Haushaltes.[58]
2. Beurteilung der finanziellen Situation des Haushaltes in einem Jahr.[59]
3. Einschätzung der konjunkturellen Entwicklung in einem Jahr.[60]
4. Erwartung der konjunkturellen Entwicklung in den nächsten fünf Jahren.[61]

[57] MCSI steht für Michigan Consumer Sentiment Index.

[58] Frage: „We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?“

[59] Frage: „Now looking ahead–do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?“ (vgl. auch Gfk-Index)

[60] Frage: „Now turning to business conditions in the country as a whole–do you think that during the next twelve months we'll have good times financially, or bad times, or what?“ (vgl. auch Gfk-Index)

[61] Frage: „Looking ahead, which would you say is more likely–that in the country as a whole we'll have continuous good times during the next five years or so, or that we will have periods of widespread unemployment or depression, or what?“

5. Beabsichtigte Käufe.[62]

Im Vergleich zum Gfk-Index wird die eingeschränkte Objektiviät und fehlende Standardisierung der Operationalisierung desselben theoretischen Konzepts deutlich. Neben einer unterschiedlichen Fragestellung und Anzahl an Indikatoren können die Teilnehmer der MCSI-Befragung frei auf die Fragen antworten. Die Gfk-Befragung hingegen bietet lediglich drei Antwortmöglichkeiten. Der Vorteil von freien Antworten ist, dass ihre Kategorisierung ggf. an mehr Anhaltspunkten motiviert werden kann. Andererseits birgt es auch ein messtechnisches Risiko. Eine objektive Wiederholbarkeit der Messung ist nicht zwingend gegeben oder ist u.U. abhängig vom Interviewer.

Für die Bildung der Indikatoren werden die Sentimente der fünf Themenbereiche bestimmt und die relative Häufigkeit der positiven Antworten berechnet. Die fünf berechneten relativen Häufigkeiten werden addiert und zu dem Basiswert von 1966 (6,7558) in Beziehung gesetzt. Die Aussage des Index steht damit in Bezug auf das Basisjahr.

$$MCSI = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{6.7558} + 2.0 \tag{5.18}$$

Alle Indikatoren gehen gleich gewichtet in den Index ein. Jedoch wird über das theoretische Konzept eine ungleiche Gewichtung der Dimensionen *aktuelle Lage* (40%) und *Erwartung* (60%) umgesetzt, da sie jeweils über eine verschiedene Anzahl an Indikatoren verfügen. So kann der MCSI auch als gewichtet-additiver Index aus den Subindizes *Index of Current Conditions* (ICC) und *Index of Consumer Expectations* (ICE) beschrieben werden. Ersterer soll die Einschätzung der aktuellen Lage bestimmen, letzterer die Erwartungen gegenüber der Zukunft.

$$MCSI = 0.4 * ICC + 0.6 * ICE \tag{5.19}$$

wobei der ICC gegeben ist durch

$$ICC = \frac{X_1 + X_5}{2.6425} + 2.0 \tag{5.20}$$

[62] Frage: „About the big things people buy for their homes–such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or bad time for people to buy major household items?“ (vgl. auch Gfk-Index)

und der ICE durch

$$ICE = \frac{X_2 + X_3 + X_4}{4.1134} + 2.0 \tag{5.21}$$

Neben der Normierung des Index auf das Basisjahr 1966 ist in Formel 5.18 die Konstante „2.0“ auffällig. In der Geschichte des Index gab es gelegentlich Änderungen im Sampling-Design, was eine Korrektur mittels einer Konstante notwendig machte. Sie wird seit 1972 zum berechneten Quotienten addiert und unterlag ihrerseits verschiedenen Anpassungen.[63] Die Nutzung einer solchen Konstante kann notwendig sein, um die Veränderungen des Stichprobendesigns abzubilden, jedoch führt sie im Allgemeinen zu einer geringeren Transparenz und Nachvollziehbarkeit. Informationen zur Konstruktion des Index, Veröffentlichungsdaten, Fragebogen und Stichprobendesign können detailliert unter UMich (2016) eingesehen werden.

Die Nähe zur Universität Michigan führt bei der Erstellung des Index zur Beachtung wissenschaftlicher Standards, was die Qualität bezüglich der Tabelle 4.15 auf Seite 156 positiv beeinflusst. Dies betrifft insbesondere die Datenerhebung und Konstruktion des Index.

Der theoretische Rahmen des Index wird nicht hergeleitet, sondern vielmehr über die gesammelten Daten selbst begründet. Das bedeutet, dass die Validierung z.B. über die Prognosefähigkeit des Konzepts stattfindet. Es existiert keine klare Definition des theoretischen Konzepts *Consumer Sentiment* und keine theoretische Begründung des Zusammenhangs der beobachtbaren Indikatoren zu diesem Konstrukt.

Die Datenerhebung wird hingegen detailliert und nach wissenschaftlichen Standards beschrieben (ebd.). Dazu gehört u.a. die wissenschaftliche Diskussion über fehlende Werte, Verzerrungen und zum Stichprobendesign (UMich, 2015). Auch über die Gewichtung und Aggregation der Daten wird ausführlich Auskunft gegeben. Damit ist die grundsätzliche Berechnungslogik des Index transparent und nachvollziehbar. Jedoch wird die Anwendung der verwendeten Formel nicht begründet. Es bleibt unklar, warum z.B. Sentimente gebildet werden und ausgerechnet eine additive Kombination der Indikatoren durchgeführt wird.

[63] Folgende Werte hatte die Konstante in der Geschichte des MCSI: Apr. 1972 bis Nov. 1981: 2,7; seit Dez. 1981: 2.

Die Darstellung und Verbreitung der Indexwerte unterliegt Restriktionen. Grundsätzlich können der Webseite des Index[64] sämtliche Daten in Tabellen- und Diagrammform entnommen werden. Die Daten werden dabei bis auf die Ebene der Indikatoren und bis zum Jahr 1978 zur Verfügung gestellt. Die Berechnung des Index kann damit ab der Ebene der Indikatoren komplett nachvollzogen werden. Grafisch werden ausschließlich Zeitreihendiagramme in Linienform angeboten.

Die Restriktion bei der Verbreitung der Daten liegt in der eingeschränkten Aktualität. Daten, die aktueller als einen Monat sind, sind nur Sponsoren zugänglich. Dazu gehören z.B. Informationsdienstleistungs-, Nachrichten- und Medienunternehmen wie Bloomberg[65].

Seit 2007 wird der Index zusammen mit der Nachrichtenagentur Reuters herausgegeben. Reuters erhält dabei einen zeitlichen Vorsprung bei der Veröffentlichung der Indexwerte. Zwischen 2007 und 2013 gab die Agentur diesen Vorsprung an elitäre Kunden weiter, was zu Beschwerden führte. Insbesondere im Hochfrequenzhandel an der Börse führt eine zeitliche Verzögerung von 2 Sekunden zu Nachteilen. (Javers, 2013)

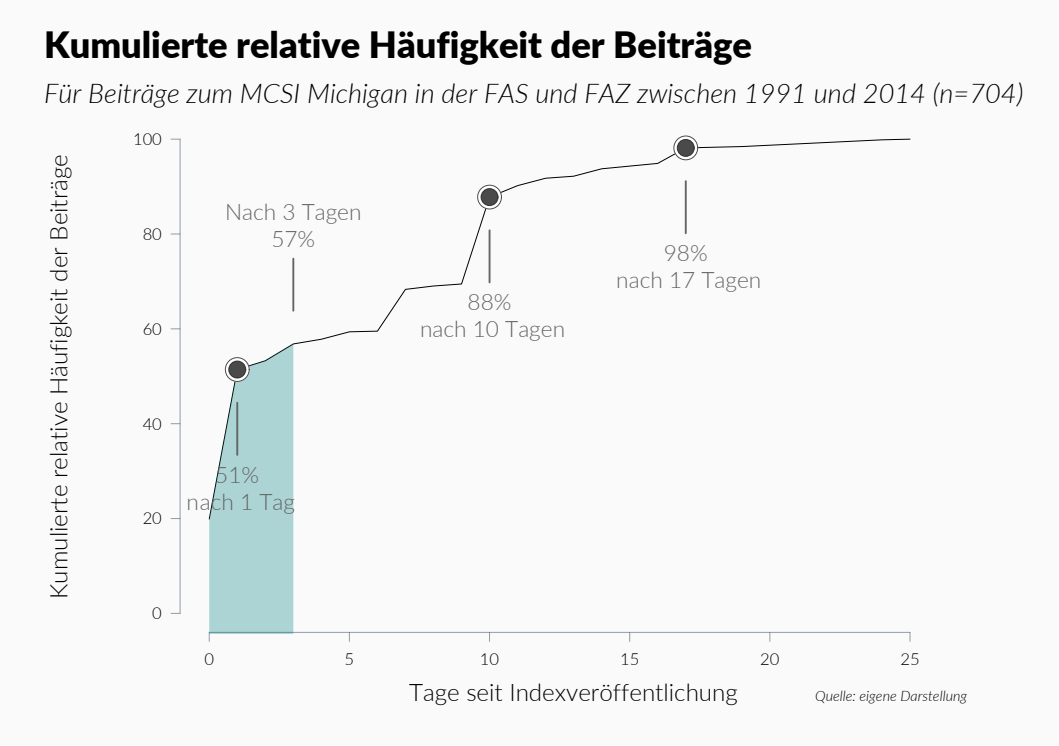
5.7.8.1 Inhaltsanalyse zum MCSI

Der MCSI gehört zu den ältesten Konsumklimaindizes. Er wird seit 1952 regelmäßig berechnet und veröffentlicht. Bis 1959 ein bis drei mal pro Jahr, zwischen 1960 und 1977 quartalsweise und ab 1978 monatlich. Obwohl er ausschließlich Aussagen über den US-amerikanischen Markt trifft, hat er in Folge der engen Verflechtung der globalen Finanzmärkte auch eine Relevanz für deutsche Märkte.

Die ersten Beiträge zum Index in der FAZ gehen auf das Jahr 1960 zurück (vgl. Abbildung 5.21). Eine regelmäßige Nennung ist jedoch erst ab 1990 festzustellen, wobei die Häufigkeit der Beiträge ab 2001 sprunghaft ansteigt. Eine mögliche Ursache ist die zu dieser Zeit einsetzende Rezession in den USA und die teils gravierenden Folgen für die europäische Wirtschaft: „Das Wachstum in Deutschland verfiel in eine Art Dauerstagnation, die Arbeitslosigkeit erreichte Höchststände“ (Rees, 2008: S. 3). Ab 2001 wurde der MCSI in einige regelmäßig erscheinende Artikelserien der FAZ aufgenommen. Im Durchschnitt wurde er seitdem jährlich in 47,2 Beiträgen erwähnt. Damit wird pro

[64] Siehe `https://data.sca.isr.umich.edu` (besucht am 2. Okt. 2015).
[65] Siehe `http://www.bloomberg.com/quote/CONSSENT:IND` (besucht am 2. Okt. 2015).

Abbildung 5.20: Aktualität der Beiträge zum MCSI



Monat im Mittel viermal auf den Index eingegangen. Zwischen 2005 und 2013 liegt die Häufigkeit der Beiträge auf einem stabilen Niveau zwischen 36 und 43 Beiträgen pro Jahr und fällt ab 2014 rapide ab. Der Grund für die Abnahme liegt in der fehlenden Nennung des Index in u.a. der Serie *Termine der Woche* ab März 2014.

Nahezu alle Beiträge zum Index erscheinen unter dem Ressort *Wirtschaft* und *Geld und Mehr* (Tabelle 5.28). Dies unterstreicht die wirtschaftliche Relevanz des Index.

Für die Bestimmung der Aktualität der Beiträge wurden sowohl die Termine der Vorabveröffentlichungen[66] als auch die der finalen Veröffentlichungen der Webseite der Universität Michigan entnommen (UMich, 2016). Sie liegen im PDF-Format für alle Monate seit 1991 vor. Der Zeitraum ab 1991 wird als aussagekräftig für die Bewertung der Aktualität eingestuft.

Abbildung 5.20 zeigt, dass der überwiegende Teil der Beiträge auf die Veröffentlichung der Universität Michigan zurückzuführen ist (57%). Der stärkere Anstieg zwischen Tag 10 und 11 ist auf die Beitragsserie *Termine der Woche* zurückzuführen, in der die Indexveröffentlichung teilweise sechs Tage im voraus angekündigt wird. Diese Beiträge werden

[66] In den Beiträgen wird auch auf die vorveröffentlichten Werte des Index eingegangen. Beispielhaft sei Reuters (27. September 2003: S. 26) genannt.

Abbildung 5.21: Anzahl Beiträge zum MCSI

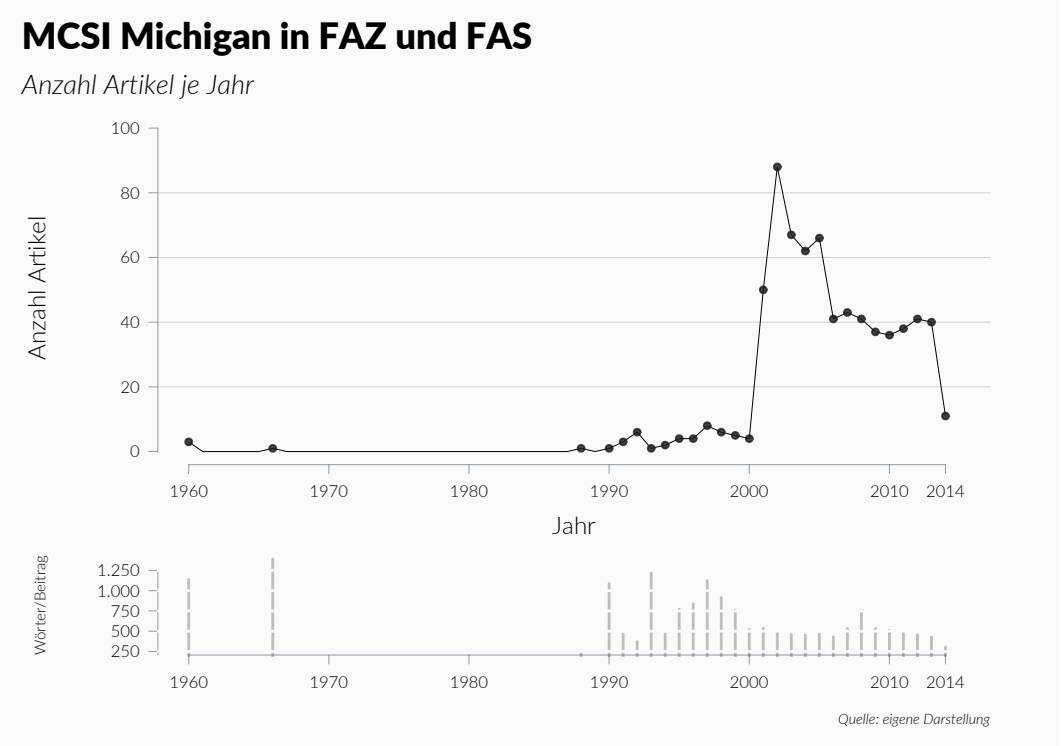


Tabelle 5.27: Numerische Kennzahlen zum MCSI

Kennzahl	Min	Median	Max	Mittelwert	m. F.
Titelseite	0	0	100	0.1	3.8
Wörter/ Beitrag	53	386	3500	515.6	383.6
Anzahl Absätze	1	5	146	36.2	46
Wörter/ Titel	1	4	11	4.3	1.7
Anzahl Indexnennung/ Beitrag	1	1	3	1.1	0.4
Indexnennung/ Titel (Prozent)	-	-	-	0.4	-
Indexnennung/ 1. Absatz (Prozent)	-	-	-	35.8	-
Indexnennung/ Beitrag (Prozent)	-	-	-	72.4	-
Satzlänge	2.7	12.8	27.1	11.9	4.8
Aktualität	0	1	25	5	5.5

gemäß der Definition in Abschnitt 5.4.1.2.9 der Vorperiode zugeschrieben, was in diesem Fall den irreführenden Anstieg erklärt. Schließlich wird in den Beiträgen der Serie explizit auf die Veröffentlichung des MCSI Bezug genommen. Insgesamt wird dem Index auf Basis dieser Ergebnisse eine hohe Aktualität zugeschrieben.

Die absolute Länge der Artikel unterliegt einer sehr großen Streuung zwischen 53 und 3500 Wörtern. Im Mittel sind die Beiträge mit 516 Wörtern jedoch verhältnismäßig kurz (vgl. Abschnitt 5.4.1.2.1).

Für eine eher geringe allgemeine Relevanz des Index spricht, dass er nur äußerst selten (0,4%) im Beitragstitel genannt wird (vgl. Tabelle 5.27). In 35,8% der Beiträge wird er

Tabelle 5.28: Ressort-Statistik zum MCSI

Ressort	rel. Anzahl	Anzahl
Wirtschaft	98.2	697
Geld Und Mehr	0.7	5
Geld Und Anlage	0.4	3
Immobilienmarkt	0.1	1
Natur Und Wissenschaft	0.1	1
Technik Und Motor	0.1	1
Wissenschaft	0.1	1
Ohne Angabe	0.1	1
Gesamt	100	710

im ersten Absatz genannt, was für die Relevanz des Index im Beitrag gewertet werden kann. Im Mittel erscheint er jedoch erst im letzten Viertel des Beitrags. Auf der Titelseite der FAZ und FAS erscheint er nahezu nie.

Die dominierende Zahl an Beiträgen (90%) ist Serien zuzuordnen. Das zeigt die manuelle Inhaltsanalyse (vgl. 5.29). Zu den regelmäßig erscheinenden Beiträgen gehört *Börse heute* (33,3%), *Termine der Woche* (30%), *Amerikanische Börsen* (16,7%) und *Renten- und Devisenmarktbericht* sowie *Rentenmarktbericht* (zusammen 10%).

Der überwiegende Teil der Beiträge wurde als *Meldung/Kurznachricht* verfasst (77%). Dieses Wert korrespondiert mit dem automatisiert erhobenen Kriterium Artikellänge, da die Beiträge im Mittel verhältnismäßig kurz sind. Die Kürze der Meldungen spiegelt sich auch in der Verwendung des MCSI im Beitrag wider. Der Großteil der Beiträge beschränkt sich auf die bloße Nennung des Index, ohne ihn argumentativ einzusetzen (94%). Jedoch reicht der Umfang des Beitrags, um den Indexwert in 60% der Fälle zeitlich einzuordnen, also mit einem Vorperiodenwert in Verbindung zu bringen. Dies findet entweder rein sprachlich (gestiegen, gefallen) oder mit der konkreten Nennung der Werte statt. Eine Besonderheit ist, dass in 20% der Beiträge der Index zusätzlich in Verbindung zu den Erwartungen von Analysten gebracht wird:

> „Das Verbrauchervertrauen in Amerika dürfte nach Messung der Universität von Michigan im Juli abermals gestiegen sein. Analysten rechnen bei den heute Nachmittag europäischer Zeit vorgelegten Daten mit einem Wert von 84 Punkten.“ (Mohr, 26. Juli 2013: S. 21)

Das bedeutet, dass hier eine doppelte Erwartungshaltung vorliegt. Analysten erwarten bestimmte Erwartungen bei den Konsumenten. Es ist fraglich, ob solche Prognosen

sinnvoll sind. Jedoch sind sie bekanntermaßen das Tagesgeschäft der Analysten. Sollten diese treffsicher sein, würde es die Existenzberechtigung des Index in Frage stellen. Denn sie würden, (a) Redundanz bedeuten und (b) je nach Art der Bestimmung durch die Analysten, die Verbindung der beobachtbaren Indikatoren zum theoretischen Konzept in Frage stellen (vgl. Abschnitt 4.3.2). In diesem Zusammenhang lässt sich die Frage stellen, was von den Analysten quantifiziert wird? Das theoretische Konzept *Konsumneigung* oder der Index. Letzteres sollte eigentlich irrelevant sein, da der Indexwert selbst nur eine Abbildung darstellt. Er existiert ausschließlich in der Theorie und sollte sämtliche quantitative Aussagen über ihn hinfällig machen. Es sei denn, der Index wirkt auf andere Systeme ein (Funktion des Index im instrumentellen Sinne). Denn dann kann aus der Analystenprognose abgeleitet werden, wie z.B. die Finanzmärkte auf die Veröffentlichung des Index reagieren werden.

Tatsächlich haben die Prognosen der Analysten in dieser Hinsicht für den Markt eine Relevanz. Es geht darum, als Erstes an Informationen zu kommen, um dem Markt einen Schritt voraus zu sein. Insofern hat an dieser Stelle der zitierte Beitrag einen zusätzlichen Wert, indem er die Veröffentlichung des Indexwertes scheinbar vorwegnehmen kann.

Überraschungen entstehen in den Beiträgen ausschließlich über Abweichungen von den erwarteten Indexwerten. Dies geschieht zu 17% in geringerem und in 10% zu größerem Maße.

> „Dagegen enttäuschte das überraschend schwächere Vertrauen der amerikanischen Verbraucher in die künftige Konjunkturentwicklung ihres Landes die Investoren. Der von der Universität Michigan errechnete Index fiel auf 87,7 Punkte von 89,3 Punkten im August.“ (Reuters, 27. September 2003: S. 26)

Ein Konflikt ist bei keinem der Beiträge festzustellen.

5.7.8.2 Der MCSI im Kontext der Kommunikation

Der MCSI operationalisiert, wie der Gfk-Index, das theoretische Konzept *Konsumneigung*. Es handelt sich dabei um ein qualitatives Merkmal, was weder fundamental noch

Tabelle 5.29: Ergebnisse der manuellen Inhaltsanalyse zum MCSI

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
29.08.2003	Die Brse heute	FAZ, 29.08.2003, Nr. 200, S. 19	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich, Sachlich	Gering	Nein
27.09.2003	Wall Street schliet nach uneinheitlichen Konjunkturdaten ein wenig leichter	FAZ, 27.09.2003, Nr. 225, S. 26	Nachricht	Nein	gro	Nein	Zeitlich	Moderat	Nein
12.05.2001	Nach Bekanntgabe hohen Verbrauchervertrauens sinken die Aktienkurse	FAZ, 12.05.2001, Nr. 110, S. 33	Nachricht	Nein	gro	Nein	Zeitlich	Gering	Ja
17.02.2006	Die Brse heute	FAZ, 17.02.2006, Nr. 41, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
14.04.2012	Die Kurse an den Brsen in Europa fallen	FAZ, 14.04.2012, Nr. 88, S. 23	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich	Gering	Nein
03.02.2012	TERMINE DER WOCHE	FAZ, 03.02.2012, Nr. 29, S. 19	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
21.11.2005	Die F.A.Z.-Termine der Woche - Wirtschaft	FAZ, 21.11.2005, Nr. 271, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
28.07.2006	Die Brse heute	FAZ, 28.07.2006, Nr. 173, S. 19	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
23.11.2005	Die Brse heute	FAZ, 23.11.2005, Nr. 273, S. 23	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
01.10.2005	Amerikanische Aktienindizes fast unverndert	FAZ, 01.10.2005, Nr. 229, S. 30	Nachricht	Nein	gering	Nein	Zeitlich	Gering	Nein

Tabelle 5.30: Ergebnisse der manuellen Inhaltsanalyse zum MCSI

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
22.10.2010	TERMINE DER WOCHE	FAZ, 22.10.2010, Nr. 246, S. 15	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
27.11.1991	Wall Street leicht verbessert	FAZ, 27.11.1991, S. 28	Nachricht	Nein	gering	Nein	Zeitlich, Sachlich	Gering	Nein
12.09.1997	Die Brse heute	FAZ, 12.09.1997, Nr. 212, S. 17	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
14.06.2003	Michigan-Index treibt Euro nach oben	FAZ, 14.06.2003, Nr. 136, S. 26	Meldung/ Kurznachricht	Nein	gro	Nein	Zeitlich, Sachlich	Gering	Ja
17.10.2008	Nachrichten aus dem Krisenherd	FAZ, 17.10.2008, Nr. 243, S. 23	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich	Gering	Nein
18.10.2003	Gewinnmitnahmen nach guten Quartalszahlen	FAZ, 18.10.2003, Nr. 242, S. 28	Nachricht	Nein	gering	Nein	Zeitlich	Gering	Nein
10.03.2008	TERMINE DER WOCHE	FAZ, 10.03.2008, Nr. 59, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
09.11.2006	Die Brse heute	FAZ, 09.11.2006, Nr. 261, S. 23	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
30.11.2012	TERMINE DER WOCHE	FAZ, 30.11.2012, Nr. 280, S. 16	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
18.01.2008	General Electric berichtet	FAZ, 18.01.2008, Nr. 15, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich, Sachlich	Gering	Nein

Tabelle 5.31: Ergebnisse der manuellen Inhaltsanalyse zum MCSI

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
27.06.2003	Die Brse heute	FAZ, 27.06.2003, Nr. 146, S. 19	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
24.12.2003	Rentenhndler warten auf neue Zahlen	FAZ, 24.12.2003, Nr. 299, S. 26	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich	Gering	Nein
27.02.2010	Anleihen und Euro wenig verndert	FAZ, 27.02.2010, Nr. 49, S. 26	Nachricht	Nein	gering	Nein	Zeitlich, Sachlich	Gering	Nein
04.12.2006	Termine der Woche	FAZ, 04.12.2006, Nr. 282, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
26.07.2013	Zuversicht in Amerika	FAZ, 26.07.2013, Nr. 171, S. 21	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich	Gering	Nein
13.03.2010	Euro auf hchstem Stand seit einem Monat	FAZ, 13.03.2010, Nr. 61, S. 24	Nachricht	Nein	gering	Nein	Zeitlich, Sachlich	Gering	Nein
25.11.2002	Die F.A.Z.-Termine der Woche	FAZ, 25.11.2002, Nr. 274, S. 14	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich	Gering	Nein
09.05.2005	Die F.A.Z.-Termine der Woche - Wirtschaft	FAZ, 09.05.2005, Nr. 106, S. 23	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich	Gering	Nein
23.07.2010	TERMINE DER WOCHE	FAZ, 23.07.2010, Nr. 168, S. 17	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich	Gering	Nein
17.01.2004	Kursgewinne in Wall Street nach neuen Konjunkturdaten	FAZ, 17.01.2004, Nr. 14, S. 28	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich	Gering	Nein

abgeleitet bestimmt werden kann. Das Konzept wird mehrdimensional über verschiedene Sentimente definiert und ist daher prädestiniert für die Indexbildung.

Der MCSI ist als Beispiel für die steigende Anforderung an Messapparaturen zu sehen. Komplexe Zusammenhänge innerhalb der Wirtschaft sollen kontrollierbar, überschaubar und kommunizierbar gemacht werden. Die Relevanz dieser Quantifizierung wird über die lange Historie von über fünfzig Jahren und der zeitnahen sowie regelmäßigen Nennung des Index in den Beiträgen der FAZ und FAS deutlich. Seit 2001 hat er einen festen Platz in regelmäßig erscheinenden Beitragsserien, wie z.B. *Termine der Woche*.

Neben der Nennung der Indexwerte in den Beiträgen wird inhaltlich nicht weiter auf ihn eingegangen. In der Regel sind ein bis drei Sätze ausreichend. Die Verbreitung und Geschichte des Index macht es anscheinend überflüssig, ihn in den Beiträgen näher zu beschreiben oder zu diskutieren.

Eine Besonderheit des MCSI sind die Prognosen von Analysten zu künftigen Indexwerten. Grundsätzlich sind Printmedien für zeit-sensible Börseninformation kein geeignetes Medium. Dies ist sicherlich ein Grund, weswegen zwei Drittel der Beiträge (*Termine der Woche*, *Börse heute*) die Veröffentlichung des Index nur ankündigen. Der Gehalt der Ankündigung wird z.T. noch über die Nennung des erwarteten Indexwertes angereichert. Eine Information, die es dem Leser erlaubt, das Ereignis in Verbindung zur Prognose zu beurteilen.

Die Glaubwürdigkeit des Index wird, neben der langen Historie, durch die umfassende Transparenz positiv beeinflusst. Wissenschaftliche Diskussionen und detaillierte Information zur Konstruktion sollten die Zurechnung von Objektivität positiv beeinflussen. Gestützt wird dieser Eindruck von der guten Reputation der Universität Michigan.

Wie z.B. auch der Gfk-Konsumklimaindex und ifo-Geschäftsklimaindex ist der MCSI zukunftsgerichtet. Somit kann seine Qualität retrospektiv anhand von Vergleichen mit z.B. der realen Entwicklung beurteilt werden. Die Glaubwürdigkeit in den Index wird steigen, wenn sich die erwartete Entwicklung bewahrheitet.

Rein wissenschaftlich gesehen, können bei dem MCSI die gleichen Einwände gebracht werden, wie z.B. beim GfK-Index (vgl. Abschnitt 5.7.5.2). Die Objektivität des theoretischen Konzepts kann bereits darüber in Frage gestellt werden, dass ähnliche Konzepte grundsätzlich anders operationalisiert werden.

5.7.9 Deutscher Aktienindex

Der *Deutsche Aktienindex* (DAX) lässt sich als Beispiel für eine ganze Reihe von Aktienindizes betrachten. Seine Prominenz und Relevanz hebt sich gegenüber den anderen Aktienindizes dadurch ab, dass „auf keinen europäischen Index mehr Finanzprodukte begeben und in keinen mehr Geld investiert [wird]“ (Mohr, 28. Juni 2013: S. 24).

Aktienindizes verdichten die Kursentwicklung eines Marktes in einer Maßzahl. „Die Gesamtbeschreibung des Marktes wäre wegen der kaum überschaubaren Fülle von Einzelinformationen sonst nur schwer zu bewerkstelligen und zu beurteilen“ (Janssen und Rudolph, 1992: S. 2). In dieser Hinsicht senkt der Index Komplexität und Mehrdimensionalität. Eine abstraktere Betrachtungsweise sieht Aktienindizes als Indikatoren, Standards oder Barometer, die die Börsenstimmung und das Börsenklima quantifizieren. (Bleymüller, 1966: S. 19)

Neben der Informationsfunktion kommt dem Aktienindex auch eine operative Funktion zu, da er z.B. auch bei börsenmäßigen Termingeschäften als Basisobjekt zugrunde gelegt wird.[67] So existiert in Deutschland ein börsenmäßiger Terminhandel auf Indexprodukte (DAX-Future).

Der Deutsche Aktienindex wurde von der *Arbeitsgemeinschaft der Deutschen Wertpapierbörsen* und der *Frankfurter Wertpapierbörse AG* gemeinschaftlich entwickelt und am 23. Juli 1988 offiziell vorgestellt. Damit ist er verglichen mit dem Aktienindex der Frankfurter Allgemeinen Zeitung – erstmals 1961 veröffentlicht – verhältnismäßig jung. Die Zielstellung des DAX ist:

> „Er soll Händlern und Anlegern einen raschen Überblick über das aktuelle Marktgeschehen in deutschen Standardaktien vermitteln, gleichzeitig aber auch als längerfristiger Vergleichsmaßstab dienen.“[68]

Beim DAX handelt es sich um die Aggregation von Kursen eines ausgewählten Aktienkorbes zum Berichtzeitpunkt bezogen auf einen Basiszeitpunkt. Die Zusammensetzung des Aktienkorbes wird regelmäßig geprüft und ggf. angepasst. Für die Aufnahme eines Aktienunternehmens (Indikator) in den DAX müssen bestimmte Kriterien erfüllt

[67] Für eine detaillierte Übersicht über die Funktionen von Aktienindizes sei auf Janssen und Rudolph (1992: S. 3f.) verwiesen.

[68] Dr. Rüdiger von Rosen, Sprecher des Vorstandes der Frankfurter Wertpapierbörse AG, im Vorwort zu Janssen und Rudolph (ebd.: S. V).

werden. Es werden nur die 30 größten und umsatzstärksten deutschen Aktienunternehmen an der Frankfurter Wertpapierbörse einbezogen. Die Auswahl erfolgt anhand von Marktkapitalisierung und Orderbuchumsatz.[69]

Der DAX wird in zwei verschiedenen Varianten veröffentlicht: als Performance- und Kursindex. In der Regel ist unter der umgangssprachlichen Bezeichnung DAX der Performance-Index zu verstehen. Seine Berechnung folgt der Formel:

$$DAX_t = K_T * \frac{\sum_i p_{it} * q_{iT} * ff_{iT} * c_{it}}{\sum_i p_{i0} * q_{i0}} * 1000$$

mit:

c_{it} = Korrekturfaktor der Gesellschaft i zum Zeitpunkt t zur Bereinigung von Dividenden u. Kapitalmaßnahmen

ff_{it} = Free Float-Faktor der Gattung i zum Zeitpunkt T

i = Gesellschaft i (i = 1,..., 30)

p_{i0} = Schlußkurs der Gesellschaft i am Handelstag vor der ersten Aufnahme in einen Index der Deutschen Börse

p_{it} = Aktienkurs der Gesellschaft i zum Zeitpunkt t

q_{i0} = Anzahl der Aktien der Gesellschaft i am Handelstag vor der ersten Aufnahme in einen Index der Deutschen Börse

q_{iT} = Anzahl der Aktien der Gesellschaft i zum Zeitpunkt T

t = Berechnungszeitpunkt des Index

K_T = Konstanter Verkettungsfaktor, gültig ab Zeitpunkt T

T = letzter Anpassungs-/Verkettangstermin

(5.22)

Vereinfacht ist der DAX ein gewichteter Index, bei dem die Einzelkurse p gemäß des Konzepts von Laspeyres[70] ins Verhältnis zum Basisjahr (1987 = 1.000) gesetzt werden. Die Gewichtung der Titel wird über die Größe des jeweiligen Aktienunternehmens festgelegt.

[69] Für einen detaillierten Überblick über alle Auswahlkriterien für die Aufnahme von Aktienunternehmen in den DAX sei auf Deutsche Börse AG (2015: S. 23ff.) verwiesen.

[70] Für weiterführende Informationen zum Preisindex nach Laspeyres sei auf Fahrmeir (1997: S. 551f.) verwiesen.

Aus dem Aktienkurs p und der Anzahl der Aktien q lässt sich die Marktkapitalisierung berechnen. Für die Berechnung des DAX wird dabei jedoch nur die Marktkapitalisierung von im Streubesitz befindlichen Aktien einbezogen (*ff*). Das bedeutet, dass die Anteile von Großaktionären nicht in die Berechnung eingehen. (vgl. Deutsche Börse AG (2015)).

Im Gegensatz zu dem Kursindex bezieht der Performance-Index die Dividenden und Kapitalmaßnahmen ein. Das bedeutet, dass bei ihm Dividendenausschüttungen keinen Indexrückgang zur Folge haben, da sie durch Wiederanlage im Index wertmäßig erhalten bleiben. Die Korrektur wird über den Faktor c_{it} in Formel 5.22 erreicht. Um bei einer langanhaltenden Kumulation der Korrekturfaktoren c_{it} Gewichtsverzerrungen zu vermeiden, werden die Korrekturfaktoren jährlich auf 1 zurückgesetzt und in den Verkettungsfaktor K_T überführt. (ebd.: S. 22ff.)

Im Jahr 2002 hat die Deutsche Börse das Indexkonzept umgestellt. Seither gehen Unternehmen nicht mehr mit ihrem kompletten Börsenwert in den Index ein. Wie oben bereits angemerkt, werden nur im Streubesitz befindliche Anteile einbezogen. Der in Formel 5.22 verwendete *Free Float-Faktor* bemisst genau die Anzahl der im Streubesitz befindlichen Aktien. Er wird von der Börse festgelegt und ist nicht frei verfügbar.

Der Index sieht sich vor allem wegen seiner Komplexität Kritik gegenüber, zumal die Berechnung seit 2002 mit der Einführung des *Free Float-Faktors* nur noch schwer nachvollziehbar ist.

> „Die Konzeption des Dax als Performance-Index mit Dividendenberücksichtigung gilt unter Indexexperten als nicht mehr zeitgemäß. Die Berücksichtigung der Korrekturfaktoren macht die Nachbildung komplizierter als bei einem reinen Kursindex. [...] Die aktuellen Free-Float-Faktoren erhalten Kunden nur gegen Bezahlung von Gebühren. Der aktuelle Dax-Stand ist seither für Außenstehende nicht mehr nachvollziehbar. [...] Moderne Indexkonzepte sehen zudem eine Gleichgewichtung aller Indextitel vor; dies ist transparenter und mindert das Risiko bei meist höherem Ertrag.“ (Mohr, 28. Juni 2013: S. 24)

Damit ergeben sich auch Auswirkungen auf die Qualität des Index mit Bezug auf Tabelle 4.15 auf Seite 156. Grundsätzlich wird der theoretische Rahmen des Index transparent

dargestellt. Die Selektion der zugrunde liegenden Aktien (Indikatoren, Dimensionen) wird über Regeln nachvollziehbar begründet. Jedoch bleibt unbegründet, ob 30 Unternehmenswerte für die Bildung des Index tatsächlich repräsentativ für den gesamten Markt sind.[71]

Die Konstruktion des DAX wird im *Leitfaden zu den Aktienindizes* der Deutsche Börse AG (2015) ausführlich dargestellt. Die Gewichtung der einzelnen Dimensionen wird anhand der Unternehmensgröße motiviert. Jedoch führt der *Free-Float-Faktor* zu Einschränkungen bei der Nachvollziehbarkeit der Berechnung, da er nicht frei verfügbar ist. Außerdem war der Index in der Vergangenheit durch diesen Faktor anfällig gegenüber Extremsituationen. So kam es 2008 durch eine heftige Short Squeeze der Volkswagen Aktie zur Situation, dass nur noch 6% ihrer Anteile in Streubesitz waren und unter einem Kurs von zeitweise 1.000 EUR das Unternehmen nach der o.g. Berechnungslogik die weltgrößte Marktkapitalisierung besaß (Currie, 2008). Die Aktie wurde folglich im Index übergewichtet.

Negativ bezüglich der Transparenz der Konstruktion sind die diversen Auswahlkriterien für die Aufnahme eines Unternehmens. Außerdem wird z.B. die Kappungsgrenze bei der Gewichtung (ein Titel darf maximal ein Gewicht von 10% im Index besitzen) nicht theoretisch begründet. Diese definierten Regeln steigern die Komplexität des Indexkonstruktes.

Die Aktualität des Index ist grundsätzlich positiv. 1992 wurde der Index noch im 60-Sekunden-Takt berechnet (Janssen und Rudolph, 1992: S. 6), aktuell bereits sekündlich Deutsche Börse AG (2015). Problematisch könnte in diesem Zusammenhang sein, dass die berechneten Indexdaten durch z.B. die Latenz der Datenleitungen nicht jedem Akteur zur selben Zeit zur Verfügung stehen. Dieser technische Umstand kann schnell zu größeren Marktnachteilen führen. Außerdem haben Printmedien Schwierigkeiten, der hohen Veröffentlichungsfrequenz nachzukommen. In der Regel kann sich nur auf eine ex post Beschreibung des Kursverlaufs und auf Schlusskurse bezogen werden.

Die Verbreitung der berechneten Werte findet über verschiedene Kanäle und Darstellungsformen statt. Die Berechnung des DAX basiert auf den Kursen des elektronischen Handelssystems Xetra. Das bedeutet, dass die „messbaren“ Elemente des Index ausschließlich von Xetra erhoben werden. Die anschließende Berechnung des DAX folgt

[71] Zum Vergleich: Der FAZ-Index beinhaltet 100 Aktientitel.

Tabelle 5.32: Anzahl Tage zwischen Nennung des DAX

Tage	Anzahl	rel. Anzahl
1	5708	93.9
2	350	5.8
3	23	0.4

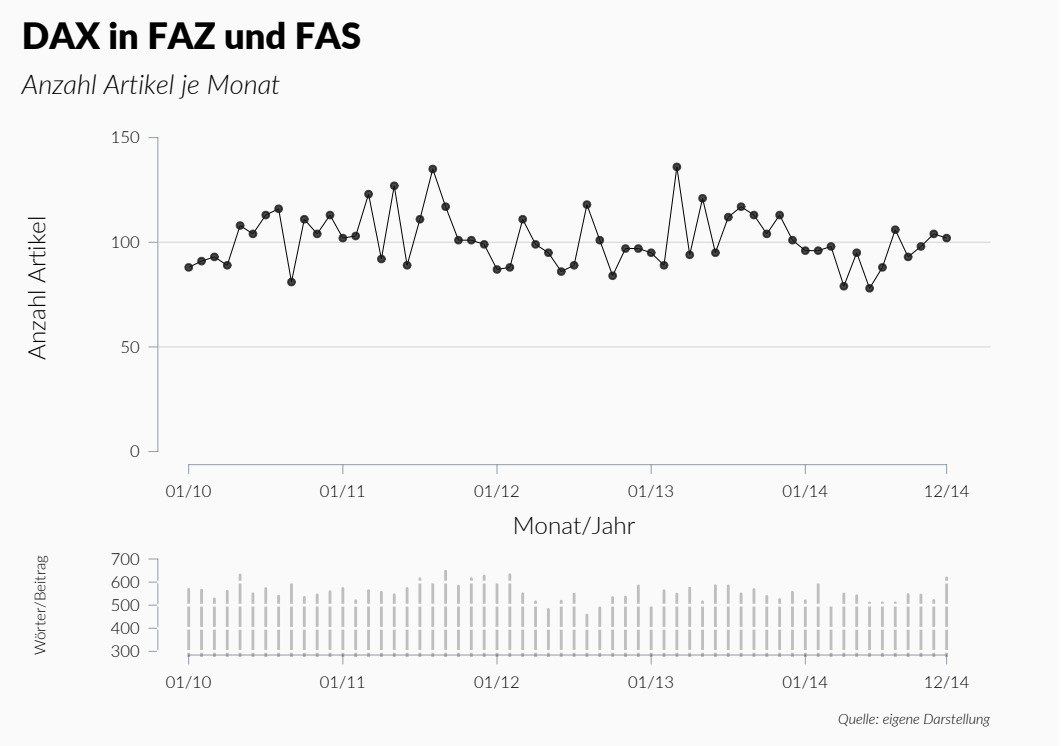
Formel 5.22 und kann grundsätzlich lückenlos nachvollzogen werden. Jedoch ist der DAX eine eingetragene Marke der Deutschen Börse AG. Seine Nutzung und Veröffentlichung unterliegt damit Einschränkungen (Deutsche Börse AG, 2015: S. 57). Die aktuellen Werte des DAX sind auf diversen Plattformen zugänglich (z.B. Webseiten von Banken, Handelsplattformen oder im Speziellen auf `www.faz.net`). In der Regel beschränkt sich die Veröffentlichung der Werte auf die bloße Nennung des aktuellen Wertes, Tabellen oder Zeitreihendiagramme in Linienform.

5.7.9.1 Inhaltsanalyse zum DAX

Der DAX ist der wichtigste Aktienindex in Deutschland und gehört auch innerhalb der FAZ und FAS zu den meistzitierten Indizes. Da der Index sekündlich berechnet und veröffentlicht wird, ist die Aktualität der Beiträge in FAZ und FAS bedingt zu interpretieren. Das im Schema für die Inhaltsanalyse (Abschnitt 5.7.1) genannte Diagramm zur Aktualität der Beiträge findet in Bezug auf den DAX keine Verwendung. Der DAX wird nahezu in jeder Ausgabe der FAZ und FAS zitiert und hat insofern bereits eine hohe Aktualität. Daher lautet die Frage in diesem Zusammenhang eher, an welchen Tagen der Index nicht zitiert wurde. In Tabelle 5.32 werden die Werte zur Beantwortung dieser Frage zusammengetragen. Demnach wird in 93,9% der Fälle der Index täglich genannt, in den restlichen 6,2% mindestens alle drei Tage. Wegen der regelmäßigen Nennung des Index ist ihm eine hohe Aktualität und große zeitliche Relevanz beizumessen.

Erwartungsgemäß erscheint der überwiegende Teil der Beiträge im Ressort *Wirschaft* und *Geld und Mehr* (zusammen 92,6%). Der DAX ist direkt mit der Börse und über die Unternehmen mit der Wirtschaft verbunden. Damit entfaltet er insbesondere in diesem Bereich eine Relevanz. Die restlichen 7,2% verteilen sich auf eine Vielzahl von Ressorts. Die Ressorts mit der nächstgrößeren Häufigkeit sind mit jeweils 1,2% *Politik* und *Verlagsbeilage*. Alle weiteren besitzen unter 1% und können Tabelle 5.34 entnommen werden.

Abbildung 5.22: Anzahl Beiträge zum DAX



Aufgrund der häufigen Nennung des DAX in FAZ und FAS wurden für diesen Index nur Beiträge zwischen Anfang 2010 und Ende 2014 einbezogen. Damit soll sichergestellt werden, dass der Verarbeitungsaufwand im Verhältnis zu den gewonnen Informationen steht. Innerhalb der genannten fünf Jahre wurde der Index jährlich zwischen 1133 und 1300 mal genannt. Um die Frequenz der Nennung besser beurteilen zu können, wurde für den DAX die Abbildung 5.22 auf Ebene des Kalendermonats erstellt. Damit wird deutlich, dass auf Jahresebene die Anzahl der Beiträge relativ konstant bleibt, zwischen den Monaten jedoch einer starken Streuung unterliegt (78 bis 136). Pro Monat wird der Index in durchschnittlich 101,4 Beiträgen zitiert. Das entspricht täglich mehr als drei Beiträgen. Aufgrund seiner hochfrequenten und regelmäßigen Nennung ist ihm eine hohe Relevanz zuzuschreiben. In keinem der betrachteten Monate wird er seltener als in 78 Beiträgen genannt. Damit kann die Vermutung aufgestellt werden, dass die Nennung des DAX nicht auf außergewöhnliche Ereignisse zurückzuführen ist, sondern von einer allgemeinen Relevanz motiviert wird.

Die Länge der Beiträge ist mit 554,6 Wörtern durchschnittlich (vgl. Verteilung in Abschnitt 5.4.1.2.1.1), schwankt jedoch in einem großen Bereich zwischen 30 und 4133 Wörtern.

Tabelle 5.33: Numerische Kennzahlen zum DAX

Kennzahl	Min	Median	Max	Mittelwert	m. F.
Titelseite	0	0	100	2.8	16.6
Wörter/ Beitrag	30	471	4133	554.6	354.9
Anzahl Absätze	1	6	150	7.1	7.5
Wörter/ Titel	1	5	17	5	2.1
Anzahl Indexnennung/ Beitrag	1	1	38	2.3	2.7
Indexnennung/ Titel (Prozent)	-	-	-	8.5	-
Indexnennung/ 1. Absatz (Prozent)	-	-	-	41.6	-
Indexnennung/ Beitrag (Prozent)	-	-	-	46.9	-
Satzlänge	1.9	15.4	101	15.7	3.3
Aktualität					NaN

Tabelle 5.34: Ressort-Statistik zum DAX

Ressort	rel. Anzahl	Anzahl
Wirtschaft	85.4	5196
Geld Und Mehr	7.2	441
Verlagsbeilage	1.2	75
Politik	1.2	74
Feuilleton	0.8	46
Beruf Und Chance	0.7	41
Redaktionsbeilage	0.6	34
Titelseite Rhein-Main-Zeitung	0.4	23
Sport	0.3	21
Geld & Mehr Spezial	0.3	18
Immobilienmarkt	0.3	18
Gesamt	100	6086

Neben der Häufigkeit und Frequenz der Beiträge sprechen weitere empirische Ergebnisse für die Relevanz des DAX. So wurde er in 2,8% der Fälle auf der Titelseite erwähnt und in 8,5% der Beiträge erscheint sein Name bereits im Titel. Der letztgenannte Wert ist im Vergleich zu allen anderen betrachteten Indizes der höchste. Das Gleiche gilt für die Nennung des Index im ersten Absatz, was auf 41,6% der Beiträge zutrifft.

Eine größere Zahl von Beiträge zum DAX lassen sich Artikelserien zuordnen. Diese erscheinen regelmäßig und sind somit z.T. auch für die regelmäßige Nennung des DAX verantwortlich. Am häufigsten erscheint der DAX in der Serie *Europäische und amerikanische Börsen* (9,3%). Danach folgt mit 3,1% die Serie *Menschen & Wirtschaft* und mit 1,8% die Serie *Die Börse heute*. Auf die Serien *Was Den Markt bewegt* und *Börsenwoche* entfallen rund ein Prozent der Beiträge.

Die Genreformen der Beiträge entsprechen im Wesentlichen der durchschnittlichen Länge der Beiträge. Der überwiegende Teil wurde als Nachricht verfasst (40%) und jeweils rund 30% als *Bericht* sowie *Meldung/Kurznachricht* (30% und 27%).

Auffällig ist, dass lediglich in jedem zweiten Beitrag der DAX im zeitlichen (47%) oder sachlichen (3%) Vergleich betrachtet wird. In dem anderen Teil der Beiträge wird er überwiegend eingesetzt, um Unternehmen oder Personen (Manager, Vorstand) zu klassifizieren, z.B. indem sie mit einem Prädikat versehen werden (hier: Leiter eines Dax-Konzerns):

> „[...] Die Schweizer Gemeinschaftsbörse wurde ein Erfolg, und als die Frankfurter Banken Anfang der Neunziger die damals noch vielen Einzelbörsen zur Deutschen Börse AG zusammenlegten, holten sie Francioni als eines der ersten Vorstandsmitglieder an den Main. Zwanzig Jahre ist das nun her, seit acht Jahren leitet der Schweizer den Dax-Konzern. [...]“ (Kanning, 11. Mai 2013: S. 42)

Ein sachlicher Vergleich entsteht, wenn Indizes miteinander verglichen werden. Im folgenden Beispiel in Verbindung mit dem Unternehmen MAN:

> „Doch die Kommunikation mit Investoren und Analysten wird für ein Unternehmen obsolet, das vor einem halben Jahr noch Dax-Konzern war, heute zum Mitglied im M-Dax degradiert ist und womöglich in Bälde überhaupt nicht mehr gelistet sein könnte.“ (Köhn, 20. Februar 2013: S. 14)

In diesem Beispiel wird auch die Transparenz des Index deutlich. Er wird zwar namentlich genannt, jedoch werden keine weiteren Informationen zum Index geliefert. Nach diesem Schema wird in 80% der Beiträge Hintergrundwissen zum DAX vorausgesetzt, was als Indiz für seine Relevanz und Popularität angesehen werden kann. Die übrigen 20% liefern spezielle Informationen zum DAX, wie z.B. im Folgenden bezugnehmend auf den Börsengang von Osram:

> „Mehr als die Hälfte der Aktien will Siemens plazieren. Damit ist Osram ein Kandidat für den Leitindex. Für einen Schnelleinstieg in den Dax im Dezember wird es jedoch wohl zu früh sein, da die Hürden für die Aufnahme hoch sind. Realistischer ist die Aufnahme im September 2012 mit der jährlichen Überprüfung des Index.“ (Köhn, 10. Juni 2011: S. 12)

Die über das Allgemeinwissen hinausgehende Information bezieht sich hierbei auf den Termin der jährlichen Überprüfung des Index. Dort wird sichergestellt, dass die aktuell im Index befindlichen Unternehmen die Kriterien noch erfüllen. Andernfalls werden sie durch andere ausgetauscht. Darüber hinaus existieren auch deutlich strengere Kriterien für einen Schnelleinstieg (*Fast-Entry*-Regel).[72]

Die Beiträge mit Bezug auf den DAX besitzen nur sehr selten eine negative Akzentuierung (17%). In den meisten Fällen beschränken sie sich auf die objektive Darstellung von Wirtschaftsdaten. Nur selten werden Sorgen oder Probleme aufgegriffen, wie z.B. Konjunktursorgen oder ein zu niedriger oder zu hoher Ölpreis.

In den Beiträgen gibt es im Wesentlichen keine (63%) oder nur eine geringe (37%) Überraschung. Diese bezieht sich hauptsächlich auf nicht erwartete Änderungen im Indexwert.

Zum überwiegenden Teil (63%) wird der DAX in den Beiträgen nicht argumentativ eingesetzt. Dabei handelt es sich in der Regel um die bloße Nennung der Werte oder des Index, beispielsweise über *DAX-Unternehmen* oder *DAX-Vorstand*.

5.7.9.2 Der DAX im Kontext der Kommunikation

Der DAX gehört zu den meistzitierten Indizes der FAZ und FAS. Sein Anspruch ist die Operationalisierung eines ganzen Marktes. Er reduziert Komplexität, indem er ein mehrdimensionales Konstrukt in einer Maßzahl zusammenfasst. Insofern kann er als Beispiel für die steigende Anforderung an Messapparaturen gesehen werden.

Im Vergleich zu anderen untersuchten Indizes (z.B. GfK-Konsumklimaindex, MCSI, HDI, HPI) basiert der DAX auf rein quantitativen Indikatoren. Das bedeutet, dass diese selbst bereits eine quantitative Struktur besitzen. Diese wird weitestgehend über das Prinzip von Angebot und Nachfrage innerhalb des Börsenhandels festgelegt. Das Grundproblem des DAX beschränkt sich damit im Wesentlichen auf die Selektion der Indikatoren sowie deren Verdichtung, nicht aber deren Quantifizierung.

[72] Für detailierte Informationen zu den Kriterien sei auf Deutsche Börse AG (2015: S. 23f.) verwiesen.

Tabelle 5.35: Ergebnisse der manuellen Inhaltsanalyse zum DAX (I)

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
09.03.2010	Das Beste aus zwei Brsenwelten	FAZ, 09.03.2010, Nr. 57, S. 11	Bericht	Nein	Nein	Nein	Nein	Gering	Ja
06.03.2010	Preisgekrnter Biotech-Dienstleister	FAZ, 06.03.2010, Nr. 55, S. 17	Bericht	Nein	Nein	Nein	Nein	Gering	Nein
20.02.2010	Amerikanische Verbraucherpreise sttzen die Aktienmrkte	FAZ, 20.02.2010, Nr. 43, S. 22	Nachricht	Nein	Nein	Nein	Zeitlich	Moderat	Ja
16.02.2010	Dax-Konzerne erreichen zehn Millionen Menschen in sozialen Medien	FAZ, 16.02.2010, Nr. 39, S. 15	Bericht	Nein	Gering	Nein	Nein	Moderat	Nein
11.05.2013	Reto Francioni	FAZ, 11.05.2013, Nr. 108, S. 42	Sonstiges	Nein	Nein	Nein	Nein	Gering	Nein
02.05.2013	Am Aktienmarkt droht in diesem Mai keine Korrektur	FAZ, 02.05.2013, Nr. 101, S. 19	Bericht	Nein	Nein	Nein	Zeitlich	Gering	Ja
21.03.2013	Lanxess und gefallene Engel	FAZ, 21.03.2013, Nr. 68, S. 17	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
20.02.2013	MAN-Finanzvorstand Lutz flieht vor dem VW-Vertrag	FAZ, 20.02.2013, Nr. 43, S. 14	Nachricht	Ja	Gering	Nein	Sachlich	Gering	Nein
17.05.2014	Es htte schlimmer kommen knnen	FAZ, 17.05.2014, Nr. 114, S. 29	Nachricht	Nein	Gering	Nein	Zeitlich	Gering	Ja
11.04.2014	Fed-Protokolle gnnen den Investoren nur kurz Ruhe	FAZ, 11.04.2014, Nr. 86, S. 28	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich	Gering	Nein

Tabelle 5.36: Ergebnisse der manuellen Inhaltsanalyse zum DAX (II)

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
07.12.2012	Die Frauen rcken auf	FAZ, 07.12.2012, Nr. 286, S. 47	Nachricht	Nein	Nein	Nein	Nein	Gering	Nein
07.05.2012	Treibende Kraft	FAZ, 07.05.2012, Nr. 106, S. 38	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
03.08.2011	Dax fllt unter die Marke von 6800 Punkten	FAZ, 03.08.2011, Nr. 178, S. 20	Nachricht	Ja	Gering	Nein	Zeitlich	Moderat	Nein
02.08.2011	Zweifel begleiten Metro	FAZ, 02.08.2011, Nr. 177, S. 19	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Ja
13.07.2011	Verhaltene Beruhigung an den Mrkten	FAZ, 13.07.2011, Nr. 160, S. 17	Nachricht	Nein	Gering	Nein	Zeitlich	Gering	Nein
10.06.2011	Siemens setzt auf den Brsengang von Osram	FAZ, 10.06.2011, Nr. 134, S. 12	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Moderat	Ja
10.05.2011	Die Hrden einer Fusion	FAZ, 10.05.2011, Nr. 108, S. 12	Nachricht	Nein	Nein	Nein	Nein	Gering	Nein
21.03.2011	Politische Unsicherheit befrdert Volatilitt	FAZ, 21.03.2011, Nr. 67, S. 20	Nachricht	Nein	Nein	Nein	Zeitlich	Gering	Nein
11.03.2011	Auf dem Weg zu neuen Ufern	FAZ, 11.03.2011, Nr. 59, S. 21	Bericht	Nein	Nein	Nein	Nein	Gering	Ja
16.01.2011	Grafik der Woche	FAS, 16.01.2011, Nr. 2, S. 38	Meldung/ Kurznachricht	Nein	Gering	Nein	Zeitlich	Gering	Ja

Tabelle 5.37: Ergebnisse der manuellen Inhaltsanalyse zum DAX (III)

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
08.12.2010	Der Dax klettert auf mehr als 7000 Punkte	FAZ, 08.12.2010, Nr. 286, S. 19	Bericht	Nein	Gering	Nein	Zeitlich	Moderat	Ja
25.09.2012	Schaeffler-Banken verkaufen Aktien von Continental	FAZ, 25.09.2012, Nr. 224, S. 17	Meldung/ Kurznachricht	Nein	Nein	Nein	Nein	Gering	Nein
01.02.2013	Bayern Mnchen auf dem Weg zur Weltmarke	FAZ, 01.02.2013, Nr. 27, S. 30	Bericht	Nein	Nein	Nein	Nein	Moderat	Nein
27.07.2014	Warum Top-Manager scheitern	FAS, 27.07.2014, Nr. 30, S. 15	Bericht	Ja	Gering	Nein	Nein	Gering	Nein
01.09.2013	Syrien	FAS, 01.09.2013, Nr. 35, S. 29	Nachricht	Ja	Nein	Nein	Zeitlich	Gering	Nein
08.07.2011	Finanzprodukte fr jedermann	FAZ, 08.07.2011, Nr. 156, S. 11	Bericht	Nein	Nein	Nein	Nein	Gering	Nein
11.08.2011	Abermals hohe Kursverluste an Aktienmrkten	FAZ, 11.08.2011, Nr. 185, S. 9	Nachricht	Nein	Gering	Nein	Zeitlich	Gering	Nein
18.12.2010	Kursverluste an Europas Aktienbrsen	FAZ, 18.12.2010, Nr. 295, S. 24	Meldung/ Kurznachricht	Nein	Nein	Nein	Zeitlich	Gering	Nein
04.10.2012	Dem Versicherer Talanx gelingt der Brsengang	FAZ, 04.10.2012, Nr. 231, S. 19	Nachricht	Nein	Gering	Nein	Zeitlich	Gering	Ja
28.04.2012	Daimler zeigt Rendite-Delle	FAZ, 28.04.2012, Nr. 100, S. 14	Nachricht	Ja	Gering	Nein	Zeitlich	Gering	Ja

Die Indikatoren des DAX sind somit als gegeben anzusehen. Für die Konstruktion des DAX ist es z.B. nebensächlich, ob die Unternehmen über den Aktienkurs realistisch bewertet sind. Schließlich stellt sich bei den Aktienkursen ohnehin die Frage, was sie eigentlich abbilden. Das Ziel des DAX ist es lediglich, die Kursentwicklung der enthaltenen Unternehmen zusammenzufassen und damit die Komplexität des mehrdimensionalen Konstruktes zu reduzieren.

Die hohe Frequenz der Nennung des DAX in FAZ und FAS, die breite Streuung über verschiedene Ressorts sowie das vorausgesetzte Hintergrundwissen zum Index unterstreichen die allgemeine Relevanz des Index. Er scheint kein Domänen-Wissen darzustellen, sondern wird in diversen Bereichen aufgegriffen. Innerhalb der Beiträge übernimmt der Index zwei Funktionen. Eine basiert auf dem Konzept des Index, die andere auf seinen Werten. Werden z.B. Aussagen über Unternehmen oder Personen mit einem Prädikat (z.B. DAX-Vorstand) versehen, so besteht inhaltlich der Bezug auf das Konzept mit seinen Selektionskriterien. Es lassen sich z.B. Aussagen zur Unternehmensgröße kurz und prägnant kommunizieren. Die Glaubwürdigkeit des DAX wird in diesem Zusammenhang von dem Vertrauen in die Objektivität bei der Selektion der Unternehmenstitel in den Index abhängen. Unabhängig davon sollte bereits die enorme Reputation des Index seine Glaubwürdigkeit und die Zurechnung von Objektivität erhöhen.

Bei der Kommunikation der Indexwerte lassen sich Parallelen zu den in Abschnitt 3.3 genannten Eigenschaften von auf Messung basierender Zahlen sehen. Denn der Begriff DAX ist über das theoretische Konzept hinreichend definiert. Aussagen werden i.d.R. ausschließlich über dieses Konzept getroffen. Das Interesse bezieht sich also im Wesentlichen auf seine quantifizierte Form. Anders als z.B. bei manchen sozialwissenschaftlichen Indizes, bei denen das Interesse neben dem wissenschaftlich stilisierten „Phänomen“ auch an seiner lebensweltlichen Erscheinung erhalten bleibt. Die Glaubwürdigkeit und Zurechnung von Objektivität der DAX-Werte erfordert in diesem Fall doppeltes Vertrauen: (1) in die Zuverlässigkeit der Verfahren, mit denen die Indikatorwerte gewonnen werden (Aktienkurse etc.) und (2) in die Zuverlässigkeit der verwendeten mathematischen Verfahren sowie deren korrekte Anwendung zur Berechnung des Index.

Obwohl es Kritik am Konzept des DAX als Performance-Index gibt, ist nicht absehbar, dass seine Relevanz in Zukunft abnehmen wird. Innerhalb der letzten fünf Jahre ist zumindest in FAZ und FAS nichts Gegenteiliges festzustellen. Er bleibt auf einem konstant

hohen Niveau bezogen auf die Anzahl der Beiträge.

Die Anforderungen an den DAX haben sich mit der schneller werdenden Welt geändert. Er wird mittlerweile im Sekundentakt berechnet und kann auf diversen Webseiten[73] in Echtzeit verfolgt werden. Die hohen Anforderung an die Aktualität der DAX-Werte hat nicht zu einer geringeren Relevanz des Index im vergleichsweise langsamen Printmedium FAZ und FAS geführt (keine untertägige Veröffentlichung möglich). Mit der Nennung von Schlusskursen und der Beschreibung des Tagesverlaufs haben die Nachrichten hier eher den Charakter einer Selektion, indem eine Vielzahl von Werten (bis zu 30.600 pro Tag) zu einer Aussage zusammengefasst wird.

5.7.10 Korruptionswahrnehmungsindex

Der *Korruptionswahrnehmungsindex* (CPI)[74] operationalisiert die in einer Nation wahrgenommene Korruption im öffentlichen Sektor. Er wurde an der Universität Passau von dem Wirtschaftstheoretiker Johann Graf Lambsdorff entwickelt und seit 1995 von *Transparency International* (TI) veröffentlicht. (Lambsdorff, 2007)

Der CPI soll für möglichst viele Nationen berechnet werden. Dafür wird sich ausschließlich öffentlich zugänglichen Datenquellen und Umfragen bedient und keine eigene Datenerhebung durchgeführt. Viele der zugänglichen Datenquellen sind nicht für alle Nationen verfügbar. Deswegen wurde das Konzept des Index über eine variable Anzahl von Indikatoren definiert. Die insgesamt zwölf Indikatoren (für den CPI 2014) entstammen jeweils einer anderen Datenquelle (TI, 2014).[75] Ist ein Indikator für eine Nation verfügbar, wird er für die Berechnung des CPI einbezogen. Die minimale Anzahl an Indikatoren für die Aufnahme in den CPI wurde statistisch fundiert auf drei festgelegt.[76]

Die verschiedenen Datenquellen basieren weitestgehend auf der Befragung und Auswertung von Experten, Geschäftspersonen oder Risikoagenturen (Lambsdorff, 2007: S. 236ff.). Die Qualität der Indikatoren ist damit nur sehr eingeschränkt kontrollierbar, weil es sich nicht um Eigenerhebungen handelt. Hinzu kommt, dass die Verbindung der Indikatoren zum theoretischen Konzept von der Verfügbarkeit der Daten abhängt. So

[73] Als ein Beispiel sei `http://www.wallstreet-online.de` genannt.

[74] Alternative Schreibweisen sind: Corruption Perceptions Index (CPI), Korruptionswahrnehmungsindex, Korruptionsindex.

[75] Die zwöf verschiedenen Datenquellen können unter `https://www.transparency.de/Verwendete-Quellen.2573.0.html` (besucht am 7. Jun. 2015) eingesehen werden.

[76] Eine Untersuchung von Saisana und Saltelli (2012: S. 13ff.) zeigte keine Verzerrungen (Bias).

kann die Anzahl der Indikatoren des theoretischen Konzepts je Land zwischen 3 und 12 schwanken und die Zusammensetzung an Indikatoren auf Basis der Kombinatorik theoretisch über alle Länder verschieden sein.

Die Verbindung der beobachtbaren Indikatoren zum theoretischen Konzept wird damit jedoch auf die Datenquellen mit ihrer jeweiligen Methodologie externalisiert.

Die Indikatoren messen statistisch betrachtet alle in gleichem Maß die wahrgenommene Korruption einer Nation.[77] Die Operationalisierung des Konzepts *Wahrgenommene Korruption* über mehrere Indikatoren gewähre einerseits eine höhere Zuverlässigkeit der „Messung“ und ermögliche andererseits die Einbeziehung von mehr Nationen (Saisana und Saltelli, 2012).

Vor der Zusammenfassung der Indikatoren wird eine Standardisierung (vgl. Abschnitt 4.3.5.1) mit Skalierung auf den Mittelwert 45 und eine Standardabweichung von 20 durchgeführt.[78] Alle Werte außerhalb des Wertebereiches [0,100] werden abgeschnitten (Lambsdorff, 2007: S. 10).

$$\text{Indik}_\text{n} = \frac{X_\text{n} - \bar{x}}{s} * sign * 20 + 45 \tag{5.23}$$

Mit der Variablen *sign* wird gesteuert, welche Richtung die jeweilige Skala hat, d.h. ob geringe Werte für eine geringe wahrgenommene Korruption stehen oder umgedreht. Im ersten Fall ist die Variable 1, im zweiten Fall -1.

Nach der Normalisierung werden die Indikatoren über eine gleich gewichtete Addition zusammengefasst.

$$CPI = \sum_{3}^{n} \frac{\text{Indik}_\text{n}}{\text{n}} \tag{5.24}$$

Die gleiche Gewichtung aller Indikatoren wurde über die Durchführung einer PCA (vgl. 4.3.5.3.1) validiert. Die Anzahl der Indikatoren schwankt über die verschiedenen Nationen, je nach Verfügbarkeit der Daten. Der Wertebereich des Index liegt zwischen null und hundert.

[77] Zwischen allen Indikatoren wurde eine hohe Korrelation festgestellt (vgl. weiterführend Saisana und Saltelli (2012: S. 10ff.)).

[78] Für die Bestimmung der Standardisierungskoeffizienten werden fehlende Werte imputiert. Genaueres ist Saisana und Saltelli (ebd.: S. 10) zu entnehmen.

Die oben beschriebene Berechnung des CPI basiert auf einer Revision in 2012. Davor folgte der Index einer anderen Berechnung. Je Datenquelle wurden Perzentile gebildet und die Länder entsprechend ihrer Werten zugeordnet. Die CPI Werte vor der Revision sind damit nicht mit den neueren vergleichbar. Im Zuge der Revision des Index informierte Lambsdorff im August 2009 TI über seinen Rückzug aus dem Projekt (Turner, 2009).

Mit dem Rückzug von Lambsdorff als Mit-Herausgeber des CPI wird auch die Qualität des Index in Bezug auf Tabelle 4.15 auf Seite 156 beeinflusst. Der theoretische Rahmen des Index wird in Lambsdorff (2007) detailliert beschrieben. Problematisch in diesem Zusammenhang ist, dass das Konzept seit der Revision zur Diskussion steht. Lambsdorff soll dazu in einer E-Mail an Transparency International folgendes geschrieben haben:

> „TI-S will try to continue somehow with a substitute for the CPI. Even though most of them are rather new to the debate, they will try to make the new product look like the old one. This is time for me to let them go their way.“ (Turner, 2009)

Der Bruch zwischen Erfinder und Herausgeber wirkt sich auf die Glaubwürdigkeit aus, da sie innere Unstimmigkeiten vermuten lässt. Auch die Revision im Allgemeinen ist ungünstig, da sie frühere Berechnungen in Frage stellt. Zumal keine Transformationen angeboten werden, um ältere mit neueren Werten vergleichbar zu machen. Die Revision ist in dieser Hinsicht zu grundlegend.

Negativ auf die Qualität des Index wirkt sich ebenfalls aus, dass die Selektion der Daten nicht an dem theoretischen Konzept motiviert wird, sondern an der Verfügbarkeit adäquater Daten. Ursache sind sicherlich die umfangreichen Kosten einer Umfrage. Da die PCA zur Zusammenfassung der Indikatoren verwendet wird, entstehen u.U. umfangreiche Einschränkungen durch die variable Zahl an Indikatoren (z.B. mehrere Modelle für ein Konzept).

Außerdem handelt es sich bei den verwendeten Daten z.T. um Befragungen. Die Daten sind daher ggf. nur eingeschränkt reliabel und reproduzierbar. Die Datenquellen selbst werden transparent genannt und lassen sich komplett nachvollziehen.

Die Normalsierungs- und Aggregationsmethoden werden umfangreich und lückenlos diskutiert und untermauert. Außerdem wird über verschiedene (statistische) Tests die Validität der Ergebnisse und Methoden geprüft und dokumentiert. Dabei wird sich eng an Nardo u. a. (2008) orientiert.

Die Verbreitung der Daten findet über die *Webseite*[79] und über eine Broschüre statt. Auf der Webseite werden Daten und Diagramme zur Verfügung gestellt. Die Visualisierungsformen konzentrieren sich auf Choroplethenkarten, eingefärbte Tabellen und Infografiken. Sie sind u.a. mit sozialen Medien wie Facebook oder Twitter verbunden.

5.7.10.1 Inhaltsanalyse zum CPI

Der Korruptionswahrnehmungsindex wird seit 1995 jährlich berechnet und veröffentlicht. Die zugehörigen Pressemitteilungen befinden sich auf der Webseite von TI [80]. Da vor dem Jahr 2000 vereinzelte Pressmitteilungen fehlen, wird sich bei der Bestimmung der Aktualität auf den Zeitraum 2000 bis 2014 beschränkt und die jeweiligen Veröffentlichungstermine der Pressemitteilungen werden zugrunde gelegt.

In Abbildung 5.23 wird deutlich, dass der größte Anstieg in den ersten drei Tagen stattfindet. Gemäß der Definition in Abschnitt 5.4.1.2.9 werden 11% der Beiträge auf die Veröffentlichung durch TI zurückgeführt. Absolut ist das ca. ein Beitrag pro Jahr (0.7). Danach verteilen sich die Beiträge weitestgehend gleichmäßig auf den Zeitraum zwischen zwei Indexveröffentlichungen. Einige Beiträge am rechten Rand der Abbildung sind inhaltlich auf die Ankündigung der Veröffentlichung des Index zurückzuführen. Dazu gehören z.B. Beiträge der Serie *Termine der Woche*. Da in diesen Beiträgen die Veröffentlichung teilweise mehrere Tage im Voraus angekündigt wird, werden sie zur vorherigen Veröffentlichung gezählt.

Dem CPI kann auf Basis von Abbildung 5.23 eine hohe Aktualität unterstellt werden. Die Abbildung legt nahe, dass sofort auf Veröffentlichungen des Index reagiert wird oder diese angekündigt werden (starker Anstieg der Kurve in den ersten drei Tagen). Danach behält der Index eine nahezu gleichbleibende Relevanz bis zur Folgeveröffentlichung aufrecht. Diese wird über den kontinuierlichen Anstieg des Graphen zwischen Tag 3 und 400 deutlich. Im Mittel wird der Index in diesem Zeitraum in 5,9 Beiträgen aufgegriffen.

[79] Siehe `http://www.transparency.org/cpi2014` (besucht am 7. Jun. 2015).
[80] Siehe `http://www.transparency.org/research/cpi` (besucht am 7. Jun. 2015).

Abbildung 5.23: Aktualität der Beiträge zum CPI

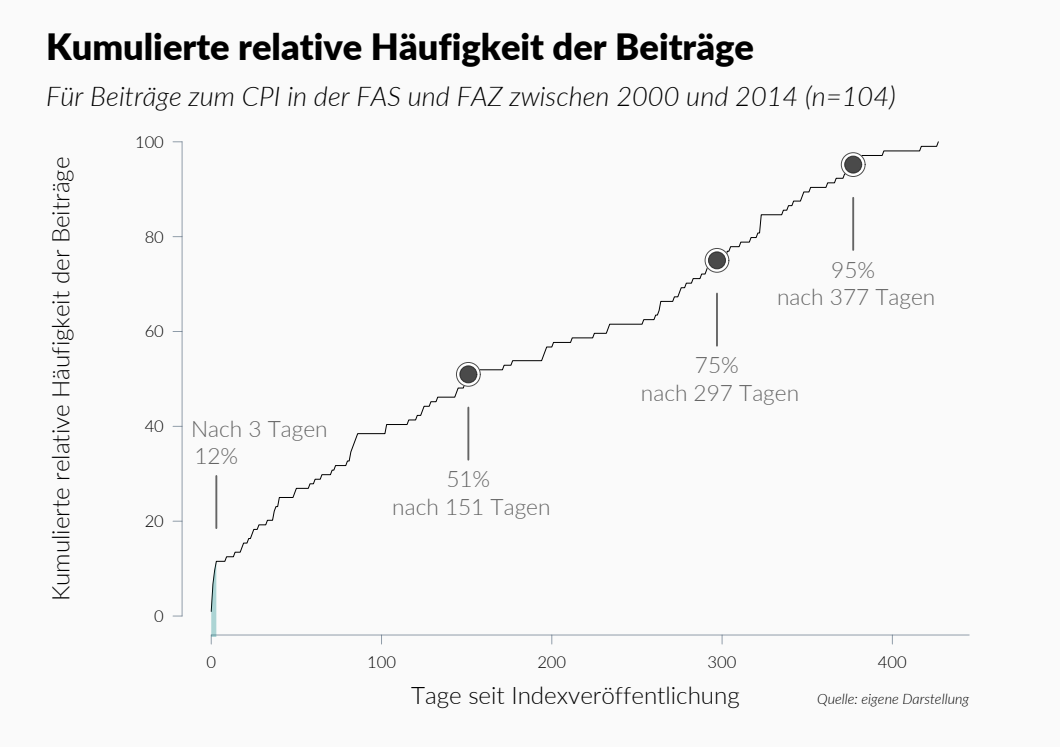


Abbildung 5.24: Anzahl Beiträge zum CPI

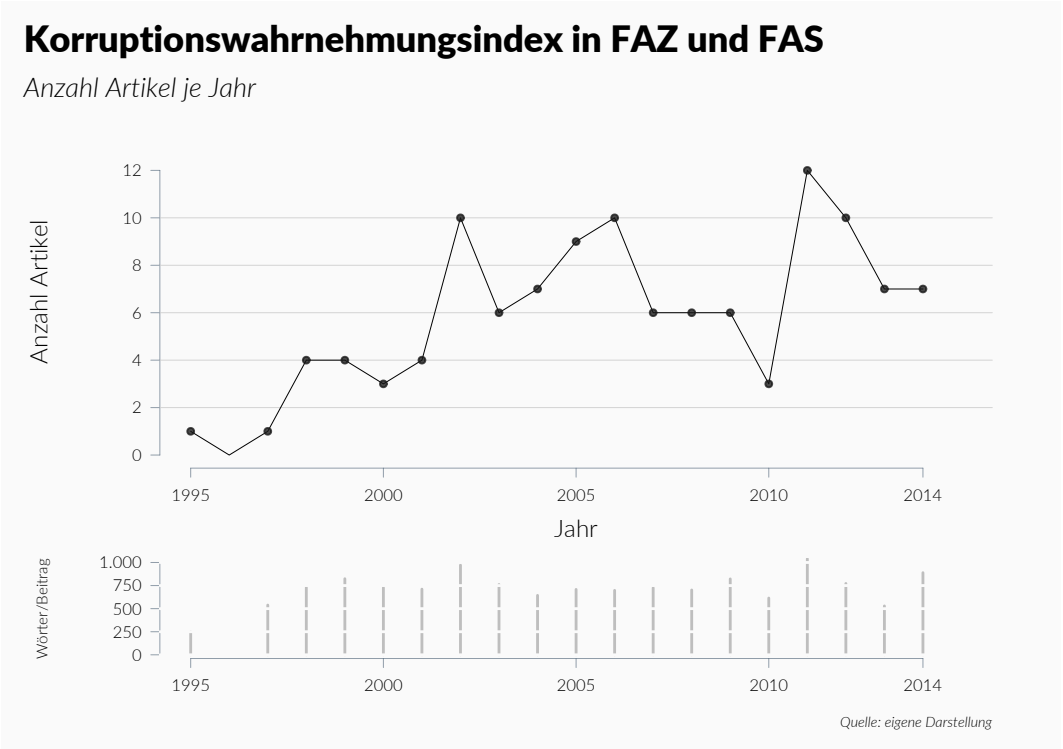


Tabelle 5.38: Numerische Kennzahlen zum CPI

Kennzahl	Min	Median	Max	Mittelwert	m. F.
Titelseite	0	0	100	2.6	15.9
Wörter/ Beitrag	74	749	1990	773.3	362.6
Anzahl Absätze	1	9	123	21.1	31.5
Wörter/ Titel	1	4	11	4.4	2
Anzahl Indexnennung/ Beitrag	1	1	6	1.1	0.5
Indexnennung/ Titel (Prozent)	-	-	-	2.6	-
Indexnennung/ 1. Absatz (Prozent)	-	-	-	8.6	-
Indexnennung/ Beitrag (Prozent)	-	-	-	52	-
Satzlänge	4.1	16.8	29.8	15.9	5
Aktualität	0	151	427	175.3	132.9

Die meisten Beiträge erscheinen im Ressort *Wirtschaft* (53,2%) und *Politik* (33,9%). Die restlichen Beiträge verteilen sich im Wesentlichen auf *Feuilleton* (2,8%), *Sport* (2,8%) und *Gesellschaft*, *Meinung* sowie *Verlagsbeilage* mit zusammen 5,4%.

Knapp jeder fünfte Beitrag (19%) im Ressort *Wirtschaft* entfällt auf die Serie *Termine der Woche*. Insgesamt macht diese Serie 10,1% der Beiträge aus. Der Inhalt besteht dabei im Wesentlichen in der Ankündigung von Veröffentlichungen.

Der erste Beitrag mit Bezug auf den Index wurde 1995 veröffentlicht. In diesem Jahr wurde der Index auch erstmals berechnet und herausgegeben. Bis auf 1996 wurde der Index seitdem mindestens einmal pro Jahr genannt, seit 2000 in der Regel deutlich öfter (vgl. Abbildung 5.24), mindestens jedoch dreimal. Die maximale Anzahl an Beiträgen pro Jahr wird 2012 erreicht. Inhaltlich ist dieses Hoch auf die Griechenland-Krise zurückzuführen, bei der u.a. auch die Korruption thematisiert wurde.

Die Länge der Beiträge schwankt zwischen 74 und 1990 Wörtern. Die durchschnittliche Länge der Beiträge liegt mit 771 Wörtern im oberen mittleren Bereich, verglichen mit den anderen Indizes (vgl. Verteilung in Abschnitt 5.4.1.2.1.1).

Die Relevanz des Index zeigt sich darin, dass 2,8% der Beiträge auf der Titelseite veröffentlicht wurden, sowie ebenfalls 2,8% den Namen des Index im Titel nennen. Im ersten Absatz wurde er in 9,2% der Beiträge genannt.

Die Bezeichnung des Index variiert über die Beiträge. Der überwiegende Teil benutzt unmissverständlichen Bezeichnungen, wie z.B. *Korruptionswahrnehmungsindex*, *Korruptionsindex* oder *Corruption Perceptions Index*. In anderen Beiträgen wird die Bezeichnung z.T. nur von dem Konzept *Index* abgegrenzt. In diesem Zusammenhang belegte Schreibweisen sind *Korruptionsskala*, *Korruptions-Rangliste*, *Korruptions-Ranking*,

Tabelle 5.39: Ressort-Statistik zum CPI

Ressort	rel. Anzahl	Anzahl
Wirtschaft	54.3	63
Politik	31.9	37
Feuilleton	3.4	4
Sport	2.6	3
Gesellschaft	1.7	2
Meinung	1.7	2
Verlagsbeilage	1.7	2
Geld Und Mehr	0.9	1
Immobilienmarkt	0.9	1
Rhein-Main · Darmstadt · Südhessen · Mainz · Offenbach	0.9	1
Ohne Angabe	0	0
Gesamt	100	116

Welt-Korruptionsbarometer. Jedoch wurden auch missverständliche Bezeichnung identifiziert, wie z.B. *Transparency-Index*. Denn neben dem CPI gibt Transparency International einen weiteren Index unter der Bezeichnung *Bribe Payer Index* heraus.[81]

Der überwiegende Teil der Beiträge wird in der Genreform *Bericht* verfasst (80%). Die restlichen 20% erscheinen als *Nachricht*. Dem Thema Korruption ist dabei sicherlich zu schulden, dass 90% eine negative Akzentuierung besitzen.

Die Nennung des Index beschränkt sich meist auf den Namen und den Herausgeber TI (80%). Nur eher selten (20%) wird detailliert auf das Konzept des Index eingegangen. Insgesamt besitzen die Beiträge damit überwiegend eine geringe Transparenz und eher selten eine hohe.

Der Index wird in allen Beiträgen für einen nationalen Vergleich eingesetzt (100%). Oftmals folgen die Aussagen dabei dem „Schema Indexname – Herausgeber – Rangplatz“:

> „Auf dem Korruptionsindex von Transparency International steht China gemeinsam mit Indien auf Platz 72 von 178 Ländern.“ (Hein, 13. Oktober 2007: S. 14)

In 10% der Beiträge wird der Index im zeitlichen Vergleich dargestellt. Bei den Aussagen wird sich dabei in der Regel nur auf eine Rangliste bezogen, in der Nationen gemäß ihrer Indexwerte eingereiht sind. Im Beitrag wird dann z.B. der Rangplatz eines Landes zwischen zwei verschiedenen Wellen verglichen.

[81] Dieser operationalisiert die Korruption in der Unternehmenswelt.

Tabelle 5.40: Ergebnisse der manuellen Inhaltsanalyse zum CPI

Datum	Titel	Quelle	Genre	Konflikt	Überraschung	Quellenangabe	Vergleich	Transparenz	Argumentation
28.10.2002	Noch ist die türkische Wirtschaft nicht fit für die Europäische Union	FAZ, 28.10.2002, Nr. 250, S. 18	Bericht	Ja	Nein	Nein	National	Gering	Ja
04.08.1998	Damit die Ehrlichen nicht die Dummen bleiben	FAZ, 04.08.1998, Nr. 178, S. 10	Bericht	Ja	Nein	Nein	National	Gering	Ja
23.09.1998	Korruptionsindex in Berlin vorgestellt	FAZ, 23.09.1998, Nr. 221, S. 7	Nachricht	Ja	Gering	Nein	National	Hoch	Ja
19.02.2011	Das Geld der Diktatoren	FAZ, 19.02.2011, Nr. 42, S. 11	Bericht	Ja	Gering	Nein	National	Gering	Ja
14.12.2012	Eine Gesellschaft stürzt ins Bodenlose	FAZ, 14.12.2012, Nr. 292, S. 31	Bericht	Ja	Gering	Nein	National	Gering	Ja
17.03.2002	Der Kölner Müll stinkt überall	FAS, 17.03.2002, Nr. 11, S. 37	Bericht	Ja	Gering	Nein	National, Zeitlich	Hoch	Ja
04.12.2013	Korruption bremst Thailands Wirtschaft	FAZ, 04.12.2013, Nr. 282, S. 10	Bericht	Ja	Gering	Nein	National	Gering	Ja
13.10.2007	Das große Kommunisten-Treffen	FAZ, 13.10.2007, Nr. 238, S. 14	Bericht	Nein	Gering	Nein	National	Gering	Ja
29.12.2011	Aquino macht Ernst im Kampf gegen Korruption auf den Philippinen	FAZ, 29.12.2011, Nr. 303, S. 17	Nachricht	Ja	Gering	Nein	National	Gering	Ja
15.02.2003	Ein wunderbarer neuer Stil	FAZ, 15.02.2003, Nr. 39, S. 5	Bericht	Ja	Gering	Nein	National	Gering	Ja

Solche Aussagen sind leicht zu deuten. Insbesondere, da der „rohe“ Indexwert ohne einen Vergleich schwer interpretierbar ist. Nachteil ist jedoch, dass die Informationen zu den Abständen zwischen den Ländern verloren gehen (vgl. Abschnitt 2.2.4.2). Außerdem ist der Vergleich zweier Zeitpunkte nur mit Einschränkungen möglich. Beispielsweise wird eine zunehmende Korruption in der Rangliste nicht auffällig, wenn sie in allen Nationen gleichermaßen zunimmt. Ebenfalls denkbar ist die Rang-Verbesserung einer Nation, durch das Verschlechtern einer anderen. Völlig unabhängig davon, wie sich die betreffende Nation bzgl. der Korruption entwickelt hat.

Der nationale Vergleich der wahrgenommenen Korruption wurde in allen untersuchten Beiträgen argumentativ eingesetzt. Eine Überraschung wiesen die Beiträge in Wesentlichen (80%) nur in geringem Maße aus. Die restlichen Beiträge enthielten keine Überraschung.

5.7.10.2 Der CPI im Kontext der Kommunikation

Für die Quantifizierung der wahrgenommenen Korruption im öffentlichen Sektor bedarf es eines theoretischen Konzepts. Weder die fundamentale noch die abgeleitete Messung kommen für die numerische Bestimmung dieses mehrdimensionalen Konstrukts in Frage. Erschwerend für die Quantifizierung von Korruption ist, dass sie sich meist nur im Hintergrund abspielt. In diesem Zusammenhang erscheint es verständlich, dass sich das Konzept des CPI auf die *wahrgenommene* Korruption bezieht. Auf diese Weise wird sich klar von der „realen“ Korruption distanziert. Das verringert vermutlich das Misstrauen in die resultierenden Indexzahl.

Korruption ist einer jener qualitativen Begriffe, die in der Evolution der Messung zur indirekten oder indikatorbasierten Messung überleiten. Die Überführung des qualitativen in einen quantitativen Begriff wird im Fall des CPI mit Hilfe von Experten, Geschäftspersonen und Risikoagenturen erreicht. Die Verdichtung der verschiedenen Facetten und Dimensionen von Korruption wird durch sie externalisiert. Das bedeutet jedoch auch, dass die Einschätzung des Grades von Korruption nicht komplett nachvollziehbar sein wird. Das Zustandekommen der Einschätzung ist individuell und nur eingeschränkt reliabel und reproduzierbar. Das theoretische Konzept umgeht mit der Experteneinschätzung die Notwendigkeit, Dimensionen von Korruption benennen zu müssen. Gleichzeitig bedeutet das aber auch, dass der Begriff Korruption unscharf bleibt und

nicht objektiv klar ist, was eigentlich quantifiziert wird. Das hat Folgen für die Objektivität und Glaubwürdigkeit, da die Ergebnisse nicht komplett reproduzierbar sind.

Im Vergleich zu den anderen Indizes innerhalb dieses Kapitels werden im Verlauf der Indexkonstruktion nicht verschiedene Dimensionen zusammengefasst. Es geht vielmehr darum, aus der Vielzahl der Datenquellen und Indikatoren den gemeinsamen latenten Faktor *wahrgenommene Korruption* zu extrahieren. Die Verbindung zwischen den beobachtbaren Indikatoren und dem theoretischen Konzept wird über statistische Methoden aus den Daten selbst abgeleitet und nicht theoretisch. Dieses Vorgehen ist für den Laien schwierig nachzuvollziehen. Da das Verfahren der PCA jedoch wissenschaftlich fundiert ist, sollte die Zurechnung von Objektivität in diesem Zusammenhang positiv gewertet werden, jedoch mit den o.g. Einschränkung hinsichtlich der Objektivität der Indikatoren.

In der Inhaltsanalyse wurde deutlich, dass der Index in zweierlei Hinsicht Relevanz besitzt. Einerseits verfügt der Index über einen Neuigkeitswert, der sich über die sofortige Resonanz in der FAZ und FAS äußert. Andererseits bietet der Index die Möglichkeit, den komplexen Begriff Korruption zu kommunizieren, z.B. indem zwei Nationen einfach hinsichtlich des Grades an Korruption miteinander verglichen werden können. In folgendem Beitrag wird die Veränderung des Indexwertes z.B. direkt in Verbindung zur Verflechtung von Politik und Wirtschaft gebracht.

> „Verflechtung, Filz und Korruption - das muß nicht die logische Folge von zuviel Nähe zwischen Wirtschaft und Politik sein. Doch es ist in Deutschland immer häufiger der Fall. So ist die Bundesrepublik auf dem Korruptionsindex der unabhängigen Organisation Transparency International vom Sommer des vergangenen Jahres von ehemals Rang 14 auf Platz 20 abgerutscht.“ (Klöpfer, 17. März 2002: S. 37)

Die Begründung mittels der Verschlechterung des Rangplatzes ist jedoch nur bedingt zutreffend, da grundsätzlich auch die Verbesserung anderer Nationen dafür verantwortlich sein kann.

Die Zurechnung von Objektivität zum CPI erfordert dreifaches Vertrauen: (1) in die Zuverlässigkeit der Messverfahren, über die die Rohdaten gewonnen wurden, (2) in die Zuverlässigkeit der Operationalisierung des theoretischen Konzepts und (3) in die Zuverlässigkeit der verwendeten statistischen Verfahren sowie deren korrekte Anwendung.

Bei der wissenschaftlichen Betrachtung der Messverfahren, der Operationalisierung und der verwendeten statistischen Verfahren treten jedoch Schwachstellen hervor. Denn die Gewinnung der Rohdaten kann mit den diversen Befragungen nur eingeschränkt nachvollzogen werden (Auswahl der Experten etc.). Die Operationalisierung des theoretischen Konzepts und des Begriffs Korruption kann ebenfalls nicht nachvollzogen werden, da die Operationalisierung von den Befragten individuell vorgenommen wird und kein Modell zugrunde liegt. Nur der dritte Punkt lässt sich wissenschaftlich umfassend nachvollziehen.

Der CPI genießt dennoch einen hohen Bekanntheitsgrad. Bestätigt wird dies von der häufigen Nennung auf der Titelseite und im Titelbereich. Vorteilhaft für die Glaubwürdigkeit ist die Reputation und das politische Engagement von *Transparency International*. Die Verfahren zur Konstruktion des Index rücken dadurch vermutlich in den Hintergrund. Dies zeigt sich auch in den Beiträgen, bei denen der Index im Wesentlichen nicht weiter erläutert wird. Anders als bei anderen untersuchten Indizes (z.B. HDI, HPI) wird beim CPI nicht weiter auf das Indexkonzept eingegangen, um z.B. zu diskutieren, was Korruption eigentlich ist und wie sie ggf. bekämpft werden kann. Ursächlich dafür ist, dass die Dimensionen von Korruption vom Indexkonzept nicht beschrieben werden.

5.7.11 Gegenüberstellung und Zusammenfassung der Untersuchungseinheiten

In der Kommunikation sind die Eigenschaften von Indizes und auf Messung basierender Zahlen z.T. sehr verschieden. Eine Leistung der untersuchten Indizes ist die Kommunikation von Komplexität. Die Voraussetzungen eines symbolisch generalisierten Kommunikationsmediums erfüllen sie jedoch i.d.R. nicht.

Alle untersuchten Indizes werden aus mehreren Indikatoren konstruiert und sind insofern mehrdimensional. Sie verdichten eine Fülle von Informationen in einer Maßzahl und ermöglichen damit, Mehrdimensionalität kommunizierbar zu machen. Teilweise bieten sie auf Basis ihres Konzept eine Grundlage zur Diskussion komplexer Begriffe.

In der manuellen Inhaltsanalyse wurde deutlich, dass das theoretische Konzept einiger Indizes thematisiert wird. Es wird entweder als Diskussionsgrundlage für Begrifflichkeiten genutzt oder um dem Rezipienten den Gegenstand der Quantifizierung zu verdeutlichen. Da das Indexkonzept die Basis für den Indexwert darstellt, wird die Diskussion oder Beschreibung des Konzepts auch die Glaubwürdigkeit der Indexzahl beeinflussen.

Eine Eigenschaft auf Messung basierender Zahlen ist, dass sie tendenziell affirmativ sind. Dies trifft auf Beiträge, die das Indexkonzept beschreiben, nicht zu. Denn in den Aussagen über das Konzept ist immer eine Ja-/Nein-Fassung enthalten. Damit beinhalten sie automatisch die Möglichkeit der Negation. Sollten Zweifel an der Beziehung zwischen theoretischem Konstrukt und Indikatoren aufkommen, wird sich dies auf die Glaubwürdigkeit der Indexzahl auswirken und auch auf die Wahrscheinlichkeit für die Annahme einer Kommunikation.

Auf die Zurechnung von Objektivität wirkt sich aus, dass die meisten untersuchten Indizes über das Beschriebene hinaus weisen. Das bedeutet, dass sich Aussagen nicht nur auf das wissenschaftlich stilisierte Phänomen (Konzept) beziehen, sondern auch ein Interesse an seiner lebensweltlichen Erscheinung erhalten bleibt.

Es zeigte sich, dass auch Indizes auf Selektionen im Luhmannschen Sinne basieren. Diese erfüllen bestimmte Voraussetzungen, welche mit einigen Einschränkungen detailliert nachvollzogen werden konnten. Ob sie jedoch für eine interpretationsfreie Beschreibung stehen können, ist schwer zu beantworten. Die Vielzahl an Revisionen und Alternativkonzepten lässt vermuten, dass keine Eindeutigkeit existiert und Objektivität schwer zu beurteilen ist.

Auch in der Kommunikation der Indexwerte – in Abgrenzung zum Indexkonzept – zeigten sich einflussreiche Unterschiede. Indexwerte wurden entweder in ihrer numerischen Form kommuniziert oder in Form von Aussagen. Diese Aussagen beziehen sich entweder auf Formulierungen des *mehr* und *weniger* oder z.B. auf die Nummer innerhalb einer Rangliste. Letztgenannte Aussagen beinhalten ebenfalls eine Ja-/Nein-Fassung und damit die Möglichkeit einer Negation. Die Eigenschaften von Zahlen als symbolisch generalisiertes Kommunikationsmedium sind unter bestimmten Voraussetzungen auch auf Indizes übertragbar. In den untersuchten Fällen erfüllten diese jedoch nur die wenigsten. Eine Erklärung dafür setzt bei den Indikatoren an, die – im Gegensatz zu naturwissenschaftlichen Messobjekten – i.d.R. keine quantitative Struktur besitzen.

Tabelle 5.41: Ressort-Statistik je Index in Prozent

Ressort	BLI	BMI	DAX	Gfk-I	HPI	HDI	ifo-I	CPI	MCSI
Wirtschaft	83.3	9.2	85.4	98.5	55.6	22.7	94.5	54.3	98.2
Politik		15.5	1.2			43.2	0.8	31.9	
Geld Und Mehr	16.7		7.2	1	33.3	6.8	2.3	0.9	0.7
Feuilleton		7.4	0.8			9.1	0.1	3.4	
Natur Und Wissenschaft		15.1				4.5			0.1
Sport		12.9	0.3					2.6	
Reise					11.1	4.5			
Wissenschaft		7.4				2.3			0.1
Gesellschaft		2.6				2.3		1.7	
Ohne Angabe		3.7					0.2		0.1
Verlagsbeilage			1.2				0.3	1.7	
Beruf Und Chance			0.7			2.3			
Frankfurt		2.6							
Technik Und Motor		2.2							0.1
Titelseite						2.3			
Sonntagszeitung									
Jugend Schreibt		2.2							
Meinung								1.7	

Anmerkung: Es werden nur Ressorts ausgewiesen, die für mindestens einen Index eine prozentuale Häufigkeit größer oder gleich eins aufweisen.

Für eine leichtere deskriptive Gegenüberstellung werden die Ergebnisse der untersuchten Indizes in Tabelle 5.42 und 5.41 zusammengefasst. Die wichtigsten Ergebnisse werden im Folgenden kurz dargestellt.

Die für die vorliegende Arbeit selektierten Indizes erschienen im Wesentlichen in den Ressorts *Witschaft*, *Geld und Mehr* und *Politik*. Auf alle anderen Ressorts fielen maximal 15,1% der Beiträge. (vgl. Tabelle 5.42)

Mit jeweils über 80% dominieren die Beiträge zum BLI, DAX, Gfk-Konsumklimaindex, ifo-Geschäftsklimaindex und MCSI im Ressort *Wirtschaft*. Gleiches gilt mit etwas über 50% für die Indizes HPI und CPI. Die Beiträge zum BMI (15,5%) und HDI (43,2%) dominieren im Bereich *Politik*. Jedoch wird der BMI mit einer ähnlichen Häufigkeit auch in den Ressorts *Politik*, *Natur und Wissenschaft* sowie *Sport* zitiert.

Im Mittel wird ein Index in fünf verschiedenen Ressorts genannt.[82] Überdurchschnittlich streuen damit die Indizes BMI (11), DAX (4), HDI (10) und CPI (7). Am unteren Ende befindet sich mit lediglich einem Ressort der MCSI.

Die höchste Anzahl an Beiträgen für das Jahr 2014 entfällt deutlich auf den DAX mit 1133 Beiträgen. Danach folgt mit weit weniger Beiträgen der ifo-Index mit 95 Beiträgen.

[82] In die Betrachtung gehen nur Ressorts ein, die mindestens 1% der Beiträge eines Index erhalten.

Er wird oft in Verbindung zum DAX gebracht, da er als Frühindikator wertvolle Informationen zur künftigen deutschen Konjunktur geben kann. An dritter Stelle der betrachteten Indizes landet der BMI mit 26 Beiträgen in 2014. (vgl. Tabelle 5.42)

Bis auf den BLI, HPI und BLI wurden alle betrachteten Indizes mindestens einmal auf der Titelseite erwähnt. Am häufigsten mit 11,4% der HDI, gefolgt vom BMI mit 8,5%. Die relative Anzahl bezieht sich dabei auf die Gesamtzahl an Beiträgen zu einem Index. Für zukünftige Betrachtungen sollte die Berechnung angepasst werden. Hintergrund ist, dass insbesondere die Zahlen zum DAX irreführend sind. Insgesamt erscheint dieser absolut betrachtet am häufigsten auf der Titelseite. Da er jedoch im Schnitt in ca. drei Beiträgen je Ausgabe erwähnt wird, ist die relative Häufigkeit deutlich geringer als bei anderen Indizes. Eine Möglichkeit wäre, die relative Häufigkeit nicht auf Ebene der Beiträge zu berechnen, sondern auf Ebene der Ausgabe.

Aussagekräftiger ist die Nennung des Index im Titel des Beitrags. Beiträge zum DAX enthalten zu 8,5% bereits den Namen des Index im Titel. Beim ifo-Index sind es noch 4,4%, beim CPI 2,6%.

Für die Relevanz eines Index steht ebenfalls das Kriterium, ob der Index im ersten Absatz oder im Teaser des Beitrags genannt wird. Dies gilt für 41,6% der Beiträge zum DAX, 35,8% der Beiträge zum MCSI, 34,3% der Beiträge zum Gfk-Index, 31,5% der Beiträge zum ifo-Index und 27,7% der Beiträge zum BMI. Alle anderen Indizes erscheinen im Wesentlichen erst im späteren Verlauf des Beitrags.

Insgesamt lässt sich aus der Analyse ableiten, dass einige Indizes eine große Relevanz in der FAZ und FAS entwickeln konnten. Diese wird u.a. beeinflusst von dem Themenbereich, der Aktualität, der Bekanntheit, ihrem Einfluss auf andere Systeme und der Reputation des Herausgebers. Die Funktion der Indizes reicht dabei von einer Grundlage zur Diskussion von Problemen oder Herausforderungen bis hin zur Kommunikation komplexer und mehrdimensionaler Begriffe. Im Gegensatz zu naturwissenschaftlichen Messungen lassen sie sich jedoch nicht als symbolisch generalisiertes Kommunikationsmedium begreifen.

Tabelle 5.42: Mittelwerte je Index

Kennzahl	BLI	BMI	DAX	Gfk-I	HPI	HDI	ifo-I	CPI	MCSI
Titelseite	0	8.5	2.8	0	0	11.4	2.9	2.6	0.1
Wörter/ Beitrag	656.2	697.8	554.6	509.6	917.9	1057.2	542.7	773.3	515.6
Anzahl Absätze	30.2	8.5	7.1	32.2	12.8	11.7	16.9	21.1	36.2
Wörter/ Titel	4.2	4.5	5	3.9	4.4	4.5	4.4	4.4	4.3
Anzahl Indexnennung/ Beitrag	1.3	1.2	2.3	1.1	1.6	1.1	1.3	1.1	1.1
Indexnennung/ Titel (Prozent)	0	0.7	8.5	0.5	0	0	4.4	2.6	0.4
Indexnennung/ 1. Absatz (Prozent)	0	27.7	41.6	34.3	11.1	2.3	31.5	8.6	35.8
Indexnennung/ Beitrag (Prozent)	53	56	46.9	56.5	63.7	100	54.6	52	72.4
Satzlänge	14.2	16.2	15.7	12	15.3	17.3	14.4	15.9	11.9
Aktualität	81.2			4.4	544.6	218.8	9.1	175.3	5
Anzahl Beiträge 2014	2	26	1133	5	0	4	95	7	11

Kapitel 6

Fazit

Eine fundamentale Leistung dieser Arbeit ist die Ausarbeitung der Beziehung zwischen dem naturwissenschaftlichen Messprozess und dem statistischen Index. Sie ist substantiell für die Beantwortung der Forschungsfrage, da die medialen Funktionen statistischer Indizes maßgeblich von deren Bezug zum Messprozess abhängen. In der Praxis existieren verschiedene theoretische Ansätze, Modelle und Zielstellungen zur Konstruktion und Konzeption von Indizes. Diese sind z.T. so divergent, dass die Diskussion, ob sie als „Messung“ zu betrachten sind, in der Wissenschaft nicht abschließend geklärt ist. In Folge dessen sind die medialen Funktionen eng an die jeweilige Indexkonstruktion gebunden und nur unter bestimmten Bedingungen verallgemeinerbar. Der Autor dieser Arbeit bezieht hier eine deutliche Position, indem er Indizes klar von Messungen abgegrenzt.

In der vorliegenden Arbeit wurden Ansätze zur Konstruktion von Indizes aus einer Vielzahl von Disziplinen zusammengebracht und bewertet. Die maßgebliche Leistung in diesem Zusammenhang ist die Entwicklung eines theoretischen Modells, mit dessen Hilfe die medialen Funktionen verschiedenartiger Indexkonzepte erstmals umfassend analysiert werden. Außerdem wirkt sie der terminologischen Verwirrung entgegen, die mit der Vielzahl an Methoden zur Indexkonstruktion und deren Stellung im naturwissenschaftlichen Messprozess einhergeht.

Mit der Einordnung der Indizes in die Evolution quantitativer Begriffe innerhalb und außerhalb der „Wissenschaft“ wird eine Grundlage erarbeitet, um die Funktion von Indizes innerhalb der Verbreitungsmedien zu untersuchen. Für den Abgleich der theoretischen Erkenntnisse mit der Empirie wurde eigens ein völlig neuartiges System entwickelt,

welches den Medienwissenschaften zukünftig weitreichende Analysemethoden bezüglich digitaler Medien bereitstellt.

Mit der Zusammenführung und Diskussion existierender Methoden und Konzepte zur Konstruktion von Indizes legte die Arbeit offen, dass einheitliche Standards fehlen und die Vielzahl heterogener Ansätze nur schwer auf einen erkenntnistheoretischen Nenner zu bringen ist. Dies wurde nicht nur zwischen, sondern auch innerhalb verschiedener Disziplinen festgestellt. Die Arbeit brachte durch den interdisziplinären Vergleich zum Vorschein, dass die Frage, *ob ein Index als „Messung" betrachtet werden kann*, nicht von allen Disziplinen aufgeworfen oder beantwortet wird und somit diesbezüglich kein Konsens besteht. Die Stellung des Index innerhalb des naturwissenschaftlichen Messprozesses ist in der Fachliteratur folglich nicht abschließend geklärt. Es besteht hier die Notwendigkeit einer Aufarbeitung, da ein Index für den Autor der Arbeit keine wissenschaftliche Messung darstellt und klar von dem Begriff „Messung" abgegrenzt werden muss.

Eine weitere Leistung der vorliegenden Arbeit ist in diesem Zusammenhang das Einbeziehen statistischer Methoden in die erkenntnistheoretische Betrachtung des naturwissenschaftlichen Messprozesses. Durch die Einführung zwei unterschiedlicher Betrachtungsmodelle – eines sieht Indizes als Resultat der messbaren Indikatoren (formatives Modell), das andere als Ursache (reflektives Modell) – konnte der Unterschied greifbar gemacht werden. Jedoch ist diese Lösung als nicht befriedigend bewertet worden, da ein einheitliches Verständnis in Form eines einzigen Konstruktionsmodells fehlt. Bezugnehmend auf die einleitend beschriebene etymologische Bedeutung eines Index, in der er ganz allgemein als eine Art „Hinweisschild" oder „Wegweiser" betrachtet werden kann, ist das formative Modell – gegenüber dem refklektiven Modell – klar zu bevorzugen. Die Funktion in diesem Sinne ist es, deutlich auf einen Sachverhalt hinzuweisen und nicht, diesen zu messen.

Im Rahmen der historischen Betrachtung zeigte die Arbeit auf, dass die Ausdifferenzierung der beiden Konstruktionsmodelle in der Mitte des letzten Jahrhunderts zu verorten ist. Die wissenschaftliche Auseinandersetzung mit Indizes im Allgemeinen ist zu Beginn des 20. Jahrhunderts zu sehen. Vornehmlich in Form der Verhältniszahlen oder Aktienindizes verdichteten Indizes zu dieser Zeit, in der Regel auf Basis eines formativen

Modells, mehrere Dimensionen zu einer Grundaussage. Die Frage nach der Stellung von Indizes im Messprozess war weitestgehend irrelevant.

Aktienindizes messen z.B. nicht, sondern fassen eine Vielzahl von Aktien (Dimensionen) zu einer Grundaussage zusammen. Damit ermöglichen sie einen leichten Überblick sowie Umgang mit einem komplexen System. Ähnlich wie der Index innerhalb eines Buches listen sie verschiedene Aktien auf, die bestimmte Kriterien erfüllen. Die Dimensionen formen den Index und der Indexwert ist die Abbildung des Indexkonzepts.

Mit der Verbreitung der indikatorbasierten „Messung“ nimmt der Bezug zur etymologischen Bedeutung des *Index* ab. Die wissenschaftshistorische Betrachtung dieser Arbeit zeigt hier vor allem zwischen 1958 und 1979 wegweisende wissenschaftliche Publikationen. Der Index basiert nach den theoretischen Betrachtungen nicht mehr nur auf *formativen* Dimensionen. Stattdessen wird das Konstrukt der *reflektiven* Indikatoren eingeführt. Nun wird der Index als „Ursache“ betrachtet und über Indikatoren, in denen er sich manifestiert, quantifiziert. Wie die Forschung ergeben hat, steht dieses Konstrukt der *Ursache* offenkundig im Widerspruch zur Bedeutung des Index als *Mittler*. Zumindest dann, wenn der Anspruch erhoben wird, das latente Konstrukt zu *messen*. Sollte ein Index tatsächlich etwas *messen*, müsste das Ergebnis korrekterweise mit einem anderen Ausdruck versehen werden. In dieser Hinsicht konnte der Autor zeigen, dass die Diskussion, ob ein Index als Messung angesehen werden kann, obsolet ist. Denn wäre das der Fall, wäre das Ergebnis kein Index und stünde auch nicht für ein mehrdimensionales Konzept.

Reflektive Modelle haben dennoch ihre Berechtigung – ebenso die Diskussion über ihre Stellung im naturwissenschaftlichen Messprozess. Wie sich gezeigt hat, ist der Begriff *Index* an dieser Stelle jedoch irreführend, da mit dem reflektiven Modell „Realität“ abgebildet und nicht nur darauf hingewiesen werden soll. Außerdem führt es zu einer „Verwirrung“, dass beide Modelle (reflektiv und formativ) für die Konstruktion von Indizes verwendet werden. Denn damit kann ein Index potentiell Ursache oder Resultat zugleich sein, was Folgen auf die Interpretation sowie Kommunikation hat. Der erkenntnistheoretische Unterschied verschwindet scheinbar hinter der gemeinsamen Bezeichnung *Index*.

Die Arbeit zeigt, dass sich der Begriff *Index* aktuell als ein Konglomerat aus genauen

und ungenauen Quantifikationen („Messungen“) im wissenschaftlichen Sinne, Aggregationen, Klassifikationen und Verhältnissen präsentiert. Die Kommunikation ist damit uneindeutig und in Folge dessen die mediale Funktion auch. So hängt z.B. die Annahme einer Kommunikation ganz wesentlich davon ab, ob die Zahl als objektive und interpretationsfreie Beschreibung der Realität angesehen werden kann. Dem Begriff *Index* fehlt es jedoch synchron und diachron an Differenzierung und Eindeutigkeit, was es u.a. ermöglicht, ihn für die Kommunikation unscharfer und vager Begrifflichkeiten zu instrumentalisieren.

Für die Kommunikation der z.T. divergenten Indexkonzepte wird in der Praxis nicht einheitlich auf die zur Verfügung stehenden Begrifflichkeiten zurückgegriffen. Indizes mit einem stärkeren Bezug zum Messprozess werden teilweise als *Quotient* oder *Kennzahl* bezeichnet. Im Gegensatz dazu erlaubt die englische Bezeichnung *Composite Indicator* eine klarere Abgrenzung vom Messprozess und die Hervorhebung des Charakters einer Zusammenfassung. Jedoch wird der Hintergrund einer Mehrdimensionalität nicht immer über den Begriff kenntlich gemacht, wie das Beispiel der Klimaerwärmung verdeutlicht.

Eine begriffliche Abgrenzung von einer klassischen Messung erscheint antagonistisch zur Kommunikation von Indizes über Zahlen. Die Veröffentlichung von Indizes geht teilweise mit der Nennung einer Zahl einher. Wie die Forschung ergeben hat, werden in der alltäglichen Erfahrung Zahlen – in Abgrenzung zu Ziffern – jedoch in der Regel als Resultat einer Messung im klassischen Sinne angesehen. Die Kommunikation von Indizes über Zahlen birgt damit die Gefahr, dass Indexwerte mit Messungen in Verbindung gebracht werden, obwohl ihr Konzept dies eigentlich nicht zulässt. Die Funktionen und Eigenschaften von Indizes innerhalb der Medien bewegen sich in diesem Fall zwischen ihrer tatsächlich immanenten Funktion und der von auf Messung basierenden Zahlen im Allgemeinen.

Die empirische Untersuchung dieser Arbeit belegte in diesem Zusammenhang, dass die Verbreitungsmedien über Methoden verfügen, mit denen Indizes von einer Messung abgegrenzt werden. So zeigte sich in den Beiträgen, dass z.T. auf die Nennung von Zahlenwerten verzichtet wird. Vornehmlich bei Indizes, deren Anspruch die Abbildung eines latenten Konstruktes ist oder deren Indikatoren keine quantitative Struktur besitzen. Im Vordergrund dieser Beiträgen stand entweder ausschließlich das Indexkonzept oder quantitative Aussagen, die sich auf das *mehr* und *minder* beschränken. Aussagen dieser

Form unterscheiden sich von der Nennung von Zahlenwerten. Denn sie beinhalten eine Ja-/Nein-Fassung und damit die Möglichkeit einer Negation.

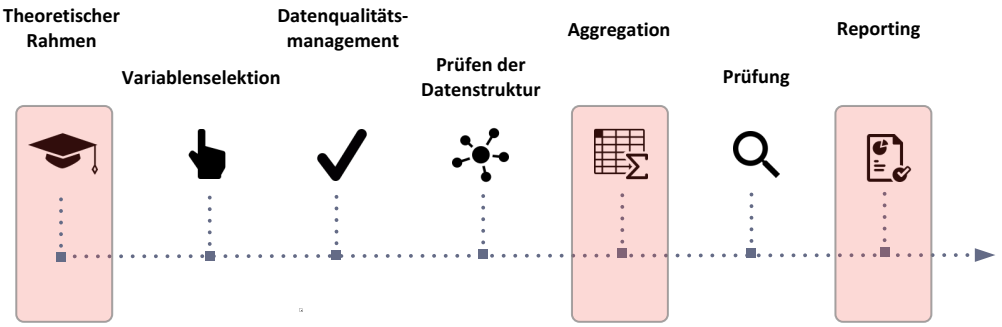
Eine weitere Erscheinungsform der Abgrenzung von einer Messung ist die explizite Formulierung des Zusammenhangs zwischen Index und Indikatoren oder die Beschreibung des Indexkonzepts. Im empirischen Teil wurde daher analysiert, welche Elemente der Indexkonstruktion in den Beiträgen thematisiert werden. Denn neben dem unklaren Bezug zum Messprozess zeigte sich, dass Aussagen auf Basis von Indizes ohne zusätzliches Wissen z.T. missverständlich sind. Dies trifft vor allem auf latente Konstrukte zu, bei denen nicht-gegenständliche Objekte quantifiziert werden. Die Konstitution eines umgangssprachlichen Begriffs (z.B. „Glück“) über eine spezifische Selektion und Konkatenation von Merkmalen ist oft subjektiv und willkürlich und nicht immer intuitiv nachvollziehbar. Dies wird um so deutlicher, da sich die Zusammensetzung der Merkmale und die Konstruktionsmethoden der untersuchten Indizes nicht selten über die Zeit geändert haben. Generell ist davon auszugehen, dass dem Leser das Konzept eines Index nicht bekannt ist, sondern Unklarheit darüber herrscht, was er abbildet oder „anzeigt“.

Für einige untersuchte Indizes wurde deutlich, dass das theoretische Konzept thematisiert wird. Es wurde entweder als Diskussionsgrundlage für Begrifflichkeiten genutzt oder zur Erklärung des zu operationalisierenden Gegenstands, z.B. um dem Rezipienten zu verdeutlichen, worin das Ziel der Quantifizierung liegt. Da das Indexkonzept die Basis für den Indexwert darstellt, wird die Diskussion oder Beschreibung des Konzepts Auswirkungen auf die Glaubwürdigkeit der Indexzahl besitzen.

Eine Eigenschaft auf Messung basierender Zahlen ist, dass sie tendenziell affirmativ sind. Dies trifft auf Beiträge, die das Indexkonzept beschreiben, nicht zu. Denn in den Aussagen über das Konzept ist immer eine Ja-/Nein-Fassung enthalten. Damit beinhalten sie automatisch die Möglichkeit der Negation. Sollten Zweifel an der Beziehung zwischen theoretischem Konstrukt und Indikatoren aufkommen, wird sich dies auf die Glaubwürdigkeit der Indexzahl auswirken und auch auf die Wahrscheinlichkeit der Annahme einer Kommunikation.

Abbildung 6.1 verdeutlicht, welche Schritte der Indexkonstruktion innerhalb der Beiträge thematisiert werden, um z.B. die Wahrscheinlichkeit von Missverständnissen zu reduzieren.

Abbildung 6.1: Schritte der Indexkonstruktion, die in Beiträgen der FAZ aufgegriffen werden (zutreffende sind farblich hinterlegt)



Quelle: Eigene Darstellung

Durch die Analyse konnte deutlich gemacht werden, dass bei wenigen Indizes der theoretische Rahmen aufgegriffen und die Verbindung der beobachtbaren Indikatoren zum Indexkonzept thematisiert wird. Das ist der wichtigste Schritt, um dem Empfänger zu vermitteln, ob der Index auf die beobachtbaren Indikatoren wirkt oder umgedreht. Außerdem verdeutlicht er, ob der Index den Anspruch erhebt, etwas zu messen oder stattdessen mehrere Dimensionen zusammenzufassen.

Teilweise wird bei der Diskussion zusätzlich auf die Aggregation der Merkmale eingegangen. Über die Thematisierung dieses Schritts innerhalb der Indexkonstruktion wird die Transparenz erhöht. Außerdem ist ein umfassenderes Verständnis vom abzubildenden Gegenstand möglich, da die Gewichtung der konstituierenden Merkmale zugänglich ist. Neben den beiden Schritten *Theoretischer Rahmen* und *Aggregation* werden in den meisten Beiträgen lediglich die Ergebnisse präsentiert und damit auf den letzten Schritt der Indexkonstruktion eingegangen. Wie die Analyse zeigt, finden die restlichen Punkte aus Abbildung 6.1 keine Beachtung.

Im empirischen Teil wurde außerdem eine umfassende Recherche zu den Konstruktionsmethoden der Indizes durchgeführt. Dabei wurde eine Vielzahl von Quellen einbezogen, der Großteil der Informationen stammt jedoch vom Herausgeber des Index selbst. Die Recherche legt dabei offen, dass nur wenige Herausgeber sämtliche Informationen veröffentlichen. Die Transparenz und Übersichtlichkeit der Konstruktionsmethoden variiert z.T. sehr stark: Zu wenigen Indizes wird explizit ein Dokument mit den Konstruktionsmethoden und dem theoretischen Rahmen geliefert – teilweise sogar mit Angabe der

Fehlerwahrscheinlichkeit und Genauigkeit. Andere hingegen nennen, ähnlich einer Blackbox, lediglich die konstituierenden Merkmale und die Ergebnisse. Die Objektivität des Index lässt sich damit nicht immer beurteilen. In der verwendeten Qualitätsmatrix zur Beurteilung der Indizes wurde dieser Faktor innerhalb der Transparenz berücksichtigt. Insgesamt zeigt sich, dass die Qualität (nach Definition der Qualitätsmatrix) der untersuchten Indizes – z.B. hinsichtlich der Transparenz oder Datenquellen – sehr unterschiedlich ist.

Über die Qualitätsmatrix wurde herausgearbeitet, dass ein Index nicht per se wissenschaftlich ist. Die Wissenschaftlichkeit ist im Wesentlichen von dem theoretischen Konstrukt abhängig, inklusive des explizierten Zusammenhangs zu den beobachtbaren Indikatoren sowie den angewendeten Konstruktionsmethoden. Der wissenschaftliche Index ist transparent, indem ausführliche Informationen zu jedem Schritt der Indexkonstruktion sowie zu den verwendeten Quelldaten bereitgestellt und begründet werden. Darüber hinaus verwendet er angemessene Visualisierungstypen innerhalb der Kommunikation.

Ein wissenschaftlicher Index sucht ferner den Vergleich zu existierenden Maßen, indem er z.B. Unterschiede und Gemeinsamkeiten identifiziert sowie seinen Mehrwert verdeutlicht. Über die Reflexion der Qualität wird die eigene Aussagekraft zur Diskussion gestellt.

Bei der Analyse der kommunikativen Eigenschaften eines Index trat in der Arbeit erneut die eingangs erwähnte Abhängigkeit zum Messprozess in den Vordergrund. Denn wird ein Index als „Messung“ betrachtet, kommen ihm spezifische kommunikative Eigenschaften zu, die es z.T. zulassen, ihn als symbolisch generalisiertes Kommunikationsmedium aufzufassen. Indem der Index dann Objektivität signalisiert und anzeigt, dass er auf Selektion beruht, erhöht er in Abwesenheit persönlicher Interaktion die Akzeptanz von Kommunikation. Außerdem wird der Anschein von Unpersönlichkeit geweckt. Es entsteht der Eindruck, er stünde für Realität an sich.

Es zeigte sich, dass auch Indizes auf Selektionen im Luhmannschen Sinne basieren. Diese erfüllen bestimmte Voraussetzungen, welche mit einigen Einschränkungen detailliert nachvollzogen werden konnten. Ob sie jedoch für eine interpretationsfreie Beschreibung stehen können, ist schwer zu beantworten. Die Vielzahl an Revisionen und Alternativkonzepten lässt vermuten, dass keine Eindeutigkeit existiert und Objektivität schwer zu beurteilen ist.

In der Analyse zeigte sich, dass die Zurechnung von Objektivität auch von der Thematisierung des theoretischen Konzepts abhängt und davon, dass Indizes z.T. über das beschriebene hinausweisen. Das heißt, dass sich Aussagen nicht nur auf das wissenschaftlich stilisierte Phänomen (Konzept) beziehen, sondern auch ein Interesse an seiner lebensweltlichen Erscheinung erhalten bleibt.

Die Vermittlung des theoretischen Konzeptes beeinflusst z.T. antagonistisch die Akzeptanz einer Kommunikation über Zahlen. Wird es nicht thematisiert, besteht die Gefahr eines Missverständnisses. Gleichzeitig wird die Zurechnung von Objektivität dann weitestgehend von der Reputation des Herausgebers abhängen. Sie führt zu der Notwendigkeit eines dreifachen Vertrauens: (1) in die Zuverlässigkeit der „Messverfahren“, über die die Rohdaten gewonnen wurden, (2) in die Zuverlässigkeit der Operationalisierung des theoretischen Konzepts und (3) in die Zuverlässigkeit der verwendeten statistischen Verfahren sowie deren korrekte Anwendung. Unter der Voraussetzung von Vertrauen wird der Index die Wahrscheinlichkeit für die Annahme einer Kommunikation erhöhen.

Wird das Indexkonzept hingegen beschrieben, verringert sich die Wahrscheinlichkeit eines Missverständnisses. Wobei die Wahrscheinlichkeit einer Ablehnung steigt, weil eine darunter getroffene Aussage bereits die Negation enthält. Bei den untersuchten Indizes offenbarte sich zudem, dass z.B. durch Alternativkonzepte oder Revisionen die signalisierte Objektivität instabil ist. Zudem beeinflussten Intransparenz und Unübersichtlichkeit die Glaubwürdigkeit.

Auf Basis der empirischen Ergebnisse ist davon auszugehen, dass die wenigsten Indizes die Eigenschaften eines symbolisch generalisierten Kommunikationsmediums teilen. Die Voraussetzungen sind schwer zu erfüllen und i.d.R. nicht mit den Zielen einer Indexkonstruktion zu vereinen. Wie eingangs vermutet, ist die Funktion *als* Medium bei Indizes in größerem Maße dahingehend zu werten, dass sie der Herausforderung nachkommen, Komplexität und Mehrdimensionalität zu kommunizieren.

Alle untersuchten Indizes basieren auf mindestens zwei Dimensionen, die über ein Konstrukt auf einen eindimensionalen Raum reduziert werden. In dieser Hinsicht vermögen es alle, Komplexität und Mehrdimensionalität zu reduzieren. Die Funktion der Indizes in den Verbreitungsmedien ist die Kommunikation dieser Mehrdimensionalität. Jedoch wird das mehrdimensionale Konzept nicht ausschließlich über den Index kommuniziert. Teilweise werden auch Einzeldimensionen oder mit dem Konzept zusammenhängende

Maßzahlen aufgegriffen. Das Aufgreifen eines Index findet damit nicht unbedingt zu Lasten einer fundamentalen oder abgeleiteten Messung statt, sondern hängt vielmehr von der Aussage im Beitrag ab sowie der Relevanz und Gewichtung einer einzelnen Dimension. Die Kommunikation von Komplexität und Mehrdimensionalität wird damit in der Regel entweder über den Index allein oder in Verbindung mit einer spezifischen Selektion (von Dimensionen oder anderer Maßzahlen) ermöglicht.

Es wurde festgestellt, dass die Funktion der Reduktion von Komplexität und Mehrdimensionalität es zulässt, die beiden eher unterschiedlichen Modelle zur Konstruktion von Indizes – d.h. zum einen das Modell, das Indizes als Resultat betrachtet und zum anderen das Modell, das Indizes als Ursache betrachtet – auf einen Nenner zu bringen. Denn, abgesehen von dem Unterschied von Ursache und Wirkung, konkatenieren beide mathematisch betrachtet in der Regel verschiedene Indikatoren und werden über einen einzigen Wert kommuniziert. Inhaltlich ergibt sich dennoch der Unterschied, dass das erstgenannte Modell tatsächlich aus mehreren Merkmalen zusammengesetzt ist, das letztgenannte hingegen lediglich in diesen manifestiert beobachtet werden und so gesehen nicht als Aggregat dieser Merkmale begriffen werden kann.

Über die Analyse der Indizes innerhalb des empirischen Teils der Arbeit wurde deutlich, dass sich die Reduktion von Komplexität z.T. auch auf die Kommunikation vager und abstrakter Begriffe wie „Glück“ oder „Vertrauen“ bezieht. Die Konstruktion des Index konstituiert einen Gegenstand, der in der Regel mit dem zu kommunizierenden Begriff mehr oder weniger Gemeinsamkeiten besitzt. Die Auswahl der konstituierenden Merkmale ist zwar z.T. arbiträr, jedoch wird über den Index der gesamtheitliche Begriffsraum definitorisch auf eine fixe Menge an Merkmalen reduziert. Diese „Neuschöpfung“ eines Begriffs wird in der Kommunikation z.T. mit einem umgangssprachlichen Begriff assoziiert. So geschehen beim einleitend beschriebenen *Happy Planet Index*, in dessen Zusammenhang ein Beitrag die Frage aufwirft, wie *glücklich* ein Land ist. In diesem Fall wird der Begriff *Glück* mit den konkatenierten Merkmalen *Lebenserwartung*, *Zufriedenheit* und *ökologischer Fußabdruck* assoziiert.

Da in einigen Beiträgen ausschließlich die Indexwerte in Zahlenform – ohne eine generelle Diskussion des Index – präsentiert werden, lässt sich davon ausgehen, dass die Zurechnung von Objektivität und Unbestreitbarkeit nicht ausschließlich an wissenschaftliche

Bedingungen gebunden ist, sondern – je nach Hintergrundwissen – über außerepistemische Maßnahmen bestimmt wird. Ein differenziertes und umfassendes Bild zu einem Index haben vermutlich die wenigsten Empfänger, da der Rechercheaufwand groß ist. Zusätzlich ist ein gewisses Maß an inhaltlichem und vor allem mathematischem Hintergrundwissen erforderlich. Zudem bieten einige Indizes schlichtweg nicht die Transparenz, um ein detailliertes Bild zu ermöglichen. Aus diesem Grund hängt die Akzeptanz auch von der Reputation des Herausgebers und dem Vertrauen in die Objektivität des Index ab.

Der Empfänger muss damit die Entscheidung über das Akzeptieren oder Ablehnen der Aussage u.U. ohne Wissen über das Indexkonzept treffen und damit abseits der immanenten Funktion des Index. Zahlen können dabei – wenn sie mit einer Messung assoziiert werden – die Kommunikation wahrscheinlicher machen. Doch auch suggestive Aussagen oder Kommentare des Autors können je nach Empfänger mehr oder weniger auf die Glaubwürdigkeit des Index einwirken.

Zudem signalisiert die Nennung eines Index in einem Beitrag Verdichtung und Relevanz. Die empirische Untersuchung erwies, dass das Aufgreifen eines Index zwar nicht immer mit der Nennung einer Zahl einhergeht. Ist dies jedoch der Fall, signalisiert die Auffälligkeit (Präsentation) des Indexwertes Relevanz. Die Zahl hebt sich rein optisch deutlich von ihrem Umfeld, das aus Wörtern besteht, ab.

Die Aktualität der Nachricht hat nur auf einen Teil der untersuchten Indizes Einfluss. Bezogen auf die Relevanz profitieren vor allem die ökonomischen Indizes von der Aktualität der Beiträge. Dabei zeigt die Untersuchung auch, dass das Printmedium insbesondere bei einer untertägigen Veröffentlichung von Indexwerten jedoch vor eine Herausforderung gestellt wird. Untertägige Indizes, zusammen mit sozialen und politischen entfalten ihre Relevanz eher über inhaltliche Aspekte oder das Thema.

Bei der empirischen Untersuchung der medialen Funktionen trat rasch zum Vorschein, dass längere Zeitreihen für aussagekräftige Ergebnisse notwendig sind. Einige Indizes werden in den Verbreitungsmedien nur wenige Male innerhalb eines Jahres aufgegriffen. Die manuelle Selektion und Durcharbeitung von Beiträgen erscheint vor dem Hintergrund eines mehrjährigen Zeitfensters nicht umsetzbar. Aus der Notwendigkeit heraus,

praktikable Alternativen zu erarbeiten, war ein weiteres wesentliches Anliegen der vorliegenden Arbeit, zu prüfen, inwieweit die Inhaltsanalyse des empirischen Teils automatisierbar ist. Im Ergebnis steht den Medienwissenschaften eine völlig neuartige Analysemethode für die digitalen Medien zur Verfügung.

Die Anbindung der Webdatenbank *DBpedia* stellte sich als äußerst nützlich heraus. Als Ausgangspunkt bietet sie die Möglichkeit automatisiert Aufgriffkritieren für die Selektion digitaler Beiträge zu erstellen. Der größte Vorteil der Datenbank besteht darin, dass auch gängige Abkürzungen, alternative Schreibweisen und „Falschschreibungen“ einbezogen werden können. Darüber hinaus besteht die Möglichkeit, die Aufgriffkriterien in mehreren Sprachen ausgeben zu lassen.

Die Inhalte der Datenbank werden kontinuierlich aktualisiert und liegen in strukturierter Form vor. Probleme, die aus mangelhafter Qualität der Inhalte resultieren, sind für die Erstellung einer Schlüsselwortliste zu vernachlässigen. Im entwickelten System standen lediglich der Titel und mögliche alternative Schreibweisen im Vordergrund. Die Inhalte selbst waren nebensächlich.

Das eigens entwickelte System stellt auf einfachste Weise die Zusammenstellung einer Liste von Aufgriffkriterien über das interne Klassifikationssystem der Datenbank zur Verfügung. Nach Eingabe eines Konzepts oder einer Klasse werden die eingruppierten Subjekte als Schlüsselwortliste zurückgeliefert.

Die Schlüsselwortliste dient als Ausgangspunkt für die Selektion digitaler Beiträge. Im konkreten Fall handelte es sich um eine Schlüsselwortliste von Subjekten (Indizes) unter dem Konzept „Index“, wobei verschiedene Schreibweisen eines Index mit „und“ sowie „oder“ Verknüpfungen verbunden werden. Verallgemeinernd kann die erarbeitete Methode ohne Weiteres auf andere Gegenstandsbereiche angewendet werden. Ferner ist auch der digitale Korpus prinzipiell frei wählbar. Die Methode kann überall eingesetzt werden, wo eine Freitextsuche über Beiträge zur Verfügung steht, das bedeutet z.B. über eine digitale Suchmaske oder das Vorliegen von Beiträgen in maschinenlesbarer Textform. Die Schlüsselwörter der Liste werden nacheinander an das Suchsystem übergeben und die zurückgegebenen Ergebnisse in Verbindung zum Schlüsselwort erfasst. Prinzipiell ist die Beziehung zwischen dem Schlüsselwort und dem Ergebnis eine 1:n-Beziehung, wobei $n \geq 0$. Über ein Schlüsselwort können also mehrere Beiträge aufgegriffen werden. Diese

werden im Folgeschritt über Funktionen in deskriptive Kennzahlen überführt, wobei sich die Funktionen auf Kriterien zur Analyse des Beitragsinhalts beziehen.

Weiterhin wurden verschiedene Kriterien entwickelt und auf Automatisierbarkeit geprüft. Diese war gegeben, wenn die Regeln und Bedingungen in maschinenlesbare Anweisungen (Funktionen) überführt werden konnten, was nicht immer der Fall war. Der Aufwand einer Automatisierung der übrigen Kriterien ist zwar durchaus denkbar, stand jedoch nicht im Verhältnis zum Ergebnis. Es ist damit zu rechnen, dass die Komplexität der Kriterien den Aufwand und die Fehleranfälligkeit der Lösung erhöhen.

Im aktuellen Entwicklungsstand entfaltet das System trotz des teilweise manuellen Aufwands einen umfangreichen Nutzen. So bietet es als Ausgangspunkt eine Struktur zur automatisierten Inhaltsanalyse, bei der Beiträge als Inputfaktoren zu deskriptiven Statistiken transformiert werden. Damit wird ein schneller und umfassender Überblick über eine große Menge an Beiträgen möglich. Außerdem lassen sich aus den Statistiken Hypothesen formulieren und explorativ prüfen. Das bereits fertig entwickelte Set an automatisierten Kriterien ermöglicht eine einfache Adaption für weitere Untersuchungen. Einerseits können die bereits definierten Kriterien benutzt werden, andererseits ermöglicht die Struktur der Lösung eine einfache Anpassung oder Erweiterung.

Die Prüfung im Rahmen dieser Forschungsarbeit legte offen, dass unter bestimmten Hypothesen eine komplette Automatisierung nicht sinnvoll, u.U. auch schwer umsetzbar ist. Um den manuellen Aufwand dabei trotzdem überschaubar zu halten, wurde als Alternative die Stichprobentheorie aufgezeigt. Gemeinsam mit der Halbautomatisierung bietet sie die Möglichkeit, trotz manueller Elemente, Hypothesen über eine größere Auswahl an Beiträgen zu prüfen.

Mit den vier äußerst nützlichen Funktionen des vorgestellten Systems: (1) automatisierte Erstellung einer Liste mit Aufgriffkriterien (2) Abfrage digitaler Medien (3) automatisierte Inhaltsanalyse und (4) Stichprobenziehung für die manuelle Inhaltsanalyse stehen den Medienwissenschaften innovative Analysemethoden für digitale Medien zur Verfügung. Der wesentliche Nutzen besteht in der Kontrolle des Aufwands sowie des Fehlers von Aussagen – dank der Möglichkeit, große Beitragsmengen einzubeziehen. Des Weiteren sind Weiterentwicklungen hinsichtlich eines kontinuierlichen Monitorings von medialen Themen unter geringem Aufwand denkbar. Vorteilhaft ist zudem, dass alle

vier Funktionen losgelöst voneinander Verwendung finden können. Die automatisierte Liste von Aufgriffkriterien kann hilfreich für die manuelle Selektion von Beiträgen sein. Außerdem kann es z.B. bei komplett manuellen Inhaltsanalysen effektiv sein, vorab die Menge potentieller Beiträge innerhalb eines Zeitraums zu bestimmen oder relevante Beiträge automatisiert zu selektieren. Und letztlich ermöglicht die Stichprobenziehung das Umdenken bei der Definition der Beitragsgrundlage, z.B. in der Form, dass ein Verbreitungsmedium nicht über drei zusammenhängende Monate hinweg untersucht wird, sondern stichprobenartig 90 Ausgaben innerhalb eines Jahres analysiert werden.

Um das entwickelte System möglichst bequem und kostenfrei einer großen Nutzerzahl zur Verfügung stellen zu können, wurde es mit der Open Source Statistik-Umgebung R entwickelt. Der entstandene Source-Code steht öffentlich auf Github zur Verfügung und kann ohne größere R-Kenntnisse als Paket installiert werden. Dort kann es von jedem Nutzer bei Bedarf angepasst und weiterentwickelt werden. Auch Verbesserungsvorschläge oder Fehlermeldungen der Nutzer werden über das Repository verwaltet und bearbeitet.

Zukünftig bleibt für statistische Indizes wünschenswert, dass sich eine einheitliche Definition über die verschiedenen Forschungsdisziplinen herausbildet und der Bezug zum naturwissenschaftlichen Messprozess abschließend geklärt wird. Die zunehmende Komplexität der Welt wird auch weiterhin Medien benötigen, die eben diese kommunizierbar machen. Mit der steigenden Vernetzung, der stetig wachsenden Menge an Daten und dem Anspruch, Entscheidungen und Argumente mit Zahlen zu begründen, ist immer noch mit einer zunehmenden Zahl an Indizes zu rechnen. Umso wichtiger wird es sein, dass die unmissverständliche Kommunikation über spezifische Maßnahmen gesichert wird. Auch in anderen Belangen hatten die Naturwissenschaften Mühe, Standards zu definieren und einzufordern. So war es z.B. lange Zeit nicht selbstverständlich, dass die Kommunikation eines Gewichts über verschiedene Regionen hinweg exakt und unmissverständlich war. Es bleibt daher die Erwartung, dass objektive Standards in absehbarer Zeit auch die Kommunikation von Indizes unmissverständlich machen.

Quantitative Begriffe waren lange Zeit eine Domäne der Naturwissenschaften. Mit Hilfe der Indizes wird es nun vor allem den Geisteswissenschaften ermöglicht, den Raum auszufüllen, wo Sprache alleine nicht mehr objektiv genug zu sein scheint und bloße Zahlen nicht genügen, um Bedeutung zu generieren. Einige Disziplinen scheinen mit

dem schnellen Einzug der quantitativen Begriffe jedoch nicht Schritt halten zu können. Die Theorie und Adaption der quantitativen Methoden ist noch im Prozess und führt in diesem Zusammenhang z.T. auch zu Produkten, die wissenschaftlich nicht von Dauer sein werden. Letztlich sollte immer die Frage nach der Sinnhaftigkeit einer Quantifikation gestellt werden und Daten oder Indizes nicht nur zum Selbstzweck erhoben werden.

Literaturverzeichnis

Abdallah, Saamah u. a. (2006) *The Happy Planet Index*. Technical Report. London. URL: http://www.neweconomics.org/page/-/files/The_Happy_Planet_Index.pdf (besucht am 3. Dez. 2015).

– . (2012) *The happy planet index: 2012 report*. Technical Report. London. URL: http://www.happyplanetindex.org/assets/happy-planet-index-report.pdf (besucht am 3. Dez. 2015).

Abela, Andrew. (2006) *Choosing a good chart*. URL: http://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html (besucht am 23. Sep. 2015).

Albers, Sonke. (2009) *Methodik der empirischen Forschung*. Wiesbaden: Gabler.

Alberti, Hans-Joachim von. (1957) *Mass und Gewicht: geschichtliche und tabellarische Darstellungen von den Anfängen bis zur Gegenwart*. Berlin: Akad.-Verl.

Allmeling, Anne und Thomas Weber. (8. Juni 2012) "Mehr Monarchie wagen!" In: *Frankfurter Allgemeine Zeitung*, Nr. 131, S. 35.

Alonso, José Antonio und M Teresa Lamata. (2006) "Consistency In The Analytic Hierarchy Process: A New Approach". In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 14.4, S. 445–459.

Anderson, Oskar. (1954) *Probleme der statistischen Methodenlehre in den Sozialwissenschaften*. Würzburg: Physica.

Apel, Detlef u. a. (2010) *Datenqualität erfolgreich steuern: Praxislösungen für Business-intelligence-Projekte*. München u.a.: Hanser.

Baier, Daniel. (2009) *Conjointanalyse: Methoden - Anwendungen - Praxisbeispiele*. Berlin u.a.: Springer.

Bandura, Romana. (2008) *A Survey of Composite Indices Measuring Country Performance: 2008 Update*. Working Paper. New York. URL: http://www.usc.ac.ir/IPPWebV1C035/TemplateFileFolder/10-8-2013/OriginalFolder/0a6e29c4-

1fa6-4496-a362-000f0851d4fc_indices_2008_bandura.pdf (besucht am 23. Apr. 2015).

Bankhofer, Udo. (1995) *Unvollständige Daten- und Distanzmatrizen in der multivariaten Datenanalyse*. Bergisch Gladbach u.a.: Eul.

Bauer, Laurie. (2003) *Introducing linguistic morphology*. Edinburgh: Edinburgh Univ. Press.

Bauer, Raymond Augustine. (1966) *Social indicators*. Cambridge, Mass.: MIT Press.

Baum, Antonia. (2. November 2014) "Kebaya und Liebe". In: *Frankfurter Allgemeine Sonntagszeitung*, Nr. 44, S. V3.

Berkeley, George. (1710/2004) *Eine Abhandlung über die Prinzipien der menschlichen Erkenntnis*. Hamburg: Meiner.

Bernau, Patrick. (11. Mai 2014) "Das Geld der anderen". In: *Frankfurter Allgemeine Sonntagszeitung*, Nr. 19, S. 20.

– . (20. September 2009) "Was die Völker der Welt glücklich macht". In: *Frankfurter Allgemeine Sonntagszeitung*, Nr. 38, S. 45.

– . (30. Oktober 2011) "Das Glück für jeden Typ". In: *Frankfurter Allgemeine Sonntagszeitung*, Nr. 43, S. 48.

Berndt, Frauke und Daniel Fulda. (2012) *Die Sachen der Aufklärung*. Hamburg: Meiner.

Besozzi, Claudio und Helmut Zehnpfennig. (1976) "Methodologische Probleme der Index-Bildung". In: *Techniken der empirischen Sozialforschung* 5, S. 9–55.

Best, Karl-Heinz. (2005) "Satzlänge". In: *Quantitative Linguistik / Quantitative Linguistics. Ein internationales Handbuch / An International Handbook*. Hrsg. von R. Köhler, G. Altmann und R. Piotrowski. Berlin, Boston: De Gruyter Mouton, S. 298–304.

Bleymüller, Josef. (1966) *Theorie und Technik der Aktienkursindizes*. Wiesbaden: Gabler.

Boarini, Romina. (2011) *Designing Your Better Life Index from a methodological perspective*. URL: http://theblogprogress.blogspot.de/2011/05/designing-your-better-life-index-from.html (besucht am 24. Nov. 2015).

Böhme, Gernot. (1974) "Über Kants Unterscheidung von extensiven und intensiven Größen". In: *Kant-Studien* 65, S. 239–258.

Böhme, Gernot. (1976) "Quantifizierung - Metrisierung: Versuch einer Unterscheidung erkenntnistheoretischer und wissenschaftstheoretischer Momente im Prozeß der Bildung von quantitativen Begriffen". In: *Zeitschrift für allgemeine Wissenschaftstheorie* 7.2, S. 209–222.

Bonjean, Charles, Richard Hill und Dale MacLemore. (1967) *Sociological measurement: An inventory of scales and indices*. San Francisco: Chandler.

Bortkiewicz, Ladislaus von. (1923) "Zweck und Struktur einer Preisindexzahl." In: *Nordisk Statistisk Tidsskrift* 2, S. 369–408.

– . (1924) "Buchbesprechung von Irving Fisher, The making of index numbers." In: *Archiv für Sozialwissenschaft und Sozialpolitik* 51.

Bortz, Jürgen. (2006) *Forschungsmethoden und Evaluation*. Berlin u.a.: Springer.

Bourier, Günther. (2013) *Beschreibende Statistik*. Wiesbaden: Springer Gabler.

Brachinger, Hans Wolfgang und Friedmann Ost. (1996) "Modelle mit latenten Variablen: Faktorenanalyse, Latent-Structure-Analyse und LISREL-Analyse". In: *Multivariate statistische Verfahren*. Bd. 2, S. 639–766.

Brosch, Achim. (31. Dezember 1991) "Theoriehunger". In: *Frankfurter Allgemeine Zeitung*, S. N3.

Bürkl, Rolf. (2009) *Methode der Erhebung und Berechnung des GfK-Konsumklimas*. URL: `http://commons.de/wp-content/uploads/GfK_Konsumklimaindex.pdf` (besucht am 3. Juli 2015).

– . (2013) *Methode der Erhebung und Berechnung des GfK Konsumklima Europa*. Nürnberg. URL: `http://www.gfk.com/de/Documents/Pressemitteilungen/2013/MethodenbeschreibungGfKEuroklima_dfin.pdf` (besucht am 3. Juli 2015).

Burzan, Nicole. (2014) "Indikatoren". In: *Handbuch Methoden der empirischen Sozialforschung*. Hrsg. von Nina Baur. Wiesbaden: Springer Fachmedien, S. 1029–1036.

Campbell, N. R. und H. Jeffreys. (1938) "Symposium: Measurement and Its Importance for Philosophy". In: *Vasa* 17, S. 121–151.

Campbell, Norman Robert. (1928) *An account of the principles of measurement and calculation*. London u.a.: Longmans, Green.

Carnap, Rudolf. (1956) "The methodological character of theoretical concepts". In: *Minnesota studies in the philosophy of science*. Hrsg. von Herbert Feigl und Michael Scriven, S. 38–76.

Cherchye, Laurens u. a. (2007) "An Introduction to 'Benefit of the Doubt' Composite Indicators". In: *Social Indicators Research* 82.1, S. 111–145.

Christophersen, Timo und Christian Grape. (2006) "Die Erfassung latenter Konstrukte mit Hilfe formativer und reflektiver Messmodelle". In: *Methodik der empirischen Forschung*. Hrsg. von Sönke Albers. Wiesbaden: Dt. Univ.-Verl., S. 115–132.

Coleman, James Samuel. (1964) *Introduction to mathematical sociology*. New York: Free Press of Glencoe u.a.

Currie, Antony. (2008) *Porsche erwischt Spekulanten auf dem falschen Fuß*. URL: `http://www.handelsblatt.com/meinung/kolumnen/vw-aktie-porsche-erwischt-spekulanten-auf-dem-falschen-fuss/3043772.html` (besucht am 18. Jan. 2016).

Daston, Lorraine. (1991) "The Ideal and Reality of the Republic of Letters in the Enlightenment". In: *Science in Context* 4.2, S. 367–386.

Daston, Lorraine und Peter Galison. (1992) "The Image of Objectivity". In: *Representations* 40, S. 81–128.

Dedekind, Richard. (1893) *Was sind und was sollen die Zahlen?* Braunschweig: Vieweg.

– . (1912) *Stetigkeit und irrationale Zahlen*. Braunschweig: Vieweg.

Dehaene, Stanislas. (1999) *Der Zahlensinn oder warum wir rechnen können*. Basel u.a.: Birkhauser.

Deutsche Börse AG. (2015) *Leitfaden zu den Aktienindizes der Deutschen Börse*. Technischer Bericht. URL: `http://dax-indices.com/DE/MediaLibrary/Document/Leitfaden_Aktienindizes.pdf` (besucht am 21. Jan. 2016).

Ebbinghaus, Heinz-Dieter. (1992) *Zahlen*. Berlin u.a.: Springer.

Eknoyan, Garabed. (2008) "Adolphe Quetelet (1796-1874) - the average man and indices of obesity". In: *Nephrology Dialysis Transplantation* 23.1, S. 47–51.

Erbrecht, Rüdiger u. a. (1999) *Das große Tafelwerk*. Berlin: Volk und Wissen Verlag.

Euclides. (3. Jh. v. Chr./2005) *Die Elemente: Bücher I - XIII*. Frankfurt am Main: Deutsch.

Eulitz, Christa-Maria, Sigrid Scheuermann und Hans-Joachim Thier. (1968) *Brockhaus abc Naturwissenschaft und Technik. 1. A - K*. Leipzig: Brockhaus.

Eurostat. (2011) *European Statistics Code of Practice*. URL: `http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15` (besucht am 17. Juli 2015).

Fahrmeir, Ludwig. (1997) *Statistik: der Weg zur Datenanalyse*. Berlin u.a.: Springer.

Finanz-Lexikon. (2015) *ifo Geschäftsklimaindex*. URL: `https://www.finanz-lexikon.de/ifogeschaeftsklimaindex_745.html` (besucht am 5. Nov. 2015).

Fisher, Irving. (1923) *The making of index numbers: a study of their varieties, tests, and reliability*. Boston u.a.: Houghton Mifflin.

Flaskämper, Paul. (1928) *Theorie der Indexzahlen: Beiträge zur Logik des statistischen Vergleichs*. Berlin u.a.: de Gruyter.

Frankfurter Allgemeine Archiv. (2015) *Frankfurter Allgemeine Archiv*. URL: `http://faz-archiv-approved.faz.net/intranet/biblionet/` (besucht am 30. Okt. 2015).

Frege, Gottlob. (1884/1990) *Die Grundlagen der Arithmetik*. Hildesheim u.a.: Olms.

Frisch, Ragnar. (1936) "Annual Survey of General Economic Theory: The Problem of Index Numbers." In: *Econometrica* 4.1, S. 1–38.

Früh, Werner, Carsten Wünsch und Pascal Klopp. (2004) "TDU-Unterhaltungsindex. Ein Instrument zur empirischen Ermittlung von Unterhaltungserleben". In: *Medien & Kommunikationswissenschaft* 52.4, S. 515–544.

Gläser, Jochen. (2006) *Wissenschaftliche Produktionsgemeinschaften: Die soziale Ordnung der Forschung*. Frankfurt u.a.: Campus.

Goffman, Erving. (1982) *Das Individuum im öffentlichen Austausch*. Frankfurt am Main: Suhrkamp.

Google. (o.J.) *Display live data on your site*. URL: `https://developers.google.com/chart/` (besucht am 30. Sep. 2015).

Grossarth, Jan. (5. Mai 2012) "Die Berechnung des Glücks". In: *Frankfurter Allgemeine Zeitung*, Nr. 82, S. 12.

Günther, Gotthard. (1968) "Kritische Bemerkungen zur gegenwärtigen Wissenschaftstheorie: Aus Anlass von Jürgen Habermas: Zur Logik der Sozialwissenschaften". In: *Soziale Welt* 19.3, S. 328.

Gustafsson, Anders. (2007) *Conjoint measurement: methods and applications*. Berlin u.a.: Springer.

Haberler, Gottfried. (1927) *Der Sinn der Indexzahlen: eine Untersuchung über den Begriff des Preisniveaus und die Methoden seiner Messung*. Tübingen: Mohr.

Härdle, Wolfgang. (2007) *Multivariate Statistics: exercises and solutions*. New York, NY: Springer.

Hartig, Johannes, Andreas Frey und Nina Jude. (2008) "Validität". In: *Testtheorie und Fragebogenkonstruktion*. Hrsg. von Helfried Moosbrugger und Augustin Kelava. Berlin u.a.: Springer, S. 135–163.

Haustein, Heinz-Dieter. (2001) *Weltchronik des Messens: Universalgeschichte von Maß und Zahl, Geld und Gewicht*. Berlin u.a.: de Gruyter.

Hein, Christoph. (13. Oktober 2007) "Das große Kommunisten-Treffen". In: *Frankfurter Allgemeine Zeitung*, Nr. 238, S. 14.

Heintz, Bettina. (2000) *Die Innenwelt der Mathematik: zur Kultur und Praxis einer beweisenden Disziplin*. Wien u.a.: Springer.

– . (2007) "Zahlen, Wissen, Objektivität: Wissenschaftssoziologische Perspektiven". In: *Zahlenwerk*. Hrsg. von Andrea Mennicken und Hendrik Vollmer. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 65–85.

Helmholtz, Hermann v. (1879/1959) *Die Tatsachen in der Wahrnehmung. Zählen und Messen, erkenntnistheoretisch betrachtet.* Darmstadt: Wiss. Buchges.

Hermberg, Paul. (1920) "Die Indexzahlen der Frankfurter Zeitung". In: *Weltwirtschaftliches Archiv* 16, S. 537–541.

– . (1930) "Statistik und Teuerung: Zur Frage der Messung der Tauschwertschwankungen des Geldes." In: *Weltwirtschaftliches Archiv* 31, S. 424–444.

Heß, Hans-Ulrich. (2010) *Das Wunder der Anwendung: Einführung in die Analysis und ihre Anwendung in Naturwissenschaft und Technik*. Berlin: Logos.

Holland, Heinrich und Kurt Scharnbacher. (2006) *Grundlagen der Statistik*. Wiesbaden: Gabler.

Ifo Institut. (2011) *ifo Geschäftsklimaindex ab Mai 2011 neu gewichtet*. URL: `http://www.cesifo-group.de/de/ifoHome/presse/Pressemitteilungen/Pressemitteilungen-Archiv/2011/Q2/press_15790620.html` (besucht am 5. Nov. 2015).

– . (2015) *Wie wird das ifo Geschäftsklima ermittelt?* URL: `https://www.cesifo-group.de/de/ifoHome/facts/Survey-Results/Business-Climate/Calculating-the-Ifo-Business-Climate.html` (besucht am 5. Nov. 2015).

International Monetary Fund. (2003) *Data Quality Assessment Framework and Data Quality Program*. URL: `http://www.imf.org/external/np/sta/dsbb/2003/eng/dqaf.htm` (besucht am 17. Juli 2015).

Jacobs, Rowena, Peter Smith und Maria Goddard. (2004) *Measuring performance: an examination of composite performance indicators: a report for the Department of Health*. Technical Paper. York. URL: `https://www.york.ac.uk/che/pdf/tp29.pdf` (besucht am 16. Okt. 2015).

Jambu, Michel. (1992) *Explorative Datenanalyse*. Stuttgart u.a.: G. Fischer.

Janßen, Birgit und Bernd Rudolph. (1992) *Der Deutsche Aktienindex DAX*. Frankfurt am Main: Knapp.

Javers, Eamon. (2013) *Thomson Reuters Gives Elite Traders Early Advantage*. URL: `http://www.cnbc.com/id/100809395` (besucht am 7. Jan. 2016).

Jones, T Colwyn und David Dugdale. (2001) "The concept of an accounting regime". In: *Critical Perspectives on Accounting* 12.1, S. 35–63.

Kanning, Tim. (11. Mai 2013) "Reto Francioni". In: *Frankfurter Allgemeine Zeitung*, Nr. 108, S. 42.

Kellerer, Hans. (1960) *Statistik im modernen Wirtschafts- und Sozialleben*. Hamburg: Rowohlt.

Kessler, Waltraud. (2007) *Multivariate Datenanalyse: für die Pharma-, Bio- und Prozessanalytik; ein Lehrbuch*. Weinheim: Wiley-VCH.

Keys, Ancel u. a. (1972) "Indices of relative weight and obesity". In: *Journal of chronic diseases* 25.6, S. 329–343.

Khalid, Malik. (2013) *Human Development Report: Technical notes*. New York. URL: `http://hdr.undp.org/sites/default/files/hdr_2013_en_technotes.pdf` (besucht am 3. Dez. 2015).

Kleist v., Heinrich. (1799/1997) "Heinrich von Kleist an Ulrike von Kleist am 12.11.1799". In: *Sämtliche Werke und Briefe: in vier Bänden. Bd. 4. Briefe von und an Heinrich von Kleist: 1793 - 1811*. Hrsg. von Klaus Müller-Salget und Ilse-Marie Barth. Frankfurt am Main: Dt. Klassiker-Verl.

Klette, Kathrin. (2. Februar 2011) "Niedrige Löhne als Lockstoff". In: *Frankfurter Allgemeine Zeitung*, Nr. 36, S. 48.

Klöpfer, Inge. (17. März 2002) "Der Kölner Müll stinkt überall". In: *Frankfurter Allgemeine Sonntagszeitung*, Nr. 11, S. 37.

Köhler, Reinhard. (2005) "Gegenstand und Arbeitsweise der Quantitativen Linguistik". In: *Quantitative Linguistik: ein internationales Handbuch*. Hrsg. von Rajmund G. Piotrowski. Berlin u.a.: de Gruyter, S. 1–16.

Köhler, Reinhard, Gabriel Altmann und Rajmund G. Piotrowski. (2005) *Quantitative Linguistik. Ein internationales Handbuch*. Berlin u.a.: de Gruyter.

Köhn, Rüdiger. (10. Juni 2011) "Siemens setzt auf den Börsengang von Osram". In: *Frankfurter Allgemeine Zeitung*, Nr. 134, S. 12.

– . (20. Februar 2013) "MAN-Finanzvorstand Lutz flieht vor dem VW-Vertrag". In: *Frankfurter Allgemeine Zeitung*, Nr. 43, S. 14.

Kolb, Steffen. (2004) "Verlässlichkeit von Inhaltsanalysedaten. Reliabilitätstest, Errechnen und Interpretieren von Reliabilitätskoeffizienten für mehr als zwei Codierer". In: *M&K, 2004*, S. 335–354.

Körte, Peter. (25. Februar 2007) "Der mündige Bürger". In: *Frankfurter Allgemeine Sonntagszeitung*, Nr. 8, S. 29.

Krantz, David, Duncan Luce und Patrick Suppes. (1989) *Foundations of measurement*. San Diego, Calif. u.a.: Acad. Pr.

Krickeberg, Walter. (1914) "Einige Neuerwerbungen der nordamerikanischen Sammlung des Königl. Museums für Völkerkunde". In: *Zeitschrift für Ethnologie* 46.4/5, S. 670–745.

Krippendorff, Klaus. (2004) "Reliability in Content Analysis". In: *Human Communication Research* 30.3, S. 411–433.

– . (2013) *Content analysis: an introduction to its methodology*. Los Angeles u.a.: Sage.

Krystek, Ulrich und Günter Müller-Stewens. (1993) *Frühaufklärung für Unternehmen: Identifikation und Handhabung zukünftiger Chancen und Bedrohungen*. Stuttgart: Schäffer-Poeschel.

Kunkel, André. (2003) "Zur Prognosefähigkeit des Ifo-Geschäftsklimas und seiner Komponenten sowie die Überprüfung der 'Dreimal-Regel'". In: *ifo Diskussionsbeiträge* 80, S. 1–30.

Lambsdorff, Johann. (2007) *The institutional economics of corruption and reform: theory, evidence, and policy*. Cambridge u.a.: Cambridge Univ. Press.

Lauf, Edmund. (2001) "Zur Reliabilitat von Inhaltsanalysen und deren Darstellung in kommunikationswissenschaftlichen Fachzeitschriften." In: *Publizistik* 46.1, S. 57.

Lazarsfeld, Paul Felix. (1959) "Problems in methodology". In: *Sociology today: Problems and prospects*. Hrsg. von Robert King Merton u. a. New York: Basic books, S. 38–78.

Lazarsfeld, Paul Felix und Morris Rosenberg. (1956) "The language of social research; a reader in the methodology of social research". In: *Journal of Philosophy* 53.7, S. 248–255.

Leibniz, Gottfried W. (1696/1996) *Philosophische Werke/Hauptschriften zur Grundlegung der Philosophie 3*. Hamburg: Meiner.

Lippe, Pete von der und Andreas Kladroba. (2004) "Messung komplexer Variablen als Summe von Punktzahlen - Eine beliebte Methode des measurement without theory Measuring Complex Variables by Sums of Scores - A Popular and Questionable

Method of Measurement Without Theory". In: *Jahrbücher für Nationalökonomie und Statistik* 1-2, S. 1–115.

Lisch, Ralf. (1978) *Grundlagen und Modelle der Inhaltsanalyse: Bestandsaufnahme und Kritik*. Reinbek bei Hamburg: Rowohlt.

Luhmann, Niklas. (1981) "Die Unwahrscheinlichkeit der Kommunikation". In: *Soziologische Aufklärung. 3. Soziales System, Gesellschaft, Organisation"*. Hrsg. von Niklas Luhmann. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 25–34.

– . (1992) *Die Wissenschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.

– . (1997) *Die Gesellschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.

Lütters, Holger und Jörg Staudacher. (2008) "Wirksame Unterstützung: Strategische Kontrolle mit dem Analytic Hierarchy Process". In: *Marketing Review St. Gallen* 25.2, S. 44–49.

Max Planck Institute for Informatics. (2014) *YAGO: A High-Quality Knowledge Base*. URL: `http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/` (besucht am 29. Okt. 2015).

Mayntz, Renate. (1978) *Einführung in die Methoden der empirischen Soziologie*. Opladen: Westdt. Verl.

o. V. (9. Oktober 2011) "Was wissen Sie über Wirtschaft?" In: *Frankfurter Allgemeine Sonntagszeitung*, Nr. 40, S. 28.

re.. (22. März 2001) "Renten prasentieren sich fester". In: *Frankfurter Allgemeine Sonntagszeitung*, Nr. 69, S. 42.

Mennen, Ann-Kristin. (2013) *Gesund für einen Tag*. URL: `http://www.spiegel.de/karriere/berufsleben/lehrer-mit-uebergewicht-verbeamtung-undurchsichtig-a-879074.html` (besucht am 9. Dez. 2015).

Mohr, Daniel. (26. Juli 2013) "Zuversicht in Amerika". In: *Frankfurter Allgemeine Zeitung*, Nr. 171, S. 21.

– . (28. Juni 2013) "Ein Quantensprung für die Börse". In: *Frankfurter Allgemeine Zeitung*, Nr. 147, S. 24.

Molenberghs, Geert u. a. (2014) "Missing Data". In: *Handbook of Epidemiology SE - 20*. Hrsg. von Wolfgang Ahrens und Iris Pigeot. New York: Springer, S. 1283–1335.

Morgenstern, Oskar. (1956) "Experiment und Berechnung großen Umfangs in der Wirtschaftswissenschaft". In: *Weltwirtschaftliches Archiv* 76, S. 179–239.

Nardo, M. u. a. (2005) *Tools for composite indicators building*. Technical Report. Ispra. URL: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC31473/EUR21682EN.pdf (besucht am 11. Mai 2016).

– . (2008) *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Technical Report. Paris. URL: http://www.oecd.org/std/42495745.pdf (besucht am 11. Feb. 2016).

Neuendorf, Kimberly A. (2002) *The content analysis guidebook*. Thousand Oaks u.a.: Sage.

Neugebauer, Otto. (1935) *Mathematische Keilschrift-Texte*. Berlin: Springer.

Odekon, Mehmet. (2015) *Booms and Busts: An Encyclopedia of Economic History from the First Stock Market Crash of 1792 to the Current Global Economic Crisis*. London: Routledge.

OECD (Organisation for Economic Co-operation and Development). (2011) *How's Life?* Technical Report. URL: http://dx.doi.org/10.1787/9789264121164-en (besucht am 18. Nov. 2015).

– . (2013a) *Excecutive Summary*. URL: http://www.oecdbetterlifeindex.org/de/media/bli/documents/Excecutive-Summary_DE.pdf (besucht am 18. Nov. 2015).

– . (2013b) *Measuring well-being and progress*. Working Paper. Paris. URL: http://www.oecd.org/std/MeasuringWell-BeingandProgressBrochure.pdf (besucht am 18. Nov. 2015).

– . (2014) *Better Life Index 2014: Definitions And Metadata*. URL: http://www.oecd.org/statistics/OECD-Better-Life-Index-2014-definitions.pdf (besucht am 18. Nov. 2015).

– . (2015a) *OECD Better Life Index*. URL: http://www.oecdbetterlifeindex.org (besucht am 30. Sep. 2015).

– . (2015b) *Was ist den Menschen weltweit am wichtigsten?* URL: http://www.oecdbetterlifeindex.org/de/antworten/ (besucht am 18. Nov. 2015).

– . (2015c) *Worum handelt es sich beim Better Life Index?* URL: http://www.oecdbetterlifeindex.org/de/about/better-life-initiative/ (besucht am 18. Nov. 2015).

O.V. (2014) *google-motion-charts-with-r*. URL: http://code.google.com/p/google-motion-charts-with-r/ (besucht am 29. Sep. 2015).

Panagiotis, Tsarchopoulos. (2007) *European Innovation Scoreboard 2006*. URL: `http://www.urenio.org/2007/02/23/european-innovation-scoreboard-2006/` (besucht am 17. Juli 2015).

Passoth, Jan-Hendrik und Josef Wehner. (2012) *Quoten, Kurven und Profile: zur Vermessung der sozialen Welt*. Wiesbaden: Springer.

Petersen, Mitchell A. (2004) *Information: Hard and soft*. Working Paper. Evanston. URL: `http://www.kellogg.northwestern.edu/faculty/petersen/htm/papers/softhard.pdf` (besucht am 16. Okt. 2015).

Pfanzagl, Johann. (1966) *Allgemeine Methodenlehre der Statistik II*. Berlin: de Gruyter.

– . (1971) *Theory of measurement*. Würzburg u.a.: Physica-Verl.

Pickel, Gert. (2013) *Indizes in der Vergleichenden Politikwissenschaft*. Wiesbaden: Springer VS.

Pickel, Susanne und Gert Pickel. (2012) "Die Messung von Indizes in der Vergleichenden Politikwissenschaft - methodologische Spitzfindigkeit oder substantielle Notwendigkeit". In: *Zeitschrift für Vergleichende Politikwissenschaft* 6, S. 1–17.

Plickert, Philip. (23. Dezember 2008) "Deutsche Wirtschaft steht vor einem langem Winter". In: *Frankfurter Allgemeine Sonntagszeitung*, Nr. 300, S. 10.

– . (7. Juli 2012) "BIP, Wachstum und Glück". In: *Frankfurter Allgemeine Zeitung*, Nr. 83, S. 11.

Polzin, Brigitte. (2007) "Die Reduktion von Daten: ein Vergleich datenreduzierender Verfahren auf ihre Leistungsfähigkeit im sozialwissenschaftlich-statistischen Umfeld am Beispiel einer Untersuchung zu den Faktoren des subjektiven Sicherheitsgefühls von Frauen und Männern". Unveröffentlichte Dissertation. Ruhr-Universität Bochum.

Porter, Theodore M. (1992) "Quantification and the Accounting Ideal in Science". In: *Social Studies of Science* 22.4, S. 633–651.

– . (1995) *Trust in numbers*. Princeton: Princeton University Press.

Power, Michael. (2004) "Counting, control and calculation: Reflections on measuring and management". In: *Human relations* 57.6, S. 765–783.

Prün, Claudia und Petra Steiner. (2005) "Quantitative Morphologie: Eigenschaften der morphologischen Einheiten und Systeme". In: *Quantitative Linguistik: ein internationales Handbuch*. Hrsg. von Rajmund G. Piotrowski. Berlin u.a.: de Gruyter, S. 227–243.

Rahlf, Thomas. (2014) *Datendesign mit R*. München: Open Source Press.

Ralph, Jeff. (2015) *A practical introduction to index numbers*. Chichester u.a.: Wiley.

Randolph, Rainer. (1979) *Pragmatische Theorie der Indikatoren - Grundlagen einer methodischen Neuorientierung*. Göttingen: Vandenhoeck & Ruprecht.

Raupp, Juliana und Jens Vogelgesang. (2009) *Medienresonanzanalyse eine Einführung in Theorie und Praxis*. Wiesbaden: VS, Verl. für Sozialwiss.

Rees, Andreas. (2008) "Die US-Wirtschaft vor der Rezession - was sollte Europa tun?" In: *ifo Schnelldienst* 7.

Reuters. (27. September 2003) "Wall Street schließt nach uneinheitlichen Konjunkturdaten ein wenig leichter". In: *Frankfurter Allgemeine Zeitung*, Nr. 225, S. 25.

Rhk. (5. April 2002) "Übergewicht erhöht Risiko bei Unfällen". In: *Frankfurter Allgemeine Zeitung*, Nr. 79, S. 10.

Ribecca, Severino. (2015) *The Data Visualisation Catalogue*. URL: `http://www.datavizcatalogue.com/search.html` (besucht am 23. Sep. 2015).

Ritter, Johannes. (13. Dezember 2012) "Das andere Leben des Hartmut Ostrowski". In: *Frankfurter Allgemeine Zeitung*, Nr. 291, S. 16.

Rödder, Simone. (19. Januar 2003) "Neue Waffe gegen Volkskrankheit Diabetes". In: *Frankfurter Allgemeine Sonntagszeitung*, Nr. 3, S. 10.

Ross-Larson, Bruce und Eileen Hanlon. (1990) *Human Development Report*. New York u.a.: Oxford University Press.

Rössler, Patrick. (2010) *Inhaltsanalyse*. Konstanz: UVK-Verl.-Ges.

Rothe, Ursula. (1988) "Polylexy and compounding". In: *Glottometrika 9*. Hrsg. von Klaus-Peter. Bochum: Brockmeyer, S. 121–134.

Rubin, Donald B. (1976) "Inference and Missing Data". In: *Biometrika* 63.3, S. 581–592.

Rüther, Tobias. (13. August 2006) "Jede ist eine Trauminsel". In: *Frankfurter Allgemeine Zeitung*, Nr. 32, S. V3.

Saint-Mont, Uwe. (2011) "Messtheorie". In: *Statistik im Forschungsprozess*. Hrsg. von Uwe Saint-Mont. Heidelberg: Physica-Verlag HD, S. 23–76.

Saisana, Michaela und Andrea Saltelli. (2012) *Corruption Perceptions Index 2012 Statistical Assessment*. Technical Report. Luxemburg. URL: `http://publications.jrc.ec.europa.eu/repository/handle/JRC77239` (besucht am 11. Jan. 2012).

Saisana, Michaela, Andrea Saltelli und Stefano Tarantola. (2005) "Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168.2, S. 307–323.

Saisana, Michaela und Stefano Tarantola. (2002) *State-of-the-art report on current methodologies and practices for composite indicator development*. Technical Report. Ispra. URL: `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.402.5612&rep=rep1&type=pdf` (besucht am 19. Juli 2015).

Schäfer, Thomas. (2010) "Explorative Datenanalyse: Muster und Zusammenhänge erkennen". In: *Statistik I*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 99–128.

Schlaudt, Oliver. (2009) *Die Quantifizierung der Natur*. Paderborn: mentis.

Schlittgen, Rainer. (2009) *Multivariate Statistik*. München: Oldenbourg.

Schnell, Rainer. (2005) *Methoden der empirischen Sozialforschung*. München u.a.: Oldenbourg.

Schramm, Helmar, Ludger Schwarte und Jan Lazardzig. (2006) *Spektakuläre Experimente: Praktiken der Evidenzproduktion im 17. Jahrhundert*. Berlin, New York: Walter De Gruyter.

Schröder, K. (1997) *Konzeption eines planungsunterstützenden Informationssystems für das Dienstleistungsunternehmen Kommune*. Berlin: Duncker & Humblot.

Schwabish, Jon. (2014) *The Graphic Continuum*. URL: `http://blog.visual.ly/graphic-continuum` (besucht am 23. Sep. 2015).

Schwiesau, Dietz. (2016) *Nachrichten - klassisch und multimedial: ein Handbuch für Ausbildung und Praxis*. Wiesbaden: Springer VS.

Seipelt, Chantal. (2010) "Konstruktion eines Index zur Bewertung der Dienstleistungsqualität in einem Call Center". Diplomarbeit. Freie Universität Berlin.

Seligmann, Siegfried. (1910) *Der böse Blick und Verwandtes: ein Beitrag zur Geschichte des Aberglaubens aller Zeiten und Völker. 1*. Berlin: Barsdorf.

Senat der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz. (2006) *Stellungnahme zum ifo Institut für Wirtschaftsforschung an der Universität München (ifo)*. Evaluationsbericht. URL: `http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Evaluierung/Senatsstellungnahmen/Senatsstellungnahme-ifo-2006.pdf` (besucht am 13. Nov. 2015).

Sherman, L.A. (1888) "Some observation upon the sentence-length in English prose". In: *University of Nebraska Studies* 1, S. 119–130.

Statistisches Bundesamt. (2008) *Statistisches Bundesamt 2008 Klassifikation der Wirtschaftszweige*. URL: `https://www.destatis.de/DE/Publikationen/Verzeichnis/KlassifikationWZ08_3100100089004.pdf?__blob=publicationFile` (besucht am 30. Sep. 2015).

Stegmüller, Wolfgang. (1970) *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band II: Theorie und Erfahrung*. Heidelberg: Springer.

Stevens, Stanley. (1946) "On the Theory of Scales of Measurement". In: *Science* 103.2684, S. 677–680.

– . (1959) "Measurement, psychophysics, and utility". In: *Measurement: Definitions and theories*, S. 18–63.

Stichweh, Rudolf. (1984) *Zur Entstehung des modernen Systems wissenschaftlicher Disziplinen: Physik in Deutschland 1740 - 1890*. Frankfurt a.M.: Suhrkamp.

Structured Solutions AG. (2011) *Leitfaden zur F.A.Z. Europa Indexfamilie*. Technischer Bericht. Frankfurt am Main. URL: `http://www.solactive.com/wp-content/uploads/2012/06/DE000SLA8EB6-FAZ_Europa_Indexleitfaden_1_2.pdf` (besucht am 18. Nov. 2015).

Suppes, Patrick und Joseph Zinnes. (1962) "Basic measurement theory". In: *Handbook of Mathematical Psychology* 1, S. 3–76.

Tarantola, Stefano u. a. (2006) "A new estimator for sensitivity analysis of model output: An application to the e-business readiness composite indicator". In: *Reliability Engineering & System Safety* 91.10-11, S. 1135–1141.

TI (Transparency International). (2014) *Corruption Perceptions Index 2014: Full Source Description*. General Report. URL: `http://files.transparency.org/content/download/1857/12438/file/CPI2014_DataBundle.zip` (besucht am 11. Jan. 2014).

Torgerson, Warren S. (1958) *Theory and methods of scaling*. New York u.a.: Wiley.

Toutenburg, Helge und Christian Heumann. (2008) *Deskriptive Statistik*. Berlin u.a.: Springer.

Trapp, Wolfgang. (1998) *Kleines Handbuch der Maße, Zahlen, Gewichte und der Zeitrechnung*. Stuttgart: Reclam.

Tukey, John Wilder. (1977) *Exploratory data analysis*. Reading, Mass. u.a.: Addison-Wesley.

Turner, Lyle. (2009) *Johann Lambsdorff Retires the Corruption Perceptions Index*. URL: `https://www.globalintegrity.org/2009/09/post-452/` (besucht am 11. Jan. 2016).

UMich (University of Michigan). (2014) *Index Calculations*. Technical Report. URL: `https://data.sca.isr.umich.edu/fetchdoc.php?docid=24770` (besucht am 11. Jan. 2014).

UMich (University of Michigan). (2015) *Papers and Miscellaneous Items*. URL: https://data.sca.isr.umich.edu/papers.php (besucht am 7. Jan. 2016).

– . (2016) *Survey Information*. URL: https://data.sca.isr.umich.edu/survey-info.php (besucht am 6. Jan. 2016).

UNDP (United Nations Development Programme). (2015) *Why is geometric mean better suited for the HDI than the arithmetic mean?* URL: http://hdr.undp.org/en/content/why-geometric-mean-better-suited-hdi-arithmetic-mean (besucht am 2. Dez. 2015).

Vollmer, Hendrik. (2003) "Bookkeeping, accounting, calculative practice: the sociological suspense of calculation". In: *Critical Perspectives on Accounting* 14.3, S. 353–381.

Wagenführ, Rolf. (1967) *Statistik leicht gemacht: Einführung in die deskriptive Statistik*. Köln: Bund-Verl.

Walsh, C. M. (1921) "The Best Form of Index Number: Discussion." In: *Quarterly Publications of the American Statistical Association* 17.133, S. 537–544.

Watkins, Kevin. (2007) *Human Development Report*. New York u.a.: Palgrave Macmillan.

Weber, Erna. (1974) *Einführung in die Faktorenanalyse*. Jena: Fischer.

Weigel, P. (1921a) "Indexziffern". In: *Zeitschrift für die gesamte Staatswissenschaft / Journal of Institutional and Theoretical Economics* 62 (117).2, S. 128–144.

– . (1921b) "Indexziffern im Inland und im Ausland". In: *Zeitschrift für die gesamte Staatswissenschaft / Journal of Institutional and Theoretical Economics* 76.4, S. 450–477.

Werner, Rudolf. (1975a) *Soziale Indikatoren und politische Planung: Einführung in Anwendungen der Makrosoziologie*. Reinbek bei Hamburg: Rowohlt.

Werner, Ulrich. (1975b) "Bemerkungen zum Problem der Index-Konstruktion in der Soziologie". Forschungsbericht Nr. 89. Wien: Institut für Höhere Studien, S. 1–30.

WHO (World Health Organization). (2015) *Body mass index - BMI*. URL: http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi (besucht am 9. Dez. 2015).

Wikipedia. (o.J.) *Happy Planet Index*. URL: https://de.wikipedia.org/wiki/Happy_Planet_Index (besucht am 2. Juli 2015).

Wilke, Jürgen. (2009) "Presse". In: *Das Fischer Lexikon Publizistik Massenkommunikation*. Hrsg. von Elisabeth Noelle-Neumann, Winfried Schulz und Jürgen Wilke. Frankfurt am Main: Fischer.

Witthöft, Harald. (1998) *Georgius Agricola (1494-1555) über Maß und Gewicht - in der Antike und in seiner Zeit*. Forschungsbericht. Chemnitz. URL: `https://www-user.tu-chemnitz.de/{~}fna/03witthoeft.pdf` (besucht am 10. Mai 2013).

Wöllenstein, Roswitha. (2008) *Rezeption der Umweltproblematik in der Betriebswirtschaftslehre: Eine empirische Rekonstruktion und strukturationstheoretische Analyse der ökologieorientierten Forschung in der Betriebswirtschaftslehre*. Mering: Rainer Hampp Verlag.

Wussing, Hans, Heinz-Wilhelm Alten und Heiko Wesemüller-Kock. (2008) *6000 Jahre Mathematik*. Berlin, Heidelberg: Springer.

Zeisel, Hans. (1970) *Die Sprache der Zahlen. Say it with figures*. Köln u.a.: Kiepenheuer & Witsch.

Anhang A

Tabellen

Tabelle A.1: Einträge in der deutschen Wikipedia zum Konzept oder Label „Index“

Name	Anzahl Wörter
(a,b)-Baum	2186
A-Square Shock Power Index	7236
ÖkoDAX	8678
Änderungsindex	1337
Índice de Precios y Cotizaciones	12315
Aarne-Thompson-Index	20653
ABX.HE	5719
AEX-Index	18258
Affenindex	7578
Afriso-Euro-Index	5024
Aktienindex	7315
Aktienindex des Statistischen Reichsamtes	35239
Aktienindex für erneuerbare Energien	6976
All Ordinaries	13055
Aminosäureindex	1142
AMX index	3069
Analytical Profile Index	1466
Apnoe-Hypopnoe-Index	4105
Arbeitsbewältigungsindex	6317

Tabelle A.1

Arbeitsklima-Index für Deutschland	5419
Arbeitsklima Index	4444
Area Mass Index	3040
Aromaindex Apfel	2749
Arts and Humanities Citation Index	829
Athex Composite Share Price Index	15492
Austrian Traded Index	17413
Autoindex	2587
Average Directional Movement Index	3041
B-Baum	21116
B*-Baum	6070
B+-Baum	8310
Börsenindex	1423
Balaban-J-Index	4269
Baltic Clean Tanker Index	6133
Baltic Dirty Tanker Index	5787
Baltic Dry Index	46599
Barthel-Index	3763
Baupreisindex	2375
Beißindex	2467
BEL20	10058
BELEX15	2087
BERI-Index	4294
Berry-Index	1408
Bertelsmann Transformation Index	24907
BET 10	3207
Bibliotheksindex	3866
Big-Mac-Index	21018
Bildindex der Kunst und Architektur	1103
Bitmap-Index	2728
Blattflächenindex	3058
Bloomberg Commodity Index	9691

Tabelle A.1

BODE-Index	2567
Body-Mass-Index	24774
Bonitätsindex	2089
Bovespa-Index	15321
Boyden-Index	1224
Brechungsindex	34082
Breidbart-Index	8596
Bribe Payers Index	810
BSE Sensex	19566
CAC 40	20916
CAC Next 20	3104
CAC Small	15458
Carbon Preference Index	1232
Case-Shiller-Index	18233
Case Mix Index	1342
CBOE Volatility Index	5274
CDAX	1903
Cdindex	57
CECEEUR	1574
CEETX	6200
Cetanindex	2706
Champernowne-Index	4205
China Traded Index	1471
Chlorose-Index	2130
Colour-Index	8125
Consumer Confidence Index	17685
Continuous Commodity Index	12909
Cook Partisan Voting Index	1755
Cost Index	3296
Credit Default Swap Index	2560
CROBEX	2858
CSI 300	3365

Tabelle A.1

Current Index to Statistics	783
Customer-Loyalty-Index	501
Czech Traded Index	2575
Dalton-Index	6891
Datenbankindex	4703
DAX	51538
DAXglobal Asia Index	1512
DBIX India Index	3717
Demokratieindex	59785
Derwent World Patent Index	2410
Deutscher Immobilienaktienindex	1297
Deutscher Rentenindex	16962
DGB-Index Gute Arbeit	5502
Digitaler Portraitindex	3396
Dispersionsindex	1957
DivDAX	4956
Diversitätsindex	621
Dow Jones Composite Average	11875
Dow Jones Global Titans	6500
Dow Jones Industrial Average	79229
Dow Jones Sustainability Index	12693
Dow Jones Transportation Average	21888
Dow Jones Utility Average	17214
Dynkin-Index	1236
Dyspnoe-Index	1529
Earth Similarity Index	5193
Economic Diversification Index	602
Economic Vulnerability Index	728
EG-Index-Nummer	3911
EGX 30 Index	11791
Einkaufsmanagerindex	1819
Elendsindex	1788

Tabelle A.1

Emergency Severity Index	7916
Empire State Index	1310
Energie-Mass-Index	7059
Energy Efficiency Design Index	3589
Entry Standard Index	5629
Environment for DeveLoping KDD-Applications	
Supported by Index-Structures	10351
Environmental Performance Index	10505
EPIX	2215
EPRA-Index	1064
Euro Currency Index	11434
Euro Effective Exchange Rate Index	12972
EURO STOXX 50	13200
FAO Food Price Index	10109
Farbindex	2204
Farbindex (Geologie)	1972
Farbwiedergabeindex	11141
FAZ-Index	2884
FHFA House Price Index	9511
Freier Androgenindex	3338
Freiheitsindex	11828
Fruchtwasserindex	2363
FT 30 Index	2999
FTSE 100 Index	26346
FTSE 250 Index	16256
FTSE 350 Index	3059
FTSE Bursa Malaysia KLCI	13947
FTSE Italia Mid Cap	3179
FTSE MIB	19994
FTSE/Athex Large Cap	5348
FTSE/JSE All-Share Index	9434
Functional Capacity Index	9582

Tabelle A.1

Gaius Iulius Vindex	5028
Gallagher-Index	8885
Gender-related Development Index	1354
German Entrepreneurial Index	2132
Geschäftsklimaindex	387
Geschwindigkeitsindex	6388
Global Challenges Index	3001
Global Competitiveness Index	8024
Global Peace Index	26769
Glykämischer Index	13577
Grasland-Feuerindex	926
GRAU-Index	36894
Gridfile	2891
Grubel-Lloyd-Index	20219
Guadagni-Little-Index	1729
H-Index	13038
Handflügelindex	2466
Hang Seng China Enterprises Index	2144
Hang Seng Index	28612
Happy Planet Index	6720
Hardgrove-Index	2500
Harmonisierter Verbraucherpreisindex	2843
HARPEX	16009
Hashtabelle	19624
HASPAX	3062
HDAX	1104
Headcount-Index	1168
Herfindahl-Index	5297
Herzindex	1550
Hitzeindex	15998
Homogenitätsindex	2688
Howe Robinson Container Index	16780

Tabelle A.1

Huglin-Index	5921
Human Assets Index	519
Human Development Index	43958
Human Poverty Index	1857
Human Sustainable Development Index	19535
Hungarian Traded Index	2135
Hunger-Index für Indien	4878
HWWI-Rohstoffpreisindex	1353
IBEX 35	20494
IBoxx	1439
IBrX-50	6244
ICOMP Index	8243
IDX Composite	9552
Ifo-Geschäftsklimaindex	7938
Immobilien Austrian Traded Index	1111
Immobilienindex	5783
Infrarotindex	2436
Instabilitätsindex	1371
Insulin-Index	2138
International Plant Names Index	1677
Internationaler Prognostischer Index	2934
ISE-100 Index	11740
ISEQ Overall Index	13054
ITraxx	5398
ITraxx SovX Western Europe Index	2410
Jungscher Typenindex	4361
Kakwani-Index	1996
Kanzerogenitätsindex	3110
Karnofsky-Index	3326
Karyopyknoseindex	2464
KASE Index	1942
Kettenindex	1705

Tabelle A.1

Keyword Efficiency Index	1213
Klimaschutz-Index	13791
Klimawandelindex	2356
Knöchel-Arm-Index	4805
Kohlenwasserstoff-Index	2318
Koinzidenzindex	11078
Kolm-Index	2522
Konditionsindex	65
Konjugierter Index	2101
Konsumklimaindex	701
Konsumklimaindex (GfK)	3504
Kontagiositätsindex	2804
Konvektiv-Index	3833
Korruptionswahrnehmungsindex	13619
KOSDAQ	956
KOSPI	17218
Kp-Index	2225
Krankenversicherungsprämien-Index	2176
KSE-30	863
Landesindex der Konsumentenpreise	3986
Late-Index	478
Lerner-Index	9767
Lesbarkeitsindex	10838
Lewis-Index	1202
Lindex	1954
Lippenstift-Index	2485
Living Planet Index	1916
Logistics Performance Index	1269
London Metal Exchange Index	1145
LQ-45	4100
Machtindex	9175
MakeIndex	2003

Tabelle A.1

Manifestationsindex	953
Mannheimer Peritonitis-Index	2315
Master Patient Index	4855
MDAX	11416
Mei Moses Fine Art Index	1033
Merck Index	1467
MERVAL	11738
Methylphenanthrenindex	549
Migrant Integration Policy Index	5655
Misery Index	4043
Monomerindex	1733
MSCI Emerging Markets Index	3213
MSCI GCC 50 Index	2932
MSCI World	13503
Multidimensional Poverty Index	1906
Multiindex	9731
NAHB/Wells Fargo Housing Market Index	7178
NASDAQ-100	14093
Nasdaq Composite	21330
Natur-Aktien-Index	2849
Nelson-Index	1813
Nemax	66
Nemax All Share	66
Nemax50	66
Nemwax	2436
Nikkei 225	48530
NISAX20	3801
Nordatlantischer Oszillationsindex	83
Normalized Differenced Vegetation Index	4185
NRW-MIX	1776
NTX (Index)	837
Numerischer Index	77

Tabelle A.1

NYSE Arca Gold BUGS Index	11682
NYSE Composite	9971
NZX 50 Index	12596
OBX Index	12299
OMX Baltic 10	1392
OMX Copenhagen 20	13280
OMX Helsinki 25	17289
OMX Iceland 15	3468
OMX Stockholm 30	13840
OMX Tallinn Index	1762
OMX Vilnius Index	1007
Operativer Index	2961
Overnight Index Swap	711
Palmer-Dürre-Index	1069
Parodontaler Screening-Index	5216
Patentindex	3607
Pathogenitätsindex	7275
Pearl-Index	3762
PeerIndex	6950
Peren-Clement-Index	8641
Periodical Contents Index	1817
Personalmanagement-Professionalisierungs-Index	25085
Petrussa-Index	1190
PFTS Index	2392
Philadelphia Fed Index	464
Philadelphia Gold and Silver Index	10835
Philippine Stock Exchange Index	13729
Philly Fed Index	1329
Photovoltaik Global 30 Index	8007
Physical Electricity Index	4164
Polish Traded Index	1971
Ponderal-Index	3200

Tabelle A.1

Populationsindex	1659
Preis-Informationsdienst	11070
Preisindex	23157
Privatverschuldungsindex	3834
Prozessfähigkeitsindex	4554
PSI 20	13801
Purchasing Managers Index	12368
R-Baum	10336
Radiative Forcing Index	1429
Randi-Index	888
Rapid Shallow Breathing Index	2087
RATZ-Index	1947
REAL-Index	5167
Relative Strength Index	2376
Relativer Temperaturindex	3604
Rentenindex	5253
Retentionsindex	3650
Retikulozytenproduktionsindex	2967
Rice-Index	788
Robin-Hood-Index	3646
Rogers International Commodity Index	11221
Rohstoffindex	9840
Rosenbluth-Index	1991
ROTX	2265
Roud Folk Song Index	5648
RTS-Index	21630
Russell 1000	8900
Russell 2000	13693
Russell 3000	4731
Russian Depositary Index	2659
Russian Traded Index	1689
S&P 100	15150

Tabelle A.1

S&P 500	81479
S&P 500 Banks Index	2461
S&P 500 Short Index	1362
S&P Africa 40 Index	4981
S&P CNX Nifty	4917
S&P Construction & Engineering Index	711
S&P Global 1200	964
S&P GSCI	10955
S&P Tiefsee Öl & Gas Index	3380
S&P/ASX 50	6662
S&P/TSX 60	4182
S&P/TSX Composite Index	17156
Sauerstoffindex	3478
Sauerstoffsättigungsindex	788
Schattenfinanzindex	14412
Scheuerindex	1563
Schlammindex	2176
Schmelzflussindex	5356
Schockindex	2429
Science Citation Index	3119
Science Impact Index	5094
SDAX	10229
Sen-Index	2587
SET Index	9003
Shanghai Stock Exchange 180 Index	10883
Shanghai Stock Exchange 50 Index	6433
Shannon-Index	3435
ShortDAX	3747
Silt Density Index	2036
Simpson-Index	1933
Single-Index-Modell	4576
SMI MID	6656

Tabelle A.1

Social Sciences Citation Index	2347
Social Security Death Index	4072
SOFIX (Aktienindex)	1943
Soil-Adjusted Vegetation Index	1992
Sokolow-Lyon-Index	2918
Solarer Radioflussindex	5393
Southern Oscillation Index	2131
Speech Transmission Index	4094
SQL Phrase Index	2208
SSE Composite Index	17394
Stand Density Index	1211
STOXX Europe 50	13604
STOXX Nordic 30	1518
Straits Times Index	14473
Stress-Strain-Index	1869
Suits-Index	5596
Sustainable Process Index	8084
Swiss Bond Index	3990
Swiss Leader Index	5794
Swiss Market Index	15633
Swiss Performance Index	8851
System Average Interruption Duration Index	2763
TA-100 Index	10639
Tadawul All-Share Index	10995
Taiwan Capitalization Weighted Stock Index	11381
TecDAX	8999
Teilindex	1486
Television Lighting Consistency Index	3638
TEPIX	8971
The Index	1730
Theil-Index	3660
Thomson Reuters/Jefferies CRB Index	17450

Tabelle A.1

TIOBE-Index	3583
To Aru Majutsu no Index	18915
TOPIX	858
Toxic 100 Index	6178
Trade Weighted US Dollar Index	13064
Tragfähigkeitsindex	7870
TSH-Index	2691
U.S. Dollar Index	11027
UB-Baum	2355
University of Michigan Consumer Sentiment Index	13796
UV-Index	10688
Value Line Composite Index	4323
Vanhanen-Index	2282
Vapour Lock Index	2594
VDAX	3070
VDAX-NEW	7707
Vegetationsindex	2216
Verbraucherpreisindex für Deutschland	11601
Verum index sui et falsi	1324
Very Easy Rodent-Oriented Netwide Index to Computerized Archives	1315
Viereckenindex	3047
Viskositätsindex	2747
Volatilitätsindex	3936
Vulkanexplosivitätsindex	8237
Welthunger-Index	52914
Wetterindex	1685
Whitebock-Index	1177
Wiener-Index	4593
WIG 20	12627
WIG30	11463
Wilshire 5000	10635

Tabelle A.1

Windows-Leistungsindex	3294
Wobbeindex	3778
Wohlfahrtsindex	4405
Wolkenkratzer-Index	4336
World Port Index	647
World Water Index	1094
Youden-Index	1336
Yules Index	572
Zürcher Fluglärm-Index	7611
ZEW-Index	1396

Quelle: Abfrage A.1, DBpedia 16.09.2015

Die Informationen von Tabelle A.1 können inklusive URL mit der SPARQL-Abfrage A.1 auf `http://de.dbpedia.org/sparql` abgerufen und ggf. um aktuelle Einträge ergänzt werden:

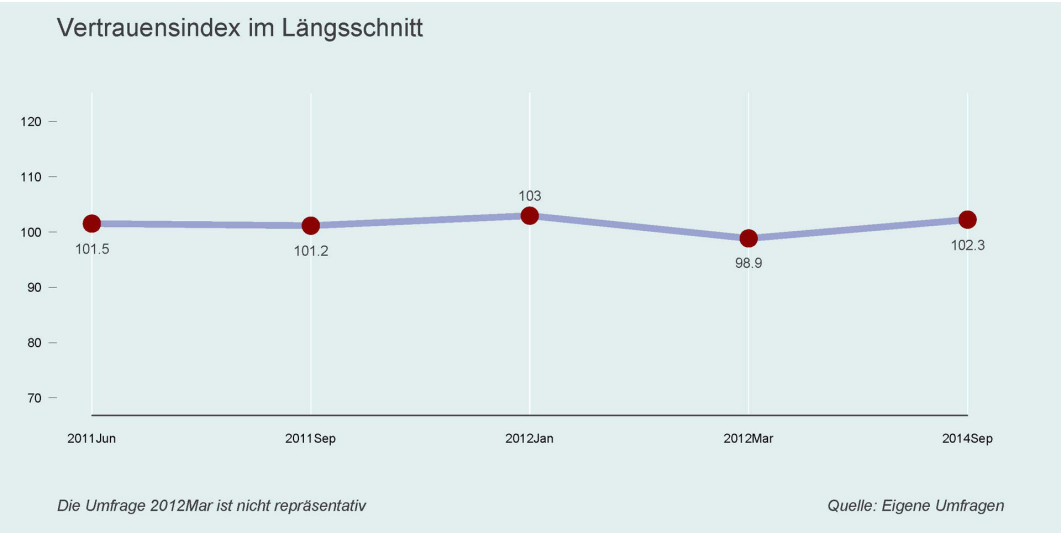
Syntax A.1: SPARQL-Code zum Abrufen aller Einträge zum Konzept oder Label „Index“

```
1  PREFIX dcterms: <http://purl.org/dc/terms/>
2  SELECT DISTINCT ?Seite ?wikiPageLength ?subject_label ?concept ?type
3                  ?comment ?abstract
4  WHERE { ?Seite dcterms:subject ?concept .
5    OPTIONAL {?Seite <http://dbpedia.org/ontology/wikiPageLength> ?wikiPageLength
     }
6    OPTIONAL {?Seite rdf:type ?type}
7    OPTIONAL {?Seite rdfs:label ?subject_label}
8    FILTER   (regex(STR(?concept), 'index') || (regex(?subject_label,
9    '.Index[^a-z]','i'))).
10   OPTIONAL {?Seite <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> ?type}
11   OPTIONAL {?Seite <http://www.w3.org/2000/01/rdf-schema#comment> ?comment}
12   OPTIONAL {?Seite <http://dbpedia.org/ontology/abstract> ?abstract}
13 }
14 LIMIT 10000
```

Anhang B

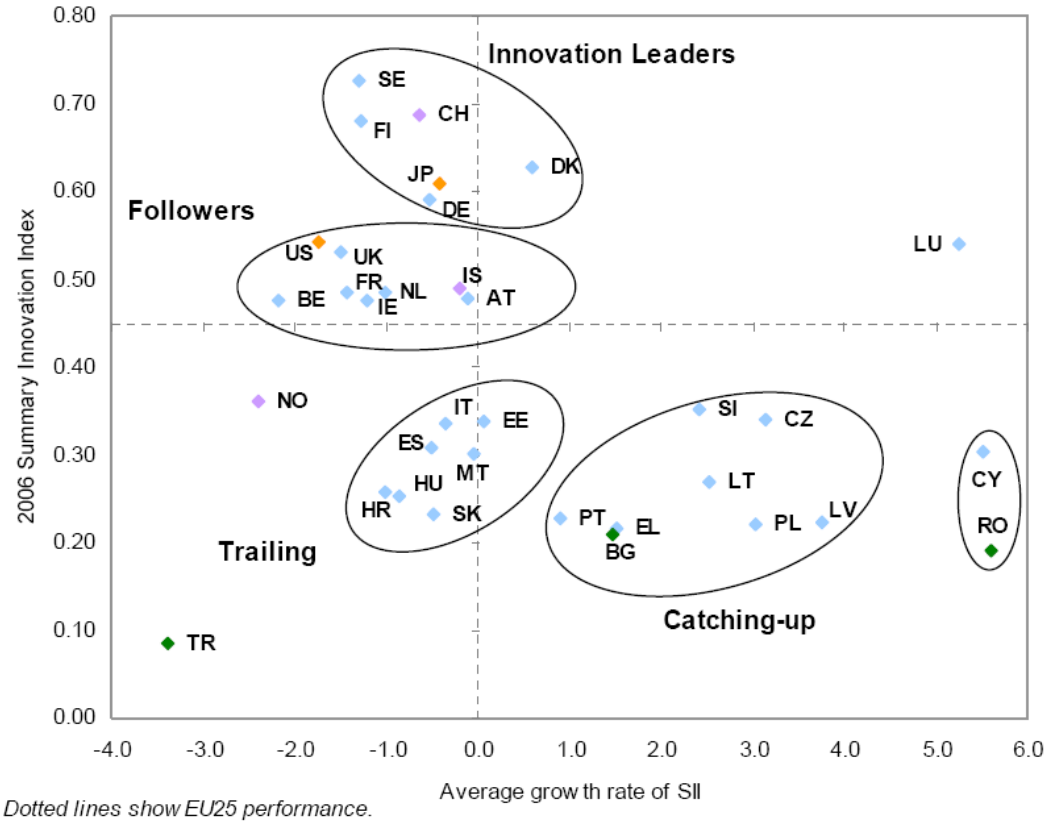
Diagramme

Abbildung B.1: Beispiel für ein Liniendiagramm



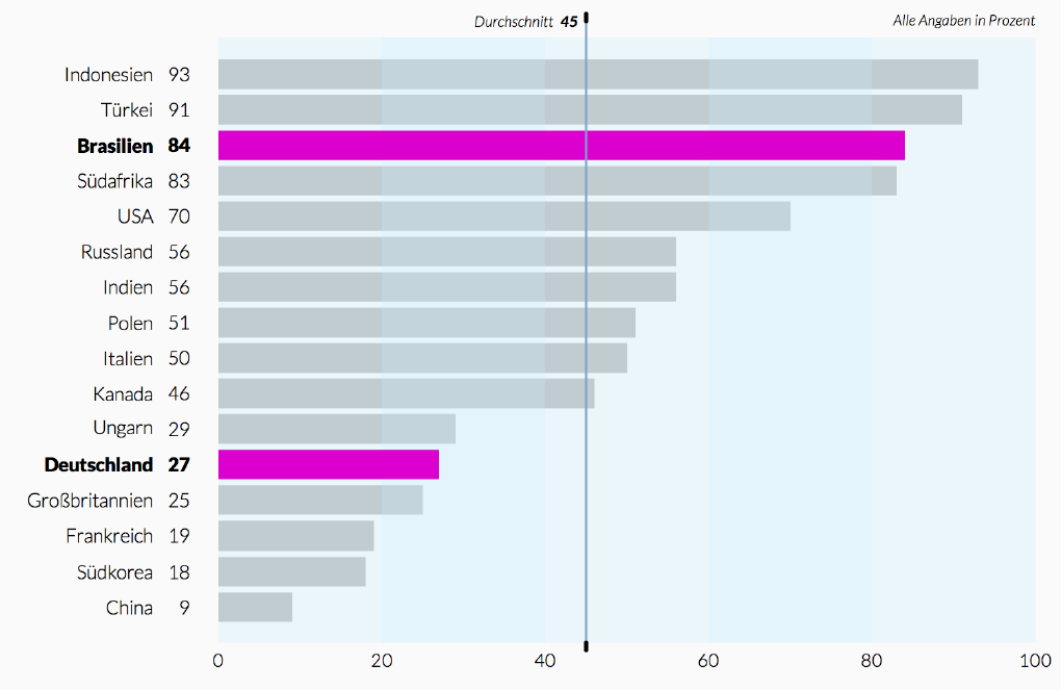
Quelle: Eigene Darstellungen und Berechnung

Abbildung B.2: Beispiel für ein Streudiagramm



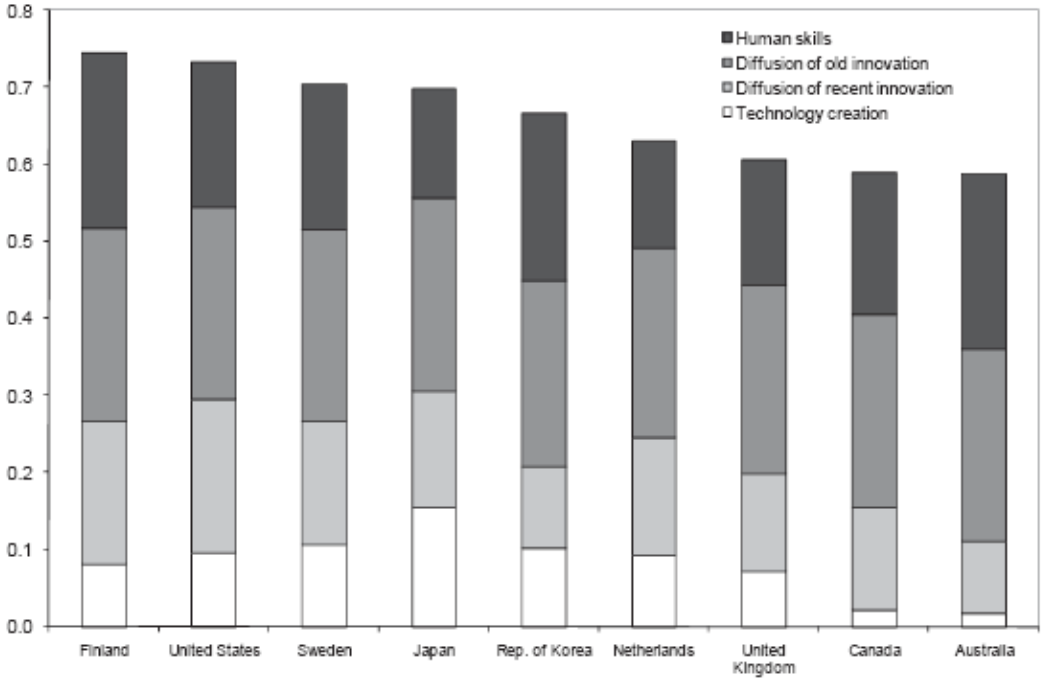
Quelle: Panagiotis (2007)

Abbildung B.3: Beispiel für ein Balkendiagramm



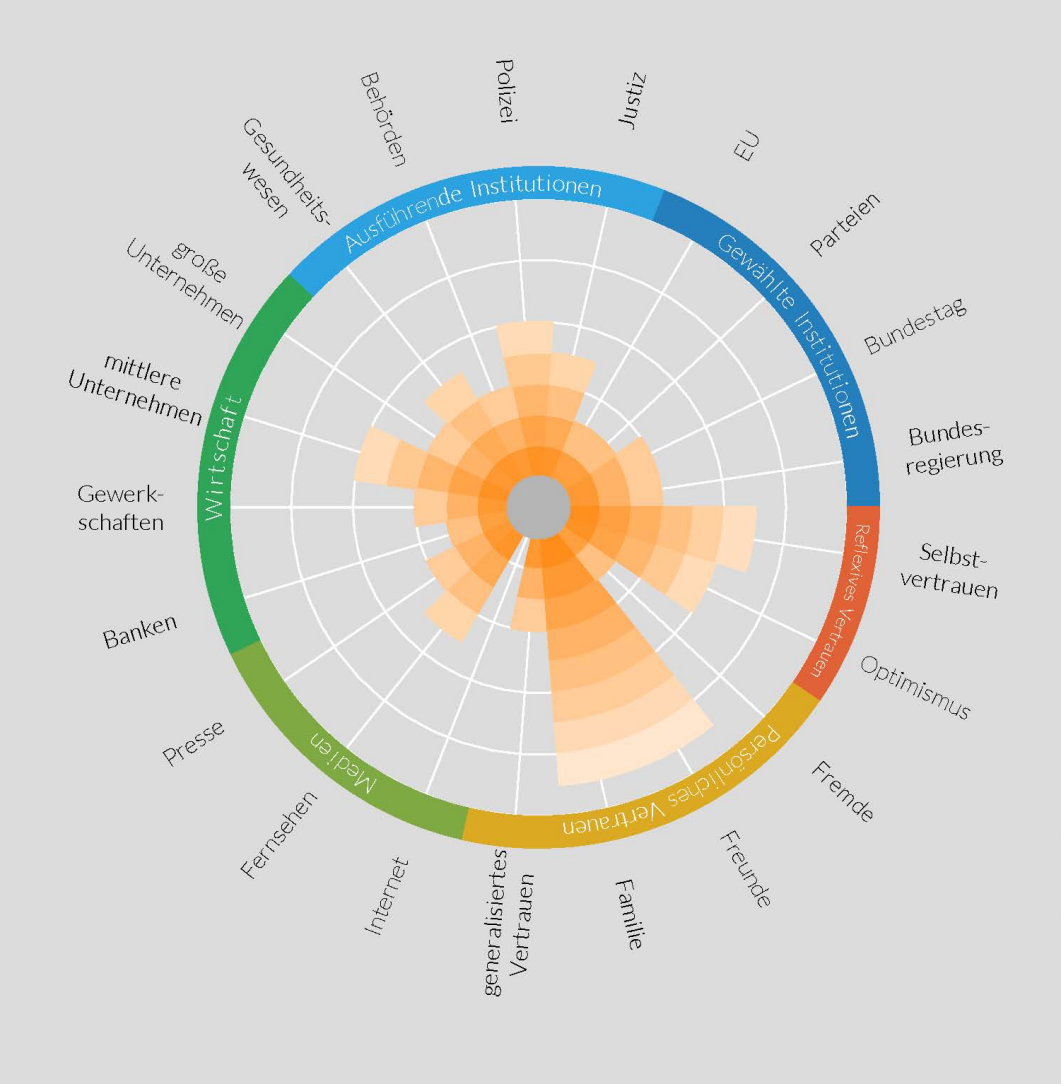
Quelle: Rahlf (2014: S. 192)

Abbildung B.4: Beispiel für ein gestapeltes Säulendiagramm



Quelle: Nardo u. a. (2008: S. 36)

Abbildung B.5: Beispiel für ein gestapeltes Radialsäulendiagramm



Quelle: Eigene Darstellung und Berechnung

Anhang C

Abfragen

Ausgangspunkt der Extraktion von Indizes aus dem Yago-SPARQL-Endpunkt ist die Taxonomie des Begriffs *Index* in *Wordnet*. In *Yago* wurde diese mit den Kategorien in *Wikipedia* verbunden. Mit der folgenden Abfrage können alle Entitäten zum Typ *Index* über `http://lod.openlinksw.com/sparql` abgefragt werden:

Syntax C.1: SPARQL-Code zum Abrufen der Entitäten zum Typ Ïndex

```
1 SELECT DISTINCT ?Seite WHERE {
2   ?Seite a <http://yago-knowledge.org/resource/wordnet_index_106639674> .
3 }
```

Als Ergebnis werden 422 Entitäten zurückgegeben.[1] Im folgenden Schritt könnte die Abfrage z.B. um Informationen zu Artikellänge in Wikipedia, zugehörige Kategorien, Datum der Etablierung und alternative Schreibweisen erweitert werden.

[1] Zeitpunkt der Abfrage: 12.05.2014

Anhang D

R-Code

Syntax D.1: Funktion „getConceptData“ aus dem R-Paket „analyseMedia“ zur Abfrage der DBpedia

```
#' Query DBpedia.
#'
#' @description Queries the DBpedia using a SPARQL endpoint.
#' @param concept name of the Wikipedia concept where to look after corresponding subjects.
#' @param subjectLabel if specified, additionally include matching subjects.
#' @param spqlEndPoint the sparcle endpoint where the query is send to.
#' @return The function returns all subjects from DBpedia which are associated to the given concept as a data frame.
#' @export
getConceptData <- function(concept="index",subjectLabel=NULL,spqlEndPoint="http://de.dbpedia.org/sparql") {
  filter <- switch (
    EXPR = length(subjectLabel)+1,
    " (regex(STR(?concept), '" %p0% concept %p0% "'))",
    " (regex(STR(?concept), '" %p0% concept %p0% "') ||
          (regex(?subject_label, '." %p0% subjectLabel %p0% "','" %p0% tolower(substring(subjectLabel,1,1)) %p0% "')))"
  )

  qry <- "PREFIX dcterms: <http://purl.org/dc/terms/>
          SELECT DISTINCT ?URL ?wikiPageLength ?subject_label
          ?concept ?type ?comment ?abstract
          WHERE { ?URL dcterms:subject ?concept .
          OPTIONAL {?URL <http://dbpedia.org/ontology/wikiPageLength>
          ?wikiPageLength}
          OPTIONAL {?URL rdf:type ?type}
          {?URL rdfs:label ?subject_label}
```

```
25            FILTER " %p0% filter %p0% ".
26            OPTIONAL {?URL <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
27            ?type}
28            OPTIONAL {?URL <http://www.w3.org/2000/01/rdf-schema#comment>
29            ?comment}
30            OPTIONAL {?URL <http://dbpedia.org/ontology/abstract> ?abstract}
31            }
32            LIMIT 10000"
33
34    sparqlData <- SPARQL(url=spqlEndPoint,
35                         query=qry,format="xml")
36    return(sparqlData$results)
37
38  }
```

Anhang E

Dokumentation

E.1 Überblick über das R-Paket „analyseMedia“

Das R-Paket analyseMedia vereinfacht die quantitative Inhaltsanalyse von digitalen Inhalten, indem es Funktionen (1) zur Selektion, (2) zum Datenabruf, (3) zur Datenverarbeitung, (4) zur Stichprobenziehung und (5) zur Erstellung von deskriptiven Statistiken bereitstellt (vgl. Abbildung 5.1 auf Seite 162). Die Funktionen sind parametrisiert, wodurch z.B. flexibel neue Datenquellen hinzugefügt werden können.

E.2 Installation

Die Installation des Paketes wird unter R ($\geq$ 3.2.4) mithilfe der *Development Tools* (devtools) und ein Repository[1] auf Github ermöglicht:

```
1 library(devtools)
2 install_github("digitaleMedien/analyseMedia")
```

Über diesen Befehl werden automatisch sämtliche abhängige Pakete installiert. Um Namenskonflikten bei den Funktionen der verschiedenen Pakete vorzubeugen, wurden bei

[1] Das Repository analyseMedia ist unter `https://github.com/digitaleMedien/analyseMedia` erreichbar (besucht am 15. Nov. 2016).

den meisten abhängigen Paketen lediglich die notwendigen Funktionen importiert. Folgende Pakete werden für die Ausführung von analyseMedia komplett geladen (Depends) oder nur teilweise importiert (Imports):

- **Depends** SPARQL ($\geq$ 1.16), dplyr ($\geq$ 0.4.3)
- **Imports** xml2 ($\geq$ 0.1.2), rmarkdown ($\geq$ 0.9.5), stringr ($\geq$ 1.0.0), rvest ($\geq$ 0.3.1), knitr ($\geq$ 1.12.3), tidyr ($\geq$ 0.4.1)

Mit der Paketinstallation werden zwei Unterordner im Paketpfad angelegt: `\resources` und `\data`. Im ersten Ordner befinden sich verschiedene Konfigurationsdateien sowie Vorlagen für die Erstellung von PDF-Dateien. Der zweite Ordner ist ein Arbeitsverzeichnis für die temporären Dateien, die z.T. von den Funktionen des Pakets angelegt werden.

E.3 Anwendung der Module

Im folgenden Beispiel soll die Verwendung eines Teils der Funktionen des Pakets „analyseMedia“ verdeutlicht werden.

E.3.1 Selektion

Im folgenden Beispiel sollen alle Subjekte aus der Webdatenbank DBpedia abgefragt werden, die innerhalb ihres Namens den Term „index“ enthalten oder die einem Konzept zugeordnet sind, welches den Term „index“ enthält.

```
1 library(AnalyseMedia)
2
3 # create list of search terms via dbpedia
4 indexKeywords <- getConceptData(concept = "index",subjectLabel = "Index")
5 indexKeywords <- cleanCharacters(indexKeywords) # remove wrong characters
6 indexKeywords <- filterResultSet(indexKeywords) # remove wrong results
7
8 # export data to file
9 write.csv2(indexKeywords, file = sysFile("data","indexKeywords.csv"))
```

Syntax E.1: Beispiel-Syntax für die Abfrage der DBpedia

Über die Zeile 4 in Syntax E.1 wird die Abfrage der DBpedia durchgeführt. Die Funktion in Zeile 5 bereinigt die Ergebnisse von ungültigen oder fehlerhaften Zeichen. In Zeile 6 werden aus dem Datensatz invalide Einträge entfernt. Der bereinigte Datensatz wird schließlich über Zeile 9 in das Unterverzeichnis `\data` des Pakets gespeichert.

E.3.2 Datenabruf

Im folgenden Beispiel sollen nun sämtliche Beiträge mit Bezug zu einem Index von `faz.net` abgerufen und verarbeitet werden. Um die Datenmenge für das Beispiel einzuschränken, werden nur Beiträge des Blog-Bereichs einbezogen. Außerdem wird nicht die unter Abschnitt E.3 erstellte Liste von Indizes verwendet, sondern lediglich das Suchwort „DAX“.

```
1  library ( AnalyseMedia )
2
3  # configuration files
4  mediaResources <- read.csv2(sysFile("resources","digitalMedia.cnf")
5                              ,quote = "\'",stringsAsFactors = F)
6  mediaNodes     <- read.csv2(sysFile("resources","digitalMediaNodes.cnf")
7                              ,quote = "\'",stringsAsFactors = F)
8
9  # media target
10 mediaTarget <- "fazBlog"
11
12 # keyword list
13 keywordList <- "dax"
14
15 # save resultpage from search engine
16 sapply(keywordList,function(x) getResultPages(mediaTarget,x))
17 # extract article - urls and article attributes from query result page
18 loadedResultPages <- list.files(sysFile("data/resultpages"))[grep(paste0(
       mediaTarget,".*",keywordList),list.files(sysFile("data/resultpages")))]
19
20 dat1 <- lapply(loadedResultPages,function(x)  analysePage(mediaTarget,x)) %>%
21   do.call(rbind,.) %>%
22   mutate_each(funs(as.character)) %>%
23   mutate(articleId = as.character(1:nrow(.)))
24
25 # get the full article - page with the url provided in resultpage
26 sapply(1:nrow(dat1),function(x) getPages(dat1$pageURL[x],mediaTarget,dat1$
       searchTerm[x],x))
27 # extract article attributes from article pages
28 loadedArticlePages <- list.files(sysFile("data/articlepages"))[grep(paste0(
       mediaTarget,".*",keywordList),list.files(sysFile("data/articlepages")))]
```

```
29 dat2 <- lapply(loadedArticlePages,function(x) analysePage(mediaTarget,x)) %>%
30   do.call(rbind,.) %>%
31   mutate_each(funs(as.character))
32
33 ## combine data and write out
34 dat <- left_join(dat2,dat1,by = c("articleId","searchTerm"))
35 # clean data
36 dat <- cleanData(dat,mediaTarget)
37 write.csv2(dat, file = paste0(sysFile("data"),mediaTarget,"_articleDatabase.csv")
38            ,row.names = FALSE,quote=1:length(dat), fileEncoding = "utf8")
```

Syntax E.2: Beispiel-Syntax für den Abruf von Beiträgen

Über die Zeilen 4 und 5 der Syntax E.2 werden die Konfigurationsdateien geladen. Sie enthalten die Strukturdaten zu den digitalen Plattformen (damit kann z.B. der Beitragstitel identifiziert werden).

In Zeile 11 wird der Suchbegriff definiert. Möglich ist an dieser Stelle auch die Definition eines Vektors mit mehreren Suchbegriffen. Werden mehrere Suchbegriffe definiert, wird die Funktion *getResultPage* in Zeile 14 pro Suchbegriff aufgerufen. Dieser wird über die Funktion an das in Zeile 8 festgelegte digitale Medium gesendet (hier: Blog-Bereich von `faz.net`). Sämtliche von der digitalen Plattform zurückgegebenen Suchergebnisse werden seitenweise in einem temporären Verzeichnis im Paket-Pfad abgelegt (`\data\resultpages`). Die zugehörigen Dateinamen werden in Zeile 16 in einem Vektor gesammelt.

In Zeile 18 werden die Seiten über die im Vektor befindlichen Dateinamen aus dem temporären Verzeichnis über die Funktion *analysePage* geladen, verarbeitet und als Datensatz zurückgegeben. Der Datensatz enthält u.a. sämtliche URLs zu den Einzelseiten der zurückgegebenen Ergebnisse.[2] Jeder Beitrag steht dabei in einer Zeile.

Der Datensatz mit den Suchergebnisse wird in Zeite 24 genutzt, um sämtliche Einzelseiten zu den Beiträgen nacheinander abzurufen und in einem temporären Verzeichnis abzulegen. Die Einzelseiten werden dann in Zeile 27 geladen, verarbeitet und als Datensatz zurückgegeben.

[2]Bei der Eingabe eines Suchbegriffes in eine Suchmaske werden im ersten Schritt alle Ergebnisse aufgelistet. Die URLs der Ergebnisse werden in eine Liste gespeichert, um sie danach nacheinander abzurufen.

In Zeile 32 werden die beiden Datensätze (Suchergebnisse und Einzelseiten) zusammengeführt. Dies hat den Vorteil, dass die Ergebnisse der Suchseite mit denen der Einzelseite verglichen werden können.

In Zeile 34 werden die Daten validiert und bereinigt, um sie danach über die Zeile 35 als CSV-Datei im Unterverzeichnis des Pakets abzulegen.

E.3.3 Datenverarbeitung und Erstellung deskriptiver Statistiken

In diesem Schritt wird als Input ein Datensatz aus einem temporären Verzeichnis des Paketpfades genutzt (vgl. Abschnitt E.3.2). Dieser wird über die Zeile 4 in Syntax E.3 geladen. Mittels der Funktion *descArticles* werden pro Beitrag Kennzahlen der quantitativen Inhaltsanalyse erhoben und zum Datensatz hinzugefügt. Zeile 11 bis Zeile 56 verdichten diese Informationen zu einem Set an Tabellen und Diagrammen und legen diese als CSV- oder PDF-Datei im Unterordner des Paketpfades ab (`\data\table` oder `\data\plot`).

```
1  library(AnalyseMedia)
2
3  # descriptive statistcs for articles
4  dat <- read.csv2(paste0(sysFile("data"),mediaTarget,"_articleDatabase.csv")
5                   ,sep=";",stringsAsFactors = FALSE,encoding = "utf8"
6                   ,quote = "\'",fileEncoding = "utf8")
7  dat <- dat %>%
8    mutate(articleDate = as.Date(articleDate))
9
10 dat <- descrArticles(dat)
11 exportTable <- statTab(dat)
12
13 # frequency by time
14 freqDat <- freqByTime(dat)
15
16 # export summary tables for each searchterm
17 # ressort table
18 ressortTable <- lapply(unique(dat$searchTerm),function(x)
19                        aggStatTab(dat,x,"aggregatedRessortStatistic")
20                  )
21 sapply(seq(1:length(ressortTable)),function(x)
22   write.csv2(ressortTable[[x]],
23              file = paste0(sysFile("data/table")
24                            ,mediaTarget,"_aggregatedRessortStatistic_"
25                            ,unique(dat$searchTerm)[x],".csv")
26              ,row.names = FALSE
```

```
              ,quote=1:length(ressortTable[[x]]))
)

# aggregated table statistic
aggregatedExportTable <- lapply(unique(dat$searchTerm),function(x)
                                  aggStatTab(dat,x)
                          )
sapply(seq(1:length(aggregatedExportTable)),function(x)
  write.csv2(aggregatedExportTable[[x]],
              file = paste0(sysFile("data/table")
                            ,mediaTarget,"_aggregatedExportTable_"
                            ,unique(dat$searchTerm)[x],".csv")
              ,row.names = FALSE
              ,quote=1:length(aggregatedExportTable[[x]]))
)

# dual plot for all searchterms in datset
sapply(unique(freqDat$searchTerm[!is.na(freqDat$searchTerm)]), function(x)
  dualPlot(
            titleMain = x,
            titleSub = "Subtitle",
            yAxisTitle = "Y-Axis",
            xAxisTitle = "X-Axis",
            refText = "Source:",
            freqDat,
            pdfName = paste0(sysFile("data/plot"),mediaTarget,"_",x,".pdf"),
            y2AxisTitle = "Y2-Axis",
            selectSearchTerm = x
  )
)
```

Syntax E.3: Verarbeitung der Beiträge und Erstellung deskriptiver Statistiken

E.3.4 Stichprobenziehung

Für die manuelle Analyse von Beiträgen wird im Folgenden beispielhaft gezeigt, wie aus der Gesamtheit der Beiträge eine Stichprobe gezogen wird. Anschließend wird jeder Beitrag strukturiert und formatiert als PDF-Dokument exportiert.

```
library(AnalyseMedia)

# minimum 10 articles, maximum 30 articles, otherwise 5 percent
sampleDat <- sampleArticles(dat)
write.csv2(sampleDat, file = paste0(sysFile("data/sampledpages"),mediaTarget,"_
    articleSample.csv"),row.names = F)
```

```
6
7  # create pdf-file for sampled articels (MikTeX must be installed)
8  render(sysFile("resources","articleSamplingDocument.Rmd"),
9         output_file = paste0(mediaTarget,"_articleSample.pdf"),
10        output_dir = sysFile("data/sampledpages"),clean = T,envir = .GlobalEnv)
```

Syntax E.4: Stichprobenziehung von Beiträgen

Als Datensatz fungiert wieder der in Syntax E.3, Zeile 4 geladene Datensatz. Aus diesem wird über die Funktion *sampleArticles* eine Stichprobe gezogen (Zeile 4) und in Zeile 5 als CSV-Datei in den Unterordner des Pakets gespeichert (data/sampledpages). Um eine leichte Lesbarkeit zu gewährleisten, werden die Informationen aus der Tabelle wieder in eine lesbare Form umgewandelt und als PDF-Dokument in den Unterordner des Pakets abgelegt (`\data\sampledpages`).

E.4 Anpassung

Das R-Paket kann angepasst und weiterentwickelt werden. Denkbar ist das Hinzufügen zusätzlicher Kriterien für die quantitative Inhaltsanalyse oder eines neuen digitalen Mediums.

Um das Paket anpassen zu können, muss das zugehörige Repository von Github geklont werden.[3] Das Repository lässt sich über `https://github.com/digitaleMedien/analyseMedia` erreichen.

Nach dem Klonen können alle Funktionen und Konfigurationsdateien des R-Projektes lokal angepasst werden. Sämtliche Funktionen des Pakets befinden sich im Unterordner `\R`, die Konfigurationsdateien befinden sich unter `\inst\resources`.[4]

[3] Einen Einstieg in das Klonen von Github-Repositories bietet `https://help.github.com/articles/cloning-a-repository/` (besucht am 11.03.2016).

[4] Bei der Erstellung des Pakets werden alle Ordner im Verzeichnis „inst“ dem Paketpfad hinzugefügt. Das bedeutet, dass im Paketpfad dann „resources“ eine Ebene höher erscheint und kein Ordner „inst“ existiert.

E.4.1 Hinzufügen eines neuen digitalen Mediums

Für das Hinzufügen eines digitalen Mediums oder eines neuen Merkmals müssen die Konfigurationsdateien `digitalMedia.cnf` und `digitalMediaNodes.cnf` im Unterverzeichnis `\inst\resources` des Pakets erweitert werden. Ihr Aufbau entspricht einer Tabelle mit mehreren Spalten. In der erstgenannten Konfigurationsdatei wird der Name des Mediums mit der zugehörigen URL zur Suchmaschine der Seite hinzugefügt. In der Datei `digitalMediaNodes.cnf` befinden sich vier Spalten:

- **media** Name des Mediums (frei wählbar).
- **target** Name des Merkmals (frei wählbar).
- **nodeText** Name des Knotenpunkten im XML-Baum, mit dem die gewünschte Kennzahl eindeutig identifiziert werden kann (abhängig von der Plattform).
- **attr** Optionales Attribut, falls z.B. statt dem Text eines Knotens, die Information in „href“ abgefragt werden soll (abhängig von der Plattform).

In der Tabelle ist jede Zeile eindeutig über *media* und *target* identifiziert. Soll ein neues Medium hinzugefügt werden, ist für jedes gewünschte Merkmal ein neuer Eintrag anzulegen und die Information zum Auffinden des Knotens hinzuzufügen. Soll ein neues Merkmal erhoben werden, müssen je Medium jeweils eine Zeile hinzugefügt und die Felder *target* sowie *nodeText* gefüllt werden.

Das Hinzufügen eines neuen Mediums ist mit der Bearbeitung der beiden Konfigurationsdateien abgeschlossen. Falls ein neues Merkmal hinzugefügt werden soll, ist zusätzlich die Funktion *analysePage* anzupassen (vgl. Abschnitt E.4.2).

E.4.2 Anpassen einer R-Funktion

Über das R-Paket analyseMedia werden insgesamt 16 Funktionen zur Verfügung gestellt (vgl. Tabelle E.1). Darüber hinaus existieren zahlreiche unsichtbare Funktionen für interne Prozesse. Eine detaillierte Beschreibung der Funktionen kann entweder über die Hilfe des Pakets eingesehen werden oder als PDF-Datei unter Github[5] (vgl. Abbildung E.1).

[5] Siehe `https://github.com/digitaleMedien/analyseMedia/blob/master/AnalyseMedia-manual.pdf`.

Tabelle E.1: Funktionen des R-Pakets „analyseMedia“

Funktion	Kurzbeschreibung
aggStatTab	Build aggregation table.
analysePage	Collect metadata of webpage.
cleanCharacters	Remove wrong characters.
cleanData	Clean article data.
descrArticles	Describe article data.
dualPlot	Combination of linechart and barplot.
filterResultSet	Remove wrong results.
freqByTime	Table with frequency by time.
getConceptData	Query DBpedia.
getPages	Get article-page html.
getResultPages	Get html of result-pages from search engine.
noParagraph	Remove paragraphs from string.
removeHtml	Remove html-code from content.
sampleArticles	Sample articles.
statTab	Arrange table.
sysFile	package path.

Sämtliche Funktionen wurden im Unterordner `\R` des Projekts abgelegt und thematisch in einzelne R-Datein aufgeteilt:

- **01_helper.R** Enthält Hilfsfunktionen, die i.d.R. von anderen Funktionen genutzt werden.
- **02_get.R** Enthält Funktionen zum Abrufen, Einlesen oder Erzeugen von Daten.
- **03_clean.R** Enthält Funktionen zum Bereinigen von Daten.
- **04_aggregate.R** Enthält Funktionen zur Zusammenfassung und Verdichtung von Daten.
- **05_export.R** Enthält Funktionen zum Exportieren von Daten, z.B. in Tabellen- oder Diagrammform.

Soll beispielsweise der Funktion *descrArticles* (vgl. Syntax E.5) eine weitere Kennzahl zur quantitativen Inhaltsanalyse hinzugefügt werden, muss die Datei `04_aggregate.R` bearbeitet werden.

```
1  descrArticles <- function(.dat) {
2    .dat %>%
3      mutate(
4        numSerchTerm      = str_count(tolower(articleContent),tolower(searchTerm)),
5        numWordsTitle     = str_count(pageHeadline,"\\s+")+1,
```

```
6       numWordsContent    = str_count(articleContent,"\\s+")+1,
7       sentenceCount      = str_count(articleContent,"\\?|\\.|\\!"),
8       searchTermInTitel  = str_detect(tolower(pageHeadline),tolower(searchTerm)),
9       numParagraph       = str_count(startStringWithLetter(articleContent)
10                                     ,"\\n[a-zA-Z]|\\t[a-zA-Z]"
11                            ),
12      searchTermFirstPg  = str_detect(tolower(substring(startStringWithLetter(
13                                                          articleContent
14                                                        ),1,
15                                                        str_locate(
16                                                          startStringWithLetter(
17                                                            articleContent
18                                                          ),"\t|\n"
19                                                        )[,1]
20                                               )
21                                     ),
22                                     tolower(searchTerm)
23                            )
24    ) %>%
25    mutate(
26      sentenceLength     = numWordsContent/sentenceCount,
27      serchTermFirstSeen= ifelse(numSerchTerm>0,
28                                 str_locate(articleContent,searchTerm)[,1]/
29                                   str_length(articleContent)*100
30                                 ,NA
31                            )
```

Syntax E.5: Funktion „descrArticles“ aus dem R-Paket „analyseMedia“ zur Erhebung der quantitativen Kennzahlen zu einem Beitrag

Falls die Berechnung der neuen Kennzahl eine komplexe Logik erfordern sollte, ist diese als Funktion in `01_helper.R` abzulegen. Die Syntax zur Bestimmung der Kennzahl (mit oder ohne Hilfsfunktion) kann z.B. zwischen Zeile 23 und 24 eingefügt werden. Die Kennzahl steht danach allen Folgefunktionen zur Verfügung.

Abbildung E.1: Dokumentation zum R-Paket „analyseMedia“

Package ‘AnalyseMedia’

November 10, 2016

Type Package

Title AnalyseMedia

Version 0.11

Author Martin Badicke

Maintainer <martin.badicke@gmail.com>

Depends R (>= 3.2.4), SPARQL (>= 1.16), dplyr (>= 0.4.3)

Imports xml2 (>= 0.1.2), rmarkdown (>= 0.9.5), stringr (>= 1.0.0), rvest (>= 0.3.1), knitr (>= 1.12.3), tidyr (>= 0.4.1)

ByteCompile TRUE

Description AnalyseMedia

License closed

Copyright (c) 2013-2016 Martin Badicke. All Rights Reserved.

RoxygenNote 5.0.1

NeedsCompilation no

R topics documented:

aggStatTab . . . 2
analysePage . . . 2
cleanCharacters . . . 3
cleanData . . . 3
descrArticles . . . 4
dualPlot . . . 4
filterResultSet . . . 5
freqByTime . . . 6
getConceptData . . . 6
getPages . . . 7
getResultPages . . . 8
noParagraph . . . 8
removeHtml . . . 9
sampleArticles . . . 9
statTab . . . 10
sysFile . . . 10

Index **12**

Quelle: Dokumentation zum R-Paket „AnalyseMedia“ auf Github