Entwicklung eines Information-Retrieval-Systems zur Unterstützung von Gefährdungs- und Risikoanalysen

vorgelegt von
Kerstin Hartig, M.Sc.

von der Fakultät IV – Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktorin der Ingenieurwissenschaften
— Dr.-Ing. —

genehmigte Dissertation

Promotionsausschuss:
Prof. Dr.-Ing. Uwe Nestmann (Vorsitzender)
Prof. Dr.-Ing. Dr. h.c. Stefan Jähnichen (Gutachter)
Prof. Dr. rer. nat. Andreas Vogelsang (Gutachter)
Prof. Dr.-Ing. Steffen Helke (Gutachter)

Tag der wissenschaftlichen Aussprache:
11.04.2019

Berlin 2019

Veröffentlichungen

Die in dieser Dissertation vorgestellte Arbeit entstand zwischen Juni 2013 und Januar 2019 an der Technischen Universität Berlin. Einige Ideen und Darstellungen sind bereits in den im Folgenden aufgelisteten Veröffentlichungen erschienen:

Konferenzbeiträge:

K. Hartig, T. Karbe: *Recommendation-Based Decision Support for Hazard Analysis and Risk Assessment*, 8th International Conference on Information, Process, and Knowledge Management (eKNOW), April 2016, Venedig, Italien, ISBN 978-1-61208-472-5, http://dx.doi.org/10.14279/depositonce-6974

K. Hartig, T. Karbe: *Semantic Network Skeleton - A Tool to Analyze Spreading Activation Effects*, 8th International Conference on Information, Process, and Knowledge Management (eKNOW), April 2016, Venedig, Italien, ISBN 978-1-61208-472-5, http://dx.doi.org/10.14279/depositonce-6973

V. Michalke, **K. Hartig**: *Explanation Retrieval in Semantic Networks - Understanding Spreading Activation based Recommendations*, 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KDIR), November 2016, Porto, Portugal, ISBN 978-989-758-203-5, http://dx.doi.org/10.14279/depositonce-7136

Journalbeitrag:

K. Hartig, T. Karbe: *Spreading Activation Simulation with Semantic Network Skeletons*, International Journal on Advances in Intelligent Systems, IARIA, Vol. 10, Nr. 1&2, Juni 2017, ISSN 1942-2679, http://dx.doi.org/10.14279/depositonce-6972

Kurzfassung

Die Entwicklung sicherheitsbezogener Systeme im Automobilbereich unterliegt speziellen Anforderungen an die funktionale Sicherheit, die in der Norm ISO 26262 beschrieben sind. Um die funktionale Sicherheit zu gewährleisten, werden zu Beginn der Systementwicklung Gefährdungs- und Risikoanalysen (G&Rs) erstellt. In G&Rs werden mögliche Fehlfunktionen in verschiedenen Betriebssituationen untersucht und ihre Risiken bewertet. G&Rs gelten aufgrund ihres experten- und erfahrungsbasierten Charakters als sehr aufwendig. Beispielsweise müssen alle potenziellen Gefährdungen erkannt und Inkonsistenzen zwischen und innerhalb von G&Rs vermieden werden. Eine systematische und automatisierte Wiederverwendung von Erfahrungswissen aus früheren G&Rs gilt als Herausforderung, bietet allerdings Potenzial, den Aufwand zu verringern und Experten zu unterstützen.

In dieser Arbeit präsentieren wir ein Information-Retrieval (IR)-System, das dazu in der Lage ist, Wissen früherer G&Rs zu archivieren, das gespeicherte Wissen zielgerichtet zu durchsuchen und bei der Erstellung neuer G&Rs zur Verfügung zu stellen. Dadurch ermöglichen wir die Wiederverwendung von Wissen, um die Konsistenz von G&Rs zu steigern und eine effizientere Durchführung zu unterstützen. Darüber hinaus generalisieren wir unseren Ansatz derart, dass der Aufbau domänenspezifischer IR-Systeme auch auf andere Anwendungsfälle jenseits der funktionalen Sicherheit im Automobilbereich übertragbar ist.

Der erste Beitrag dieser Arbeit besteht in der Entwicklung eines Rahmenwerks für den Aufbau domänenspezifischer assoziativer IR-Systeme mit generischen und konfigurierbaren Lösungen für die einzelnen Komponenten. Dazu gehören Techniken zum automatisierten Aufbau semantischer Netze sowie zu deren Verfeinerung. Dabei reduzieren wir den manuellen Aufwand, der üblicherweise bei der Wissensrepräsentation entsteht, indem wir auf das vorhandene Datenmodell fokussieren. Des Weiteren konzeptualisieren wir die Aktivierungsausbreitung als semantisches Suchverfahren und erweitern es unter anderem um eine neuartige Technik zur direkten Steuerung des Ausbreitungsverhaltens in den semantischen Netzen. Um die Nutzerakzeptanz zu erhöhen, stellen wir außerdem ein Verfahren zur automatisierten Erklärung der Ergebnisse des IR-Systems vor.

Als zweiten Beitrag präsentieren wir eine aus vier Schritten bestehende Methode für den Aufbau eines domänenspezifischen IR-Systems aus dem generischen Rahmenwerk. Damit trifft der Entwickler nach initialer Analyse von Daten und Domäne sukzessive Entscheidungen für die Auswahl und Konfiguration der Komponenten des IR-Systems sowie für deren Optimierung. Wir unterstützen den Entwickler zudem mit Techniken, die insbesondere die anfängliche Datenanalyse und Konfiguration der Aktivierungsausbreitung vereinfachen. Zum einen stellen wir das semantische Netzwerkskelett als Abstraktion semantischer Netze vor, dessen Visualisierung große semantische Netze wesentlich überschaubarer macht. Zum anderen ermöglichen wir die Simulation der Aktivierungsausbreitung auf diesem Netzwerkskelett, um frühzeitig ungeeignete Konfigurationen auszuschließen.

Mit dem dritten Beitrag zeigen wir die Machbarkeit der Methode und verwenden das generische Rahmenwerk beim Aufbau eines domänenspezifischen assoziativen IR-Systems zur Unterstützung von G&Rs. Dafür verwenden wir Daten aus der industriellen Praxis. Ein Prototyp ermöglicht eine Evaluation, die zum einen eine gute Qualität der Ergebnisse des IR-Systems für G&Rs aufzeigt und zum anderen Anhaltspunkte für geeignete und ungeeignete Konfigurationen liefert. Wir präsentieren unter anderem, wie sich Konfigurationen durch Anwendung evolutionärer Algorithmen optimieren lassen. Diese Vorgehensweise kann vom Entwickler domänenspezifischer IR-Systeme adaptiert werden und erweitert daher sein Handwerkszeug bei der Umsetzung der vorgestellten Methode.

Abstract

Automotive systems engineering must adhere to the functional safety standard ISO 26262. According to the standard, the hazard analysis and risk assessment (HARA) is one of the first safety activities during the development of safety-related systems. In this safety activity, experts examine potential malfunctions and their consequences in different situations and specify safety goals in order to reduce risks to an acceptable level. Performing HARAs is a time-consuming and expensive activity, because it requires extensive experience and domain knowledge. For example, it is essential to identify all potential hazards and to prevent inconsistencies both within and between HARAs. Thus, safety engineers would benefit from decision support that allows the reuse of approved knowledge from previous analyses. However, knowledge reuse and its automation are considered to be challenging tasks.

In this thesis, we present an information retrieval system that represents the results from previous HARAs in a semantic network and searches it for useful recommendations during a new HARA. The presented approach aims at making optimal use of the reuse potential and, therefore, increases the consistency of HARAs as well as the efficency of their development. Moreover, we generalize the approach to make it applicable to other domains and artifacts.

First, we present a framework for developing domain-specific associative information retrieval systems that comprises both generic and configurable solutions for each component. The framework includes techniques to automatically create and refine semantic networks. We reduce manual efforts that usually come with knowledge representation by focussing on the data's structure provided by its data model. Furthermore, we generalize and formalize spreading activation as semantic search method and extend it by a new technique to control the spreading behavior. We additionally provide a method to create meaningful explanations for the semantic search results, which is crucial for the user's acceptance and trust.

The second contribution of this thesis is a four-step method that describes how to create a domain-specific information retrieval system using the aforementioned framework. The method addresses the developers of information retrieval systems. We additionally support developers with techniques that facilitate challenging tasks such as the initial data analysis or spreading activation configuration. That includes the semantic network skeleton, an abstract view on a semantic network that summarizes its basic structural information, which makes it easier to comprehend. In addition, we present an approach for spreading activation simulation of semantic networks utilizing their semantic network skeletons. Estimating the expected spreading activation behavior allows the detection of undesirable effects and therefore supports the early elimination of unsuitable configuration settings.

Third, we demonstrate the feasibility of the method by creating a domain-specific information retrieval system to support HARAs using the presented framework. All concepts and techniques are implemented in a prototype to facilitate their evaluation. The evaluation outcome reveals good quality of the retrieval results and indicates what configuration settings are advantageous. Moreover, we show how configuration settings can be optimized with evolutionary algorithms, which extends the developer's tool set.

Inhaltsverzeichnis

1 Einleitung **1**

2 Grundlagen und verwandte Arbeiten **9**
2.1 Information-Retrieval-Systeme . . . 9
2.1.1 Grundbegriffe . . . 9
2.1.2 Arbeitsweise und Aufbau . . . 11
2.1.3 Evaluation im Information Retrieval . . . 15
2.1.4 Zusammenfassung und Bezug zur Arbeit . . . 25
2.2 Repräsentation durch semantische Netze . . . 26
2.2.1 Grundbegriffe . . . 26
2.2.2 RDF-basierte Repräsentation . . . 26
2.2.3 Datenmodellbasierte Transformationen . . . 30
2.2.4 Zusammenfassung und Bezug zur Arbeit . . . 31
2.3 Retrieval durch Aktivierungsausbreitung . . . 32
2.3.1 Ablauf der Aktivierungsausbreitung im Information Retrieval . . . 32
2.3.2 Aktivierungsverteilungsfunktionen . . . 33
2.3.3 Beschränkungen . . . 34
2.3.4 Zusammenfassung und Bezug zur Arbeit . . . 35
2.4 Funktionale Sicherheit in der automobilen Systementwicklung . . . 36
2.4.1 Grundbegriffe . . . 36
2.4.2 Sicherheitslebenszyklus . . . 37
2.4.3 Gefährdungs- und Risikoanalyse . . . 40
2.4.4 Zusammenfassung und Bezug zur Arbeit . . . 44
2.5 Verwandte Arbeiten . . . 45
2.5.1 Konsistenzsicherung bei FuSi-Aktivitäten . . . 45
2.5.2 Aktivierungsausbreitung in semantischen Netzen . . . 49

3 Aufbau eines domänenspezifischen IR-Systems für G&Rs **53**
3.1 G&R-Unterstützung als IR-Problem . . . 53
3.1.1 Wiederverwendungspotenziale in der Praxis . . . 53
3.1.2 Lösungsstrategie . . . 57
3.1.3 Zusammenfassung und Beitrag . . . 62
3.2 Aufbau domänenspezifischer semantischer Netze . . . 63
3.2.1 Datenmodell als Basis für die Wissensrepräsentation . . . 63
3.2.2 Aufbau des G&R-Basisnetzes . . . 66
3.2.3 Semantische Anreicherungen . . . 73
3.2.4 Semantische Reduktionen . . . 85
3.2.5 Anfragerepräsentation . . . 87
3.2.6 Zusammenfassung und Beitrag . . . 89
3.3 Aktivierungsausbreitung in semantischen Netzen . . . 90
3.3.1 Aktivierungsverteilungsfunktionen . . . 90
3.3.2 Beschränkungen . . . 94
3.3.3 Algorithmus . . . 95
3.3.4 Steuerung des Ausbreitungsverhaltens durch Ausbreitungsmodi . . . 97
3.3.5 Zusammenfassung und Beitrag . . . 105
3.4 Automatische Aufbereitung der Ergebnisse . . . 106
3.4.1 Suchzielspezifische Filterung . . . 106

3.4.2 Generierung von Erklärungen für Empfehlungen . . . 108
3.4.3 Zusammenfassung und Beitrag . . . 114
3.5 Prototypische Umsetzung . . . 115
3.5.1 Nutzersichten . . . 115
3.5.2 Aufbau . . . 116

4 Evaluation **119**
4.1 Evaluationsplan . . . 119
4.1.1 Komparative Evaluationsdurchführung . . . 119
4.1.2 Bestimmung der objektiven Relevanz . . . 123
4.1.3 Berechnung der Systemrelevanz und Distributionen . . . 129
4.1.4 Evaluationsfragen . . . 129
4.2 Urteilerübereinstimmung . . . 130
4.2.1 Brennan-Prediger- und Gwet-AC_1-Koeffizient . . . 130
4.2.2 Konkordanzkoeffizient nach Kendall . . . 133
4.2.3 Übereinstimmung der zusammengefassten objektiven Relevanz . . . 135
4.3 Explorative Evaluation des IR-Systems für zufällige Konfigurationen . . . 137
4.3.1 Durchführung . . . 137
4.3.2 Auswertung . . . 138
4.3.3 Fazit . . . 144
4.4 Konfigurationsoptimierung mit evolutionären Algorithmen . . . 145
4.4.1 Durchführung . . . 145
4.4.2 Auswertung der anfrageübergreifenden Optimierung . . . 146
4.4.3 Auswertung der anfragespezifischen Optimierung . . . 156
4.4.4 Relevanz und Nützlichkeit der Ergebnisse des IR-Systems . . . 170
4.4.5 Auswahl geeigneter Konfigurationen . . . 172
4.5 Fazit und Bewertung . . . 174

5 Semantisches Netzwerkskelett **175**
5.1 Syntax und Semantik . . . 175
5.2 Erstellen semantischer Netzwerkskelette . . . 178
5.3 Verwendung als Hilfsmittel . . . 179
5.4 Simulation der Aktivierungsausbreitung auf dem Netzwerkskelett . . . 181
5.4.1 Vorgehen . . . 181
5.4.2 Experimente . . . 185
5.4.3 Analysepotenziale der Simulationsergebnisse . . . 188
5.5 Zusammenfassung und Beitrag . . . 189

6 Zusammenfassung und Ausblick **191**
6.1 Zusammenfassung der Ergebnisse . . . 191
6.2 Ausblick . . . 194

Literaturverzeichnis **199**

Abkürzungsverzeichnis **211**

Abbildungsverzeichnis **214**

Tabellenverzeichnis **216**

A Transformationen **217**

B Zusätzliche Ergebnisse der Datenanalyse **218**
B.1 Auswertung der Redundanz im G&R-Datensatz 218
B.2 Auswertung der Tokenanzahl in Tripeln mit Literalen 219

C Zusätzliche Evaluationsergebnisse **221**
C.1 Urteilerübereinstimmung 221
C.1.1 BP- und Gwet-Koeffizienten für nicht reduzierte Poolgröße 221
C.1.2 BP- und Gwet-Koeffizienten für reduzierte Poolgröße 223
C.1.3 Konkordanzkoeffizient nach Kendall für reduzierte Poolgröße 225
C.2 Paarweise Übereinstimmung zwischen zusammengefasster objektiver Relevanz und Raterbewertung 226
C.3 Auswertung der Evaluationsergebnisse für zufällig generierte Konfigurationen 228
C.3.1 Wahl des kombinierten Bewertungsmaßes 228

1 Einleitung

Die Systementwicklung im Automobilbereich zeichnet sich durch wachsende Komplexität aufgrund einer ständigen Zunahme von softwarebasierten Fahrzeugfunktionen und einer steigenden Individualisierung für den Nutzer bei immer kürzer werdenden Produktionszyklen aus [Bro06]. Verstärkt wird dies durch hohe und sogar steigende Anforderungen an die funktionale Sicherheit [Win13]. Die Komplexität durchzieht alle Ergebnisse des Systementwicklungsprozesses, die sogenannten Entwicklungsartefakte, beispielsweise Anforderungsdokumente, Testspezifikationen oder Sicherheitsanalysen. Hohe Komplexität erzeugt zumeist hohen Entwicklungsaufwand. Dennoch existieren Wiederverwendungspotenziale für Entwicklungsartefakte auf System-, Komponenten- und Funktionsebene. Zum einen entstehen sie über Baureihen hinweg aufgrund der Ähnlichkeit wiederkehrender und weiterentwickelter Systeme. Zum anderen ergeben sie sich auch innerhalb von Baureihen aufgrund der Ähnlichkeit der Systeme. Ziel einer effizienten Systementwicklung ist die Reduzierung des Aufwands für wiederkehrende Aufgaben und daher die Wiederverwendung von Artefakten oder Teilen von Artefakten, was idealerweise auch zu einer Erhöhung der Konsistenz führt und mit stärkerer Automatisierung einhergeht.

Eine besondere Herausforderung stellt die Automatisierung stark erfahrungs- und expertenbasierter Aufgaben dar, beispielsweise durch die Wiederverwendung von Erfahrungswissen, welches in Entscheidungen geflossen ist. Hierbei spielen semantische Zusammenhänge im Wissen eine tragende Rolle und beeinflussen dessen Wiederverwendbarkeit. Entwicklungsartefakte aus dem Bereich der funktionalen Sicherheit basieren besonders stark auf dem Wissen von Experten. Die Gefährdungs- und Risikoanalyse (G&R) gilt als aufwendige Sicherheitsanalyse, die am Anfang jeder Entwicklung von sicherheitskritischen Systemen steht. Der Aufwand ist zwar auch im Umfang der G&Rs mit ihren vielen Teilschritten begründet, vorrangig aber in ihrem stark erfahrungsbasierten Charakter. Eine Studie zur funktionalen Sicherheit in der Automobilbranche [MSB13] ermittelt Erfahrung mit 44% als das am meisten verwendete Hilfsmittel für Ergebnisse der G&R. Wissen und Erfahrung aus vergangenen G&Rs spielen eine wesentliche Rolle. Dieses Wissen ist zwar in den jeweiligen Artefakten dokumentiert, allerdings sind die Zusammenhänge durch die enorme Menge an dokumentiertem Wissen für die Experten zunehmend schwerer zu erfassen. Hier existiert ein Mangel an automatisierter Hilfestellung. Mit der vorliegenden Arbeit unterstützen wir die Durchführung von G&Rs durch Empfehlungen auf Basis des vorhandenen Domänenwissens.

Im Folgenden beleuchten wir den Entwicklungsprozess automobiler Systeme mit besonderem Fokus auf die Anforderungen zur funktionalen Sicherheit. Ein Schwerpunkt liegt auf den Potenzialen für eine effizientere und konsistentere Durchführung von G&Rs. Nach Darlegung der speziellen Herausforderungen, die mit der Nutzung dieser Potenziale einhergehen, stellen wir die Beiträge sowie den Aufbau dieser Arbeit vor.

Funktionale Sicherheit in der automobilen Systementwicklung

Im Jahr 2011 wurde die Norm ISO 26262 [ISO26262] „Road Vehicles – Functional Safety“ veröffentlicht. Sie adressiert Gefährdungen, die durch Fehlfunktionen sicherheitsrelevanter elektrischer und elektronischer Systeme in Straßenfahrzeugen entstehen können. Zudem beschreibt sie Methoden, Analysen und Vorgehen zur Berücksichtigung der funktionalen Sicherheit (FuSi) bei der Systementwicklung. Um sicherheitsrelevante Systeme gemäß dem aktuellen Stand der Technik zu entwickeln, ist die Anwendung der Norm verpflichtend. FuSi-Aktivitäten gliedern sich in die verschiedenen Systementwicklungsphasen ein und werden

somit auf verschiedenen Abstraktionsebenen durchgeführt. Die hohen und sogar steigenden Anforderungen an die Sicherheit von elektronischen Systemen im Fahrzeug haben großen Einfluss auf die Entwicklungsprozesse und -werkzeuge [SZ16, Win13].

G&R Am Anfang jeder FuSi-konformen Systementwicklung steht die G&R. Ihre Ausgangsbasis ist die Beschreibung eines zuvor definierten Untersuchungsgegenstandes auf System-, Funktions- oder Komponentenebene mittels vorhandener Entwicklungsunterlagen, beispielsweise Schnittstellenbeschreibungen und Architekturentscheidungen. Daraus werden mögliche Fehlfunktionen identifiziert, deren Auswirkungen in verschiedenen Betriebssituationen untersucht und deren Risiken bewertet werden. Aus diesen Bewertungen ergeben sich das entsprechende Automotive Safety Integrity Level (ASIL) und notwendige Sicherheitsziele können abgeleitet werden. Diese sind Basis für eine spätere Bestimmung von technischen Maßnahmen zur Risikominimierung. Die G&R ist eine unverzichtbare, aber aufwendige FuSi-Aktivität. Der hohe Aufwand entsteht, weil die Analyse aus vielen Teilschritten besteht, zudem stark erfahrungsbasiert ist und daher weitestgehend manuell durch Domänenexperten durchgeführt wird [MSB13]. Eine Steigerung des Aufwands ergibt sich oftmals durch die vielen Anpassungszyklen in den Spezifikationen im Rahmen eines iterativen Systementwicklungsprozesses, was die mehrfache Ausführung beziehungsweise Anpassung der G&R für veränderte Systeme erfordert [MSB13].

Wiederverwendungspotenziale Potenziale zur Automatisierung existieren meist da, wo sich Arbeitsschritte wiederholen, im Falle der G&R bei den Entscheidungsprozessen für oftmals ähnliche Szenarien. Wiederverwendungspotenziale existieren dort, wo für ein zu bewertendes Szenario das Wissen über Entscheidungen in ähnlichen Szenarien verfügbar und nutzbar ist. Allerdings können kleine Unterschiede oder Veränderungen in der Ausgangssituation zu komplexen Veränderungen in der Betrachtung und damit zu anderen Entscheidungen führen. Dennoch sind in der Praxis Ähnlichkeiten in den Teilergebnissen aufgrund von Gemeinsamkeiten in der Ausgangssituation beobachtbar. Alleine bei den Hybridantrieben existieren für verschiedene Baureihen verschiedene Antriebskonfigurationen wie Front-, Hinterrad- oder Einzelradantrieb. Hier sind beispielsweise erneute Durchführungen der G&R für die verschiedenen Antriebskonfigurationen erforderlich, die dennoch einige Gemeinsamkeiten besitzen. Aber auch innerhalb einer Baureihe gibt es Gemeinsamkeiten in der Entscheidungsfindung, beispielsweise durch querschnittliche Gefährdungen wie Brand. Komplett unterschiedliche Systeme wie das Antriebssystem und die Standheizung können bezüglich Gefährdungen durch Brand Gemeinsamkeiten in den Teilergebnissen der Analysen aufweisen, beispielsweise in den Effekten, erwarteten Reaktionen oder Sicherheitszielen.

Informationsbedürfnis Ziel ist es, das Wissen, welches bereits in die Bewertung ähnlicher Situationen geflossen ist, nutzbar zu machen. Besonderes Interesse gilt hierbei den Entscheidungen, die Experten schon einmal basierend auf dem Zusammenspiel komplexer Umstände getroffen haben. Das Informationsbedürfnis liegt demnach nicht in der konkreten Abfrage von Daten mit einer erwarteten Lösung, die entweder richtig oder falsch ist, sondern vielmehr in der Extraktion von nützlichen beziehungsweise relevanten Daten für eine zu lösende Problemstellung. Solche relevanten Informationen können bei der Durchführung von G&Rs in jedem Schritt hilfreich sein. Die Bewertung, ob vermeintlich relevante Teilergebnisse aus abgeschlossenen G&Rs für die konkrete Problemstellung anwendbar sind, muss dennoch die Person treffen, welche die Analyse durchführt. Ausschlaggebend für die Qualität einer automatisierten Unterstützung ist nicht die Ausgabe definitiver Entscheidungen, sondern die Versorgung der Experten mit relevanten Informationen, die sie befähigen, eine Entscheidung gut und effizient zu treffen. Als Konsequenz verstehen wir das Informationsbedürfnis als Bedarf an Empfehlungen zur Entscheidungsunterstützung.

Lösungsidee und Herausforderungen

domänenspezifisches IR-System

Empfehlungen als Ergebnisse von Empfehlungssystemen finden unter anderem Anwendung im World Wide Web, wo sie inzwischen fester Bestandteil der meisten Online-Verkaufsplattformen sind. Hauptsächlich werden hier Produkte empfohlen. Dafür werden Produkt- und Nutzerdaten gesammelt und oftmals kollaborativ ausgewertet. Darauf basierend werden mittels spezieller Algorithmen zur Informationsgewinnung vermeintlich relevante Produkte vorgeschlagen. Solche Systeme profitieren meist von der Größe der verfügbaren Datenmengen. Im Gegensatz dazu fehlt für Empfehlungen in kleinen bis mittleren Domänen wie bei G&Rs die große Anzahl an Nutzern, deren Verhalten und Feedback für die Auswertungen genutzt werden können. Außerdem fällt das vorhandene Domänenwissen ungleich kleiner aus als die Fülle an Daten, die Online-Empfehlungssystemen zur Verfügung steht. Zwei Herausforderungen stehen im Fokus dieser Arbeit: Zum einen, wie sich ein solches domänenspezifisches Erfahrungswissen mit den semantischen Zusammenhängen gut repräsentieren lässt, damit zum anderen eine geeignete Informationsrückgewinnung aus diesen Daten erfolgen kann, so dass die Durchführenden von G&Rs in ihrem Entwicklungsprozess gut unterstützt werden. Diese beiden Herausforderungen sind auch die Kernaufgaben von IR-Systemen [Hen08]. In dieser Arbeit entwickeln wir daher ein domänenspezifisches IR-System zur Unterstützung von G&Rs während der Systementwicklung im Automobilbereich.

Wissensrepräsentation

Um das Wissen nutzbar zu machen, muss es in eine Form gebracht werden, welche dessen Verwendung überhaupt erst ermöglicht, in eine sogenannte Wissensrepräsentation. Konzepte zur Wissensrepräsentation unterscheiden sich unter anderem in ihrer Ausdrucksstärke und sollten zur hier geplanten Anwendung passen. Im Gegensatz zum Daten-Retrieval, also der Abfrage von Daten oder Fakten aus einer Datenbasis, beispielsweise aus einer relationalen Datenbank, ist die Zugehörigkeit zur Lösungsmenge beim IR aufgrund vager, unsicherer und unvollständiger Informationen nicht eindeutig bestimmbar und auch nicht notwendig. Stattdessen sind auch Teiltreffer oder bestmögliche Treffer als Ergebnis einer Abfrage interessant, denn der Fokus liegt auf der Extraktion von nützlichen und relevanten Informationen [Hen08]. Daher ist der Begriff Relevanz ein Schlüsselbegriff im IR [BYR99]. Semantische Netze erlauben die Repräsentation komplexen Wissens und semantischer Zusammenhänge. Formal ist ein semantisches Netz ein gerichteter Graph, bei dem die Knoten die Konzepte, also Begriffe, repräsentieren und die Kanten die Relationen zwischen den Konzepten.

Informationsextraktion

Eine weitere Herausforderung ist der potenzielle Informationsverlust bei der Wissensrepräsentation. Das Erfahrungswissen steckt in den Köpfen der Experten und fließt in ihre Entscheidungen zu den einzelnen Teilschritten der G&R ein. In dem Softwarewerkzeug, mit dem die Analysen erstellt werden, sind die Entscheidungen in Form strukturierter Daten abgelegt. Das in die Entscheidung geflossene Wissen lässt sich allerdings nur sehr schwer wieder vollständig aus den abgelegten Informationen extrahieren. Hier entsteht immer ein Informationsverlust, sowohl auf dem Weg vom Kopf zur Tastatur, als auch auf dem Weg von strukturierten Daten in spezielle Wissensrepräsentationen wie semantische Netze. Zudem erfordert die Wissensrepräsentation meist aufwendigen manuellen Einsatz von Domänenexperten, sowohl in der initialen Erstellung als auch in der Wartung und Pflege. Allerdings gilt ein erheblicher Mehraufwand in der Praxis oftmals als nicht akzeptabel. Daher verfolgen wir in dieser Arbeit einen Ansatz, der sich das Datenmodell des Entwicklungswerkzeugs zunutze macht. Dadurch ermöglichen wir zum einen die vollständige Automatisierbarkeit beim Erstellen der semantischen Netze und zum anderen die Integrierbarkeit in das Entwicklungswerkzeug.

geeignete Suchalgorithmen

Neben der geeigneten Repräsentation der Informationen ist auch die Art der Identifikation relevanter Informationen im IR von Bedeutung. Eine Möglichkeit zur Relevanzbewertung in semantischen Netzen bietet die graphbasierte semantische Suche durch Aktivierungsausbreitung (*engl.:* spreading activation). Sie unterstützt das semantische Durchsuchen der Netze und liefert Informationen über die gegenseitige Relevanz von Knoten bezüglich der spezifischen Suchziele. Aktivierungsausbreitung ist ein assoziatives Retrievalverfahren. Es markiert wellenartig das mit einem oder mehreren Startknoten in Zusammenhang stehende und damit relevante Teilnetz und berechnet für die darin enthaltenen Knoten sogenannte Aktivierungswerte, die deren Relevanz widerspiegeln. Es existieren viele Varianten der Aktivierungsausbreitung, die sich vor allem in der Formulierung der Aktivierungsfunktion und der Art der Terminierung unterscheiden und beispielsweise mit verschiedenen Beschränkungen (*engl.:* constraints), Knoten- und Kantengewichten, bevorzugten Pfaden und Schwellwerten arbeiten [Cre97]. In dieser Arbeit gehen wir der Frage nach, wie das Konzept der Aktivierungsausbreitung in einem IR-System für G&Rs sinnvoll eingesetzt werden kann.

Konfigurierbarkeit

Die Qualität der Ergebnisse hängt stark von der jeweiligen Konfiguration des Suchalgorithmus ab sowie von den Strukturen des Netzes, auf dem der Algorithmus operiert. Die Einflüsse können einzeln und in Kombination zu beobachtbaren Effekten während der Aktivierungsausbreitung führen. So stoppt die Aktivierung möglicherweise in sogenannten Senken, wenn die Konfiguration sehr restriktiv bezüglich bestimmter Netzwerkeigenschaften ist. Die Analyse dieser Einflüsse und Effekte ist aufgrund der Vielzahl an Einflussfaktoren sehr komplex. Die Identifikation optimaler Konfigurationen ist daher sehr schwer. Eine umfangreiche, qualitative Evaluation diverser Einflussfaktoren liefert Erkenntnisse zu deren Auswirkungen auf die Qualität der Ergebnisse. Da semantische Netze aufgrund ihrer Größe oftmals unübersichtlich sind, entwickeln wir als eine abstrakte, reduzierte Art der Darstellung das semantische Netzwerkskelett. Zudem stellen wir in dieser Arbeit ein Verfahren zur Simulation der Aktivierungsausbreitung vor, das eine Basis für die automatisierte Konfiguration der Aktivierungsausbreitung legt.

generischer Ansatz

Der Bedarf an einer Unterstützung durch IR-Systeme und die damit zusammenhängenden, beschriebenen Herausforderungen bei der Wiederverwendung von Wissen existieren nicht nur bei der Anfertigung von G&Rs und sind nicht auf die Systementwicklung im Automobilbereich beschränkt. Stattdessen gibt es in anderen Domänen ähnliche Herausforderungen. Diese Domänen können ebenfalls von domänenspezifischen IR-Systemen profitieren. Daher generalisieren wir die Lösungen in einem Rahmenwerk und machen sie durch eine Methode zum Aufbau domänenspezifischer IR-Systeme nutzbar.

Beiträge im Überblick

Die Abbildung 1.1 beschreibt die drei Hauptbeiträge dieser Arbeit. Beitrag A besteht in der Entwicklung eines Rahmenwerks für den Aufbau domänenspezifischer IR-Systeme, was auf der linken Seite der Abbildung dargestellt ist. Dieses Rahmenwerk enthält generische Lösungen für den Aufbau der einzelnen Komponenten domänenspezifischer assoziativer IR-Systeme. Dazu gehören zunächst Techniken zum automatisierten Aufbau semantischer Netze sowie zur semantischen Anreicherung oder zur Reduktion. Zudem nehmen wir eine Konzeptualisierung der semantischen Suche durch Aktivierungsausbreitung vor und erweitern sie um Ausbreitungsmodi, eine neuartige Technik zur direkten Steuerung des Ausbreitungsverhaltens in den semantischen Netzen. Um die Nutzerakzeptanz zu erhöhen, stellen wir außerdem ein neuartiges Verfahren zur automatisierten Erklärung der Ergebnisse des IR-Systems vor.

Als Beitrag B beschreiben wir in einer aus vier Schritten bestehenden Methode die Aktivitäten

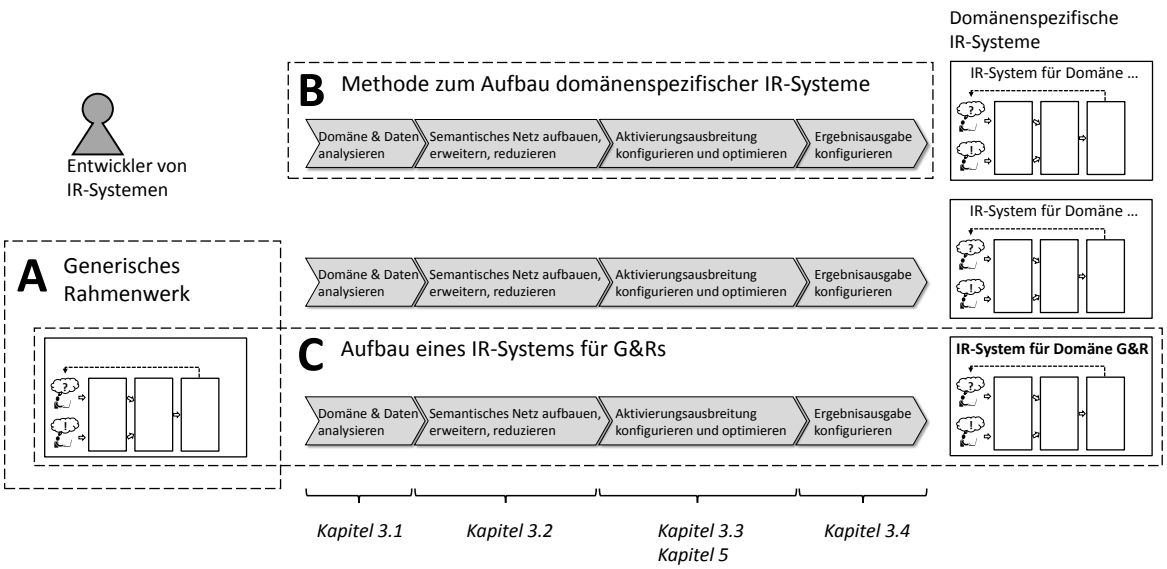


Abbildung 1.1: Beiträge dieser Arbeit im Überblick

beim Aufbau eines domänenspezifischen IR-Systems aus dem generischen Rahmenwerk. Der Entwickler eines domänenspezifischen IR-Systems trifft nach initialer Analyse von Daten und Domäne sukzessive Entscheidungen für die Auswahl und Konfiguration der Komponenten sowie deren Optimierung. Der generische Rahmen vereinfacht den Aufbau neuer IR-Systeme erheblich. Beispielsweise wird durch die Fokussierung auf das vorhandene Datenmodell der manuelle Aufwand, der üblicherweise bei der Wissensrepräsentation entsteht, deutlich reduziert. Wir unterstützen den Entwickler domänenspezifischer IR-Systeme zudem mit Techniken, die insbesondere die anfängliche Datenanalyse und Konfiguration der Aktivierungsausbreitung vereinfachen. Zum einen stellen wir das semantische Netzwerkskelett als Abstraktion semantischer Netze vor. Die Visualisierung des Netzwerkskelettes macht große semantische Netze wesentlich überschaubarer. Zum anderen ermöglichen wir die Simulation der Aktivierungsausbreitung auf diesem Netzwerkskelett, um frühzeitig ungeeignete Konfigurationen auszuschließen.

Als Beitrag C zeigen wir die Machbarkeit der Methode und verwenden das generische Rahmenwerk beim Aufbau eines domänenspezifischen, assoziativen IR-Systems zur Unterstützung von G&Rs. Dafür verwenden wir reale Daten aus der industriellen Praxis. Ein Prototyp ermöglicht eine umfangreiche Evaluation, die zum einen eine gute Qualität der Ergebnisse des IR-Systems für G&Rs zeigt und zum anderen Anhaltspunkte für geeignete und ungeeignete Konfigurationen und Netzvarianten liefert. Wir zeigen unter anderem, wie sich Konfigurationen durch Anwendung evolutionärer Algorithmen optimieren lassen. Diese Vorgehensweise kann vom Entwickler domänenspezifischer IR-Systeme adaptiert werden und erweitert daher dessen Handwerkszeug bei der Umsetzung der als Beitrag B vorgestellten Methode.

Die Beiträge dieser Arbeit werden im Folgenden kurz zusammengefasst.

A | Entwicklung eines Rahmenwerks mit generischen und konfigurierbaren Lösungen für den Aufbau domänenspezifischer IR-Systeme

- Aufbau semantischer Netze und Techniken zu deren Anreicherung und Reduktion
- Konzeptualisierung durch Vereinheitlichen, Formalisieren und Generalisieren der Aktivierungsausbreitung
- Erweiterung des Forschungsstandes um neuartige Techniken zur direkten Steuerung des Ausbreitungsverhaltens durch Ausbreitungsmodi
- Verfahren zur automatisierten Erklärung der Ergebnisse

B | Methode zum Aufbau domänenspezifischer IR-Systeme unter Verwendung des Rahmenwerks

- Erhebliche Reduktion des Aufwands beim Aufbau neuer domänenspezifischer IR-Systeme durch Fokussierung auf das Datenmodell statt manuellem Aufbau
- Vereinfachung beim Aufbau durch Visualisierung großer unübersichtlicher Netze durch deren Abstraktion als semantisches Netzwerkskelett
- Frühzeitiger Ausschluss ungeeigneter Konfigurationen durch Simulation der Aktivierungsausbreitung auf dem Netzwerkskelett

C | Aufbau eines IR-Systems für die Domäne der Gefährdungs- und Risikoanalysen

- Analyse und Aufbereitung industrieller Daten
- Prototypische Umsetzung durch Implementierung eines eigenständigen IR-Systems
- Evaluation
  - Explorative Analyse zufälliger Konfigurationen
  - Optimierung der Konfigurationen mittels evolutionärer Algorithmen
  - Qualitative Bewertung der Ergebnisse des IR-Systems

Aufbau dieser Arbeit

Die vorliegende Arbeit hat sechs Kapitel. Die zentralen Kapitel und Abschnitte dieser Arbeit sind den Methodenschritten in Abbildung 1.1 zugeordnet.

In Kapitel 2 führen wir die Kernthemen dieser Arbeit ein. Dafür stellen wir zunächst IR-Systeme vor und erläutern insbesondere deren grundsätzlichen Aufbau und Funktionsweise. Da der Fokus dieser Arbeit auf assoziativen IR-Systemen liegt, folgt eine detaillierte Vorstellung der für die Wissensrepräsentation verwendeten semantischen Netze sowie der Aktivierungsausbreitung als Retrievalverfahren. Anschließend geben wir einen Überblick über die funktionale Sicherheit nach ISO 26262 und stellen insbesondere die Gefährdungs- und Risikoanalyse und deren Herausforderungen in der Praxis vor. Abschließend betrachten wir verwandte Arbeiten genauer.

Kapitel 3 widmet sich dem Aufbau des domänenspezifischen IR-Systems für G&Rs und der Entwicklung des generischen Rahmenwerks. In Abschnitt 3.1 analysieren wir dafür zunächst, inwiefern die besonderen Herausforderungen bei der Durchführung von G&Rs als IR-Problem verstanden werden können. Wir extrahieren konkrete Anforderungen an das domänenspezifische IR-System und entwickeln daraus einen grundsätzlichen generischen Rahmen. Die darauffolgenden Abschnitte zeigen, wie wir semantische Netze automatisiert aufbauen, anreichern und reduzieren (Abschnitt 3.2), wie vorhandene und neuartige Konzepte zur Aktivierungsausbreitung als Retrievalverfahren eingebunden werden (Abschnitt 3.3) und die Ergebnisse aufbereitet und erklärt werden können (Abschnitt 3.4). Hier beschreiben wir parallel die generischen Lösungen und die konkrete Umsetzung im domänenspezifischen IR-System für G&Rs. Wir stellen außerdem den Bezug zum entsprechenden Schritt der Methode her, die den Aufbau domänenspezifischer IR-Systeme aus dem generischen IR-System beschreibt.

Kapitel 4 widmet sich einer umfangreichen Evaluation, in die wir Experten aus dem industriellen Umfeld einbeziehen. In Abschnitt 4.2 analysieren wir zunächst die Urteilerübereinstimmung und werden zeigen, dass ein ausreichendes Maß an Einigkeit zwischen den Domänenexperten besteht. Das erlaubt uns, die Ergebnisse des IR-Systems in der darauffolgenden qualitativen Evaluation mit diesen Beurteilungen abzugleichen. In Abschnitt 4.3 beginnen wir mit einer explorativen Analyse zufällig generierter Konfigurationen. Sie erlaubt einen ersten Überblick und zeigt deutlich die Notwendigkeit einer geeigneten Parametrisierung der Konfigurationen und Auswahl der Netze. In Abschnitt 4.4 suchen wir mittels evolutionärer Algorithmen nach optimalen Konfigurationen und analysieren, unter welchen Bedingungen gute Ergebnisse erreicht werden können. Wir werden beispielsweise zeigen, dass die von uns eingeführten Ausbreitungsmodi zur zusätzlichen Steuerung der Aktivierungsausbreitung sowie die semantischen Anreicherungen und Reduktionen der Netze essenziell für qualitativ gute Ergebnisse des IR-Systems sind. Die optimierten Konfigurationen verwenden wir zur Bewertung der Qualität unseres IR-Systems für G&Rs. Wir werden zeigen, dass es mit geeigneten Konfigurationen relevante und nützliche Ergebnisse liefert.

In Kapitel 5 stellen wir das semantische Netzwerkskelett vor, dessen kompakte Darstellungsform die Anwendung der Methode vereinfacht. Außerdem zeigen wir, wie die Simulation der Aktivierungsausbreitung auf dem Netzwerkskelett zum frühzeitigen Ausschluss ungünstiger Konfigurationen genutzt werden kann.

In Kapitel 6 fassen wir die Ergebnisse der Arbeit zusammen und geben einen Ausblick auf weiterführende Arbeiten.

2 | Grundlagen und verwandte Arbeiten

Dieses Kapitel erläutert grundlegende Kernthemen dieser Arbeit. Insbesondere stellen wir Information-Retrieval (IR)-Systeme und ihre zentralen Aufgaben vor, die Repräsentation und Rückgewinnung von Informationen. Den Schwerpunkt der Arbeit bilden assoziative IR-Systeme. Sie verfolgen graphbasierte Ansätze zur Lösung der IR-Aufgaben. Daher betrachten wir semantische Netze als Repräsentationsform und Aktivierungsausbreitung als Rückgewinnungsverfahren für IR-Systeme genauer. Zudem schaffen wir einen Überblick über das praktische Anwendungsfeld dieser Arbeit, welches in der funktionalen Sicherheit im Automobilbereich mit Fokus auf den Gefährdungs- und Risikoanalysen liegt. Abschließend stellen wir Forschungsarbeiten vor, die eine inhaltliche Nähe zu der vorliegenden Arbeit aufweisen.

2.1 Information-Retrieval-Systeme

Computerbasierte Informationsrückgewinnung (*engl.:* information retrieval) hat ihre Anfänge in den späten 1940er Jahren [SC12]. Mooers [Moo50] beschreibt im Jahr 1950 eine neue Herausforderung der Zeit, nämlich, dass gespeicherte (*engl.:* recorded) Informationen nur dann nutzbar sind, wenn deren Existenz und Ablageort in Informationssammlungen auch entdeckt werden können. Mooers kommentiert damit Arbeiten von Wise und Perry [WP50] am *Rapid Selector* der 1940er Jahre, eine der ersten Maschinen zur Kodierung und Speicherung bibliographischer Informationen auf Mikrofilm sowie zur Identifikation der kodierten Informationen. Nach Sanderson und Croft [SC12] wurde damit der neue Forschungsbereich Information-Retrieval erstmals benannt. Mit steigender Rechenleistung und Speicherkapazität sowie der damit einhergehenden Zunahme digitalisierter Daten steigen die Bedarfe, aber auch die Anforderungen an Retrievalverfahren [SC12]. Beschränkte sich der Bedarf anfangs hauptsächlich auf die Informationsgewinnung bibliographischer Daten, so veränderten sich mit zunehmender Technologisierung und Datenmenge auch die Art und Menge von Informationsbedürfnissen. Moderne IR-Systeme ermöglichen durch die Nutzung intelligenter Technologien eine immer bessere Suche in verschiedensten Dokumentenarten wie Nachrichten, Bildern, Videos, Musik oder Webseiten [SC12, MRS08].

2.1.1 Grundbegriffe

Information-Retrieval

Das *Information-Retrieval* zielt darauf ab, in einer *Informationssammlung* jene Informationen zu identifizieren, die für eine konkrete Problemlösung benötigt werden, also zur Befriedigung eines konkreten Informationsbedürfnisses eines Nutzers beitragen [Hen08, MRS08]. Salton [SM83] bezeichnet *IR-Systeme* als Informationssysteme, die sich nicht nur mit der Verarbeitung, Rückgewinnung und Verteilung von Informationen befassen, sondern auch mit Speicherung und Repräsentation von Informationen. IR-Systeme sind Werkzeuge zum IR [Cre97].

Dokument

Die Informationssammlung wird oftmals auch als Dokumentensammlung bezeichnet. Der Begriff *Dokument* steht allerdings für verschiedenartige Elemente. Grundsätzlich existieren keine Einschränkungen bezüglich der Art der Informationsgegenstände, die in einer Dokumentensammlung enthalten sind [SM83, SC12]. Während der Hauptanwendungsbereich 1983 nach Salton [SM83] noch in der Informationsgewinnung aus textuellen Daten in Textdokumenten und bibliographischen Einträgen lag, wird IR inzwischen auf unterschiedlichsten Datengrundlagen eingesetzt. Üblicherweise agieren IR-Systeme auf zumeist großen Datensammlungen

von unstrukturierten oder semistrukturierten Daten [SC12, MRS08].

Informationsbedürfnis

Das *Informationsbedürfnis* entspricht dem subjektiven Bedarf eines Nutzers nach Information. Ein Informationsbedürfnis muss so formuliert werden, dass das IR-System eine diesbezügliche Suche starten kann. Daher wird das Informationsbedürfnis als (Such-)Anfrage (*engl.:* query) in eine maschinenlesbare Form gebracht und so vom IR-System verarbeitet. Eine *Anfrage* ist daher oftmals eine vage Repräsentation eines Informationsbedürfnisses, da das tatsächliche Informationsbedürfnis nicht verlustfrei beschrieben werden kann [Hen08]. Eine detaillierte Auseinandersetzung mit dem Informationsverlust beim Wissenstransfer in IR-Systemen findet in Abschnitt 2.1.2.4 statt.

Anfrage

Die Kernidee des IR liegt nicht in der Gewinnung, sondern in der Rückgewinnung von Informationen: Im Gegensatz zu wissensschaffenden Technologien sind IR-Systeme Informationssysteme, welche den Nutzer über die Existenz oder Nicht-Existenz von Informationen in einer bestehenden Informationssammlung informieren und diese aus der Sammlung extrahieren, dabei aber kein neues Wissen generieren [Lan79]. In diesem Punkt unterscheiden sich IR-Systeme von Frage-Antwort-Systemen oder Expertensystemen, die beispielsweise Inferenz zur Ableitung neuen Wissens einsetzen [Lan79, SM83]. Ebenso wenig steht die direkte Abfrage strukturierter Daten aus Datenbanksystemen im Fokus des IR. Vielmehr identifizieren IR-Systeme innerhalb einer bestehenden Menge an Informationen die relevanten Informationen bezüglich der konkreten Anfrage eines Nutzers [SC12].

Relevanz

Relevanz ist „eine Eigenschaft der Beziehung zwischen der Anfrage und einem einzelnen Element der Antwortmenge“ [Fuh06] und spiegelt die subjektive Nützlichkeit einer Antwort für den Anfragenden wider. Relevant ist das, was der Nutzer als relevant bezüglich seines Informationsbedürfnisses empfindet [MRS08]. Relevanz ist zumeist nicht binär. Auf Basis vager, unsicherer und unvollständiger Informationen und ebenso vager Anfragen kann beim Versuch der Extraktion bedeutsamer Informationen nicht zwangsläufig und ausschließlich in relevante und nicht relevante Ergebnisse unterteilt werden. Vielmehr kann es ein Spektrum an Antworten mit unterschiedlicher Relevanz geben. Daher werden im IR vornehmlich *Best-Match*-Methoden verwendet, deren Ergebnisse nach Relevanz sortiert werden. Das Bewerten der Relevanz und das entsprechende Sortieren der Ergebnisse wird auch als *Ranking* bezeichnet. Im Gegensatz dazu geben *Exact-Match*-Methoden binäre Ergebnisse aus und selektieren die exakte Antwortmenge zu einer Anfrage. Dies entspricht allerdings nicht der Kernidee des IR und wird daher nur von wenigen IR-Modellen beschrieben und in IR-Systemen zumeist nur als Teilmethode oder Erweiterung eingesetzt. Die Relevanz der Antworten ist eine Eigenschaft, nach welcher sich die Qualität eines IR-Systems bewerten lässt. Zur Einschätzung wird zumeist die von einem oder mehreren Nutzern bewertete Relevanz mit der vom IR-System berechneten Relevanz eines Ergebnisobjektes bezüglich des Informationsbedürfnisses verglichen.

Best-Match

Ranking

IR-Systeme erlauben die Bearbeitung unterschiedlicher Aufgaben. Henrich [Hen08] unterscheidet die folgenden vier Aufgaben: Anfragebearbeitung, Klassifikation, Information Browsing und Information Filtering. Den Kern dieser Arbeit bildet ein IR-System, welches dem Aufgabengebiet der Anfragebearbeitung zugeordnet werden kann. *Anfragebearbeitung* stellt die Erfüllung eines konkreten Informationsbedürfnisses in den Vordergrund. Dieser Aufgabenbereich deckt ab, was oftmals intuitiv als IR verstanden wird. Basierend auf dem Informationsbedürfnis wird eine Anfrage an das System gestellt, für welche das System relevante Dokumente zurückgibt. Eine klassische Anwendung dieses Aufgabenbereiches sind Suchmaschinen, welche relevante Internetseiten bezüglich eingegebener Suchbegriffe ausgeben [Hen08, CC04].

Anfragebearbeitung

2.1.2 Arbeitsweise und Aufbau

Abbildung 2.1 zeigt die allgemeine Arbeitsweise eines IR-Systems bei der Anfragebearbeitung.

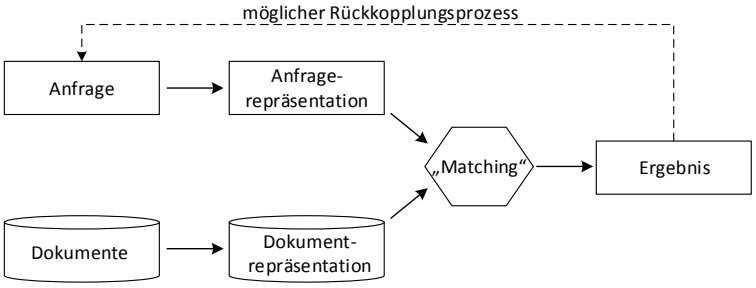


Abbildung 2.1: Allgemeine Arbeitsweise eines IR-Systems nach [Hen08]

Datengrundlage ist eine Sammlung von Dokumenten. Die Dokumente werden in eine Repräsentation überführt, die den Inhalt der Dokumente widerspiegelt. Dieser Schritt wird auch als Informationsextraktion bezeichnet [MBA17]. Ebenso wird die Anfrage zur Befriedigung eines Informationsbedürfnisses in eine geeignete Repräsentation gebracht. Ziel des IR-Systems ist es, für diese Anfrage die relevantesten Dokumente in der Dokumentensammlung zu finden, welche dann als Ergebnis ausgegeben werden. Realisiert wird die eigentliche Suche durch Anwendung einer *Retrievalstrategie*, welche in dieser Abbildung als *Matching* bezeichnet wird. Die Erweiterung um einen Rückkopplungsprozess ist optional. Dieser erlaubt eine Anpassung der Anfrage durch direktes Feedback des Anfragenden zur Relevanz der Ergebnisse. Relevanzfeedback zielt durch Konkretisierung der zumeist vage formulierten Anfragen auf die Verbesserungen der Ergebnisse ab. Aufgrund des Mehraufwands einer zusätzlichen Interaktion mit der Feedback-Komponente besteht die Gefahr geringer Nutzerakzeptanz [Fuh13]. Die Entscheidung, was *geeignete* Repräsentationen und Retrievalstrategien für eine Anwendung sind, ist elementarer Bestandteil der Konzeption von IR-Systemen.

Retrieval-strategie

IR-Modelle spezifizieren, wie zu einer gegebenen Anfrage die Antwortdokumente aus einer Dokumentenkollektion bestimmt werden [Fuh13]. Sie unterscheiden sich grundsätzlich in den Repräsentationen der Dokumente und Anfragen sowie der *Retrievalfunktion* zur Umsetzung der Retrievalstrategie, welche gemeinsam das mathematische Fundament des Modells bilden [Fuh13] und daher aufeinander abgestimmt sind [CC04]. Zudem gehört zu einem IR-Modell eine Vorschrift zur Ableitung der Repräsentationen aus Dokumenten und Anfragen [BYR99, Fuh13]. Jedes IR-Modell beantwortet damit drei grundsätzliche Fragen.

Retrieval-funktion

1. Informationsextraktion: Was wird wie extrahiert?
2. Repräsentationsform: Wie wird das Extrahierte repräsentiert?
3. Retrievalstrategie: Wie wird das Ranking durchgeführt?

Eine klassische Unterteilung nach mathematischem Fundament kann in *mengentheoretische*, *algebraische* und *probabilistische* Modelle vorgenommen werden [BYR99, Kur04]. Hinzu kommen assoziative IR-Modelle [Cre97], die graphbasierte Ansätze verfolgen und im Fokus dieser Arbeit stehen. Im Folgenden werden die drei Komponenten von IR-Modellen kurz beleuchtet.

2.1.2.1 Informationsextraktion

Informationsextraktion dient der Identifikation von Informationen in den Dokumenten für deren anschließende Umwandlung in eine strukturierte beziehungsweise semistrukturierte Repräsentationsform [MBA17]. Die Informationsextraktion sollte an die Anforderungen der jeweiligen Anwendung angepasst sein. Hierbei ist entscheidend, in welcher Detailliertheit und mit welchem semantischen Gehalt die Informationen aus den Dokumenten extrahiert werden. Idealerweise basiert das Konzept zur Informationsextraktion auf einer Abwägung zwischen der gewünschten Aussagekraft und dem Aufwand, der während der Extraktion und der anschließenden Anwendung der Retrievalfunktion entsteht.

Bei textbasierten Dokumenten wird oftmals *Indexierung* angewendet. Dafür werden gezielt Schlüsselwörter, sogenannte *Indexterme*, aus dem Textkorpus extrahiert und den Dokumenten zugeordnet [CEE+10]. Das Ausmaß der Indexierung hängt von den Anwendungen ab und kann stark variieren. So kann eine Sequenz von Wörtern beispielsweise von allen enthaltenen Wörtern, nur von einigen vermeintlich aussagekräftigen oder häufig vorkommenden Wörtern oder sogar nur durch eine das Dokument beschreibende Kategorie repräsentiert werden. Zudem erlauben erweiterte Indexierungsverfahren beispielsweise das Gewichten und die probabilistische Betrachtung von Indextermen [BYR99]. Die automatische Indexierung natürlichsprachlicher Texte wird durch linguistische Verfahren des Natural Language Processing (NLP) unterstützt. Diese bauen oftmals aufeinander auf und werden daher in sogenannten Pipelines aneinandergereiht. Eine typische Pipeline besteht nach Maynard et al. [MBA17] aus den folgenden Komponenten, welche auch in dieser Arbeit verwendet werden.

Tokenisierung (*engl.:* tokenization) bedeutet die Zerlegung eines Textes in einzelne Einheiten, welche als Token bezeichnet werden. Token sind nicht ausschließlich Wörter, sondern auch Satzzeichen oder Formeln. Die Abtrennung der Token erfolgt oftmals an Stelle von Leerzeichen, ist aber sprachabhängig [MSB+14].

Satzzerlegung (*engl.:* sentence splitting) hingegen zerlegt Texte in die enthaltenen Sätze.

Beim **Part-of-Speech (PoS)-Tagging** werden den Token Wortarten zugeordnet, wobei dafür nicht nur das jeweilige Token, sondern auch die Umgebung betrachtet wird. Das Stuttgart-Tübingen-Tagset (STTS) [STST99] stellt den De-facto-Standard für das Tagging von Wortarten in deutscher Sprache dar. Es enthält eine Zuordnung von 48 hierarchisch angeordneten Haupt- und Unterwortarten zu einheitlich verwendbaren Tags, beispielsweise Gattungsnamen (mit dem Tag NN), Eigennamen (NE) oder attributierte Adjektive (ADJA).

Morphologische Analysen zielen auf die Identifikation und Klassifikation der linguistischen Segmente eines Wortes ab [MBA17]. Ein Wort wird sprachabhängig in die Wortbasis und dessen Affixe, beispielsweise Präfix, Infix oder Suffix, aufgeteilt. Morphologische Analysen sind notwendig bei der Stammformreduktion (*engl.:* stemming), bei der Affixe entfernt und Wörter so auf ihre Basis zurückgeführt werden. Bei der Lemmatisierung werden zusätzlich potenzielle Flexionen entfernt.

Bei der **Stoppworteliminierung** werden in einer Menge von Termen jene identifiziert und entfernt, welche für den semantischen Inhalt vermutlich nicht relevant sind. Zu diesen sogenannten Stoppwörtern zählen üblicherweise Funktionswörter wie Artikel oder Präpositionen und andere unspezifische Wörter [CEE+10].

Über die vorgestellten Methoden zur Indexierung hinaus existieren weiterführende Methoden des NLP, welche sich unter anderem mit der Extraktion von Relationen zwischen Termen und deren Verlinkung beschäftigen [MBA17]. In dieser Arbeit finden sie keine Verwendung und werden daher nicht näher erläutert.

2.1.2.2 Repräsentationsform

In mengentheoretischen IR-Modellen werden extrahierte Schlüsselwörter als Menge oder Fuzzy-Menge repräsentiert, beispielsweise im Booleschen Retrievalmodell [Fuh13]. Weitaus größere Verbreitung finden allerdings die diversen vektorbasierten Modelle, die in der Regel den algebraischen Modellen zugeordnet werden. Deren Grundidee besteht darin, einen Vektorraum aufzuspannen, in dem jeder Term der Dokumentensammlung eine Dimension darstellt. Dokumente und Anfragen werden als Vektoren im entsprechenden Vektorraum repräsentiert [Sal68]. Eine Alternative sind assoziative IR-Modelle, die graphbasierte Repräsentationen zur Abbildung von Zusammenhängen beziehungsweise Assoziationen nutzen. Extrahierte Informationen werden oftmals in Form von Netzwerken dargestellt, in denen Informationsgegenstände, beispielsweise Schlüsselwörter, durch Knoten und die Assoziationen zwischen den Informationsgegenständen als Kanten zwischen den Knoten repräsentiert werden [Cre97]. Zu den assoziativen Repräsentationen gehören die in dieser Arbeit verwendeten semantischen Netze, welche in Abschnitt 2.2 gesondert vorgestellt werden.

2.1.2.3 Retrievalstrategie

Mengentheoretische Verfahren verwenden Mengenoperationen für die Rückgewinnung von Informationen aus Dokumenten, welche als Mengen abgebildet sind. Eine Anfrage kann beispielsweise ein Boolescher Ausdruck sein, der Terme mittels Boolescher Operatoren wie AND, OR oder NOT verknüpft [Fuh13]. Algebraische Verfahren nutzen zur Repräsentation Vektoren, Matrizen oder Tupel und setzen bei der Rückgewinnung auf die Berechnung ihrer gegenseitigen Ähnlichkeit. Beispielsweise wird als Retrievalfunktion für das Vektorraummodell, welches zu den algebraischen Modellen gehört, die Ähnlichkeit zwischen Anfrage- und Dokumentenvektor berechnet, im einfachsten Fall mittels des Skalarproduktes [Sal68]. Probabilistische Verfahren beziehen Wahrscheinlichkeiten und probabilistische Theoreme in die Ähnlichkeitsberechnungen ein [Kur04]. Assoziative Retrievalverfahren nutzen graphbasierte Strukturen zur Berechnung von Ähnlichkeitsmaßen auf Graphen. Zu diesen Verfahren gehört die in dieser Arbeit verwendete Aktivierungsausbreitung, welche in Abschnitt 2.3 genauer erläutert wird.

2.1.2.4 Wissenstransfer

Das IR kann als *Prozess des Wissenstransfers* bezeichnet werden [Hen08]. Der Wissenstransfer stellt eine Herausforderung dar, weil er anfällig für Informationsverlust ist. Daher werden im Folgenden die Begriffe Daten, Wissen und Information, welche in engem Zusammenhang stehen, genauer betrachtet. Anschließend wird der Wissenstransfer in IR-Systemen mit Fokus auf potenzielle Informationsverluste beschrieben.

Wissen, Daten, Informationen

Daten sind gemäß Kompakt-Lexikon der Wirtschaftsinformatik „zum Zweck der Verarbeitung zusammengefasste Zeichen“ [FKL+13]. Die einfache Aneinanderreihung von Zeichen besitzt keine Semantik, hat also vorerst keine Bedeutung [Uri08]. Daten existieren, ohne dass bekannt ist, in welchem Zusammenhang und in welcher Umgebung, also in welchem *Kontext* sie gelten. Daten können allerdings „aufgrund bekannter oder unterstellter Abmachungen Informationen [...] darstellen“ [FKL+13]. *Informationen* entstehen aus Daten genau dann, wenn über den Kontext ein Verständnis über die Beziehungen zwischen Daten oder Datensammlungen

Daten

Information

entsteht und diese in einen Zusammenhang mit anderen Informationen gebracht werden und eine Bedeutung erhalten [Uri08]. Informationen sind interpretierte Daten, abhängig vom Kontext, aber ohne Verständnis über Ursachen, Gründe, mögliche Veränderungen oder Entwicklungen [Uri08]. Dieses zusätzliche Verständnis darüber, warum etwas so ist wie es ist, sowie das Erkennen von Mustern zwischen Daten und Informationen und deren Bedeutung definiert den Begriff *Wissen*. Wissen existiert nicht nur in einem Kontext, sondern erzeugt auch neuen Kontext [Uri08]. Dabei können sich Muster dynamisch verändern. Das Verständnis darüber ermöglicht die Vorhersagbarkeit und Schlussfolgerung und damit die Ableitung und Erzeugung neuen Wissens [Uri08].

Wissen

Es lassen sich grundsätzlich zwei Typen von Wissen unterscheiden [Uri08], implizites (*engl.:* tacit) Wissen und explizites Wissen. Als *implizites* Wissen wird persönliches, in den Köpfen vorhandenes Wissen bezeichnet. Es ist das Ergebnis von Lernprozessen und Erfahrungen und kann sich durch weitere Beobachtungen verändern. Da das implizite Wissen subjektive Wahrnehmungen enthält, gilt es als kontextspezifisch. Dadurch ist es schwer vollständig in ganzer Komplexität weiterzugeben. *Explizites* Wissen hingegen ist jenes Wissen, welches in irgendeiner Form gespeichert und damit für andere Personen verfügbar ist, beispielsweise in Dokumenten oder Datenbanken [Uri08].

implizites Wissen

explizites Wissen

Als *Domänenwissen* wird das Wissen auf einem speziellen Gebiet, der *Domäne*, bezeichnet [BH08]. Wer über umfangreiches Domänenwissen verfügt, wird *Experte* des Gebietes genannt.

Domänenwissen

Informationsverlust beim Wissenstransfer im IR-System

Betrachtet man die Arbeitsweise eines IR-Systems in Abbildung 2.1 (Seite 11) mit Fokus auf den Wissenstransfer, so sind zusätzlich zwei Akteure involviert: ein Experte als Wissensproduzent, der die Dokumente erstellt, und ein Betrachter, dessen implizites Informationsbedürfnis durch die Anfrage befriedigt werden soll. Beim Erstellen eines Dokumentes durch einen Wissensproduzenten wird implizites Wissen explizit gemacht, wobei *Unvollständigkeit* entsteht, weil die Komplexität des Wissens zumeist nur begrenzt abgebildet wird [Pas03, Hen08]. Gerade bei komplexen Sachverhalten spielt der Kontext eine große Rolle. Bei Abbildung des eigenen subjektiven Wissens kann allerdings schwer bestimmt werden, welche kontextbehafteten Informationen einer anderen Person zum gleichen Verständnis verhelfen würden, zumal diese ebenso ein eigenes implizites Wissen mitbringt. *Unsicherheit* (*engl.:* uncertainty) hingegen entsteht durch unscharfe Formulierung von Aussagen. Zudem tragen *Ungenauigkeiten* (*engl.:* imprecision) und *Mehrdeutigkeiten* (*engl.:* ambiguity) zu einem von der ursprünglichen Intention abweichenden Verständnis bei. *Inkonsistenzen* zwischen verschiedenen Dokumenten können beispielsweise auftreten, wenn verschiedene Personen mit unterschiedlichen Vorstellungen Wissen beitragen. Dem unsicheren und unvollständigen explizit gemachtem Wissen stehen auf der anderen Seite implizite Informationsbedürfnisse und vage Anfragen gegenüber [Pas03, Hen08]. Die *Vagheit* von Anfragen im IR entsteht, weil es anders als im Daten-Retrieval keine spezielle Erwartung an die Antwort gibt [Hen08]. Es geht um die Identifikation der relevantesten Elemente bezüglich des subjektiven Informationsbedürfnisses, ohne zu wissen, ob überhaupt geeignete, relevante Elemente existieren und ohne diese direkt in Form eines Daten-Retrievals abzufragen, wie das beispielsweise in einer Datenbank möglich und üblich wäre. Auch die Ergebnisse sind von Unsicherheit und Vagheit betroffen. Ergebnisdaten werden durch Interpretation des Informationsnachfragenden zu Informationen. Durch den Kontext und mithilfe seines eigenen impliziten Wissens können die Informationen zu Wissen werden.

2.1.3 Evaluation im Information Retrieval

In diesem Abschnitt stellen wir die Besonderheiten bei der Evaluation von IR-Systemen vor. Insbesondere beleuchten wir die Aspekte, die für die Evaluation des IR-Systems für G&Rs in dieser Arbeit von Bedeutung sind. Wir beschreiben daher zunächst den komparativen Evaluationsaufbau, der es grundsätzlich ermöglicht, die Qualität von Ergebnissen zu vergleichen. Für die Beurteilung sind allerdings keine Testdaten verfügbar, welche die Erwartungen an die Ergebnisse widerspiegeln. Daher beschreiben wir, wie die Einschätzungen von Experten aus dem industriellen Umfeld trotz großer Datenmengen mit vertretbarem Aufwand einbezogen werden können. Anschließend stellen wir die Bewertungsmaße vor, die in dieser Arbeit zum Einsatz kommen, um zum einen die Qualität der Ergebnisse zu bewerten und zum anderen eine ausreichende Übereinstimmung der Experten zu garantieren.

2.1.3.1 Kriterien

Bei der Evaluation von IR-Systemen unterscheidet man grundsätzlich die zwei Kriterien Effektivität und Effizienz [Fuh06]. *Effizienz* bezeichnet einen möglichst schonenden Umgang mit Systemressourcen wie Speicherplatz, CPU-Zeit, Anzahl benötigter Operationen oder Antwortzeiten [Fuh06]. Die weitaus wichtigere Rolle nimmt im IR die Untersuchung der *Effektivität* ein. Diese wird von Fuhr als Verhältnis zwischen einerseits Zeitaufwand und Belastung des Nutzers und andererseits dem Nutzen durch die Qualität des IR-Systems beschrieben [Fuh06]. Im Zentrum der Effektivitätsbewertung liegt bei IR-Systemen üblicherweise die Betrachtung der Qualität der Ergebnisse.

Effizienz

Effektivität

Die Qualität eines Ergebnisses eines IR-Systems bezogen auf ein Informationsbedürfnis zeichnet sich wegen des zumeist verwendeten Best-Match-Ansatzes nicht durch Korrektheit aus, sondern vielmehr durch dessen Relevanz. Relevanz kann als binäre Eigenschaft interpretiert werden, also ein Ergebnis ist entweder relevant oder nicht. Intuitiver ist allerdings die Betrachtung von Relevanz auf einer Skala, auf der Elemente einer Antwortmenge mehr oder weniger relevant sein können. Der in Abschnitt 2.1.1 vorgestellte Relevanzbegriff wird im Kontext der Evaluation von IR-Systemen nach Fuhr [Fuh06] um zwei Aspekte erweitert. Als *Systemrelevanz* wird demnach die vom IR-System geschätzte beziehungsweise durch dessen Retrievalfunktion berechnete Relevanz eines Elementes bezüglich einer Anfrage verstanden. Die Systemrelevanz stellt also das Ergebnis eines IR-Systems dar. Als *objektive Relevanz* wird hingegen die von einem oder mehreren Beobachtern, sogenannten *Ratern*, beurteilte Relevanz bezeichnet. Die Rater bewerten dafür die Beziehung zwischen dem geäußerten Informationsbedürfnis und jedem Element der Antwortmenge. Diese Bewertungen, im Folgenden auch als Ratings bezeichnet, sollen zwar Objektivität gewährleisten, sind allerdings grundsätzlich subjektiv. Damit sie als objektiv gelten können, muss sichergestellt werden, dass ein ausreichendes Maß an Einigkeit bezüglich der Bewertungen erreicht wird. Ist diese sogenannte Interrater-Reliabilität nicht gegeben, liegt also keine ausreichende Urteilerübereinstimmung vor, so sind die Ratings für die Evaluation nicht geeignet.

Systemrelevanz

objektive Relevanz

Rater

2.1.3.2 Durchführung

Heutige Ansätze zur Evaluation von IR-Systemen basieren auf dem Cranfield-Paradigma, welches auf die Cranfield-Experimente [Cle67] der 1960er Jahre zurückgeht. Demnach basiert die Evaluation der Effektivität eines IR-Systems üblicherweise auf einer sogenannten *Testkollektion*. Diese besteht in der Regel aus drei Zutaten: einer Dokumentensammlung beziehungsweise einem Datensatz, einer Menge an Informationsbedürfnissen, die als Anfragen

Testkollektion

an das zu testende System gegeben werden können, und objektive Relevanzbeurteilungen für jede Kombination aus Dokumentensammlung und Anfrage, welche die Wahrheit (*engl.:* ground truth) darstellen sollen und auch als Goldstandard bezeichnet werden [MRS08]. Das allgemeine Vorgehen ist in Abbildung 2.2 dargestellt. Die Grundidee ist, den vom IR-System berechneten Systemrelevanzen die erwartete Wahrheit in Form der objektiven Relevanz gegenüberzustellen und mittels geeigneter Maße zu bewerten.

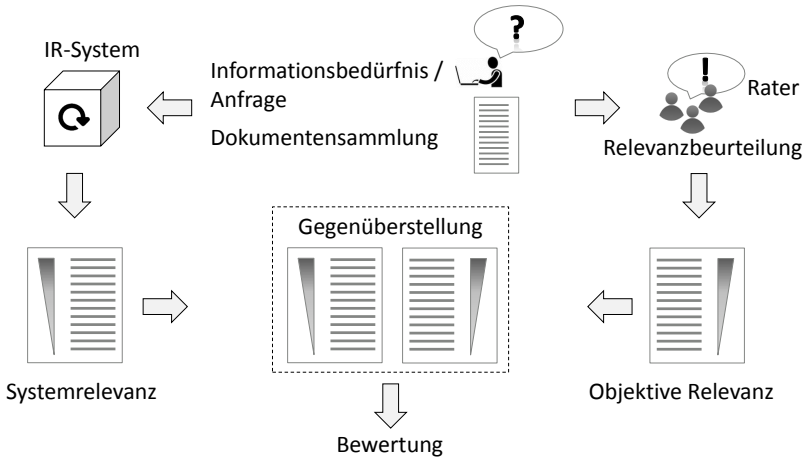


Abbildung 2.2: Grundsätzlicher Evaluationsaufbau im IR

Komparative Evaluation

Mit diesem Aufbau lässt sich nicht nur die Qualität eines einzelnen IR-Systems analysieren, sondern im Rahmen einer komparativen Evaluation auch ein Vergleich verschiedener Komponenten oder Konfigurationen eines IR-Systems oder sogar ein Vergleich verschiedener IR-Systeme vornehmen. Das ist die Grundidee von IR-Wettbewerben wie Text Retrieval Conference (TREC) [TREC] und Conference and Labs of the Evaluation Forum (CLEF) [CLEF], bei denen neue IR-Methoden gegeneinander antreten, indem IR-Systeme auf Basis einer vorgegebenen Testkollektion unterschiedliche Anfragen bearbeiten und mittels standardisierter Qualitätsmaße miteinander verglichen werden [Har13]. Abbildung 2.3 zeigt den Aufbau einer zweidimensionalen komparativen Evaluation, bei der die Ergebnisse verschiedener IR-Systeme für verschiedene Informationsbedürfnisse miteinander verglichen werden können. Jeder Bewertungsquader enthält die Gegenüberstellung einer objektiven Relevanz und einer Systemrelevanz sowie berechnete Bewertungsmaße.

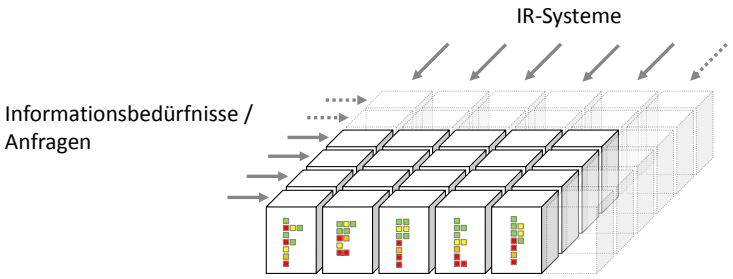


Abbildung 2.3: Dimensionen der komparativen Evaluation im IR

Relevanzbeurteilung großer Dokumentensammlungen

Die Relevanzbeurteilung gilt als sehr zeitaufwendiger und daher als ein teurer Prozess, da hierfür menschliche Experten einbezogen werden müssen [MRS08]. In der Regel sind die Dokumentensammlungen sehr groß, so dass eine vollumfängliche Relevanzbewertung aller Elemente der Sammlung nicht praktikabel ist. Gemäß der ursprünglichen Idee müssen für jedes Informationsbedürfnis alle Elemente des Datensatzes nach Relevanz beurteilt werden. Eine Möglichkeit zur Umgehung dieses Problems ist, für jedes Informationsbedürfnis lediglich eine Teilmenge des Datensatzes zu beurteilen. Diese Teilmenge wird üblicherweise durch Pooling gebildet [MRS08, Fuh06]. Beim *Pooling* wird jede Kombination aus Dokumentensammlung und Informationsbedürfnis von verschiedenen IR-Systemen bearbeitet. Aus deren nach Relevanz sortierten Ergebnislisten werden jeweils die *top-k* Elemente gewählt, wobei k die Pooltiefe darstellt. Taucht ein Element also in der top-k Auswahl eines IR-Systems auf, so wird es aufgrund seiner potenziellen Relevanz in den Pool aufgenommen und später einer objektiven Relevanzbeurteilung unterzogen. Angereichert werden diese Listen oftmals durch zusätzliche Quellen, beispielsweise eine Schlagwortsuche oder die Befragung von Experten. Dadurch können zusätzlich wichtige Elemente in den Pool aufgenommen werden. In der Regel existieren bei den so selektierten Elementen der Ergebnislisten Überschneidungen, so dass die Größe des Pools mit zunehmender Pooltiefe nicht linear ansteigt, sondern die Anzahl der neu aufzunehmenden Elemente sinkt [Zob98]. Auch wenn ein Pool eine unvollständige Liste an Antwortmöglichkeiten darstellt, so basiert Pooling dennoch auf der Annahme, dass der selektierte Pool bei ausreichender Pooltiefe und einer passenden diversifizierenden Auswahl an IR-Systemen beziehungsweise Konfigurationen annähernd alle potenziell relevanten Elemente einer Dokumentensammlung enthält. Elemente der Dokumentensammlung, welche nicht im Pool sind, also in der Poolingphase in keinem Durchlauf der verwendeten IR-Systeme unter den top-k Elementen waren, werden als nicht relevant betrachtet. Welche Pooltiefe ausreichend ist, lässt sich nicht pauschal festlegen. In größeren IR-Experimenten wie TREC sind Pooltiefen von 50, 100 oder 200 üblich.

Pooling

Die Frage, ob IR-Systeme, welche nicht an der Erstellung des Pools einer Testkollektion beigetragen haben, dennoch mit dieser Kollektion zuverlässig evaluiert werden können, wird von Zobel [Zob98] untersucht. Er kommt in einer umfangreichen nachträglichen Analyse von TREC-Daten zu dem Ergebnis, dass neue IR-Systeme mit dieser Testkollektion bei ausreichender Pooltiefe ebenso zuverlässig evaluiert werden können.

Gegenüberstellung von Systemrelevanz und objektiver Relevanz

Für jedes Element liegt zum einen die vom IR-System berechnete Systemrelevanz vor und zum anderen die von den Ratern bewertete objektive Relevanz. Zur Bewertung der Ergebnisse von IR-Systemen werden Systemrelevanz und objektive Relevanz einander gegenübergestellt, so dass verschiedene Systemrelevanzen durch Bewertungsmaße verglichen werden können. In Abbildung 2.4 ist ein Beispiel für das Zusammenführen von Systemrelevanz und objektiver Relevanzbewertung dargestellt.

Beide Relevanzen werden in einer sogenannten Distribution folgendermaßen miteinander kombiniert. Die Elemente werden zunächst gemäß der berechneten Systemrelevanz des IR-Systems sortiert und so in eine Rangordnung gebracht. Im Beispiel belegt Element 1 den ersten Rang, die Elemente 2 bis 4 hingegen teilen sich aufgrund ihrer identischen berechneten Relevanz von 8,2 den zweiten Rang. Die zehn Elemente des Beispiels verteilen sich entsprechend auf sieben Ränge. Auf der anderen Seite enthält die objektive Relevanzbewertung für jedes

Distribution

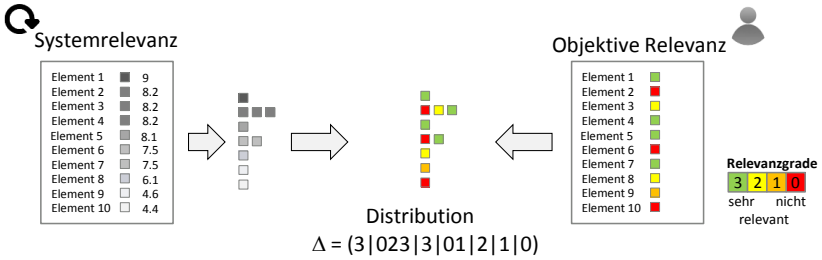


Abbildung 2.4: Beispiel für das Zusammenführen von Systemrelevanz und objektiver Relevanz

Element eine Zuordnung zur erwarteten Relevanz, welche durch die Rater unabhängig von der Systemrelevanz bestimmt wird. In der Abbildung werden dafür vier farblich markierte Relevanzgrade vergeben. Die objektive Relevanz jedes Elements wird den entsprechenden Elementen der in Rangordnung gebrachten Systemrelevanz zugewiesen. Auf Rang 1 wird dem Element 1 beispielsweise der Relevanzgrad 3 zugeordnet, die Elemente auf Rang 2 erhalten ihre entsprechende objektive Relevanz mit den Graden 0, 2 und 3. Das Ergebnis dieser Zusammenführung ist eine Abstraktion des Retrieval-Ergebnisses, die Distribution genannt wird. Sie enthält nicht mehr die berechneten Relevanzwerte des IR-Systems, sondern lediglich eine Rangfolge von berechneten Relevanzen mit Zuordnung der objektiv erwarteten Relevanz. Die gekürzte Schreibweise der dargestellten Distribution ist $\Delta = (3|023|3|01|2|1|0)$, wobei die Ränge durch vertikale Striche voneinander getrennt werden.

Für Distributionen mit zwei Relevanzgraden wird in der Regel als Notation für relevante Elemente (+) und nicht relevante Elemente (–) verwendet, z.B. $\Delta = (+|-++)$. Zudem ist im Beispiel eine schwache Rangordnung dargestellt. Grundsätzlich lässt sich unterscheiden, ob die berechnete Systemrelevanz in linearer, totaler Ordnung oder in schwacher Ordnung der einzelnen Elemente vorliegt. Bei linearer Ordnung besitzen alle Elemente unterschiedliche Relevanzwerte. Bei schwacher Ordnung können für unterschiedliche Elemente gleiche Relevanzwerte berechnet werden, was dazu führt, dass mehrere Elemente einen Rang teilen. Dieser Umstand ist wichtig bei der Wahl geeigneter Bewertungsmaße.[1]

2.1.3.3 Bewertungsmaße

Im Folgenden stellen wir Bewertungsmaße für das Hauptkriterium der Effektivität vor, die in dieser Arbeit Verwendung finden. Zur Bewertung der Effektivität beschreiben wir Maße für die Genauigkeit und die Trefferquote, die sich zur Bewertung von Retrievalverfahren eignen, deren Systemrelevanzen in schwacher Ordnung vorliegen. Ferner stellen wir ein Maß zur Beurteilung der Nützlichkeit vor, welches eine Zuordnung von mehr als zwei Relevanzgraden bei objektiven Relevanzen berücksichtigt.

Precision und Recall

Zur Bewertung der Effektivität werden im IR üblicherweise die zwei Maße Genauigkeit (*engl.:* precision) und Trefferquote (*engl.:* recall) verwendet. Die Genauigkeit beschreibt das Verhältnis der Anzahl gefundener relevanter Elemente zur Anzahl gefundener Elemente. Die

[1] Häufig wird lineare Ordnung extra hergestellt, da die Anwendung von Bewertungsmaßen für schwache Ordnung in der Regel komplizierter ist. In TREC wird schwache Ordnung beispielsweise nicht berücksichtigt, sondern lineare Ordnung hergestellt, indem die Elemente eines Ranges zufällig angeordnet werden. Nach [RBJ89] und [Fuh06] ergeben sich durch solch ein Vorgehen allerdings nicht reproduzierbare und unzuverlässige Ergebnisse.

Trefferquote hingegen beschreibt das Verhältnis der Anzahl relevanter gefundener Elemente zur Anzahl aller relevanten Elemente der Elementmenge. Zur Bewertung von Ranglisten sind solche absoluten Werte für Genauigkeit und Trefferquote für die gesamte Elementmenge oftmals wenig aussagekräftig, da sie nicht berücksichtigen, zu welchem Zeitpunkt ein relevantes Element gefunden wurde. Für die Bewertung von Ranglisten interessiert vielmehr die Genauigkeit beim Finden der einzelnen relevanten Elemente, der sogenannten Recall-Punkte. Bei n relevanten Elementen wird die Genauigkeit für jedes gefundene relevante Element an den Recall-Punkten 1/n ... n/n berechnet. Die Entwicklung von Genauigkeit in Bezug zur Trefferquote kann dann in sogenannten Recall-Precision-Graphen dargestellt werden. Um einen einzelnen Genauigkeitswert für einen Recall-Precision-Graphen zu erhalten, kann die Average Precision (AP), das arithmetische Mittel der berechneten Genauigkeit jedes Recall-Punktes, berechnet werden.

$$\mathrm{AP} := \frac{1}{\mathrm{n}} \cdot \sum_{\mathrm{k}=1}^{\mathrm{n}} \mathrm{Prec}(\mathrm{k}) \tag{2.1}$$

Damit verschiedene Recall-Precision-Graphen miteinander verglichen werden können, selbst wenn diese eine unterschiedliche Anzahl an relevanten Elementen und damit eine unterschiedliche Anzahl an Recall-Punkten besitzen, wird üblicherweise eine 11-Punkt-Interpolation vorgenommen. Dabei werden bisherige Recall-Punkte durch die neuen elf festgelegten Recall-Punkte im Bereich [0 ... 1] mit festgelegtem Recall-Intervall 0,1 ersetzt. Die von Salton [SM83] beschriebene Salton-Interpolation begradigt die entstehende Kurve zusätzlich, so dass jeder interpolierte Genauigkeitswert an einem Recall-Punkt der höchsten Genauigkeit aller Recall-Punkte $\mathrm{k}' \geq \mathrm{k}$ entspricht [MRS08].

$$\mathrm{Prec}_{\mathrm{interpol}}(\mathrm{k}) := \max_{\mathrm{k}' \geq \mathrm{k}} \mathrm{Prec}(\mathrm{k}') \tag{2.2}$$

Probability of Relevance (PRR)

Da die Ergebnislisten in dieser Arbeit eine schwache Ordnung aufweisen, müssen rangabhängige Genauigkeitswerte berechnet werden. Diese berücksichtigen, dass Elemente innerhalb eines Ranges keine Reihenfolge besitzen und daher auch keine Reihenfolge bei der Berechnung bevorzugt werden darf. Beispielsweise sollte die Bewertung der Distributionen $\Delta = (+-|+|-)$ und $\Delta = (-+|+|-)$ zu gleichen Werten für Genauigkeit und Trefferquote führen. Raghavan et al. [RBJ89] untersuchen Methoden zur Bewertung schwach geordneter Distributionen und kommen zu dem Ergebnis, dass bei schwacher Ordnung eine probabilistische Interpretation der Maße sinnvoll ist.

Ein rangabhängiges Genauigkeitsmaß ist nach Raghavan et al. die Wahrscheinlichkeit der Relevanz (*engl.:* probability of relevance). Mit dem PRR-Wert wird die Genauigkeit interpretiert als die Wahrscheinlichkeit, dass ein gefundenes Dokument relevant ist. Für einen Rang lf sei j die Anzahl irrelevanter Dokumente in den Rängen $\{1, ..., \mathrm{lf} - 1\}$, s die Anzahl relevanter Dokumente, die aus Rang lf gezogen werden, r die Anzahl relevanter Dokumente in Rang lf, i die Anzahl irrelevanter Dokumente in Rang lf sowie n die Anzahl aller relevanten Dokumente, so wird PRR für NR relevante Dokumente mit $0 < \mathrm{NR} \leq \mathrm{n}$ durch die folgende Formel berechnet [RBJ89, Fuh06].[2]

$$\mathrm{PRR}_{\mathrm{NR}} := \frac{\mathrm{NR}}{\mathrm{NR} + \mathrm{j} + (\mathrm{s} \cdot \mathrm{i})/(\mathrm{r} + 1)} \tag{2.3}$$

[2] Die Herleitung der Formeln und der möglichen Interpretationen sind in [RBJ89, Fuh06] detailliert beschrieben.

Expected Precision (EP)

Eine weitere Interpretation des Genauigkeitsmaßes bietet die erwartete Genauigkeit (*engl.:* expected precision). Die EP gibt an, welche Genauigkeit an den Recall-Punkten erwartet werden kann. Das entspricht der erwarteten Genauigkeit, mit der jedes relevante Dokument gefunden wird. Dafür werden die Wahrscheinlichkeiten der möglichen Anordnungen innerhalb der Ränge berücksichtigt [RBJ89, Fuh06].[3] Sei weiterhin t_r die Anzahl relevanter Dokumente in den Rängen $\{1, ..., lf - 1\}$, so berechnet sich die EP für NR relevante Dokumente wie folgt [RBJ89, Fuh06].

$$EP_{NR} := \sum_{k=s}^{r+i} \frac{\binom{r}{s-1}\binom{i}{k-s}}{\binom{r+i}{k-1}} \cdot \frac{r - s + 1}{r + i - k + 1} \cdot \frac{s + t_r}{k + t_r + j} \tag{2.4}$$

Sowohl für PRR als auch für EP lassen sich analog zu Formel 2.1 (Seite 19) aus den Genauigkeitswerten für jeden Recall-Punkt ein einzelner Genauigkeitswert berechnen. Dieser entspricht im Fall von EP dem arithmetischen Mittel der berechneten EP-Werte aller Recall-Punkte.

Normalized Discounted Cumulated Gain (NDCG)

Wurden für die objektive Relevanz Relevanzgrade vergeben, so ist ein weiteres Ziel bei der Bewertung der Güte von Ergebnissen eines IR-Systems, dass sehr relevante Elemente möglichst am Anfang der Ergebnisliste stehen, da das Ergebnis in dem Fall als nützlicher wahrgenommen wird. Das von Järvelin und Kekäläinen [JK02] vorgestellte NDCG ist ein Maß für die Nützlichkeit von Ergebnissen eines IR-Systems, welches die Relevanzgrade der einzelnen Elemente berücksichtigt. Deren Grundidee ist, den Nutzen für jedes Element einer Distribution über dessen Relevanzgrad zu bestimmen und über die Ränge zu kumulieren. Das Ergebnis wird logarithmisch über die Ränge abgeschwächt (*engl.:* discounted), da der Nutzen eines relevanten Elementes am Anfang einer Ergebnisliste höher ist als an deren Ende. Über diese Abschwächung lassen sich zudem Nutzererwartungen abbilden, indem beispielsweise für den ungeduldigen Nutzer der Logarithmus mit der Basis 2 und für den persistenten Nutzer die Basis 10 gewählt wird. Sei rel_i der Relevanzgrad des i-ten Elementes einer linear geordneten Distribution mit p Elementen (und demnach auch Rängen), so wird das Discounted Cumulated Gain (DCG) für einen ungeduldigen Nutzer wie folgt berechnet.

$$DCG := rel_1 + \sum_{i=2}^{p} \frac{rel_i}{\log_2 i} \tag{2.5}$$

Als Beispiel ergeben sich für zwei Distributionen $\Delta_{C1} = (3|2|3|0|3)$ und $\Delta_{C2} = (2|3|3|3|0)$ und deren ideale Anordnung $\Delta_{Ideal} = (3|3|3|2|0)$ die folgenden Werte für den DCG:
$DCG(\Delta_{C1}) \approx 8{,}18$, $DCG(\Delta_{C2}) \approx 8{,}39$ und $DCG(\Delta_{Ideal}) \approx 8{,}89$.
Mit der Normalisierung wird ein Nützlichkeitswert in Relation zur idealen Anordnung gesetzt. Die Berechnung der normalisierten Nützlichkeitswerte
$NDCG(\Delta_{C1}) = \frac{DCG(\Delta_{C1})}{DCG(\Delta_{Ideal})} \approx 0{,}92$ und $NDCG(\Delta_{C2}) = \frac{DCG(\Delta_{C2})}{DCG(\Delta_{Ideal})} \approx 0{,}94$ zeigt, dass die Distribution Δ_{C2} im Vergleich zu Δ_{C1} als nützlicher bewertet werden kann.

[3] Die Herleitung der Formeln ist ebenfalls in [RBJ89, Fuh06] beschrieben.

NDCG für schwache Ordnung

Die Berechnung des NDCG erfordert Distributionen linearer Ordnung. Die Anwendung auf Distributionen schwacher Ordnung ist nicht vorgesehen. Um die Nützlichkeit in dieser Arbeit auch für schwach geordnete Distributionen verwenden zu können, wird folgender Ansatz vorgeschlagen. Für jede Distribution kann ein minimaler $\mathrm{NDCG}_{\mathrm{min}}$ und ein maximaler $\mathrm{NDCG}_{\mathrm{max}}$ sowie deren arithmetisches Mittel $\mathrm{NDCG}_{\mathrm{avg}}$ berechnet werden. Zur Berechnung des $\mathrm{NDCG}_{\mathrm{max}}$ werden alle Elemente jedes Ranges in die bestmögliche Anordnung gebracht. Analog dazu werden die Elemente jedes Ranges zur Berechnung von $\mathrm{NDCG}_{\mathrm{min}}$ schlechtmöglichst angeordnet. Die Betrachtung von $\mathrm{NDCG}_{\mathrm{min}}$ und $\mathrm{NDCG}_{\mathrm{max}}$ hat den Vorteil, dass die Berechnung der Werte reproduzierbar bleiben, was bei zufälliger Anordnung nicht der Fall ist.

2.1.3.4 Bewertung der Urteilerübereinstimmung

Die Bewertungen der Rater sind grundsätzlich subjektiv, werden in der Evaluation allerdings als objektive Relevanz verwendet. Damit die Bewertungen der verschiedenen Rater als objektiv gelten können, sollen diese zu einem gewissen Grad übereinstimmen. Die Wahl der Maße zur Bewertung der Urteilerübereinstimmung ist von verschiedenen Faktoren des gewählten Evaluationsaufbaus abhängig. Daher präsentieren wir vorab einige Besonderheiten des Evaluationsplans, der erst in Abschnitt 4.1 detailliert vorgestellt wird. Anschließend stellen wir die Maße zur Bewertung der Urteilerübereinstimmung vor, die in dieser Arbeit Verwendung finden.

Anzahl der Rater > 2: Viele Maße zur Urteilerübereinstimmung eignen sich lediglich für die Beurteilung von zwei Ratern. Werden objektive Relevanzbewertungen von mehr als zwei Ratern durchgeführt, kommen spezielle Maße zum Einsatz. In dieser Arbeit werden drei bis vier Rater zur Beurteilung jeder Anfrage eingesetzt.

Ordinalskala: Für die Ratings wird eine Ordinalskala von 0 bis 3 verwendet. Im Gegensatz zu Elementen, welche mit Nominalskala bewertetet werden, ist neben der tatsächlichen Übereinstimmung der Bewertungen auch die Übereinstimmung der Rangfolge von Bedeutung.

Freie Randverteilung: Bei der Wahl der Maße muss zudem berücksichtigt werden, dass die Ratings nicht gleichmäßig verteilt sind. Zum einen bedeutet das, dass wesentlich mehr Bewertungselemente einer Anfrage als nicht relevant und oftmals nur sehr wenige Elemente als relevant bewertet werden. Dieser Umstand kommt daher, weil in jedem Szenario eine vollständige Betrachtung aller Elemente des gleichen Typs vorgenommen wird. Zum anderen werden keine Vorgaben zur Anzahl der Bewertungselemente gemacht, welche als relevant markiert werden sollen. Dadurch entstehende unterschiedliche Randverteilungen erschweren die Vergleichbarkeit und müssen daher bei der Wahl der Maße berücksichtigt werden.

Rangbindung: Da in dieser Arbeit eine hohe Anzahl an Elementen mit nur vier Möglichkeiten auf der Ordinalskala bewertet werden, entstehen in den Ratings große Überlappungen auf den Rängen. Mehrere Bewertungselemente teilen sich einen Rang. Dieses Phänomen wird Rangbindung genannt.

Zufallskorrigierte Übereinstimmungskoeffizienten

Die Grundidee zufallskorrigierter Übereinstimmungskoeffizienten ist es, die tatsächlich erreichte Übereinstimmung p_a zwischen Ratern mittels erwarteter zufälliger Übereinstimmung p_e zu korrigieren und zu normieren. Das bekannteste Übereinstimmungsmaß ist der von Cohen [Coh60] definierte Kappa-Koeffizient. Dessen allgemeine Definition ist

$$\kappa = \frac{p_a - p_e}{1 - p_e}. \tag{2.6}$$

Die Wahrscheinlichkeiten p_a und p_e wurden von Cohen für zwei Rater und nominalskalierte Daten mit gleicher Randverteilung definiert. Für den Einsatz in vielfältigeren oder spezielleren Situationen sowie zur Überwindung bekannter Paradoxien [BP81, CF90, Gwe08] existieren verschiedene angepasste Koeffizienten mit korrigierten Wahrscheinlichkeiten. Der Fleiss-Kappa-Koeffizient [Fle71] ist beispielsweise eine Verallgemeinerung für die Bewertung von mehr als zwei Ratern. Der Brennan-Prediger-Koeffizient [BP81] sieht eine Korrektur von p_e für eine ungleiche Randverteilung vor. Gwet definiert den AC_1-Koeffizient als paradoxresistente Variante sowie dessen gewichtete Variante AC_2 für nicht nominalskalierte Daten [Gwe08, Gwe14].

Berechnung: In Anlehnung an Formel 2.6 sind der Brennan-Prediger-Koeffizient und der Gwet-AC_1-Koeffizient definiert[4] als

$$\kappa_{BP} = \frac{p_a - p_{e_{BP}}}{1 - p_{e_{BP}}} \tag{2.7}$$

und

$$AC_1 = \frac{p_a - p_{e_{AC_1}}}{1 - p_{e_{AC_1}}}. \tag{2.8}$$

Die Wahrscheinlichkeit p_a entspricht der Übereinstimmungswahrscheinlichkeit von Fleiss [Fle71] für mehr als zwei Rater. Genau genommen entspricht p_a der Wahrscheinlichkeit, dass ein anderer zufällig gewählter Rater für ein zufällig gewähltes Bewertungselement die gleiche Kategorie wählt. Für N Bewertungselemente, welche von m Ratern durch die Zuordnung einer von c Kategorien bewertet werden, entspricht dies der Wahrscheinlichkeit

$$p_a = \frac{1}{N} \cdot \sum_{i=1}^{N} \sum_{k=1}^{c} \frac{m_{ik}(m_{ik} - 1)}{m(m-1)}. \tag{2.9}$$

Der Brennan-Prediger-Koeffizient definiert $p_{e_{BP}}$ als die einfache Wahrscheinlichkeit, eine der möglichen Kategorien zu selektieren.

$$p_{e_{BP}} = \frac{1}{c} \tag{2.10}$$

Der Gwet-AC_1-Koeffizient definiert $p_{e_{AC_1}}$ als

$$p_{e_{AC_1}} = \frac{1}{c-1} \cdot \sum_{k=1}^{c} \pi_k(1 - \pi_k), \tag{2.11}$$

[4] Die hier vorgestellten Formeln zur Berechnung der Koeffizienten sind eine Zusammenfassung aus den originalen Veröffentlichungen [Fle71, BP81, Gwe08] und der Anwendungsempfehlung von Gwet [Gwe14].

wobei π_k dem Anteil der Ratings, denen die Kategorie von k zugeordnet wurde, von der Gesamtanzahl aller Ratings entspricht.

$$\pi_k = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{m_{ik}}{m} \tag{2.12}$$

Interpretation: Bei perfekter Übereinstimmung nehmen die vorgestellten Koeffizienten den Wert 1 an. Der Wert 0 zeigt an, dass die Übereinstimmung der erwarteten zufälligen Übereinstimmung entspricht. Ist der Wert kleiner als 0, so ist die Übereinstimmung sogar schlechter als zufällig erwartet. Zur Bewertung der relativen Stärke der Übereinstimmung existieren Empfehlungen wie die von Landis und Koch [LK77], welche beispielsweise in Tabelle 2.1 abgebildet ist.[5] Derartige Tabellen liefern einen ersten Anhaltspunkt zur Beurteilung der Urteilerübereinstimmung.

Kappa	**Stärke der Übereinstimmung**
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Tabelle 2.1: Interpretation der Kappa-Werte nach Landis und Koch [LK77]

Zur Überprüfung, ob ein berechneter Koeffizient signifikant größer als Zufall ist, kann ein rechtsseitiger Einstichproben-t-Test durchgeführt werden. Die Nullhypothese H_0 besagt in dem Fall, dass der berechnete Übereinstimmungswert kleiner oder gleich dem Zufall ist, also 0. Dafür dient als Prüfwert t für einen berechneten Übereinstimmungswert κ mit Standardfehler se_κ gemäß Bortz [BLB08]

$$t = \frac{\kappa}{se_\kappa}. \tag{2.13}$$

Als Signifikanzgrenze wird für das gewählte Signifikanzniveau α, N Bewertungselemente und die jeweiligen Freiheitsgrade $Fg = N - 1$ das entsprechende Quantil der t-Verteilung $t_{1-\alpha,N-1}$ verwendet. Die Nullhypothese kann abgelehnt werden, wenn der Prüfwert t größer oder gleich der Schranke $t_{1-\alpha,N-1}$ ist.

Zur Bewertung der Aussagekraft des berechneten Wertes, also zur Beantwortung der Frage, wie gut die Schätzung des Übereinstimmungswertes ist, können mit beidseitigem Einstichproben-t-Test Konfidenzintervalle berechnet werden. Das Konfidenzintervall gibt den Wertebereich an, der den wahren Koeffizienten in der Population mit Wahrscheinlichkeit von 95% enthält.

Kendalls Konkordanzkoeffizient W

Aus den bewerteten Elementlisten der Bewertungsszenarien durch verschiedene Rater mit den möglichen Ratings einer Ordinalskala ergeben sich rangskalierte Datenreihen. Der Konkordanz-

[5] Mit Kappa sind in der Tabelle von Landis und Koch die verschiedenen, zur Zeit ihrer Veröffentlichung bekannten zufallskorrigierten Übereinstimmungskoeffizienten gemeint, unter anderem von Cohen [Coh60] und Fleiss [Fle71]. In dieser Arbeit verwenden wir diese Tabelle auch als Grundlage zur Bewertung der von uns verwendeten Koeffizienten von Gwet [Gwe08, Gwe14] sowie von Brennan und Prediger [BP81], die beide erst später entstanden sind.

koeffizient Kendalls W [KS39, KG90] ist ein Maß zur Quantifizierung der Übereinstimmung für solche rangskalierte Datenreihen, welche durch zwei oder mehrere Rater bewertet und welche nicht zwangsläufig normalverteilt sind. Zudem erlaubt er eine Korrektur bei vorhandenen Rangbindungen.

Berechnung: Für eine von $j = 1...m$ Ratern bewertete Liste von $i = 1...N$ Bewertungselementen entspricht die Rangsumme T_i eines Bewertungselementes i der Summe seiner Ränge beziehungsweise seiner Bewertungen.[6]

$$T_i = \sum_{j=1}^{m} R_{ij} \tag{2.14}$$

Die Quadratsumme der Rangsummen QSR entspricht der empirischen Varianz der Rangsummen und wird wie folgt berechnet.

$$QRS = \sum_{i=1}^{N} T_i^2 - \frac{(\sum_{i=1}^{N} T_i)^2}{N} \tag{2.15}$$

Kendalls Konkordanzkoeffizient resultiert auf dem Verhältnis der empirischen Varianz der Rangsummen (QSR) zu der maximalen Varianz. Kendalls für Rangbindungen korrigierter Konkordanzkoeffizient W wird berechnet durch

$$W = \frac{12 \cdot QSR}{m^2 \cdot (N^3 - N) - m \cdot \sum_{j=1}^{m} V_j}, \tag{2.16}$$

mit der Bindungskorrektur

$$V_j = \sum_{k=1}^{s_j} (v_k^3 - v_k), \tag{2.17}$$

wobei s_j die Anzahl der Rangbindungen in einer Reihe j und v_k die Länge der k-ten Rangbindung darstellt.

Interpretation: Der berechnete Konkordanzkoeffizient ergibt einen Wert zwischen 0 und 1, wobei 1 perfekte Konkordanz und 0 perfekte Diskordanz bedeutet. Die statistische Bedeutsamkeit des Ergebnisses von Kendalls Konkordanzanalyse wird mit dem Friedman-Test überprüft [BLB08, S.468f]. Die Prüfgröße χ_r^2 der Friedman-Rangvarianzanalyse steht in Beziehung zu W und wird ebenfalls für Rangbindung korrigiert.

$$\chi_r^2 = m \cdot (N - 1) \cdot W = \frac{12 \cdot QSR}{m \cdot N \cdot (N + 1) - \frac{1}{N-1} \cdot \sum_{j=1}^{m} V_j} \tag{2.18}$$

Gemäß Bortz et al. [BLB08, S.468] wird mit der Nullhypothese H_0 geprüft, ob die N Bewertungselemente identische Durchschnittsränge besitzen. Diese Nullhypothese sei nach Bortz

[6] Die hier vorgestellten Formeln zur Berechnung von Kendalls W und dessen statistischer Überprüfung sind der Anwendungsempfehlung eines Statistiklehrbuches [BLB08, S. 465-470] entnommen. Diese beziehen sich auf die von Kendall [KS39, KG90] präsentierten und weiterentwickelten Methoden.

et al. äquivalent zu der H_0 fehlender Konkordanz der Rangreihen. Das bedeute weiterhin, dass die H_0 verworfen werden kann, wenn der χ^2_r-Wert überzufällig groß ist. Als Signifikanzgrenze wird für das gewählte Signifikanzniveau α und die jeweiligen Freiheitsgrade $Fg = N - 1$ das entsprechende Quantil der χ^2-Verteilung $\chi^2_{1-\alpha,N-1}$ verwendet. Auf Konkordanz kann also geschlossen werden, wenn der χ^2_r-Wert größer ist als die kritische Schranke $\chi^2_{1-\alpha,N-1}$.

2.1.4 Zusammenfassung und Bezug zur Arbeit

Dieser Abschnitt schaffte einen Überblick über die allgemeine Arbeitsweise und den Aufbau von IR-Systemen sowie die Evaluation solcher Systeme. Zur Bewertung der Qualität unseres domänenspezifischen IR-Systems für G&Rs haben wir Maße zur Berechnung von Genauigkeit und Nützlichkeit der Ergebnisse vorgestellt. Da wir in der Evaluation die Meinung mehrerer Experten einbeziehen, müssen wir die Verwendbarkeit dieser Meinungen überprüfen. Daher stellten wir zudem spezielle Bewertungsmaße für die Urteilerübereinstimmung vor. Die Tabelle 2.2 fasst die vier wichtigsten Bewertungsmaße zusammen, die wir in der Evaluation dieser Arbeit verwenden.

Evaluationszweck	**Bewertungsmaß**	**Berücksichtigung der Relevanzgrade**	**Kurzbeschreibung**
Effektivität	Expected Precision EP	nein	kombiniertes Maß für erwartete Genauigkeit und Trefferquote bewertet die Reihenfolge gefundener Elemente (relevante Elemente sollten vor nicht relevanten Elementen gefunden werden)
	Normalized Discounted Cumulated Gain NDCG	ja	Maß für Nützlichkeit bewertet die Reihenfolge gefundener Elemente (Elemente mit hohem Relevanzgrad sollten vor Elementen mit niedrigem Relevanzgrad gefunden werden),
Urteilerübereinstimmung	Brennan-Prediger-Koeffizient κ_{BP} und Gwet-AC_1-Koeffizient	ja	bewerten die Übereinstimmung der Ratings von mehr als zwei Ratern im Verhältnis zu einer zufälligen Übereinstimmung
	Kendalls Konkordanzkoeffizient W	ja	bewerten die Übereinstimmung der Rangreihenfolgen der Ratings von mehr als zwei Ratern

Tabelle 2.2: Überblick über die vier wichtigsten Bewertungsmaße für die Evaluation in dieser Arbeit

IR-Systeme unterscheiden sich je nach Anwendung hauptsächlich in der Art der Wissensrepräsentation und der Retrievalstrategie, wobei graphbasierte Ansätze assoziativer IR-Systeme im Zentrum dieser Arbeit stehen. Die erste dieser beiden Komponenten, die assoziative Wissensrepräsentation durch semantische Netze, wird im folgenden Abschnitt 2.2 detaillierter beschrieben.

2.2 Repräsentation durch semantische Netze

Im Fokus dieser Arbeit stehen semantische Netze für die graphbasierte Repräsentation, welche genau genommen auf RDF-Graph, der abstrakten Syntax RDF-basierter Sprachen, basieren. Wir stellen daher, im Anschluss an eine allgemeine Begriffsklärung, die Beschreibungssprachen Resource Description Framework (RDF), RDF-Schema und Web Ontology Language (OWL) vor. Insbesondere erläutern wir RDF-Graphen. Da der manuelle Aufbau semantischer Netze durch Domänenexperten als sehr aufwendig gilt, machen wir uns das Datenmodell des Entwicklungswerkzeugs und die entsprechend strukturierten Daten zunutze. Daher stellen wir abschließend Techniken zur datenmodellbasierten Transformation vor, die einen automatisierten Aufbau von RDF-Graphen ermöglichen

2.2.1 Grundbegriffe

semantisches Netz

Ein *semantisches Netz* ist eine graphbasierte Form der Wissensrepräsentation, bei der sogenannte Konzepte durch Knoten und Beziehungen als Kanten zwischen den Knoten dargestellt werden [Den12]. Im Gegensatz zu assoziativen Netzen, in denen die Art der Beziehung auch undefiniert sein kann, besitzen die Kanten in semantischen Netzen eine definierte Bedeutung [Cre97]. Semantische Netze dienen der Visualisierung und Strukturierung von Wissen und dem „besseren Verständnis der Zusammenhänge zwischen Dingen in bestimmten Wissensdomänen“ [Den12]. Nach Dengel ist der Konzeptbegriff sehr offen. *Konzepte* können „anfassbare sowie imaginäre Dinge, Ereignisse oder Prozesse“ [Den12] sein. Weiterhin definiert Dengel eine *Kategorie* als generisches Konzept, das je nach Kontext auch als Typ oder Klasse bezeichnet wird. Objekte, die zu einer Kategorie gehören, werden *Individuen* oder *Instanzen* genannt. Verknüpfungen zwischen Konzepten sind *Relationen*, Eigenschaften von Objekten ihre *Attribute*. Ein Grundprinzip semantischer Netze ist die Objektidentität [Rei10]. *Objektidentität* bedeutet, dass jedes Objekt in einem semantischen Netz nur ein Mal existiert. Eine redundante Verteilung von Objekten im Netz wird vermieden. Stattdessen sollen alle Informationen zu einem Objekt auch bei diesem zusammenlaufen, was über die Verknüpfung durch Relationen umgesetzt wird [Rei10]. Die verschiedenen Arten von semantischen Netzen unterscheiden sich stark in der Ausdrucksmächtigkeit: von der einfachen hierarchischen Ordnung von Kategorien und Individuen in Taxonomien bis zu inferenzfähigen Netzen, den Ontologien. Semantische Netze können beispielsweise ausführbar (Petrinetze), lernend (neuronale Netze), probabilistisch (Bayessche Netze) oder aussagenbasiert (propositionale Netze) sein [Den12].

Konzept
Kategorie
Individuum, Instanz
Relation, Attribut
Objekt-identität

2.2.2 RDF-basierte Repräsentation

RDF ist eine Sprache zur Repräsentation und zum Austausch von Informationen. Sie wurde ursprünglich für den Einsatz im World Wide Web (WWW) entwickelt, findet aber auch in anderen informations- und wissensbasierten Anwendungen ihren Einsatz [HKRS08, Den12]. RDF ist Teil einer Sprachfamilie, zu der auch RDF-Schema und die verschiedenen Ausprägungen von OWL gehören. RDF-basierte Sprachen besitzen die gemeinsame, abstrakte Syntax RDF-Graph: ein graphbasiertes Datenmodell, welches ein semantisches Netz darstellt. Da in dieser Arbeit RDF-Graphen als semantische Netze verwendet werden, welche Sprachkonstrukte von RDF, RDF-Schema und OWL enthalten, folgt ein Überblick über relevante Sprachkonstrukte. Die vollständige Beschreibung findet sich in den entsprechenden

Dokumentationen.[7]

RDF und RDF-Schema

Grundsätzlich folgen RDF und RDF-Schema dem Prinzip der Trennung von Struktur und Inhalt, also von Schema- und Datenschicht, die im Folgenden auch als Modell- und Instanzebene bezeichnet werden. RDF dient der Beschreibung von Konzepten und deren Beziehungen auf Instanzebene. Mit RDF-Schema wird das RDF-Vokabular um Konstrukte zur Beschreibung von terminologischem Wissen auf Modellebene erweitert.

Konzepte werden in RDF *Ressourcen* genannt. Syntaktisch werden Ressourcen in RDF entweder von einem eindeutigen Identifizierer, hier Internationalized Resource Identifier (IRI)[8], oder einem ebenfalls eindeutig identifizierbaren Literal repräsentiert. *Literale* besitzen einen Datentyp und repräsentieren Werte wie Zahlen oder Zeichenketten. Eine Eigenschaft einer Ressource wird *Property* genannt. Sie wird von einem *Prädikat* repräsentiert, das wiederum selber ein IRI ist. Ein IRI kann durch einen Namensraum-Präfix abgekürzt werden. In Tabelle 2.3 sind die in dieser Arbeit verwendeten Präfixe für das vordefinierte Vokabular von RDF, RDF-Schema und OWL abgebildet.

Ressource

Literal

Property

Präfix	**IRI**
`rdf:`	`http://www.w3.org/1999/02/22-rdf-syntax-ns#`
`rdfs:`	`http://www.w3.org/2000/01/rdf-schema#`
`owl:`	`http://www.w3.org/2002/07/owl#`

Tabelle 2.3: Abkürzung der IRI durch Präfixe

Die Strukturierungselemente erlauben die Beschreibung der semantischen Beziehungen in RDF-Dokumenten, unter anderem von Klassen (durch das Sprachkonstrukt `rdfs:Class`), Eigenschaften (`rdf:Property`), den hierarchischen Zusammenhängen zwischen Klassen oder Eigenschaften (`rdfs:subClassOf`, `rdfs:subPropertyOf`) sowie Typzugehörigkeiten zu beispielsweise Klassen oder Eigenschaften (`rdf:type`).

OWL

OWL erweitert RDF-Schema und ermöglicht die Beschreibung komplexer semantischer Beziehungen. Als neues Konzept werden *Rollen* zur Unterscheidung von Eigenschaften eingeführt. Verbindet ein Property zwei Individuen miteinander, so handelt es sich um eine abstrakte Rolle (`owl:ObjectProperty`). Im Gegensatz dazu verbindet die konkrete Rolle Individuen mit Zahlenwerten (`owl:DatatypeProperty`). Zudem lassen sich für Eigenschaften unter anderem Restriktionen für den Definitionsbereich (`owl:onProperty`) und Wertebereich (z.B. `owl:allValuesFrom`) festlegen.

[7] Die folgenden Ausführungen beziehen sich vor allem auf Beschreibungen des World Wide Web Consortium (W3C), veröffentlicht als W3C-Empfehlungen zu Konzepten, abstrakter Syntax, Semantik von RDF [W3C14a, W3C14c], zur Beschreibung von RDF-Schema [W3C14d], zur Beschreibung von OWL [W3C12d, W3C12b] sowie dem Mapping von OWL zu RDF-Graph [W3C12c]. Im Folgenden werden lediglich die für diese Arbeit relevanten Teile vorgestellt. Eine vollständige Beschreibung aller Sprachkonstrukte findet sich in den genannten W3C-Empfehlungen.

[8] IRI ist spezifiziert im RFC 3987 [DS05] und erweitert Uniform Resource Identifier (URI) auf die vollständige Abdeckung des Unicode-Zeichensatzes [DS05].

OWL gibt es in drei Sprachvarianten, die sich in ihrer Expressivität unterscheiden: von der einfachsten Sprache OWL Lite über OWL DL bis zur ausdrucksstärksten Sprache OWL Full [HKRS08]. Sie unterscheiden sich zum einen in der Unterstützung der OWL-Sprachkonstrukte und zum anderen in Restriktionen bei deren Verwendung. OWL basiert auf einer eingeschränkten Prädikatenlogik. OWL Lite und OWL DL genügen einer speziellen Form der Prädikatenlogik erster Stufe, der Beschreibungslogik[9], die Entscheidbarkeit garantiert. Die Einschränkungen von OWL Full hingegen sind weniger restriktiv, allerdings ist Entscheidbarkeit nicht garantiert [W3C04].

RDF-Graph

Die abstrakte Syntax für RDF-basierte Sprachen, zu denen auch OWL gehört, beruht auf dem Datenmodell RDF-Graph. Die Definition des abstrakten Datenmodells ist in Abbildung 2.5 dargestellt.

```
Graph            ::= Set(Triple)
Triple           ::= (Subject, Predicate, Object)
Subject          ::= IRI | BlankNode
Predicate        ::= IRI
Object           ::= IRI | BlankNode | Literal
BlankNode        ::= RDF blank node
Literal          ::= PlainLiteral | TypedLiteral
PlainLiteral     ::= lexicalForm | (lexicalForm, languageTag)
TypedLiteral     ::= (lexicalForm, IRI)
IRI              ::= RDF URI-reference as subsequently restricted by SPARQL
lexicalForm      ::= a Unicode String
```

Abbildung 2.5: RDF-Graph - Definition des abstrakten Datenmodells [W3C12a]

Tripel

Ein RDF-Graph ist eine Menge von RDF-Tripeln. Ein *Tripel* besteht aus drei Komponenten: einem Subjekt, einem Prädikat und einem Objekt[10]. Die Aussage eines Tripels wird als *Statement* bezeichnet. Beim Aufbau eines Tripels gilt, dass ein Literal kein Subjekt, sondern immer nur Objekt eines Tripels sein kann. Daher sind in einem Statement keine Aussagen über Literale möglich. Ein Tripel lässt sich wie in Abbildung 2.6 graphisch darstellen, indem Subjekt und Objekt von Knoten und das Prädikat von einer gerichteten Kante repräsentiert werden. Ein RDF-Graph ist demnach ein gerichteter Graph, der sich durch die enthaltenen Tripel aufspannt. Alle Knoten sind im Graphen eindeutig identifizierbar. Damit erfüllen RDF-Graphen die oben beschriebene Anforderung semantischer Netze an Objektidentität.

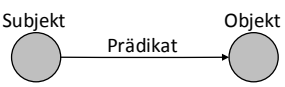


Abbildung 2.6: RDF-Tripel

anonymer Knoten

Eine besondere Art von Knoten sind sogenannte *anonyme Knoten* (*engl.:* blank nodes), die lediglich strukturelle Information enthalten. Sie sind Ressourcen ohne IRI und stattdessen

[9] Der Unterschied zur Prädikatenlogik erster Stufe besteht in der Einschränkung, dass Beschreibungslogiken nur ein- und zweistellige Prädikate erlauben [Den12].

[10] Ein Objekt im RDF-Graphen entspricht nicht dem in Abschnitt 2.2.1 vorgestellten Objektbegriff.

nur lokal über Knoten-Identifizierer identifizierbar. Als Hilfsknoten ermöglichen sie die Beschreibung mehrwertiger Beziehungen zwischen mehr als zwei Ressourcen [HKRS08] und kommen daher oftmals bei der Beschreibung von Restriktionen zum Einsatz.

Für RDF-basierte Sprachen existieren verschiedene etablierte konkrete Syntaxen, beispielsweise RDF/XML, Turtle, N3 oder N-Triples. Auf deren nähere Erläuterung wird an dieser Stelle verzichtet, da in dieser Arbeit zumeist die abstrakte Syntax des RDF-Graphen verwendet wird. Für die spätere Darstellung von Beispielen wird N-Triples [W3C14b] als eine einfache zeilenbasierte Syntax genutzt. Mit N-Triples wird jedes Tripel in der folgenden Form dargestellt:

N-Triples-Syntax

`<subject> <predicate> <object>.`

Beispiel

Die Abbildung 2.7 zeigt als Beispiel den Ausschnitt eines RDF-Graphen, wie er auch im Abschnitt 3.2 für die Wissensrepräsentation der G&R-Dokumente verwendet wird. Daher besitzen die domänenspezifischen Identifizierer den Namensraum-Präfix `<hara:>`.[11] Die Abbildung ist unterteilt in die Modellebene, in der das terminologische Wissen dargestellt ist, und die Instanzebene, welche das Wissen über die Instanzen enthält.

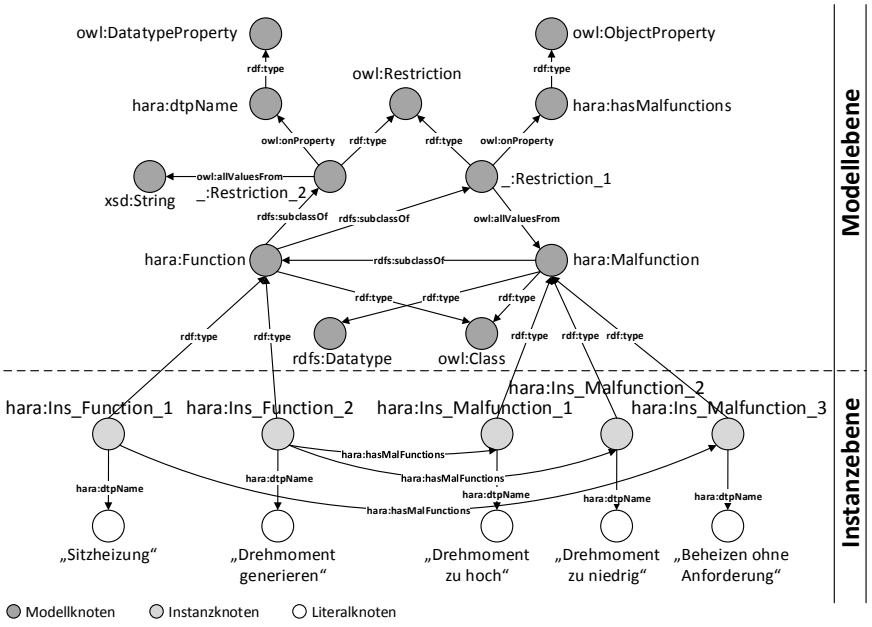


Abbildung 2.7: Beispiel eines RDF-Graphen

Im Beispiel sind auf Modellebene zunächst zwei Klassen beschrieben, wobei `hara:Malfunction` zudem Subklasse von `hara:Function` ist, beschrieben durch das Tripel `<hara:Malfunction> <rdfs:subclassOf> <hara:Function>.` Außerdem werden zwei Eigenschaften beschrieben, zum einen `hara:dtpName` vom Typ `owl:DatatypeProperty` und zum anderen `hara:hasMalfunction` vom Typ `owl:ObjectProperty`. Für jede der Eigenschaften sind Restriktionen festgelegt, die durch anonyme Knoten vom Typ `owl:Restriction` umgesetzt werden, da hier Beziehungen zwischen mehr als zwei Ressourcen beschrieben werden. Beispielsweise besagt `_:Restriction_1`,

[11] Der Präfix steht für die Abkürzung der englischen Bezeichnung für G&Rs, *hazard analysis and risk assessment* (HARA).

dass die Eigenschaft `hara:hasMalfunction` nur für Instanzen der Klasse `hara:Function` gilt und als Objekt ausschließlich Instanzen der Klasse `hara:Malfunction` verwendet werden können.

Auf Instanzebene ist erkennbar, dass die Eigenschaft `hara:hasMalfunction` ausschließlich Instanzen der Klasse `hara:Function` (als Subjekt) mit Instanzen der Klasse `hara:Malfunction` (als Objekt des Tripels) verbindet und daher die Restriktion der Modellebene erfüllt ist. Entsprechend verbindet `hara:dtpName` ausschließlich Instanzen der Klasse `hara:Function` oder deren Subklasse `hara:Malfunction` mit Literalen vom Typ `xsd:String`, beispielsweise im Tripel `<hara:Ins_Function_1> <rdfs:dtpName> <"Sitzheizung">.`

2.2.3 Datenmodellbasierte Transformationen

Die Entwicklung von Ontologiemodellen und Wissensbasen für eine Domäne ist ein aufwendiger und langwieriger Prozess, in dem Domänenexperten Begriffe, Eigenschaften und Beschränkungen in einer Wissensdomäne zum einen identifizieren und zum anderen in eine valide Form bringen müssen [AS11]. Die von Domänenexperten erstellten Ontologiemodelle sind zudem wartungsintensiv, da Änderungen in der Domäne fortlaufend auch in den Ontologien aktualisiert werden müssen. Teile des Wissens werden allerdings bereits beim Design von Unternehmensanwendungen und Datenbanken berücksichtigt. Daher hat sich gemäß Albarrak und Sibley die Verwendung der Datenmodelle solcher Anwendungen oder Datenbanken als eine pragmatische Herangehensweise etabliert. Eine W3C-Empfehlung [W3C12a] hält ein direktes Mapping relationaler Daten zu RDF und somit auf RDF-Graphen bereit. Werkzeuge wie die D2RQ-Plattform [D2RQ] implementieren das W3C-Mapping und erlauben den Zugriff und sogar SPARQL Protocol and RDF Query Language (SPARQL)-Abfragen auf virtuellen RDF-Graphen von relationalen Datenbanken, teilweise ohne diese replizieren zu müssen.

Die direkte Transformation von relationalen Datenbanken zu OWL und damit zu den dazugehörigen RDF-Graphen ist praktisch. Allerdings ist ein direkter Zugriff auf die Datenbank nicht immer möglich. In dieser Arbeit wird ebenfalls ein datenmodellbasierter Ansatz verfolgt, um bei der Wissensrepräsentation manuellen Aufwand zu reduzieren und die Automatisierbarkeit zu erhöhen. Der untersuchte Datensatz basiert auf einem relationalen Datenmodell, liegt aufgrund eines Datenexports allerdings basierend auf der Extensible Markup Language (XML) vor. Für die Transformation XML-basierter Daten zu RDF oder OWL existiert keine W3C-Empfehlung. Dennoch gibt es verschiedene Ansätze für derartige Transformationen. Die Grundidee ist zumeist, eine Abbildung von syntaktischen XML- und XML Schema Definition (XSD)-Elementen auf die entsprechenden Sprachkonstrukte von RDF, RDF-Schema und OWL bereitzustellen. Dies basiert darauf, das XML-Schema in ein OWL-Modell und die XML-Daten entsprechend in OWL-Instanzen zu konvertieren. Diese Grundidee wird unter anderem von Bohring und Auer [BA05] für ihre Strategie zur automatischen Generierung von OWL-Ontologien aus XML-Daten verwendet und ist in Abbildung 2.8 dargestellt. Zudem präsentieren Bohring und Auer ein Vorgehen zur automatisierten Ableitung des XML-Schemas aus XML-Dokumenten, unter der Annahme, dass diese auf relationalen Strukturen basieren. Für die praktische Durchführung der Transformationen stellen sie eine Implementierung mittels Extensible Stylesheet Language Transformations (XSLT) zur Verfügung. Die unterschiedlichen Ansätze teilen diese Grundidee, unterscheiden sich aber in der Art der Trennung von RDF/OWL-Modell und Instanzen, den konkreten Mappings und der Verfügbarkeit von Werkzeugunterstützung.

Die Java-Bibliothek *Ontmalizer*[12] setzt ebenfalls ein Mapping [Yük13, S. 45f.] von XML-Schema und XML-Daten zu OWL um, welches in Abbildung A.1 in Anhang A (Seite 217)

[12] Die Ontmalizer-Bibliothek [Ontmalizer] ist Ergebnis des europäischen Forschungsprojektes Salus [Salus].

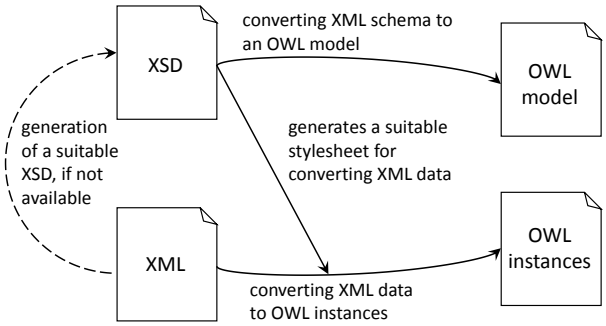


Abbildung 2.8: Ablauf der Transformation von XML/XSD zu OWL [BA05]

abgebildet ist. Angelehnt an das Vorgehen in Abbildung 2.8 wird auf Basis des Datenmodells mittels *XSD2OWL*-Transformationen die OWL-Modellebene erstellt. Darauf aufbauend wird eine Instanziierung der Daten mittels *XML2OWL*-Transformationen vorgenommen. Die Ontmalizer-Bibliothek bildet im Kern das Mapping des kommerziellen TopBraid Composers [TopBraid] nach, hat allerdings den Vorteil, dass sie frei verfügbar ist. Daher wird sie bei der Erstellung der semantischen Netze in dieser Arbeit verwendet.

2.2.4 Zusammenfassung und Bezug zur Arbeit

In diesem Abschnitt wurde die graphbasierte Wissensrepräsentation durch semantische Netze vorgestellt, die auf RDF-Graphen basieren. Dafür haben wir einen Überblick über die RDF-Sprachfamilie gegeben und Sprachkonstrukte vorgestellt, die wir im weiteren Verlauf der Arbeit bei der Erstellung semantischer Netze verwenden. In dieser Arbeit verfolgen wir das Ziel, die semantischen Netze möglichst automatisiert zu erzeugen, um den manuellen Aufwand von Domänenexperten zu verringern. Da wir zudem ein integriertes IR-System im jeweiligen Entwicklungswerkzeug anstreben, machen wir uns das Datenmodell des Entwicklungswerkzeugs und die entsprechend strukturierten Daten zunutze. Der strukturierte, tripelbasierte Aufbau von RDF-Graphen eignet sich für die Repräsentation derartig strukturierter Daten. Daher haben wir Techniken zur datenmodellbasierten Transformation vorgestellt, die RDF-Graphen aus relationalen Datenbanken oder XML-/XML-Schema-Dokumenten erzeugen. In unser prototypischen Umsetzung kommt die vorgestellte Ontmalizer-Bibliothek zum Einsatz.

IR-Systeme unterscheiden sich hauptsächlich in der Art der Wissensrepräsentation und der Retrievalstrategie. Im folgenden Abschnitt 2.3 beschreiben wir die assoziative Retrievalstrategie Aktivierungsausbreitung als zweite dieser beiden Komponenten detaillierter.

2.3 Retrieval durch Aktivierungsausbreitung

Aktivierungsausbreitung hat ihren Ursprung im Forschungsfeld der Psychologie und Psycholinguistik. In frühen Arbeiten [Qui68, CL75, And83] ist Aktivierungsausbreitung ein theoretisches Modell, welches die Prozesse veranschaulichen soll, die bei der Auswahl eines Wortes im Gehirn beziehungsweise Gedächtnis ablaufen. Die Annahme ist, dass Begriffe in einem Begriffsnetz über die assoziierten Zusammenhänge miteinander verbunden sind. Wird ein Begriff ausgewählt, also aktiviert, so werden andere mit diesem Begriff assoziierte Begriffe über die Assoziationen ebenso aktiviert. In dem Modell geht man üblicherweise davon aus, dass über die Distanz die Intensität der Aktivierung eines Knotens abnimmt, also weniger stark zusammenhängende Knoten weniger relevant sind. Die Mechanismen dieses theoretischen Modells wurden im Laufe der Zeit auf unterschiedliche Anwendungen übertragen, beispielsweise in den Bereichen der Kognitionswissenschaft, Datenbanken, Künstlichen Intelligenz, Biologie und Informationsgewinnung [Cre97]. Im Bereich des IR wird die Aktivierungsausbreitung als Verfahren für die Rückgewinnung von Informationen aus Informationssammlungen beziehungsweise für die Relevanzbewertung in assoziativen oder semantischen Netzen verwendet. Aktivierungsausbreitung bildet in dieser Arbeit die Ausgangsbasis für diverse Konzeptualisierungen und Erweiterungen, welche im IR-System zur Unterstützung von G&Rs zum Einsatz kommen und deren Qualität in der Evaluation analysiert werden. Daher beschreiben wir in diesem Abschnitt die grundlegende Anwendung von Aktivierungsausbreitung im IR.

2.3.1 Ablauf der Aktivierungsausbreitung im Information Retrieval

Aktivierungsausbreitung ist ein assoziatives Retrievalverfahren, welches auf graphbasierten Repräsentationen durchgeführt wird. Die Anfrage wird repräsentiert durch einen oder mehrere Startknoten, welche zu Beginn des Retrievalprozesses einen initialen Aktivierungswert zugewiesen bekommen und dadurch als einzige Knoten im Graphen einen aktivierten Status einnehmen. Abbildung 2.9 stellt den grundsätzlichen Ablauf des Retrievalverfahrens nach Crestani [Cre97] dar.

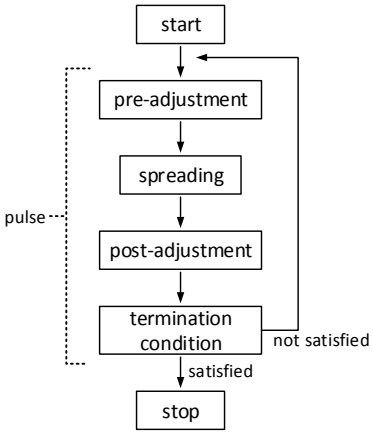
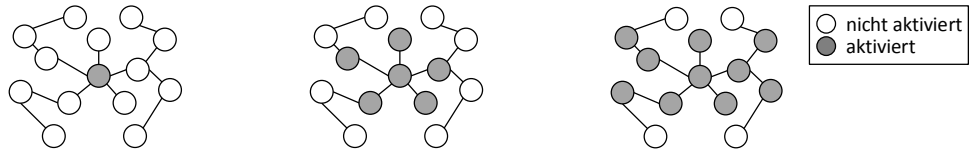


Abbildung 2.9: Ablauf des Retrievals durch Aktivierungsausbreitung nach [Cre97]

Demnach besteht der Retrievalprozess aus mehreren Iterationen, den sogenannten Pulsen (*engl.:* pulses). Jeder Puls besteht aus drei Phasen: einer Vorverarbeitungsphase (*engl.:* pre-adjustment), einer Aktivierungsausbreitungsphase (*engl.:* spreading) und einer Nachbear-

beitungsphase (*engl.:* post-adjustment). Vorverarbeitungs- und Nachbearbeitungsphase sind optional und erlauben die Integration zusätzlicher Maßnahmen, beispielsweise zur Steuerung der Ausbreitung im Graphen oder der Aktivierungswertberechnung. In der Aktivierungsausbreitungsphase werden, ausgehend von bereits aktivierten Knoten, wellenartig deren benachbarte Knoten als aktiviert markiert. In jedem Puls werden die Aktivierungswerte aller Knoten neu berechnet. Daher befindet sich der Graph am Ende jedes Pulses in einem aktualisierten Aktivierungszustand. Wird am Ende eines Pulses ein Abbruchkriterium erreicht, so endet der Algorithmus. Nach Anwendung des Algorithmus sind allen Knoten des betrachteten Graphen Aktivierungswerte zugeordnet, welche als Maß für deren Relevanz zu den Startknoten gelten. Die Interpretation des Ergebnisses ist anwendungsabhängig, grundsätzlich bedeuten allerdings höhere Aktivierungswerte eine höhere Relevanz. Abbildung 2.10 zeigt das Grundprinzip der Aktivierungsausbreitung in einem vereinfachten Ablauf zweier Pulse.



(a) Initial (b) Puls 1 (c) Puls 2

Abbildung 2.10: Schematischer Ablauf der Aktivierungsausbreitung: Aktivierungszustand des Netzes a) vor Beginn der Aktivierungsausbreitung mit einem aktivierten Startknoten, b) nach einem Puls und c) nach zwei Pulsen

2.3.2 Aktivierungsverteilungsfunktionen

Die Aktivierung wird gesteuert durch drei Aktivierungsverteilungsfunktionen: die Output-Funktion, die Input-Funktion und die Aktivierungsfunktion.[13] Mit den Funktionen werden die Output-, Input- und Aktivierungswerte in jedem Aktivierungsschritt berechnet. Grundsätzlich wird zwischen der puren, unbeschränkten Aktivierungsausbreitung (*engl.:* pure spreading activation) und der beschränkten Aktivierungsausbreitung (*engl.:* constrained spreading activation) unterschieden. Bei der puren Aktivierungsausbreitung wird ein Netz unbeschränkt geflutet, wie auch in Abbildung 2.10 dargestellt. Sie ist die einfachste Art der Aktivierungsausbreitung. Die beschränkte Aktivierungsausbreitung integriert Beschränkungen (*engl.:* constraints) in die Aktivierungsverteilungsfunktionen, welche die Berechnung von Aktivierungswerten und die Ausbreitung gezielt steuern. Üblicherweise sind Aktivierungswerte und Kantengewichte laut Crestani [Cre97] aus dem Bereich der reellen Zahlen. Dennoch ist es grundsätzlich gemäß anwendungsspezifischen Anforderungen möglich, dass beispielsweise binäre Werte (0 oder 1) oder hemmende beziehungsweise stimulierende Werte (+1, -1) verwendet werden [Cre97].

Die **Output-Funktion** bestimmt die Output-Aktivierung eines Knotens, also den Wert, der über die ausgehenden Kanten zu den Nachbarknoten transportiert wird. Bei der puren Aktivierungsausbreitung wird der aktuelle Aktivierungswert des Knotens als Output-Aktivierung ausgeleitet.

[13] Häufig werden die drei Funktionen unter dem Oberbegriff Aktivierungsfunktionen zusammengefasst. Zur begrifflichen Trennung zwischen der Teilfunktion und der Gesamtheit der drei Aktivierungsfunktionen werden die Funktionen in dieser Arbeit unter dem Begriff Aktivierungsverteilungsfunktionen zusammengefasst.

Die **Input-Funktion** bestimmt die Input-Aktivierung eines Knotens, also die eingehende Aktivierung in einen Knoten über alle eingehenden Kanten. Diese entspricht im einfachsten Fall den Output-Aktivierungen, welche über die Kanten transportiert werden. Da die Kanten in assoziativen Netzen, abgesehen von der schlichten Existenz einer Verbindung zwischen den Knoten, keine explizite Semantik tragen, werden alle Kanten meist gleichwertig betrachtet. In semantischen Netzen, in denen Kanten aufgrund ihrer Bedeutung unterschiedlich wichtig sind, können Kantengewichte die Unterschiede betonen. Dann entspricht die Input-Aktivierung eines Knotens der Aggregation aller gewichteten, eingehenden Aktivierungen des Knotens.

Die **Aktivierungsfunktion** bestimmt am Ende des Pulses den neuen Aktivierungszustand jedes Knotens. Dabei werden die bereits vorhandene und die eingehende Aktivierung berücksichtigt. Da Aktivierungswerte unter günstigen Bedingungen schnell und stark wachsen können, ist eine Aufgabe der Aktivierungsfunktion die anwendungsspezifische Steuerung und Begrenzung der Aktivitätslevel der Knoten. Das wird üblicherweise durch Anwendung spezieller Funktionen wie Logarithmus- oder Sigmoidfunktionen erreicht.

Die konkrete Definition der Funktionen variiert je nach Anwendung. Die Grundidee ist aber, dass die drei Funktionen in jedem Puls der Aktivierungsausbreitung in einer spezifizierten Reihenfolge ausgeführt werden. Ein neuer Puls startet mit der Berechnung der Output-Aktivierungswerte jedes Knotens durch die Output-Funktion, basierend auf dem Aktivierungszustand der Knoten des vorherigen Pulses (sofern vorhanden). Anschließend werden die Input-Aktivierungswerte jedes Knotens bestimmt, welche wiederum für die Berechnung der neuen Aktivierungszustände der Knoten genutzt werden.

2.3.3 Beschränkungen

Beschränkungen steuern die Aktivierungsausbreitung zusätzlich und werden in die Aktivierungsverteilungsfunktionen integriert. Beschränkungen beeinflussen unter anderem die Berechnung der Aktivierungswerte, aber auch direkt oder indirekt den Ausbreitungsverlauf sowie die Terminierung der Aktivierungsausbreitung. Im Folgenden werden einige bekannte Beschränkungen, welche auch in dieser Arbeit eine Rolle spielen, gemäß Crestanis Beschreibungen [Cre97] kurz erläutert.

Aktivierungsbeschränkung (*engl.:* activation constraint) integriert eine Steuerung des Aktivierungsprozesses durch Schwellwerte (*engl.:* thresholds). Knoten müssen, um selber Aktivierung verteilen zu können, zuerst einen festgelegten Schwellwert erreichen. Dadurch kann verhindert werden, dass Knoten mit geringer Aktivierung und daher einer vermutlich geringen semantischen Bedeutung am Verteilungsprozess teilnehmen. Schwellwerte können statisch festgelegt werden oder sich dynamisch pulsweise in Abhängigkeit zum Gesamtaktivierungszustand des Netzes anpassen. Sie können global oder knotenindividuell gelten, beispielsweise in Relation zur semantischen Bedeutung.

Dämpfung (*engl.:* decay) bezeichnet die Abschwächung von Aktivierungswerten. Mit Dämpfung können beispielsweise die Aktivierungswerte von Knoten verringert werden, wenn diese nicht kontinuierlich aktiviert werden, um so nachlassendes Interesse (*engl.:* loss of interest) nachzubilden. Dämpfung kann aber auch pulsweise erfolgen, so dass mit jedem Aktivierungsschritt weniger Aktivierung weitergegeben wird.

Distanzbeschränkung (*engl.:* distance constraint) dient ähnlich der Dämpfung zur Abschwächung von Aktivierungswerten. Allerdings wird die Aktivierung von Knoten gedämpft oder sogar gestoppt je weiter sie von den Startknoten entfernt sind. Dieser Beschränkung liegt die Annahme zugrunde, dass die semantische Distanz zweier Knoten bei steigender

räumlicher Distanz (Pfadlänge) ebenso zunimmt.

Verzweigungsbeschränkung (*engl.:* fan-out-constraint) ist die Bestrafung sich stark verzweigender Knoten durch Reduzierung ihrer Output-Aktivierung. Grund hierfür ist, dass solchen Knoten eine sehr breite semantische Bedeutung zugeschrieben wird und im Falle einer starken Verteilung in viele Richtungen die Gefahr der Verwässerung des Ergebnisses besteht.

Pfadbeschränkung (*engl.:* path constraint) bedeutet die Bevorzugung semantisch bedeutsamer Pfade sowie die Bestrafung semantisch weniger bedeutsamer Pfade. Sie wird üblicherweise durch Kantengewichtungen oder kantenspezifische Aktivierungsverteilungsfunktionen gesteuert.

Terminierung der Aktivierungsausbreitung erfolgt üblicherweise nach einer maximalen Anzahl an Pulsen. Alternativ kann die Aktivierungsausbreitung beispielsweise terminiert werden, sobald eine maximale Anzahl an Knoten oder ein festgelegter Anteil des Netzes aktiviert ist.

Nachteile unbeschränkter Aktivierungsausbreitung

Die fehlenden Steuerungsmöglichkeiten bei der unbeschränkten Aktivierungsausbreitung führen zum einen dazu, dass ein Netz bereits innerhalb weniger Pulse komplett aktiviert werden kann und die Aktivierungswerte durch die kontinuierliche gegenseitige Aktivierung über die Pulse exponentiell wachsen [Cre97]. Thiel [Thi12] beweist in seiner Dissertation, dass der Einsatz der puren, unbeschränkten Aktivierungsausbreitung immer gegen einen anfrageunabhängigen Fixpunkt konvergiert. Crestani [Cre97] empfiehlt zur Steuerung der Aktivierungsausbreitung die Nutzung der Semantik, welche möglicherweise in den Gewichtungen oder Beschriftungen von Kanten und Knoten enthalten ist. Beschränkte Aktivierungsausbreitung, so Crestani, ermöglicht die kantenspezifische Behandlung beim Transport von Aktivierung, basierend auf deren Semantik. Für die Qualität der Ergebnisse ist daher die Konfiguration des Basisalgorithmus durch gezielten Einsatz von Beschränkungen unverzichtbar.

2.3.4 Zusammenfassung und Bezug zur Arbeit

Dieser Abschnitt enthält die Grundprinzipien der Aktivierungsausbreitung bei der Verwendung im IR. Kern der Aktivierungsausbreitung ist die pulsweise Markierung von benachbarten Knoten sowie die Vergabe von Aktivierungswerten, die ein Maß für gegenseitige Relevanz und damit auch semantische Zusammenhänge bieten. Es werden insbesondere der generische Ablauf, die Verwendung von Aktivierungsverteilungsfunktionen und eine Auswahl etablierter Beschränkungen vorgestellt, die üblicherweise in die Aktivierungsverteilungsfunktionen integriert werden. Allerdings existieren weder einheitlich definierte Aktivierungsverteilungsfunktionen noch gibt es Leitfäden für deren geeignete Definition. Stattdessen unterscheiden sich konkrete Anwendungen häufig in der Formulierung der Aktivierungsverteilungsfunktionen. Daher sind die vorgestellten Grundprinzipien Ausgangsbasis für eine Konzeptualisierung der Aktivierungsverteilungsfunktionen und für eine Erweiterung um Ausbreitungsmodi, die eine zusätzliche, direkte Steuerung des Ausbreitungsverhaltens erlauben. Sie ermöglichen die reproduzierbare und konfigurierbare Verwendung der Aktivierungsausbreitung in IR-Systemen.

Die Aktivierungsausbreitung verwenden wir als Retrievalverfahren für das in dieser Arbeit entwickelte domänenspezifische IR-System, welches die Durchführung von G&Rs unterstützt. Daher widmet sich der folgende Abschnitt 2.4 den Grundlagen der funktionalen Sicherheit in der Systementwicklung mit besonderem Fokus auf die G&R als zentrale Sicherheitsaktivität.

2.4 Funktionale Sicherheit in der automobilen Systementwicklung

Aufgrund des zunehmenden Einsatzes von elektronischen und elektrischen Systemen wächst die Bedeutung der funktionalen Sicherheit seit Jahren [DINEN61508-1]. Im Jahr 1998 wurde die Norm IEC 61508 veröffentlicht, die branchenübergreifende Richtlinie zur funktionalen Sicherheit für sicherheitsbezogene elektrische, elektronische und programmierbare elektronische Systeme. Diese Grundnorm liegt inzwischen auch als europäische und deutsche Norm DIN EN 61508 vor. Branchenspezifische Normen konkretisieren die allgemeine Norm für unterschiedliche Anwendungsbereiche, z.B. die Norm ISO 26262 [ISO26262] für den Automobilbereich. Diese Norm wird angewendet für sicherheitsbezogene elektrische und elektronische Systeme (E/E-Systeme) (inklusive programmierbare elektronische Systeme) in serienmäßig hergestellten Personenfahrzeugen bis 3,5t. Sie wurde mit Beteiligung internationaler Automobilhersteller und großer Zulieferer der Branche erarbeitet und im November 2011 veröffentlicht. Sie beschreibt den aktuellen Stand der Technik. Ihre Anwendung ist freiwillig. Im Schadensfall wird sie allerdings relevant. Dann müssen Hersteller gemäß Produkthaftungsgesetz [ProdHaftG] nachweisen, dass ein Produkt, welches den Schaden verursacht hat, nach aktuellem Stand der Technik[14] hergestellt wurde. Daher empfiehlt sich die Anwendung der Norm nicht nur, sondern ist branchenüblich.

Dieser Abschnitt gibt einen Überblick über die Grundbegriffe und Ausführungen in der Norm ISO 26262 zur Integration von Sicherheitsaktivitäten in die Systementwicklung. In dieser Arbeit entwickeln wir ein domänenspezifisches IR-System, welches die Durchführung von G&Rs unterstützt. Daher beschreiben wir insbesondere den Ablauf der G&R als zentrale Sicherheitsaktivität und welche Anforderungen die Norm an sie stellt.

2.4.1 Grundbegriffe

Funktionale Sicherheit

In der Norm ISO 26262 [ISO26262-1] ist *Funktionale Sicherheit (FuSi)* (*engl.:* functional safety)[15] definiert als die Abwesenheit von unvertretbarem Risiko durch Gefährdungen (*engl.:* hazards) aufgrund des Fehlverhaltens (*engl.:* malfunctioning behaviour) von E/E-Systemen. Um funktionale Sicherheit zu gewährleisten, bedarf es einer sorgfältigen Entwicklung. Verbaute Systeme werden im Zuge dieser Entwicklung in Betrachtungseinheiten untersucht. Eine

[14] Helmig [Hel13] beschreibt aus juristischer Sicht, dass *Stand der Technik* oder *neuester Stand von Wissenschaft und Technik* unbestimmte Rechtsbegriffe sind. Genau genommen kann nach Helmig die branchenspezifische Normgebung wie die ISO 26262 „nie auf der Höhe der Zeit sein“, sondern folgt letztendlich eher der Branchenüblichkeit. Damit unterstützt er eine Position, die auch der Bundesgerichtshof im sogenannten Airbag-Urteil [BGH] verkündet hat, nämlich, dass der „maßgebliche Stand der Technik nicht mit Branchenüblichkeit gleichgesetzt werden“ darf, da branchenübliches Vorgehen „durchaus hinter der technischen Entwicklung und damit den rechtlich gebotenen Maßnahmen zurückbleiben“ [BGH] kann. Zur Vermeidung von Gefahren reichen dennoch „nach gesichertem Fachwissen der einschlägigen Fachkreise praktisch einsatzfähige Lösungen“ [BGH], so wie sie in der Norm ISO 26262 niedergeschrieben sind. Neueste, aber noch nicht serienreife technische Möglichkeiten müssen nicht verpflichtend angewendet werden. Da Risiken nach dem aktuellen Stand der Technik nicht immer vermeidbar sind, wird im Urteil die Wichtigkeit der „Abwägung von Art und Umfang der Risiken, der Wahrscheinlichkeit ihrer Verwirklichung und des mit dem Produkt verbundenen Nutzens“ [BGH] betont. Dies unterstreicht die Bedeutung der G&R in der frühen Konzeptphase der Systementwicklung.

[15] Oftmals weichen die Definitionen der Begrifflichkeiten zur FuSi in Normen, Literatur und Gebrauch geringfügig voneinander ab. Zur Begriffsklärung wird in dieser Arbeit hauptsächlich die aktuelle Fassung der Norm ISO 26262 in englischer Sprache berücksichtigt. Die hier verwendeten Übersetzungen der Begrifflichkeiten entsprechen dem üblichen Gebrauch, beziehungsweise den häufigen Übersetzungen in der Sekundärliteratur (beispielsweise [Bör15][Ros14][LPP10]). Definitionen übergeordneter Begriffe werden teilweise der aktuellen deutschen Fassung der branchenübergreifenden Norm DIN EN 61508 [DINEN61508-Bbl, DINEN61508-1] entnommen.

Betrachtungseinheit (*engl.:* item) implementiert auf Fahrzeugebene eine Fahrzeugfunktion, auf welche die Norm ISO 26262 angewendet wird. Ein *System* ist nach dieser Norm eine Menge von Elementen (*engl.:* elements), die mindestens einen Sensor, Controller und Aktuator verbindet. Außerdem wird ein *Element* als ein System oder Teil eines Systems definiert, inklusive Komponenten, Hardware, Software, Hardware-Teilen und Software-Einheiten. In der Praxis werden die Begriffe Betrachtungseinheit und System oftmals synonym verwendet.

Betrachtungseinheit

System

Wenn eine Betrachtungseinheit versagt oder nicht intendiertes Verhalten zeigt, wird dies in der Norm [ISO26262-1] als *Fehlverhalten* bezeichnet. Das Fehlverhalten einer Betrachtungseinheit kann *Schaden* (*engl.:* harm) verursachen, wobei die Norm ISO 26262 ausschließlich auf die Vermeidung von Schäden für die Gesundheit von Menschen abzielt. Eine *Gefährdung* ist eine potenzielle Schadensquelle. Ein *Risiko* wird beschrieben als Verbindung der Auftretenswahrscheinlichkeit eines Schadens und dessen Schwere (*engl.:* severity). Die *Schwere* ist wiederum die Abschätzung des Schadensausmaßes für ein oder mehrere Individuen, welches in einer potenziellen *Gefährdungssituation* (*engl.:* hazardous event) auftreten kann [ISO26262-1]. Als *unvertretbar* gilt in diesem Kontext, was gemäß geltender gesellschaftlicher moralischer Konzepte als unakzeptabel beurteilt wird.

Fehlverhalten

Schaden

Gefährdung

Risiko

Schwere

Gefährdungssituation

Funktioniert beispielsweise trotz Bremsung die Bremsbeleuchtung nicht, spricht man von Fehlverhalten. Die Kollision, die dadurch eintreten könnte, bezeichnet man als Gefährdung. Im Falle ihres Eintretens kann die Kollision einen Schaden verursachen. Funktionale Sicherheit ist gewährleistet, wenn das Restrisiko einer Kollision, also das „trotz Schutzmaßnahmen verbleibende Risiko“ [Bör15], vertretbar bleibt.

Als *sicherheitsbezogen* (*engl.:* safety-related) werden Elemente, Funktionen oder Systeme bezeichnet, welche das Potenzial besitzen, das Erreichen eines Sicherheitsziels (*engl.:* safety goal) zu beeinflussen. *Sicherheitsziele* sind gemäß Norm Top-Level-Sicherheitsanforderungen an eine Betrachtungseinheit. Sie sind das Ergebnis der systematischen Identifikation und Kategorisierung von Gefährdungen und Gefährdungssituationen der Betrachtungseinheiten, deren Risikobewertung sowie deren Einstufung der Sicherheitsintegrität. Die *Sicherheitsintegrität* ist ein Schlüsselkonzept der FuSi. Sie wird bestimmt über „die Wahrscheinlichkeit einer zufriedenstellend ausgeführten Sicherheitsfunktion“, wobei *Sicherheitsfunktionen* „besondere Funktionen [sind], die sicherstellen sollen, dass Risiken auf einer vertretbaren Stufe gehalten werden“ [DINEN61508-Bbl]. Das bedeutet, „je höher die Stufe der Sicherheitsintegrität ist, desto geringer ist die Wahrscheinlichkeit eines gefahrbringenden Ausfalls“ des sicherheitsbezogenen Systems, da „höhere Stufen der Sicherheitsintegrität [...] eine strengere Entwicklung [erfordern]“ [DINEN61508-Bbl]. Das Sicherheitsintegritätslevel wird gemäß DIN EN 61508 mittels Safety Integrity Level (SIL)-Einstufung und automobilspezifisch mittels ASIL-Einstufung bestimmt. Im Automobilbereich existieren vier ASIL-Stufen, von der kleinsten Stufe A bis zur höchsten und damit strengsten Stufe D. Sie spezifizieren die notwendigen Anforderungen durch die Norm ISO 26262 und anzuwendende Sicherheitsmaßnahmen zur Gewährleistung funktionaler Sicherheit.

Sicherheitsziel

Sicherheitsintegrität

2.4.2 Sicherheitslebenszyklus

Die Norm ISO 26262 beschreibt Anforderungen an die Entwicklung sicherheitsbezogener Systeme. Ein zentraler Aspekt im Standard ist der *Sicherheitslebenszyklus*. Er wurde vom allgemeiner beschriebenen Sicherheitslebenszyklus für E/E-Systeme der IEC 61508 abgeleitet und auf die Systementwicklung im Automobilbereich angepasst [Ros14]. Insbesondere wurde dem automobilen Sicherheitslebenszyklus das V-Modell als branchenübliches Vorgehensmodell zugrunde gelegt [LPP10].

Sicherheitslebenszyklus

Der Sicherheitslebenszyklus einer Betrachtungseinheit beschreibt die sicherheitsbezogenen Aktivitäten, hier auch Sicherheitsaktivitäten, für alle Phasen, vom Konzept bis zur Außerbetriebnahme [ISO26262-1]. Dabei sind die Sicherheitsaktivitäten eng verflochten mit den nicht explizit sicherheitsbezogenen funktions- und qualitätsorientierten Entwicklungsaktivitäten und deren Arbeitsprodukten [ISO26262-1]. Sie ergänzen die Aktivitäten des Entwicklungsprozesses. Dabei entstehen teilweise eigenständige FuSi-Artefakte, wie die Dokumentation der G&R. Teilweise werden Ergebnisse aber auch in bekannte Artefakte der Systementwicklung integriert. Beispielsweise erweitern die im funktionalen Sicherheitskonzept erarbeiteten Sicherheitsanforderungen die auch ohne FuSi bestehenden Anforderungen in den Systemlastenheften. Diesen engen Verflechtungen wird in der Norm ISO 26262 mit dem Sicherheitslebenszyklus Rechnung getragen. Abbildung 2.11 zeigt den Sicherheitslebenszyklus mit dessen Phasen und Teilphasen.

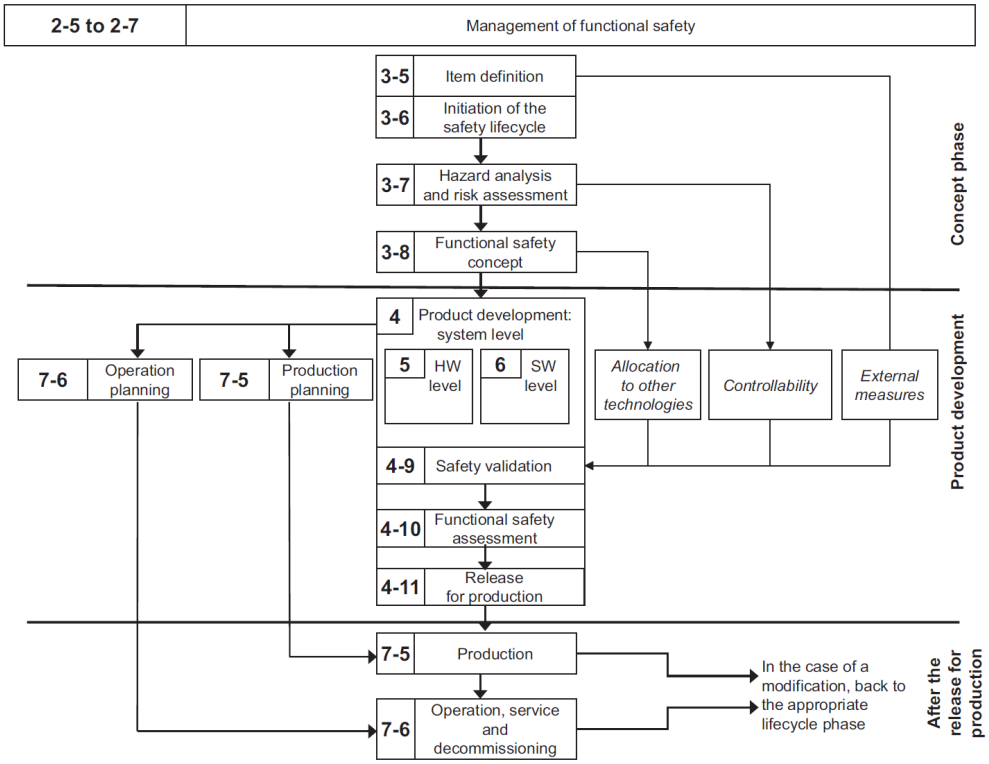


Abbildung 2.11: Die drei Phasen des Sicherheitslebenszyklus mit übergeordnetem FuSi-Management [ISO26262-2, Fig. 2]. Die vorangestellten Nummerierungen entsprechen den jeweiligen Abschnitten in der Norm.

Im Sicherheitslebenszyklus werden den drei elementaren Phasen die Haupt-Sicherheitsaktivitäten zugeordnet. Für die Sicherheitsaktivitäten und deren Arbeitsprodukte hält die Norm ISO 26262 wiederum Anforderungen und Empfehlungen bereit. Die Systementwicklung beginnt mit der Konzeptphase (*engl.:* concept phase), gefolgt von der Produktentwicklung (*engl.:* product development) und der Phase nach Produktionsfreigabe (*engl.:* after the release for production) [ISO26262-2]. Übergeordnet gehört zum Sicherheitslebenszyklus das FuSi-Management, welches vor allem die Planung, Koordination und Dokumentation der Sicherheitsaktivitäten während der verschiedenen Phasen umfasst und darüber hinaus allgemeine Anforderungen an die Sicherheitskultur einer Organisation beschreibt [ISO26262-2]. Im Folgenden werden die drei Phasen des Sicherheitslebenszyklus kurz erläutert.

Konzeptphase

Am Anfang der Systementwicklung steht die Konzeptphase. Zu ihr gehören Sicherheitsaktivitäten zur Erkennung der Gefährdungen und Festlegung der konkreten Sicherheitsanforderungen. Sorgfältige und korrekte Einschätzungen zu Beginn sollen vor Fehlern schützen, die nachträglich kostenintensiv behoben werden müssen.

Eine Betrachtungseinheit muss zunächst möglichst genau definiert werden (*engl.:* item definition). Gemäß Norm [ISO26262-3] sollen insbesondere Zweck und Funktionalität sowie Systemgrenzen, Abhängigkeiten und Schnittstellen zu anderen Systemen und Umgebungsbedingungen definiert werden. Auch Annahmen über erwartetes Verhalten oder bereits bekannte Gefährdungen sowie mögliche juristische Anforderungen sollen, soweit möglich, vorab beschrieben werden. In dieser Teilphase werden alle verfügbaren und relevanten Informationen zusammengetragen, die zu einem angemessenen Verständnis der Betrachtungseinheit beitragen. Für die definierte Betrachtungseinheit kann der Sicherheitslebenszyklus gestartet werden (*engl.:* initiation of safety lifecycle). Ausgehend von der Beschreibung der Betrachtungseinheit wird für diese eine Gefährdungs- und Risikoanalyse (*engl.:* hazard analysis and risk assessment) durchgeführt. Mit dieser Sicherheitsaktivität werden Gefährdungen aufgrund von Fehlfunktionen (*engl.:* malfunction) der Betrachtungseinheit erkannt, Gefährdungssituationen bewertet und Sicherheitsziele zur Verhinderung oder Eindämmung dieser Gefährdungssituationen und zur Vermeidung unvertretbarer Risiken festgelegt. Mit dem Funktionalen Sicherheitskonzept (*engl.:* functional safety concept) werden aus den Sicherheitszielen detaillierte Sicherheitsanforderungen für die Betrachtungseinheit abgeleitet.

Phase der Produktentwicklung

Die *Produktentwicklungsphase* umfasst alle an die Konzeptphase anschließenden Aktivitäten bis zur Freigabe der Produktion. Aus den funktionalen Sicherheitsanforderungen der Konzeptphase werden bei Erstellung eines technischen Sicherheitskonzeptes die technischen Sicherheitsanforderungen abgeleitet. Diese werden beim Systemdesign der Hardware und Software zugeordnet [ISO26262-4]. Sowohl für die Hardware- als auch für die Softwareentwicklung werden eigene an das V-Modell angelehnte Produktionszyklen gestartet, welche dann die für sie geltenden technischen Sicherheitsanforderungen umsetzen [Geb13]. Die Norm ISO 26262 beschreibt, welche Maßnahmen zur Erfüllung der Anforderungen getroffen werden sollen, angefangen vom Systemdesign bis hin zur Testplanung und -durchführung auf verschiedenen Teststufen. Außerdem wird sowohl die Integration von Hard- und Software innerhalb der Betrachtungseinheit als auch deren Integration auf Fahrzeugebene geplant und durchgeführt, so dass FuSi für das Gesamtfahrzeug gewährleistet werden kann [ISO26262-4]. Damit eine Freigabe zur Serienproduktion erfolgen kann, muss eine Sicherheitsvalidierung (*engl.:* safety validation) der integrierten Betrachtungseinheit durchgeführt werden. Bei dieser soll zum einen die Angemessenheit und Übereinstimmung von Sicherheitszielen und Sicherheitskonzepten nachgewiesen werden. Zum anderen soll gezeigt werden, dass die Sicherheitsziele korrekt, vollständig und auf Fahrzeugebene erfüllt sind und angemessene Sicherheitsmaßnahmen verwendet wurden [ISO26262-4]. Das geschieht auf Basis des Sicherheitsnachweises aus der Konzeptphase (*engl.:* safety case).

Phase nach Produktionsfreigabe

Die dritte Phase umfasst die verbleibenden Sicherheitsaktivitäten *nach Produktionsfreigabe*. Dazu gehören sowohl Produktion, Betrieb und Wartung als auch die Außerbetriebnahme.

Hier wird in erster Linie die Einhaltung der Produktions- und Betriebsplanung überwacht, insbesondere auch das Testequipment und die verwendeten Werkzeuge. Im Rahmen des Betriebs und der Wartung werden notwendige Änderungen in den Sicherheitslebenszyklus zurückgespielt und entsprechende Änderungen nachgezogen. Der Sicherheitslebenszyklus endet erst nach Außerbetriebnahme, da potenzielle Sicherheitsaspekte einer Betrachtungseinheit bei deren Stilllegung und Entsorgung berücksichtigt werden müssen [ISO26262-7].

2.4.3 Gefährdungs- und Risikoanalyse

Im Sicherheitslebenszyklus besitzen die G&R und die darin erarbeiteten Sicherheitsziele eine besondere Bedeutung. Sie dienen der Vermeidung beziehungsweise Abschwächung von Gefährdungen. Die G&R kann als „zentraler Schritt für alle weiteren sicherheitsbezogenen Aktivitäten betrachtet werden“ [LPP10]. Jede folgende Sicherheitsaktivität ist abhängig von korrekt und sorgfältig formulierten Sicherheitszielen. Anhand ihrer Erfüllung wird beurteilt, ob funktionale Sicherheit für ein System erreicht wird. Während in der Grundnorm DIN EN 61508 keine konkreten Methoden für die Durchführung der G&R vorgeschrieben sind, wird in der Norm ISO 26262 explizit eine normative Methode zur Erstellung beschrieben [LPP10]. Zudem werden Kriterien und Ziele vorgegeben, welche implizit den Rahmen für G&Rs vorgeben. Da das in dieser Arbeit entwickelte IR-System die Durchführung von G&Rs unterstützt, wird in diesem Abschnitt die G&R detailliert betrachtet.

2.4.3.1 Ablauf der G&R

Im Folgenden stellen wir die in der Norm ISO 26262 [ISO26262-3] beschriebene Methode in ihren Teilschritten vor.

Initialisierung: Die G&R benötigt die zuvor angefertigte Definition der Betrachtungseinheit. Diese enthält alle für die Sicherheitsbetrachtungen relevanten Informationen. Aus dieser Definition können die zu überprüfenden Funktionen abgeleitet werden.

Situationsanalyse und Identifikation von Gefährdungen: Die Bestimmung der Gefährdungen soll gemäß Norm systematisch mittels angemessener Techniken erfolgen. Was angemessen ist, ist nicht genau definiert. Beispiele für potenzielle Techniken reichen vom Brainstorming über die systematische Fehlermöglichkeits- und Einflussanalyse (*engl.:* failure mode & effects analysis) (FMEA) bis zu Feldstudien. Gefährdungen werden dabei als beobachtbares Verhalten auf Fahrzeugebene definiert. Sie ergeben sich aus dem Fehlverhalten der untersuchten Funktionen, den sogenannten Fehlfunktionen. Eine Fehlfunktion führt dabei nicht zwangsläufig zu einer Gefährdung. Vielmehr entstehen Gefährdungen in konkreten Betriebssituationen eines Fahrzeugs. Daher fordert die Norm ISO 26262 die Identifikation jener Betriebssituationen, in denen das Auftreten eines Fehlverhaltens einer Betrachtungseinheit zu einer Gefährdung führt. Diese Kombination aus der Gefährdung und der Betriebssituation, in der sie auftritt, wird als Gefährdungssituation bezeichnet.

Betriebssituation

Klassifikation der Gefährdungssituationen: Jede identifizierte Gefährdungssituation wird bezüglich dreier Parameter klassifiziert: Schadensausmaß (*engl.:* severity), Kontrollierbarkeit (*engl.:* controllability) und Auftretenswahrscheinlichkeit (*engl.:* exposure). Dieser Teil stellt die eigentliche Risikobewertung dar und beurteilt den potenziellen Schaden der

Personen, die im und außerhalb des Fahrzeugs möglicherweise den Gefährdungssituationen ausgesetzt sind.

Schadensausmaß

Abbildung 2.12 zeigt die Schadensausmaßklassen der Norm ISO 26262 mit den Einordnungskriterien auf. Die Schadensausmaßklasse steigt von S0, bei der ausschließlich materieller Schaden ohne Personenverletzungen zu erwarten ist, bis hin zur Klasse S3, in der lebensbedrohliche und tödliche Verletzungen zu erwarten sind. Darüber hinaus wird im informativen Teil unter anderem die Verwendung von Unfallstatistiken, Verletzungsgradtabellen wie Information Severity Score (ISS) oder Verletzungsskalen wie Abbreviated Injury Scale (AIS) und Maximum Abbreviated Injury Scale (MAIS) empfohlen. Diese ermöglichen eine international vergleichbare und aktuelle Einschätzung der Schadensschwere potenzieller Verletzungen.

	Class			
	S0	**S1**	**S2**	**S3**
Description	No injuries	Light and moderate injuries	Severe and life-threatening injuries (survival probable)	Life-threatening injuries (survival uncertain), fatal injuries

Abbildung 2.12: Schadensausmaßklassen [ISO26262-3]

Kontrollierbarkeit

Die Kontrollierbarkeit einer Gefährdungssituation durch Personen innerhalb oder außerhalb des Fahrzeuges wird in Klassen eingeteilt. Kontrollierbarkeitsklassen bewerten, ob der Fahrer oder andere potenziell gefährdete Personen die Möglichkeit haben, den Schaden abzuwenden. Abbildung 2.13 zeigt die Kontrollierbarkeitsklassen der Norm ISO 26262 mit den Kriterien zur Einordnung. Die Kontrollierbarkeitsklassen reichen von C0, bei der die Situation als generell kontrollierbar gilt, bis hin zur Klasse C3 für sehr schwer beziehungsweise nicht kontrollierbare Situationen.

	Class			
	C0	**C1**	**C2**	**C3**
Description	Controllable in general	Simply controllable	Normally controllable	Difficult to control or uncontrollable

Abbildung 2.13: Kontrollierbarkeitsklassen [ISO26262-3]

Auftretenswahrscheinlichkeit

Die Auftretenswahrscheinlichkeit bezieht sich auf die jeweilige Betriebssituation. Mit diesem Parameter wird berücksichtigt, dass das Gesamtrisiko sinkt, wenn bestimmte Betriebssituationen nur sehr selten auftreten. Abbildung 2.14 zeigt die Auftretenswahrscheinlichkeitsklassen der Norm ISO 26262 mit den Kriterien zur Einordnung. Die Parameter steigen von E0 für nahezu unmögliche Situationen bis hin zu E4 für Situationen mit hoher Auftretenswahrscheinlichkeit. Der Verband der Automobilindustrie (VDA) erkannte das Wiederverwendungspotenzial wiederkehrender Betriebssituationen und die Notwendigkeit einer einheitlichen Klassifikation. Als Ergebnis eines Einigungsprozesses von Vertretern der Automobilindustrie veröffentlichte er eine VDA-Empfehlung [VDA]. Diese enthält einen Situationskatalog mit Empfehlungen für deren Einschätzung des Parameters E für die Auftretenswahrscheinlichkeit.

	Class				
	E0	**E1**	**E2**	**E3**	**E4**
Description	Incredible	Very low probability	Low probability	Medium probability	High probability

Abbildung 2.14: Auftretenswahrscheinlichkeitsklassen [ISO26262-3]

ASIL-Bestimmung: Die ASIL-Bestimmung für die Gefährdungssituationen erfolgt auf Basis der drei zuvor bestimmten Parameter S, C und E und kann im sogenannten Risikographen abgelesen werden. Es existieren vier ASIL-Stufen, von der kleinsten Stufe A bis zur höchsten Stufe D. Abbildung 2.15 stellt die Tabelle zur ASIL-Bestimmung dar. Die Bezeichnung QM steht für Qualitätsmanagement und stellt keine ASIL-Stufe dar. Sie bedeutet, dass keine FuSi-spezifischen Anforderungen bestehen, die das übliche Qualitätsmanagement übertreffen.

Severity class	Probability Class	Controllability class		
		C1	C2	C3
S1	E1	QM	QM	QM
	E2	QM	QM	QM
	E3	QM	QM	A
	E4	QM	A	B
S2	E1	QM	QM	QM
	E2	QM	QM	A
	E3	QM	A	B
	E4	A	B	C
S3	E1	QM	QM	A
	E2	QM	A	B
	E3	A	B	C
	E4	B	C	D

Abbildung 2.15: Risikograph zur ASIL-Bestimmung [ISO26262-3]

Ableitung der Sicherheitsziele: Die Norm fordert für jede Gefährdungssituation, der eine ASIL-Stufe zugewiesen wurde, auch die Bestimmung eines Sicherheitsziels, wobei ein Sicherheitsziel auch für mehrere Gefährdungssituationen verwendet werden darf. Dabei sind Sicherheitsziele als sogenannte Top-Level-Sicherheitsanforderungen die Ausgangsbasis für die spätere Ableitung der notwendigen funktionalen Sicherheitsanforderungen und enthalten keinerlei Hinweise auf technische Lösungen, sondern lediglich eine funktionale Zielstellung. Das Sicherheitsziel übernimmt die ASIL-Einstufung der Gefährdungssituation, bei Zuordnung zu mehreren Gefährdungssituationen gilt deren höchste zugeordnete ASIL-Stufe.

Beispiel: Abbildung 2.16 zeigt einen Ausschnitt aus einer G&R, wie sie in der Praxis denkbar wäre, hier am Beispiel einer Betrachtungseinheit für das System Sitzheizung. Die in der Norm beschriebenen Teilschritte sind auf der rechten Seite abgebildet und den Teilartefakten einer G&R auf der linken Seite zugeordnet. Die Systembeschreibung enthält relevante Informationen über das System, für die Sitzheizung beispielsweise technische Details über die Heizleistung. Für das System soll die gleichnamige Funktion Sitzheizung überprüft werden. Eine Fehlfunktion ist das unaufgeforderte Beheizen. In Kombination mit der Betriebssituation *Parken in der Garage* entsteht eine Gefährdungssituation. Die Gefährdung besteht in einem Wohnungsbrand, der durch Überhitzen des Sitzes und Übergreifen des Brandes auf benachbarte Wohnräume entstehen kann. Parken in der Garage hat eine hohe Auftretenswahrscheinlichkeit (E4). Die Gefährdungssituation wird als leicht kontrollierbar (C1) bewertet, allerdings werden im Falle eines Wohnungsbrandes lebensbedrohliche Verletzungen (S3) erwartet. Daraus ergibt sich die ASIL-Stufe B. Ein abgeleitetes Sicherheitsziel kann beispielsweise sein, dass ein Überhitzen des Sitzes zu vermeiden ist.

Das Beispiel vereinfacht an dieser Stelle stark. Üblicherweise besteht eine G&R aus mehreren Funktionen, für welche jeweils mehrere Fehlfunktionen untersucht werden. Jede Fehlfunktion ist wiederum Teil unterschiedlicher Gefährdungssituationen. Zudem besteht die Ausarbeitung

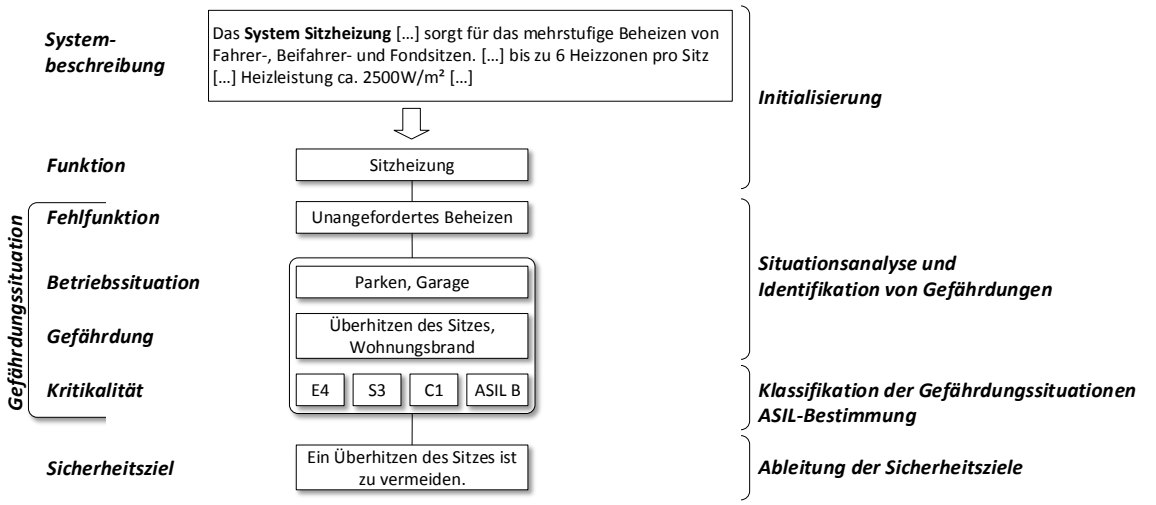


Abbildung 2.16: Beispielhafter Ausschnitt einer G&R

einer Gefährdungssituation aus wesentlich mehr Teilartefakten wie beispielsweise Effekten, erwarteten Reaktionen oder Kommentaren zu den Entscheidungen. Diese sind in diesem Beispiel nicht Teil der Betrachtung.

2.4.3.2 Anforderungen an die G&R

Es ist essenziell, dass die G&Rs sorgfältig und korrekt durchgeführt werden. Daher wird in der Norm ISO 26262 auch eine Verifikation durch ein Review solcher Analysen gefordert. Das Review überprüft zum einen die Konformität zur Systembeschreibung und zum anderen die Vollständigkeit der untersuchten Gefährdungssituationen. Die Norm adressiert hier aber auch die Gefahr der Inkonsistenzen zwischen G&Rs aufgrund subjektiver Einschätzungen. Gemäß [ISO26262-3, 7.4.5.1c] wird die Konsistenz verwandter (in der Norm ISO 26262 *related*) G&Rs gefordert. Diese Anforderung stellt in der Praxis eine große Herausforderung dar, die wir kurz erläutern. Eine Studie zur funktionalen Sicherheit in der Automobilbranche des Fraunhofer-Instituts IPA [MSB13] zeigt, dass zwei Drittel der FuSi-Projekte von den befragten Unternehmen als komplex beziehungsweise eher komplex bewertet werden und dass Erfahrung mit 44% das wichtigste Hilfsmittel bei der ASIL-Bestimmung ist. Die informativen Normvorschläge im Annex von Teil 3 der Norm ISO 26262 [ISO26262-3], welche die normativen Anforderungen ergänzen, stellen mit 43% das zweitwichtigste Hilfsmittel dar. Dieser informative Teil versorgt die Durchführenden von G&Rs mit Anhaltspunkten, Beispielen und zusätzlichen Tabellen, welche insbesondere die Klassifikation der Gefährdungssituationen durch die drei Parameter erleichtern soll. Diese Abschätzungen sind dennoch anfällig für subjektive Entscheidungen. Erfahrungen aus der Anwendung zeigen, dass die in der Konzeptphase vorgestellten Risikoanalysen keine Verfahren mit eindeutigen und reproduzierbaren Ergebnissen sind [Pau07]. Stattdessen gibt es häufig Interpretationsschwierigkeiten und das Phänomen, dass unterschiedliche Menschen das gleiche System unterschiedlich bewerten [Pau07]. Das widerspricht allerdings der ebenso vorhandenen Anforderung nach Konsistenz und Transparenz der G&Rs und unterstreicht die Wichtigkeit einer Beschäftigung mit dem Thema. Daher ist ein Ziel dieser Arbeit, die Durchführung der G&R derart zu unterstützen, dass die Anforderung an Konsistenz leichter erfüllt wird. Dafür entwickeln wir ein domänenspezifisches IR-System, welches verwandte Szenarien zum einen identifizieren kann und zum anderen eine Wiederverwendung oder Modifikation ermöglicht.

2.4.4 Zusammenfassung und Bezug zur Arbeit

Dieser Abschnitt lieferte einen Überblick über die funktionale Sicherheit im Automobilbereich gemäß der Norm ISO 26262. Einen besonderen Schwerpunkt legten wir auf die Beschreibung von G&Rs, die am Anfang jeder FuSi-konformen Systementwicklung stehen, sowie deren Herausforderungen und Anforderungen aus der Norm. G&Rs gelten als sehr aufwendig, da sie sehr komplex sind und stark auf Erfahrungswissen basieren. Eine Herausforderung ist in dem Zusammenhang, die Anforderung der Norm an Konsistenz der G&Rs zu erfüllen und subjektive Einschätzungen möglichst objektiv zu gestalten. In dieser Arbeit stellen G&Rs den anwendungsspezifischen Rahmen dar, weil wir durch den Aufbau eines domänenspezifischen IR-Systems die Wiederverwendungspotenziale erfahrungsbasierter Entscheidungen nutzbar machen und dadurch die Sicherung interner und externer Konsistenz der G&Rs erleichtern möchten.

Mit der Herausforderung der Objektivierung erfahrungsbasierter Entscheidungen und der Konsistenzsicherung bei FuSi-Aktivitäten beschäftigen sich auch andere Forschungsarbeiten. Sowohl diese Arbeiten als auch verwandte Arbeiten zur Anwendung von Aktivierungsausbreitung im Information-Retrieval betrachten wir im folgenden Abschnitt 2.5 genauer.

2.5 Verwandte Arbeiten

Die dargelegten Herausforderungen sind zum Teil Thema in aktuellen Forschungsarbeiten. Daher stellen wir in diesem Abschnitt Ansätze solcher Arbeiten vor.

2.5.1 Konsistenzsicherung bei FuSi-Aktivitäten

Die Objektivierung erfahrungsbasierter Entscheidungen zur Sicherung der von der ISO 26262 [ISO26262-3, 7.4.5.1c] geforderten Konsistenz zwischen verwandten G&Rs stellt eine große Herausforderung in der Praxis dar (siehe Abschnitt 2.4.3.2). Ein Trend in der Forschung ist der Einsatz modellbasierter Ansätze, weil die Semantik in Modellen eindeutiger beschrieben werden kann als in textbasierten Dokumenten. Neben modellbasierten Ansätzen stellen wir vertragsbasierte Ansätze und die Verwendung von Variantenmanagement zur Konsistenzsicherung vor.

2.5.1.1 Modellbasierte Ansätze

EmMORI-Methode: Ständer stellt in seiner Dissertation [Stä10] die modellbasierte Methode *EmMORI* zur Objektivierung der Risikoanalyse gemäß ISO 26262 vor. Laut Ständer resultieren insbesondere die zumeist subjektiven und zudem oftmals konservativen Experteneinschätzungen der drei Parameter zur Auftretenswahrscheinlichkeit E, zum Schadensausmaß S und zur Kontrollierbarkeit C häufig in überdimensionierten und dadurch unnötig kostenintensiven Sicherungsfunktionen. EmMORI ist eine modellbasierte Erweiterung der in der ISO 26262 beschriebenen G&R, welche auf die Objektivierung des Parameters E abzielt und gemäß Ständer auf die anderen beiden Parameter übertragbar sei. Die Grundidee ist, die Einflussfaktoren eines Parameters zu identifizieren und mittels Petrinetzen zu modellieren. Für den Parameter E sind das die beiden Faktoren *Umgebungszustand* und *Fahrzeugzustand*. Die modellierten Einflussfaktoren werden zu Zustandsmodellen kombiniert, welche wiederum zu Fahrszenarien zusammengefügt werden können. Ständer beschreibt diese Szenarien als den Kontext, in welchen die zu analysierenden Fahrzeugfunktionen eingesetzt werden. Im Vergleich mit der konservativen ASIL-Bestimmung konnte Ständer zeigen, dass mit EmMORI ebenfalls korrekte Einschätzungen getroffen werden und sogar in einigen Fällen eine Reduktion der ASIL-Stufe erreicht werden kann. Als nachteilig benennt er den relativ großen Aufwand in der Modellierung, die dafür notwendige Expertise und Akzeptanzprobleme unter anderem aufgrund fehlender Werkzeugunterstützung. Die Methode wird als Erweiterung zusätzlich und unabhängig von der eigentlichen Systementwicklung und G&R ausgeführt. Das hat den Vorteil, dass sie keine modellbasierte Systementwicklung erfordert, um zur Anwendung zu kommen. Das bedeutet allerdings einen höheren Aufwand, da die Methode nicht die Schritte der G&R ersetzt. Diese muss dennoch ausgeführt werden, um den Anforderungen der ISO 26262 zu genügen. Im praktischen Einsatz erfordert die Methode, dass die Durchführenden einer G&R nicht nur Expertise im Bereich der Funktionssicherheit benötigen, sondern zudem Experten in der Modellierung und im Umgang mit Petrinetzen sind.

SAHARA-Methode: Einen anderen modellbasierten Ansatz verfolgt Kemmann in seiner Dissertation [Kem15]. Sein strukturierter Ansatz für G&Rs namens *SAHARA* zielt ebenso auf die Erhöhung der Konsistenz und zudem auf die Verringerung der Abhängigkeit von Experten ab. Mehrdeutigkeiten, wie sie in der von G&Rs üblicherweise genutzten natürlichen Sprache häufig vorkommen, sollen durch eine formale Modellbildung abgeschafft werden. Kem-

mann entwirft eine modellbasierte Repräsentation der G&R, welche in einen modellbasierten Systementwicklungsprozess integriert werden kann. Dafür verwendet er einen semiformalen Ansatz zur Modellierung, welcher das Vier-Variablen-Modell von Parnas [PM95] um Interaktion erweitert. Vereinfacht ausgedrückt werden die Elemente Mensch, System und Umgebung in einzelnen Vier-Variablen-Modellen modelliert, deren Interaktion durch die vier Relationen des Vier-Variablen-Modells formuliert werden können. Dieser ganzheitliche Prozess ist die Grundlage für die darauf aufbauenden Methoden zur Konsistenzverbesserung während einer G&R. Ähnlich zu Ständer wählt auch Kemmann als wiederzuverwendende Objekte Betriebssituationen und deren Parameter für die Auftretenswahrscheinlichkeit. Die Idee ist, mithilfe einer Ontologie Einflussfaktoren und Domänenwissen derart abzubilden, dass Betriebssituationen, deren Auftretenswahrscheinlichkeit und die dazugehörigen Begründungen modelliert werden können. Die Wiederverwendungsstrategie basiert auf einer Auswertung der Ähnlichkeit einer neu modellierten Betriebssituation zu anderen bereits bekannten und in einer Falldatenbank verfügbar gemachten Betriebssituationen. Für die neue Betriebssituation wird dann mittels fallbasierten Schließens derjenige Parameter E ausgewählt, mit dem die ähnlichste Situation klassifiziert wurde. Alternativ können bereits vorhandene Kombinationen aus Betriebssituation und ihrer Auftretenswahrscheinlichkeit mithilfe einer ebenso von Kemmann vorgestellten Methode einer Konsistenzüberprüfung unterzogen werden.

Bezug zu dieser Arbeit: Die Arbeiten von Ständer und Kemmann basieren auf einer ähnlichen Motivation wie die vorliegende Arbeit, welche in der Objektivierung von Entscheidungen und der Erhöhung von Konsistenz in dem FuSi-Artefakt G&R liegt. Ein Nachteil ist, dass sowohl EmMORI als auch SAHARA nicht die komplette G&R begleiten. Beide Verfahren beschränken sich auf die Bestimmung der Parameter während der Risikobewertung, genauer der Betriebssituation und ihrer Auftretenswahrscheinlichkeit E. Insbesondere deren Wiederverwendungspotenzial und die Notwendigkeit einer einheitlichen Klassifikation wurde inzwischen von der Automobilindustrie erkannt (siehe Abschnitt 2.4.3.1). Als Ergebnis eines Einigungsprozesses zwischen Vertretern der Automobilindustrie wurde 2015 eine VDA-Empfehlung [VDA] veröffentlicht, welche einen Situationskatalog mit Empfehlungen für die Einschätzung des Parameters E enthält. Beide Verfahren können somit eher als eine Methode zur punktuellen Unterstützung im Ablauf der G&R gesehen werden, die zudem vorrangig für die Betrachtung außergewöhnlicher[16] Betriebssituationen geeignet ist. Beide Ansätze erfordern Mehraufwand. In Ständers Verfahren werden parallel zur eigentlichen G&R Modelle entwickelt. Kemmanns Verfahren zur Wiederverwendung gliedert sich zwar in einen Gesamtprozess ein, dennoch gilt auch hier, dass die FuSi-Experten zusätzlich mit der modellbasierten Entwicklung vertraut sein müssen. Damit ist der wiederverwendungsbasierte Anteil vermutlich nur dann sinnvoll, wenn die gesamte modellbasierte Methode zur Erstellung von G&Rs zur Anwendung kommt. Insbesondere der kreative Prozess wird allerdings von den Entwicklern weiterhin ohne Unterstützung und ohne Automatisierung erfahrungsbasiert erledigt.

Im Gegensatz dazu zielen wir in dieser Arbeit auf die Unterstützung aller Teilschritte und Teilartefakte einer G&R ab. Zudem entwickeln wir eine in das Entwicklungswerkzeug integrierte Lösung, die keinen Mehraufwand für die FuSi-Ingenieure erfordert, sondern stattdessen eine zusätzliche Quelle von Informationen darstellt, die auch den kreativen Prozess unterstützt.

[16] Außergewöhnlich sind in dem Fall Situationen, die nicht im Situationskatalog enthalten sind.

Weitere modellbasierte Ansätze: Ein weiteres Verfahren zur Sicherstellung von Konsistenz in G&Rs wurde von Beckers et al. [BHFH13] entwickelt und in einer späteren Studie [BHCH17] evaluiert. Beckers et al. beschreiben ein Unified Modeling Language (UML)-basiertes Verfahren zum Erstellen von G&Rs, welches eine nachträgliche Konsistenzüberprüfung durch Object Constraint Language (OCL) erlaubt. Dafür stellen sie automatisierte Transformationen zwischen den üblicherweise text- und tabellenbasierten Analysen und den UML-Modellen bereit, was die Nutzbarkeit des Verfahrens wesentlich erhöht. Der kreative Prozess der Erstellung einer G&R wird dabei allerdings nicht unterstützt. Der Schwerpunkt ihrer Konsistenzprüfung liegt anschließend auf der vollständigen und ISO 26262-konformen Betrachtung einer G&R, beispielsweise ob alle notwendigen Begründungen vorliegen, alle Anforderungen der Systembeschreibung berücksichtigt wurden und die Analyse konsistent zur ISO 26262 angefertigt wurde. Im Gegensatz dazu steht im Fokus der vorliegenden Arbeit die Anforderung nach Konsistenz innerhalb einer G&R und zwischen verschiedenen G&Rs, wie sie in der ISO 26262 [ISO26262-3, 7.4.5.1c] gefordert ist. Dafür unterstützen wir die Experten mit Empfehlungen zur Wiederverwendung bereits vorhandenen Wissens aus abgeschlossenen G&Rs.

Ein weiterer Trend sind kompositionale Ansätze [PMSH01, KLM03, GTS04, PTFG], welche auf die Wiederverwendung bei Sicherheitsanalysen wie der Fehlerbaumanalyse (*engl.:* fault tree analysis) (FTA) und der FMEA und damit indirekt auf die Unterstützung von G&Rs abzielen. Sie basieren zumeist auf einem Systemdesign mit UML, beispielsweise mit Komponentendiagrammen oder Verteilungsdiagrammen. Die Grundidee ist, komplexe Artefakte durch Dekomposition in kleinere Komponenten zu zerlegen, die zum einen leichter handhabbar sind und zum anderen größeres Wiederverwendungspotenzial besitzen. Beispielsweise können so Fehlerbäume in Komponenten-Fehlerbäume überführt werden [KLM03]. Komponenten werden mit ihren Schnittstellen beziehungsweise Ports, ihren gegenseitigen Abhängigkeiten und Konnektoren modelliert und zusätzlich durch mögliche Abweichungen, Fehler und Ausfälle sowie deren Effekte angereichert. Die Fehlerpropagation bei der FTA basiert auf boolescher Logik und kann nun nicht nur innerhalb einer Komponente, sondern auch über die Verbindungen zwischen den Komponenten durchgeführt werden. Solche Verfahren zielen auf die Wiederverwendung der Komponenten und der darin vorhandenen Informationen über potenzielles Fehlverhalten ab, was fehleranfälliges Kopieren vermeidet [KLM03]. Dadurch erleichtern sie G&Rs, die bei Bedarf auf Sicherheitsanalysen wie FTA und FMEA als Hilfsmittel zurückgreifen, indirekt.

Nyberg und Westman verwenden für die Fehlerpropagation formale Sicherheitsverträge [NW15]. In Sicherheitsverträgen werden in der Regel Annahmen (*engl.:* assumptions) und Anforderungen an die Umgebung festgelegt sowie Garantien (*engl.:* guarantees) spezifiziert, die vom System erfüllt werden sollen. Einige kompositionale Ansätze verwenden Sicherheitsverträge, um sicherheitsrelevante Komponenten und Sicherheitsartefakte systematisch wiederverwenden zu können. Im Fokus der Dissertation von Šljivo [Šlj15] steht beispielsweise eine Methode, welche die Wiederverwendung von Teilartefakten bei der Erstellung des Sicherheitsnachweises erlaubt. Ein ebenfalls vertragsbasierter Ansatz wird von Westman et al. [WNT13, WN18] verfolgt, um die interne und externe Konsistenz[17] und die Vollständigkeit von Sicherheitsanforderungen zu gewährleisten.

[17] Gemäß ISO 26262 soll sich eine einzelne Sicherheitsanforderung nicht widersprechen (interne Konsistenz [ISO26262-8, 6.4.2.4c]), es soll aber auch keine Widersprüche zwischen verschiedenen Sicherheitsanforderungen geben (externe Konsistenz [ISO26262-8, 6.4.3.1d]).

Bezug zu dieser Arbeit: Der Vorteil kompositionaler Ansätze liegt in der Konsistenz zwischen Systemdesign und den Sicherheitsartefakten des Systems wie den Sicherheitsanalysen FMEA und FTA. Damit unterstützen die Ansätze zwar FuSi-Aktivitäten, die G&R erleichtern sie allerdings lediglich indirekt. Eine Voraussetzung für die Anwendung der Methoden ist das Vorhandensein der jeweiligen UML-Modelle aus dem Systemdesign. In der Konzeptphase ist das Systemdesign oftmals noch nicht abgeschlossen, so dass die G&Rs zumeist ohne derartige Modelle erstellt werden. Daher spielt die Konsistenz zwischen den Komponenten für das Ziel dieser Arbeit keine große Rolle.

2.5.1.2 Wiederverwendung durch Variantenmanagement

Eine weitere Möglichkeit zur Wiederverwendung bietet das Variantenmanagement. Dafür wird in der automobilen Systementwicklung die Variabilität einer Produktlinie zumeist mittels sogenannter Merkmalsbäume (*engl.:* feature trees) beschrieben. Aus der modellierten Variabilität sollen idealerweise Artefakte wie Lastenhefte, Modelle oder Programmcode für jede Variante automatisiert abgeleitet werden. Schulze et al. [SMB13] beschäftigen sich mit der Frage, inwiefern funktionale Sicherheit und Variantenmanagement zusammengebracht werden können. Sie zeigen technische Lösungen auf, die sie in ein Entwicklungswerkzeug zur funktionalen Sicherheit, ANSYS medini analyze, integrieren. Zusätzlich zu einem Merkmalsbaum wird ein Artefaktmodell erstellt. Des Weiteren entwickeln sie ein Referenzmodell, welches die Referenzen zwischen Merkmalen und Artefakten verwaltet. Die Grundidee ist, dass durch die Auswahl von Merkmalen Varianten gebildet werden, für die wiederum über das Referenzmodell die passenden Artefakte gewählt werden können, so dass variantenspezifische FuSi-Artefakte ausgeleitet werden können. Schulze et al. zeigen anhand eines Prototyps die technische Machbarkeit ihres Konzeptes und beschreiben die Herausforderungen für variantenspezifische FuSi-Artefakte. Insbesondere stellen sie heraus, dass eine Wiederverwendung von FuSi-Artefakten immer im Kontext geschehen muss. Der Kontext ist allerdings sehr komplex und oftmals nicht vollständig bekannt, was ihn schwer modellierbar macht. Zudem seien Sicherheitsanalysen wie die FMEA und die FTA nicht auf Variabilität ausgelegt und erfordern Anpassungen. Baumgart und Fröberg untersuchen in einer systematischen Mapping Study [BF16] Forschungsarbeiten zur funktionalen Sicherheit in der Produktlinienentwicklung. Sie bestätigen, dass sich die meisten Arbeiten zum Variantenmanagement in der Funktionssicherheit auf die Anpassung einzelner FuSi-Artefakte fokussieren, insbesondere auf die strukturierten, semiformalen Techniken wie beispielsweise die FTA [BLR14], die Software-Fehlerbäume [DL09] oder, als Erweiterung des zuvor beschriebenen kompositionalen Ansatzes, für Komponenten-Fehlerbäume [GLS10, SSA13]. Darüber hinaus existieren diverse Ansätze beispielsweise für das funktionale Sicherheitskonzept [dBM+18], den Sicherheitsnachweis [dBM+15] oder allgemeine Sicherheitsanalysen [GT06].

Bezug zu dieser Arbeit: Die Forschungsarbeiten zur Verwendung von Variantenmanagement im Bereich der funktionalen Sicherheit verfolgen, wie auch wir, das Ziel einer systematischen Wiederverwendung von FuSi-Artefakten. Die Durchführung der G&R ist sehr komplex, so dass für die feingranulare Variantenmodellierung ein hoher zusätzlicher Aufwand von Experten erforderlich ist, damit tatsächlich Artefakte ausgeleitet werden können. In unserem untersuchten industriellen Umfeld ist keine Variantenmodellierung verfügbar. Dadurch sind Wiederverwendungspotenziale durch Variantenmanagement im Rahmen dieser Arbeit nicht verwendbar. Die vorliegende Arbeit zielt auf die Entwicklung eines IR-Systems ab, welches ohne zusätzlichen Aufwand genutzt werden kann.

2.5.2 Aktivierungsausbreitung in semantischen Netzen

Verschiedene Arbeiten beschäftigen sich mit dem Einsatz von Aktivierungsausbreitung als semantisches Suchverfahren. Sie unterscheiden sich beispielsweise in den Anwendungsgebieten, der Art der Wissensrepräsentation sowie der Interpretation und Konfiguration der Aktivierungsausbreitung als Algorithmus.

Anwendungsgebiete: Aktivierungsausbreitung findet als semantisches Suchverfahren in verschiedenen Bereichen Verwendung. Die Bandbreite reicht vom inhaltsbasierten Image-Retrieval (*engl.:* content-based image retrieval) [KMMK+] über die Suche nach relevanten Bildannotationen [ZWL12], die Web-Suche [CL00, GSV+07], das Retrieval von Tourismusinformationen [BDM03], den Einsatz bei der Erweiterung des WordNet [PRK], Empfehlungen in sozialen Netzwerken [MG11] bis zur Anwendung in der Robotersteuerung [LSL13]. Einige Arbeiten beschäftigen sich mit technischen Herausforderungen wie der Verringerung des Zeitaufwandes bei den Berechnungen der Aktivierungsausbreitung [GHS10, DKL+10], so dass das Suchverfahren auch Anwendung im Themenbereich Big Data [SGHP15, SSZ+17] finden kann. Im Folgenden stellen wir drei Arbeiten genauer vor, die Aktivierungsausbreitung als Suchverfahren im IR benutzen und eine besondere Nähe zur vorgestellten Arbeit besitzen.

OntoSpread: Álvarez entwickelte im Rahmen seiner Dissertation [Á12] ein Framework für die Entwicklung, Konfiguration, Anpassung und Ausführung von Aktivierungsausbreitung auf graphbasierten Strukturen, insbesondere auf RDF-Graphen und Ontologien, welches er als Java-Bibliothek *OntoSpread* [OntoSpread] veröffentlichte. Álvarez et al. [ÁPJ+11] nutzen es zur Suche in großen, heterogenen Wissensbasen zum Information-Retrieval und Dokumenten-Retrieval. Sie sehen den Anwendungsbereich vor allem im medizinischem Wissensmanagement, beispielsweise in der Empfehlung von Konzepten zum Taggen von Krankenakten sowie in der Entscheidungsunterstützung bei der klinischen Diagnose. Álvarez et al. setzen dafür die von Crestani [Cre97] beschriebenen grundlegenden Techniken zur Anwendung von Aktivierungsausbreitung in der assoziativen Informationsgewinnung um. Beispielsweise verwenden sie bekannte Beschränkungen zur pulsweisen Dämpfung und zur Dämpfung über die Distanz oder die Belohnung von Knoten, die durch mehrere Pfade erreicht werden.[18] Als Erweiterung ermöglichen sie beim Zusammenspiel verschiedener Ontologien die Beschränkung auf die Aktivierungsausbreitung innerhalb ausgewählter Ontologien, die sie als Kontext bezeichnen. Zudem sind sie restriktiv bezüglich der Anzahl an Konzepten, die sich an der Aktivierungsausbreitung beteiligen und nutzen entsprechende Schwellwerte als Abbruchkriterium. In Experimenten verwendeten sie die zwei medizinischen Ontologien *GALEN* [OpenGALEN] und *SNOMED CT* [SNOMED] und zeigten anhand statistischer Auswertungen Ergebnisse der Aktivierungsausbreitung.

Grad-Gyenges Knowledge-Graph-basierter Ansatz: Grad-Gyenge et al. verwenden Aktivierungsausbreitung in hybriden Empfehlungssystemen, die sowohl inhaltsbasierte Informationen über Eigenschaften von Empfehlungselementen als auch kollaborative Informationen wie Ratings und das Verhalten von verschiedenen Benutzern verwerten. Für die Repräsentation des Wissens definieren sie einen Knowledge-Graph [GGFW15], der an Ontologien angelehnt ist und entsprechend heterogene Informationen abbilden kann [GGF16]. Ihre Repräsentation entspricht nicht den Spezifikationen der RDF-Sprachfamilie, ist aber ebenfalls stark

[18] Bei Álvarez et al. [ÁPJ+11] wird dies als *converging paths reward* bezeichnet, was den *intersections* von Quillian und Collins [Qui68, CL75] entspricht.

typbasiert, da den Knoten und Kanten jeweils Knoten- und Kantentypen zugeordnet werden. Grad-Gyenge et al. verwenden eine von üblichen Interpretationen leicht abweichende Variante der Aktivierungsausbreitung. Beispielsweise nutzen sie die Schwellwerte zum Zurücksetzen kleiner Aktivierungswerte auf den Wert 0, anstatt diese bei der Aktivierungsverteilung zu ignorieren. Außerdem verzichten sie auf bekannte Beschränkungen und verwenden stattdessen zwei globale Parameter, die bestimmen, welchen Anteil an Aktivierung ein Knoten behält (*activation relax*) und weitergibt (*spreading relax*). Sie zeigen in einer Evaluation, dass mit Aktivierungsausbreitung Ratings von Nutzern für Filme mindestens so gut vorhergesagt werden können wie das vergleichsweise angewandte kollaborative Filterverfahren ohne Aktivierungsausbreitung. Grad-Gyenge et al. [GGWF16] zeigen außerdem die Anwendbarkeit in einem Empfehlungssystem für Bücher durch die Evaluation eines Email-Remarketing-Szenarios. Darin werten Grad-Gyenge et al. das Klick- und Kaufverhalten der Nutzer aus, welche individualisierte Empfehlungen für Bücher über einen Newsletter erhielten. Sie identifizierten, wie auch wir, die optimale Konfiguration der Aktivierungsausbreitung als Herausforderung. In einer weiteren Studie [GG17] verglichen sie daher verschiedene Kombinationen von Werten für die beiden Parameter spreading relax und activation relax und kamen zu dem Ergebnis, dass in ihrer Anwendung extreme Werte die Qualität der Empfehlungen verringern.

Schumachers hybrider Ansatz zum Fakten- und Dokumenten-Retrieval: Schumacher präsentierte in ihrer Dissertation [Sch16] ein hybrides semantisches Suchverfahren für unterschiedlich stark strukturierte Datenmengen, welches Faktensuche mit semantischer Dokumentensuche kombiniert. Dafür verwendet sie unter anderem die Aktivierungsausbreitung als graphbasiertes Suchverfahren. Damit soll die Qualität der Suchergebnisse verbessert werden, insbesondere weil die Suche nicht auf die Abfrage von Fakten beschränkt bleibt, sondern weil Zusammenhänge zu allen verfügbaren Informationen bewertet werden können. Für die Aktivierungsausbreitung werden eigens semantische Netze erstellt, deren Knoten und Kanten sich aus den Konzepten und Properties ergeben. Literale werden hierbei ignoriert, da sie nicht zur Identifikation von Zusammenhängen beitragen würden. Schumacher verwendet eine an Crestanis [Cre97] Ausführungen angelehnte beschränkte Aktivierungsausbreitung für die assoziative Informationsgewinnung, die allerdings in einzelnen Details abweicht. Beispielsweise verwendet sie die Verzweigungsbeschränkung, um die Verteilung über viele ausgehende Kanten zu unterbinden, anstatt diese abzuschwächen, oder die Distanzbeschränkung, um eine Verteilung ab einer festgelegten Distanz abzubrechen, anstatt diese zu verringern. Aufgrund der heterogenen Datenbasis eignet sich das Verfahren insbesondere für das persönliche Wissensmanagement, beispielsweise für den semantischen Desktop. Die Grundidee ist, verschiedenartige Anwendungen und Inhalte auf Desktop-Rechnern zu verknüpfen und dadurch das persönliche Wissensmanagement zu unterstützen [SKSB09]. Schumacher et al. [SSS08] verwenden für die hybride Suche des semantischen Desktops die Aktivierungsausbreitung mit Beschränkungen für Dämpfung, Verzweigung und Schwellwerten ohne Erweiterungen, wobei detaillierte Informationen zur Konfiguration der Aktivierungsausbreitung nicht bekannt sind.

Bezug zu dieser Arbeit: Wir verfolgen ebenfalls den grundsätzlichen Ansatz, die Aktivierungsausbreitung auf semantischen Netzen anzuwenden, um ein IR-Problem zu lösen. Allerdings ist eine direkte Übertragung auf unseren Anwendungsfall nicht möglich. Die vorgestellten Arbeiten lösen Teilprobleme, liefern aber kein vollumfängliches Konzept für ein im Entwicklungswerkzeug integriertes IR-System für G&Rs, das die Nutzer ohne großen Mehraufwand unterstützt.

Álvarez et al. setzen voraus, dass die Wissensrepräsentation bereits als RDF-Graph zur

Verfügung steht und machen keine Aussagen darüber, wie diese RDF-Graphen aufgebaut werden. Auch Grad-Gyenge et al. legen nicht offen, wie ihr Knowledge-Graph erzeugt wird. Im Gegensatz zu Schumacher halten wir insbesondere die Literale für semantisch bedeutend und daher die Exklusion der Literale aus dem semantischen Netz für nicht sinnvoll.

Außerdem untersuchen wir in der vorliegenden Arbeit ein IR-Problem, für dessen Lösung lediglich kleine bis mittelgroße domänenspezifischen Datensätze zur Verfügung stehen. Die Arbeit von Grad-Gyenge et al. ist auf große Datenmengen von Empfehlungssystemen ausgelegt und profitiert von den vorhandenen nutzerbezogenen Daten. Der Ansatz von Schumacher et al. ist zugeschnitten auf unterschiedlich stark strukturierte Datenmengen, die sogar aus unterschiedlichen Dokumententypen bestehen. Die Vorzüge des hybriden Ansatzes von Schumacher können daher in unserem Anwendungsfall nicht ausgeschöpft werden, da unser Datensatz derartige Unterschiede nicht aufweist.

Die vorgestellten Arbeiten machen keine Angaben zur Konfiguration der Aktivierungsausbreitung. Der Umgang mit der Konfiguration der Aktivierungsausbreitung wird zwar häufig als Problem erkannt, aber meist lediglich für das konkrete Problem angepasst. Wir beschäftigen uns in dieser Arbeit mit dieser Herausforderung. Außerdem interpretieren die Arbeiten die Aktivierungsausbreitung unterschiedlich. Daher konzeptualisieren und formalisieren wir übliche Vorgehen bei der Aktivierungsausbreitung und machen diese dadurch reproduzierbar und parametrisierbar.

In diesem Kapitel haben wir einen Überblick über die Kernthemen dieser Arbeit gegeben. Insbesondere haben wir IR-Systeme und ihre Bestandteile vorgestellt. Einen besonderen Fokus haben wir dabei auf assoziative Systeme gelegt und daher semantische Netze als Repräsentationsform und Aktivierungsausbreitung als Rückgewinnungsverfahren genauer beleuchtet. Außerdem haben wir die funktionale Sicherheit im Automobilbereich mit einem Schwerpunkt auf Gefährdungs- und Risikoanalysen als praktisches Anwendungsfeld dieser Arbeit vorgestellt. Assoziative IR-Systeme bilden das Grundgerüst für das domänenspezifische IR-System, welches wir in dieser Arbeit zur Unterstützung von G&Rs entwickeln, um Erfahrungswissen aus abgeschlossenen Analysen wiederverwenden zu können. Abschließend haben wir Forschungsarbeiten vorgestellt, die eine inhaltliche Nähe zu der vorliegenden Arbeit aufweisen. Im folgenden Kapitel präsentieren wir unsere Konzepte und Techniken zum Aufbau eines domänenspezifischen IR-Systems für G&Rs.

3 Aufbau eines domänenspezifischen IR-Systems für G&Rs

Dieses Kapitel widmet sich dem Aufbau des domänenspezifischen IR-Systems für G&Rs und der Entwicklung eines generischen Rahmenwerks, welches die Übertragung auf andere Domänen ermöglicht. Dafür stellen wir in Abschnitt 3.1 zunächst den zu lösenden Anwendungsfall anhand der üblichen Abläufe in der G&R-Praxis vor. Wir generalisieren das Problem und entwickeln eine Methode, die den Aufbau domänenspezifischer IR-Systeme aus den generischen Lösungen des Rahmenwerks in vier Schritten beschreibt. In den darauffolgenden Abschnitten präsentieren wir für die einzelnen Schritte der Methode sowohl generische Lösungen als auch die konkrete Umsetzung im domänenspezifischen IR-System für G&Rs. In Abschnitt 3.2 widmen wir uns der Wissensrepräsentation und stellen Techniken vor, die den automatisierten Aufbau semantischer Netze und deren semantische Anreicherung und Reduktion ermöglichen. Auf diesen Netzen verwenden wir ein assoziatives Retrievalverfahren, die Aktivierungsausbreitung, das wir in Abschnitt 3.3 zur reproduzierbaren Verwendung konzeptualisieren und formalisieren. Insbesondere erweitern wir die Aktivierungsausbreitung um eine neuartige Technik zur direkten Steuerung des Ausbreitungsverhaltens. Anschließend beschreiben wir in Abschnitt 3.4, wie die Ergebnisse des Retrievalverfahrens für die Nutzer des IR-Systems aufbereitet werden können und präsentieren ein Verfahren, welches Erklärungen für die Ergebnisse generiert. Alle Techniken und Konzepte haben wir in einem eigenständigen Prototyp umgesetzt, den wir abschließend in Abschnitt 3.5 beschreiben.

3.1 G&R-Unterstützung als IR-Problem

Information-Retrieval zielt auf die Rückgewinnung von Informationen aus vorhandenen Daten ab. Daher beleuchten wir zunächst die Wiederverwendungspotenziale in der von uns untersuchten Domäne anhand eines Beispiels und stellen das industrielle Umfeld und den G&R-Datensatz vor. Wir präsentieren eine Studie über das Ausmaß von Redundanz in diesen G&R-Dokumenten. Diese unterstützt unsere Vermutung, dass Wiederverwendungspotenziale in G&Rs existieren und bereits genutzt werden. Darauf aufbauend entwickeln wir eine Lösungsstrategie, um Wiederverwendungspotenziale automatisiert zu identifizieren. Da der Anwendungsfall und die Lösungsstrategie auf andere Domänen übertragbar sind, entwickeln wir generische Lösungen für die Komponenten von IR-Systemen als Rahmenwerk, welche sich domänenspezifisch anpassen lassen. Wir präsentieren zudem eine Methode, die das Vorgehen beim Aufbau domänenspezifischer IR-Systeme aus den generischen Lösungen des Rahmenwerks beschreibt.

3.1.1 Wiederverwendungspotenziale in der Praxis

In den vorherigen Kapiteln haben wir bereits die Bedeutung der G&R im Sicherheitslebenszyklus und die Besonderheiten bei der Durchführung solcher Analysen erläutert. Ausgehend von den initial verfügbaren Entwicklungsunterlagen, wie z.B. der Systembeschreibung, stehen bis zu den Sicherheitszielen als finales Ergebnis viele Teilschritte, die sorgfältig ausgearbeitet und dokumentiert werden müssen (siehe Abschnitt 2.4.3). Die Durchführung erfolgt stark erfahrungsbasiert und erfordert manuellen Aufwand von Domänenexperten.

Diesem Aufwand stehen Potenziale zur Wiederverwendung von Teilen abgeschlossener Analysen gegenüber, da Entscheidungen für ähnliche Szenarien in einem ähnlichen Kontext aufgrund

semantischer Zusammenhänge oftmals zu ähnlichen Ergebnissen führen. Wiederverwendungspotenziale existieren zum einen über Baureihen hinweg, zum anderen aber auch innerhalb von Baureihen. So kann bei der G&R für ein Außenlichtsystem einer neuen Baureihe sicherlich die Betrachtung des Vorgängersystems von Vorteil sein. Aber auch innerhalb einer Baureihe ist die Existenz von Zusammenhängen wahrscheinlich, beispielsweise zu anderen Lichtsystemen.

3.1.1.1 Relevante Zusammenhänge in G&Rs am Beispiel

In Abbildung 3.1 sind mögliche Zusammenhänge am Beispiel dreier Systeme dargestellt.

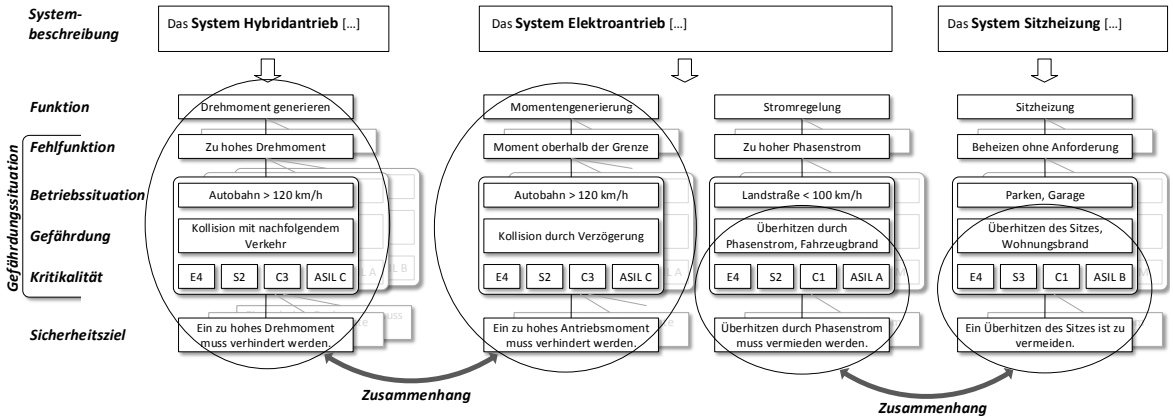


Abbildung 3.1: Beispiel für Wiederverwendungspotenziale

Deren Systembeschreibungen geben den Kontext für die jeweilige G&R vor. Sie enthalten detaillierte Informationen über das System, oftmals mit Schnittstellen- und Architekturbeschreibungen sowie Angaben zu dem Umfang der Betrachtungen. Abgebildet sind für jedes Beispielsystem jeweils eine Funktion mit einem Ausschnitt einer möglichen Gefährdungssituation und einem abgeleiteten Sicherheitsziel.[19] Sowohl für das System *Hybridantrieb* als auch für das System *Elektroantrieb* wird eine Funktion zur Generierung von Momenten untersucht. Trotz der etwas unterschiedlichen Formulierung ist mit *Momentengenerierung* und *Drehmoment generieren* semantisch das Gleiche gemeint. In der Betrachtung der identischen Betriebssituation auf der Autobahn erfolgt im Beispiel in beiden Fällen eine Bewertung mit den gleichen Kritikalitätsparametern durch die FuSi-Ingenieure. Zudem ist das formulierte Sicherheitsziel semantisch gleich. Beide Ausschnitte stehen in einem starken semantischen Zusammenhang. Die Existenz eines Zusammenhanges zwischen zwei Antriebssystemen entspricht den intuitiven Erwartungen. Aber auch hier gibt es nicht nur Gemeinsamkeiten. Kleine Unterschiede oder Veränderungen in der Ausgangssituation können zu komplexen Veränderungen in der Betrachtung und damit zu anderen Entscheidungen führen. Vermeintlich unterschiedliche Systeme wie das System *Elektroantrieb* und das System *Sitzheizung* können Gemeinsamkeiten in den Teilergebnissen der Analysen aufweisen. Auf der rechten Seite der Abbildung ist solch ein Zusammenhang bezüglich Gefährdungen durch Überhitzen und der damit verbundenen Brandgefahr dargestellt. Die Schadensschwere wird hier aufgrund der verschiedenen Betriebssituationen zwar unterschiedlich bewertet, Ähnlichkeiten existieren

[19] Das Beispiel vereinfacht an dieser Stelle stark. Üblicherweise besteht eine G&R aus mehreren Funktionen, für welche jeweils mehrere Fehlfunktionen untersucht werden. Jede Fehlfunktion ist wiederum Teil unterschiedlicher Gefährdungssituationen. Zudem besteht die Ausarbeitung einer Gefährdungssituation aus wesentlich mehr Teilartefakten wie beispielsweise Effekten, erwarteten Reaktionen oder Kommentaren zu den Entscheidungen. Diese sind in diesem Beispiel nicht Teil der Betrachtung.

dennoch in den dargestellten Sicherheitszielen, den nicht dargestellten Effekten oder erwarteten Reaktionen. Da an G&Rs die Anforderung an gegenseitige Konsistenz besteht, ist die Identifikation sowohl offensichtlicher als auch nicht offensichtlicher Zusammenhänge in der Praxis wichtig. Sie markieren potenzielle Kandidaten für Wiederverwendung.

3.1.1.2 Datensatz und industrielles Umfeld

Die in dieser Arbeit beschriebenen Beispiele und Hintergründe basieren auf Beobachtungen und Gesprächen im Arbeitsumfeld der FuSi in der Firma Daimler AG. Aus Geheimhaltungsgründen sind diese verändert und zur besseren Verständlichkeit vereinfacht dargestellt. Im besagten Team werden aktuell G&Rs mit dem Entwicklungswerkzeug für funktionale Sicherheit ANSYS medini analyze [Medini] erstellt. Für die Untersuchungen in dieser Arbeit verwenden wir einen daraus exportierter Rohdatensatz, welcher 155 abgeschlossene G&Rs enthält. Jede G&R entspricht der Analyse einer Betrachtungseinheit, beziehungsweise eines Systems.[20] Der Datensatz umfasst nahezu alle Analysen einer aktuellen Baureihe, darüber hinaus einige baureihenübergreifenden Analysen. Sie entstammen der Entwicklung aller vier im Praxisumfeld üblichen Teildomänen: Powertrain, Fahrdynamik, Telematik und Innenraum. Die mengenmäßige Zuordnung der analysierten Systeme zu den Teildomänen ist in Tabelle 3.1 aufgelistet.

Teildomäne	**Anzahl**
Powertrain	33
Fahrdynamik	52
Telematik	8
Innenraum	62
Gesamt	155

Tabelle 3.1: Kategorisierung der Systeme im Datensatz

Der Datensatz enthält mehr als 600 untersuchte Funktionen, fast 1.700 Fehlfunktionen, mehr als 4.200 Gefährdungssituationen und 540 Sicherheitsziele. Benennungen und Strukturierung der Daten entsprechen aufgrund des Datenexports dem Datenmodell des Entwicklungswerkzeugs, dessen Rolle beim Aufbau semantischer Netze in Abschnitt 3.2.1 detaillierter vorgestellt wird. In Abbildung 3.2 nehmen wir eine Zuordnung der Entitäten (Typen) und Attribute (Prädikate) des Datenmodells zu den Teilartefakten einer G&R vor. Beispielsweise verteilen sich die Informationen wie die Systembeschreibung (`dtpDescription`) und die Systemgrenzen (`dtpBoundaries`), die wir dem Teilartefakt Item zuordnen, auf unterschiedliche Typen. Diese Gruppierung dient vor allem der Vereinfachung in den Ausführungen und wird beispielsweise in der folgenden Studie zum Ausmaß der Wiederverwendung genutzt.

3.1.1.3 Ausmaß der Wiederverwendung in der Praxis

In der industriellen Praxis werden G&Rs nicht für jedes System vollständig neu erstellt. Stattdessen werden Wiederverwendungspotenziale intuitiv genutzt. Wiederverwendung durch Delta-Analysen, in denen explizit nur die Differenzen analysiert werden, sowie Copy-Paste-Modify oder Clone-and-Own sind übliche Vorgehen in der Praxis [SMB13, OSP14]. Mit der

[20] Wie bereits in Abschnitt 2.4.1 erläutert, werden die Begriffe Betrachtungseinheit und System in diesem Kontext häufig synonym verwendet, da die Systemgrenzen in der Regel auch den Abgrenzungen der Betrachtungseinheiten entsprechen.

Teilartefakte	Typ	Prädikat
Item	DefaultPlainItemStereotype	dtpAssumptions
		dtpBoundaries
		dtpEnclosingSystem
	Interface	dtpTextContent
	PlainItem	dtpDescription
		dtpMediniIdentifier
		dtpName
Funktion	ItemFunction	dtpDescription
		dtpId
		dtpMediniIdentifier
		dtpName
	ItemFunctionGroup	dtpMediniIdentifier
		dtpName
	ItemFunctionModel	dtpMediniIdentifier
		dtpName
Fehlfunktion	Malfunction	dtpDescription
		dtpId
		dtpMediniIdentifier
		dtpName
Betriebssituation	OperationalSituation	dtpDefaultExposure
		dtpDefaultExposureComment
		dtpItemUsage
		dtpLocation
		dtpRoadConditions
Gefährdungssituation	DefaultHazardousEventStereotype	dtpComment
		dtpRequiredReaction
		dtpSafetyGoalImport
	Hazard	dtpDescription
	HazardAnalysisModel	dtpMediniIdentifier
		dtpName
	HazardousEvent	dtpEffect
		dtpId
		dtpMediniIdentifier
ASIL	AsilInformation	dtpAsil
	IsoAsil	dtpControllability
		dtpControllabilityComment
		dtpExposure
		dtpExposureComment
		dtpMediniIdentifier
		dtpSeverity
		dtpSeverityComment
Sicherheitsziel	DefaultSafetyGoalStereotype	dtpAdoptedASIL
	SafetyGoal	dtpAsil
		dtpDescription
		dtpIdentifier
		dtpMediniIdentifier
		dtpName
		dtpSafeState
	SafetyRequirementsModel	dtpMediniIdentifier

Abbildung 3.2: Zuordnung der Typen und Prädikate des Datenmodells zu den Teilartefakten einer G&R

Frage, in welchem Ausmaß Duplikate in dem vorliegenden G&R-Datensatz enthalten sind, beschäftigt sich die Bachelorarbeit von Christopher Seidel [Sei18], die im Rahmen dieser Dissertation betreut wurde. Mittels Klonerkennung identifiziert Seidel in dem Datensatz Redundanz auf syntaktischer Basis. Dafür untersucht er in einer explorativen Studie alle verfügbaren Textinhalte mit dem Continuous Quality Assessment Toolkit (ConQAT) [ConQAT]. Im Folgenden fassen wir unsere Ergebnisse bezüglich Ausmaß und Arten der Redundanz kurz zusammen.

Ausmaß der Redundanz: Eine Klonklasse repräsentiert einen redundanten Text, der in der Studie aus mindestens vier Token besteht. Jedes Vorkommen dieses redundanten Textes wird entsprechend Klon dieser Klonklasse genannt. Unsere Studie zeigt, dass 73,5% aller Token des Datensatzes Teile von Klonen sind, das Gesamtausmaß an redundanten Textpassagen also relativ hoch ist. Die Anzahl der Klone je Klonklasse sinkt nichtlinear. In den meisten Fällen (2.146 Mal) existieren genau zwei Klone einer Klonklasse. Dem stehen wenige Klonklassen mit hoher Klonanzahl gegenüber. Bemerkenswert ist, dass im Datensatz einzelne Klonklassen bis zu 412 Klone besitzen, in dem Fall eine Textpassage also 412 Mal

vorkommt. Auch die Anzahl der Klonklassen sinkt nichtlinear bei zunehmender Klonlänge[21]. Etwa 92% aller Klonklassen bestehen aus 4 bis 16 Token. Den extremsten Fall bilden zwei Klonklassen der Klonlänge 146.

Arten der Redundanz: In der Studie identifizieren wir Teilartefakte, Typen und Prädikate, welche besonders häufig dupliziert werden, sowohl einzeln als auch in Kombination miteinander. Die drei Teilartefakte mit den höchsten Klonanteilen sind ASIL (34%), Gefährdungssituation (33%) und Fehlfunktion (13%). Die Prädikatarten mit den höchsten Klonanteilen sind Kommentare[22] (37%), Effekte und erwartete Reaktionen (28%) sowie Beschreibungen (13%). Insbesondere Beschreibungen und dem Teilartefakt Item zugehörige Prädikate heben sich zudem von den anderen Klonen ab, da sie besonders häufig keine hundertprozentige Übereinstimmung besitzen, sondern eine sogenannte Lücke (*engl.:* gap) aufweisen, was ein Anzeichen für das Kopieren und anschließende Modifizieren sein kann (siehe Abbildung B.1 in Anhang B.1, Seite 218). Die Analyse der Kombinationen von Prädikaten, welche besonders häufig Teil einer gemeinsamen Klonklasse sind, ergibt, dass insbesondere Funktionsnamen und Funktionsbeschreibungen gemeinsame redundante Textinhalte besitzen, die 16% dieser kombinierten Klone ausmachen (siehe Abbildung B.2 in Anhang B.1, Seite 218). Zudem tauchen der Kommentar zur Schadensschwere und der Effekt einer Gefährdung (9,6%) sowie der Kommentar zur Kontrollierbarkeit und die erwartete Reaktion bei einer Gefährdung (6,8%) häufig in gemeinsamen Klonklassen auf.

Die Ergebnisse unserer Studie unterstützen die Beobachtungen aus dem Praxisumfeld und sind ein starker Indikator für die praktische Anwendung von Copy-Paste-Modify oder Clone-and-Own als Methoden zur Wiederverwendung. Sie unterstützen zudem die Annahme der vorliegenden Arbeit, dass Wiederverwendungspotenziale im Datensatz vorhanden und nutzbar sind, und dass ihre Identifikation ein IR-Problem darstellt.

3.1.2 Lösungsstrategie

Aus den Beobachtungen der aktuellen Praxis und den identifizierten Herausforderungen ergibt sich der in Abbildung 3.3 dargestellte Anwendungsfall, der im Fokus dieser Arbeit steht.

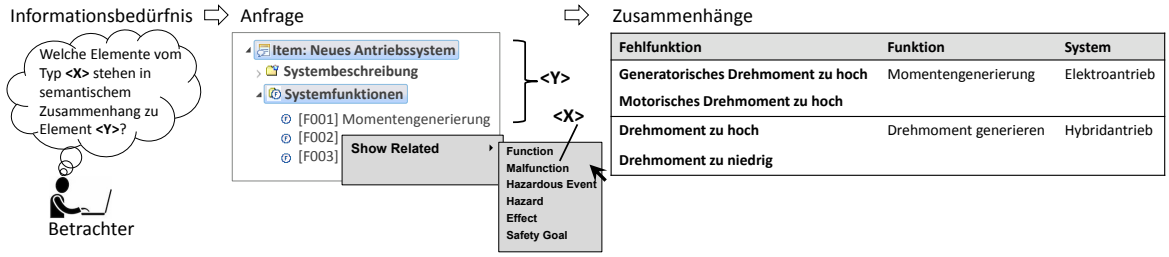


Abbildung 3.3: Anwendungsfall - Zusammenhänge erkennen

Ein Nutzer soll die Möglichkeit zur Äußerung und Befriedigung eines Informationsbedürfnisses erhalten. In einer generischen Form kann das die Frage *Welche Elemente vom Typ <X> stehen in semantischem Zusammenhang zu Element <Y>?* sein. Im Beispiel entspricht das dem folgenden Informationsbedürfnis. *Welche <Fehlfunktionen> (X) stehen in semantischem*

[21] Die Klonlänge ist die Anzahl der Token einer Klonklasse.

[22] Mehrere Prädikate wurden aufgrund ihres gemeinsamen Kommentarcharakters zusammengefasst, unter anderem dtpSeverityComment und dtpExposureComment.

Zusammenhang zur <Funktion Momentengenerierung des Systems Neues Antriebssystem> (Y)? Dieses Informationsbedürfnis wird beispielsweise über ein Kontextmenü als Anfrage formuliert. Die automatisiert bereitgestellten Ergebnisse tragen zum einen dazu bei, Überblick und Erkenntnisse zu gewinnen. Zum anderen ermöglicht dies die Identifikation und damit Empfehlung potenzieller Kandidaten für die Wiederverwendung in den verschiedenen Teilschritten einer G&R, da die Zusammenhänge oftmals als Basis für potenzielle Modifikationen und Übernahme verwendet werden können. Dadurch können Entscheidungen unterstützt werden. In Abbildung 3.3 werden beispielsweise Fehlfunktionen der Funktion *Momentengenerierung* des Systems *Elektroantrieb* und Fehlfunktionen der Funktion *Drehmoment generieren* des Systems *Hybridantrieb* als semantisch relevant identifiziert. Die Zusammenhänge und die Relevanz werden im Hintergrund durch das Retrievalverfahren erkannt und sollten sich das komplexe Wissen zunutze machen, damit die Ergebnisse auch für den Nutzer relevant und nützlich sind.

3.1.2.1 Anforderungen aus der Praxis

Funktionale Sicherheit wird in Unternehmen aufgrund ihrer besonderen Bedeutung oftmals konservativ behandelt. Veränderungen an vorhandenen Prozessen für die Aktivitäten im Sicherheitslebenszyklus dürfen die Qualität der Ergebnisse nicht gefährden. In dem Spannungsfeld zwischen Potenzialen neuer automatisierter, oftmals intelligenter Technologien und einer etablierten, konservativen, zumeist erfahrungsbasierten Praxis existieren in Unternehmen klare Erwartungen an neue Methoden. Die Nutzerakzeptanz spielt daher bei der Konzeption einer Methode zur Entscheidungsunterstützung in Sicherheitsanalysen eine wesentliche Rolle. Im Folgenden stellen wir die Hauptanforderungen vor.

Integration in verwendete Analysewerkzeuge: Für eine einfache Anwendung und einen größtmöglichen Nutzen soll konzeptuell auf die Anreicherung der existierenden Werkzeugkette um zusätzliche Entwicklungswerkzeuge verzichtet werden. Stattdessen ist eine Integrierbarkeit des Unterstützungssystems in vorhandene und verwendete Analysewerkzeuge zu bevorzugen. Für den Nutzer soll die Verwendung der Entscheidungsunterstützung keinen spürbaren Mehraufwand verursachen.

Automatisierbarkeit: Für die Nutzerakzeptanz einer G&R-Unterstützung ist deren einfache Bedienbarkeit unverzichtbar. Erheblicher Mehraufwand gilt in der Praxis als nicht akzeptabel. Allerdings erfordert insbesondere Wissensrepräsentation oftmals aufwendigen manuellen Einsatz von Domänenexperten, sowohl in der initialen Erstellung als auch in der Wartung und Pflege. Zudem ist der Erfolg semantischer Suchalgorithmen oftmals stark abhängig von Anpassungen und Konfiguration. Die Notwendigkeit solcher manuellen Anpassungen soll mit Blick auf die weitestgehende Automatisierbarkeit eingesetzter Methoden vermieden werden.

Wissensnutzung statt Wissensgenerierung: Entscheidungen sollen vom Werkzeug unterstützt, aber nicht für die Entwickler getroffen werden, da die Kontrolle über Entscheidungen bei ihnen verbleiben muss. Das bedeutet, dass trotz Automatisierung eine aktive Akzeptanz zur Übernahme vorgeschlagener Elemente in die Analyse erfolgen muss. Zudem soll ein System zur G&R-Unterstützung verstärkt auf die Wiederverwendung des vorhandenen Wissens setzen und auf die Generierung neuen Wissens verzichten. Das hat den Hintergrund, dass bereits getroffene Entscheidungen in abgeschlossenen Analysen als wahr gelten können. Die Prüfung

identifizierter Zusammenhänge zu wahren Aussagen auf Eignung einer Übernahme in den aktuellen Kontext wird als einfacher und sicherer empfunden als eine Überprüfung generierten Wissens auf Plausibilität und Korrektheit. Dies erfordert ein wesentlich höheres Maß an Verständnis.

Transparenz: Das Verständnis darüber, was zu bestimmten Empfehlungen geführt hat, kann das Vertrauen in das System fördern. Daher ist die Integration von Nachvollziehbarkeit der eingesetzten Algorithmen in das Konzept von Vorteil.

3.1.2.2 Vorteile und Nutzen auf einen Blick

Die automatisierte Identifikation syntaktischer und semantischer Zusammenhänge zielt auf die Nutzbarmachung verschiedener Vorteile ab. Diese adressieren zum einen Verbesserungen bei der Durchführung der Analysen. Zum anderen beziehen sie sich auf die qualitative Verbesserung von G&Rs.

Transparenz: Implizite Zusammenhänge erschließen sich nicht allen Experten oder bleiben sogar unentdeckt. Die automatisierte Identifikation semantischer Zusammenhänge soll Transparenz in die G&R-Landschaft und den Nutzern Überblick und Erkenntnisgewinn bringen. Die Nutzergruppe ist divers. Zu ihr gehören Experten mit unterschiedlichem Erfahrungsgrad (vom neuen Teammitglied zum erfahrenen Teamleiter) und unterschiedlichen Spezialisierungen, sowie Nutzer mit unterschiedlichen Rollen und Motivationen wie Entwickler oder Reviewer, die sich ebenso einen Überblick verschaffen wollen. Das Wissen abgeschlossener Analysen soll allen Nutzern, auch den Nicht-Experten, zugänglich sein.

Vereinfachung durch Entscheidungsunterstützung: In erster Linie sollen Zusammenhänge aufgedeckt werden, deren Existenz nicht offensichtlich ist. Eine kontextbasierte Auswertung soll jene Elemente in abgeschlossenen Analysen identifizieren, welche einen Beitrag zur aktuellen Analyse leisten können. Nimmt man diese identifizierten Zusammenhänge als Basis für Modifikationen oder lässt diese sogar in ihrem Originalzustand, so können sie auch als Empfehlungen dienen. Dadurch werden die Experten in der Gesamtheit ihrer Entscheidungen unterstützt. Ein schneller Überblick über die Zusammenhänge in den abgeschlossenen Analysen kann eine unabhängigere Bewältigung verschiedener Aufgaben fördern, da aufwendige Identifikation durch Diskussion in der Gruppe oder Konsultation der spezifischen Experten verringert wird. Das kann zudem zur Entlastung der Experten durch Entscheidungsunterstützung für Nicht-Experten beitragen. Ein vereinfachter Prozess erlaubt eine effizientere Bearbeitung der G&Rs.

Konsistenz und Vereinheitlichung: In der Norm ISO 26262 [ISO26262-3, 7.4.5.1c] wird die Konsistenz zwischen verwandten beziehungsweise ähnlichen[23] G&Rs gefordert. Das erfordert jedoch, dass eben jene G&Rs oder G&R-Teile, die in einem solchen Zusammenhang stehen, identifizierbar sind. Auf semantischer Ebene unterstützt zunehmende Konsistenz die gewünschte Objektivierung erfahrungsbasierter Entscheidungen. Neben der inhaltlichen Konsistenz zwischen G&Rs in all ihren Teilschritten zielt die Wiederverwendung auch auf eine Vereinheitlichung auf syntaktischer Ebene ab. Sie unterstützt die Etablierung einheitlicher Benennungen.

[23] In der ISO 26262 [ISO26262-3, 7.4.5.1c] *related*.

3.1.2.3 Generisches Rahmenwerk

Die Abbildung 1.1 auf Seite 5 zeigt einen Gesamtüberblick über die Lösungsstrategie in dieser Arbeit. Diese besteht darin, ein generisches Rahmenwerk zu entwickeln und mit Lösungen auszustatten, auf die ein Entwickler beim Aufbau eines domänenspezifischen IR-Systems zugreifen kann. Die vorgestellten Anforderungen aus der Praxis, zum Beispiel die Anforderungen nach Automatisierbarkeit, Integration im Entwicklungswerkzeug und Transparenz, stellen Herausforderungen dar und erfordern spezielle Lösungen. Daher entwickeln wir ein Rahmenwerk für domänenspezifische IR-Systeme, die diese Anforderungen teilen, und statten es im Verlauf dieser Arbeit mit Lösungen für diese Herausforderungen aus. In Abbildung 3.4 zeigen wir die Arbeitsweise dieser domänenspezifischen IR-Systeme in einem generischen IR-System zur Anfragebearbeitung. Dieses verfolgt einen assoziativen Ansatz zur Repräsentation und Informationsrückgewinnung. Außerdem orientiert es sich stark am Datenmodell des Entwicklungswerkzeugs, um die Herausforderungen der Automatisierbarkeit und Integrierbarkeit zu meistern. Wir stellen die einzelnen Bestandteile im Folgenden kurz vor.

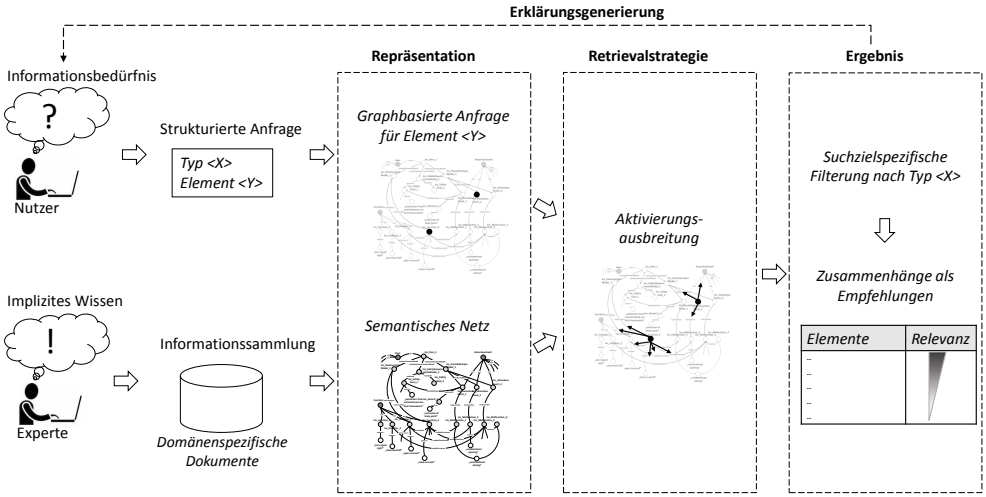


Abbildung 3.4: Generisches IR-System zur Anfragebearbeitung

Informationssammlung und Anfrage: Die Informationssammlung stellt die Datengrundlage des IR-Systems dar. Sie enthält das domänenspezifische Wissen in strukturierter Form, beispielsweise in Form relationaler Daten aus einer Datenbank. Das Informationsbedürfnis des Nutzers des IR-Systems wird als strukturierte Anfrage formuliert.

Repräsentation: Die vorhandenen Informationen aus den strukturierten Daten werden in die graphbasierte Repräsentation eines semantischen Netzes gebracht, welches die Darstellung von Zusammenhängen und Aussagen ermöglicht. Eine Besonderheit ist, dass wir uns bei Extraktion und Repräsentation an dem Datenmodell des Entwicklungswerkzeugs orientieren. Dadurch wird die Automatisierbarkeit der Methode gewährleistet, da die Informationen ohne Aufwand von Domänenexperten aus einer Datenbank mittels spezieller Transformationen in das semantische Netz überführt werden. Außerdem erleichtert die Adaption an das Datenmodell die Anforderung an Integrierbarkeit in das Entwicklungswerkzeug. Die Anfrage wird passend zur Informationssammlung ebenfalls graphbasiert repräsentiert.

Retrievalstrategie: Wir verwenden im generischen IR-System passend zur assoziativen Repräsentation eine assoziative Retrievalstrategie, die Aktivierungsausbreitung. Aktivierungsausbreitung ist eine bekannte Methode zur gegenseitigen Relevanzbewertung von Knoten eines Graphen. In dieser Arbeit wird Aktivierungsausbreitung zum einen für den vorgestellten Anwendungsfall nutzbar gemacht, zum anderen aber auch methodisch durch sogenannte Ausbreitungsmodi erweitert, so dass die Ausbreitung der Aktivierung exakter gesteuert werden kann.

Ergebnis und Erklärungsgenerierung: Als Ergebnis entsteht ein semantisches Netz, deren einzelne Knoten während der Aktivierungsausbreitung ein Maß an Aktivierung erhalten haben, welches Aufschluss über deren Relevanz zur gestellten Anfrage liefert. Zur Auswertung dieser Ergebnisse werden die Knoten nach Relevanz sortiert und nach Typen gefiltert, um die Anfrage bestmöglich zu beantworten. Diese suchzielspezifische Filterung wird insbesondere durch den struktur- und datenmodellbasierten Ansatz möglich. Das Ergebnis zeigt jene Zusammenhänge, nach denen in der Anfrage gesucht wurde. Es kann nun als Basis zur modifizierten oder sogar unveränderten Wiederverwendung dienen und als Empfehlung interpretiert werden. Zudem ist das IR-System durch ein Verfahren zur Erklärungsgenerierung angereichert, welches ebenfalls einen assoziativen Ansatz verfolgt. Mit diesem erfüllt das IR-System die Anforderung an Transparenz.

3.1.2.4 Methode zum Aufbau domänenspezifischer IR-Systeme

Für das Vorgehen beim Aufbau domänenspezifischer IR-Systeme mithilfe des Rahmenwerks definieren wir eine Methode, die in Abbildung 3.5 dargestellt ist. Die vier Schritte widmen sich einer Vorab-Analyse der Domäne und der Daten und den drei Bestandteilen des IR-Systems, die in Abbildung 3.4 als Repräsentation, Retrievalstrategie sowie Ergebnis und Erklärungsgenerierung benannt sind. Die folgenden Abschnitte dieses Kapitels widmen sich dem detaillierten Vorgehen in jedem Schritt. Daher folgt an dieser Stelle lediglich eine kurze Beschreibung.



Abbildung 3.5: Methode zum Aufbau domänenspezifischer IR-Systeme

In einer Vorab-Analyse sichtet der Entwickler des domänenspezifischen IR-Systems die Daten und Domäne. Es kann beispielsweise sinnvoll sein, zunächst zu prüfen, ob Wiederverwendungspotenziale vorhanden sind und ob überhaupt ein IR-Problem vorliegt. Außerdem kann es erforderlich sein, die Daten aufzubereiten. Im zweiten Schritt werden die Daten automatisiert in ein semantisches Netz überführt, welches die Strukturen des Datenmodells übernimmt. Der Entwickler des IR-Systems entscheidet dabei, mit welchen semantischen Anreicherungen oder Reduktionen das semantische Netz aufgewertet werden soll. Im dritten Schritt konfiguriert der Entwickler die Aktivierungsausbreitung. Wie wir in der Evaluation sehen werden, ist eine geeignete Konfiguration essenziell, um relevante und nützliche Ergebnisse zu erhalten. Daher ist es empfehlenswert, weitere Analysen zur Identifikation optimaler Konfigurationen durchzuführen. Im vierten Schritt kann der Entwickler festlegen, wie die Ergebnisse präsentiert werden sollen, beispielsweise den Umfang der Ergebnislisten. Zudem kann er hier entscheiden, ob Erklärungsgenerierung das IR-System bereichern soll, und diese gegebenenfalls konfigurieren.

3.1.3 Zusammenfassung und Beitrag

In diesem Abschnitt widmeten wir uns der ausführlichen Betrachtung der Herausforderungen einer G&R-Unterstützung im industriellen Umfeld und entwickelten eine Lösungsstrategie für diese Arbeit. Dafür analysierten wir zunächst in einer Studie das Ausmaß von Redundanz in dem von uns verwendeten G&R-Datensatz und zeigten, dass Wiederverwendungspotenziale in G&Rs existieren und bereits genutzt werden. Damit zeigten wir auch, dass die Unterstützung von G&Rs als IR-Problem betrachtet werden kann. Wir präsentierten einen Rahmen für den generischen Aufbau und die generische Arbeitsweise eines IR-Systems, welches die identifizierten Herausforderungen und Anforderungen aus der Praxis auch domänenübergreifend erfüllen kann. Dieses basiert auf assoziativem Retrieval und datenmodellbasierter Informationsextraktion. In den folgenden Abschnitten dieses Kapitels werden konfigurierbare Lösungen für die Bestandteile des generischen IR-Systems vorgestellt, die in das generische Rahmenwerk einfließen. Wir haben zudem eine Methode entwickelt, die das Vorgehen beim Aufbau domänenspezifischer IR-Systeme aus den generischen Lösungen in vier Schritten beschreibt. Genau genommen haben wir in diesem Abschnitt bereits am Beispiel der Domäne G&R gezeigt, was den ersten Schritt der Methode zum Aufbau domänenspezifischer IR-Systeme ausmacht. In einer Vorab-Analyse werden Domäne und Daten vom Entwickler des IR-Systems gesichtet, beispielsweise um festzustellen, ob es sich bei dem zu lösenden Herausforderung tatsächlich um ein IR-Problem handelt und ob sich die Anwendung der Methode demzufolge überhaupt eignet.

Die verbleibenden drei Schritte befassen sich mit der Repräsentation, der Retrievalstrategie und den Ergebnissen. Sie werden in den folgenden Abschnitten dieses Kapitels separat betrachtet. Der folgende Abschnitt 3.2 widmet sich dem zweiten Schritt, dem Aufbau der semantischen Netze.

3.2 Aufbau domänenspezifischer semantischer Netze

In diesem Abschnitt widmen wir uns dem Aufbau der domänenspezifischen semantischen Netze sowie der Repräsentation der Anfragen. Dafür werden wir zunächst vorstellen, wie wir einen automatisierten Aufbau der Netze ermöglichen, indem wir uns an den Strukturen des Datenmodells orientieren. Dadurch reduzieren wir den Aufwand beim Aufbau neuer domänenspezifischer IR-Systeme erheblich und überwinden eine der großen Hürden bei der Verwendung semantischer Netze. Anschließend beschreiben wir den Aufbau einer Basisvariante der Netze. Wir werden zudem Techniken für semantische Anreicherungen und Reduktionen der Netze präsentieren, die das Ziel einer semantischen Aufwertung verfolgen. Wir werden parallel dazu zeigen, wie wir diese Techniken auf dem G&R-Datensatz anwenden und unsere domänenspezifischen Netze erstellen. Abschließend stellen wir ein passendes Konzept für die strukturierte Repräsentation der Anfragen vor.

3.2.1 Datenmodell als Basis für die Wissensrepräsentation

Die in Abschnitt 3.1.2.1 vorgestellten Anforderungen aus der Praxis zeigen die Notwendigkeit der Automatisierbarkeit der Methode sowie der Integrierbarkeit in vorhandene Entwicklungswerkzeuge auf. Aus diesem Grund nutzen wir das Datenmodell des Entwicklungswerkzeugs als Fundament für das IR-System.

3.2.1.1 Grundidee

Abbildung 3.6 visualisiert die Grundidee zur Integration des IR-Systems. Das Datenmodell, rechts oben abgebildet, gibt Aufschluss über die im Entwicklungswerkzeug verwalteten Datenobjekte, deren Beziehungen, Hierarchien und Attribute. Bestehende Funktionalitäten des Entwicklungswerkzeugs (oben links im Bild) werden über das vorhandene Datenmodell und die entsprechend gespeicherten Daten realisiert. Die IR-Komponente erweitert das Entwicklungswerkzeug durch neue Funktionalität und erfordert für das assoziative Retrievalverfahren die graphbasierte Datenstruktur eines semantischen Netzes. Da dies ohnehin die Semantik der abgelegten Daten repräsentieren soll, wird es mittels automatisierter Transformationen direkt aus Datenmodell und Daten abgeleitet. Modifikationen zur Aufwertung des semantischen Netzes erfolgen ebenfalls automatisiert.

Die in den primären Daten vorhandene Trennung von Schema- und Datenschicht wird auf das Netz übertragen und entspricht damit dem Grundgedanken RDF-basierter Sprachen, die ebenfalls auf die Trennung von Struktur und Inhalt setzen. Die im Datenmodell beschriebene Strukturierung entspricht dem terminologischen Wissen, während die Daten das assertionale Wissen darstellen und implizit bereits Aussagen der Form Subjekt-Prädikat-Objekt enthalten (siehe Abschnitt 2.2.2). Daher eignet sich als Zielformat die graphbasierte Wissensrepräsentation eines RDF-Graphen.

Die im abgeleiteten semantischen Netz abgebildeten Informationen korrespondieren zu den Daten der primären Datensicht, verändern diese aber nicht. Die IR-Komponente wendet ihre semantischen Suchalgorithmen auf den semantischen Netzen dieser sekundären Datenhaltung an. Die integrative Lösung muss allerdings sicherstellen, dass korrespondierende Elemente beider Repräsentationen in der jeweils anderen Repräsentation identifizierbar sind. Das ist wichtig, damit beispielsweise die Eingabe des Ausgangspunktes der semantischen Suche von der IR-Komponente und die Ausgabe der Ergebnisse der Suche wiederum von den Funktionalitäten des Entwicklungswerkzeugs interpretiert werden können.

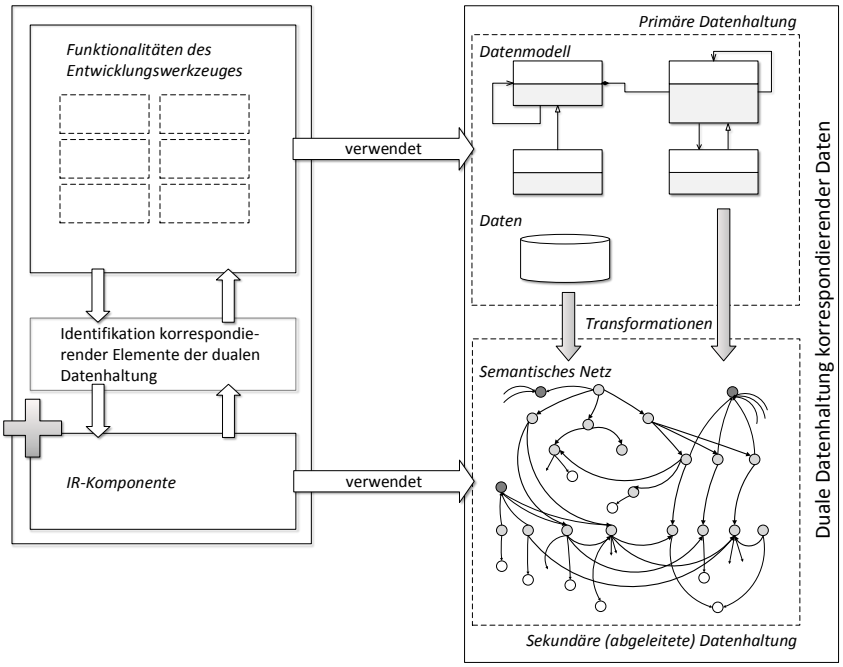


Abbildung 3.6: Integration der IR-Komponente und Verwendung einer dualen Datenhaltung

3.2.1.2 Vorteile dieses Vorgehens

Aufwand: Durch die Nutzung wertvoller Informationen, die durch das Datenmodell zur Verfügung stehen, kann manueller Aufwand zur Wissensrepräsentation durch Domänenexperten entfallen beziehungsweise gering gehalten werden. Die vollständige Automatisierbarkeit, welche durch den Fokus auf das Datenmodell möglich ist, stellt eine Besonderheit des hier vorgestellten Ansatzes dar. Durch sie wird die aufwendige Wissensrepräsentation als eine der großen Hürden bei der Verwendung semantischer Netze und Ontologien überwunden.

Gültigkeit bei Instanziierung: Aufgrund der Ableitung des grundlegenden schematischen Aufbaus des semantischen Netzes aus dem Datenmodell unterliegt es während der Instanziierung gewissen Regeln. Dadurch ist zum einen eine Validierung der Daten möglich. Außerdem ermöglicht die Transparenz der Datenstrukturen die Formulierung strukturierter Anfragen beim IR.

Objektidentifizierbarkeit: Die Transformationen werden bereits so durchgeführt, dass die Elemente des semantischen Netzes nachträglich in der primären Datenhaltung identifiziert werden können. Durch die gemeinsame Struktur kann beispielsweise gewährleistet werden, dass Elemente des primären Datensatzes im semantischen Netz identifiziert werden können und umgekehrt. Das wiederum ist notwendig für die Integration der IR-Funktionalität in das Entwicklungswerkzeug. Die Verwendung RDF-basierter Sprachen hat den Vorteil, dass für die gegenseitige Identifikation der Elemente Abfragesprachen wie SPARQL genutzt werden können.

Integrierbarkeit: Die Datenstrukturen bleiben erhalten. Das unterstützt die Integrierbarkeit in das verwendete Entwicklungswerkzeug und vereinfacht die Erfüllung der Anforderung

nach Anwendbarkeit. Durch die Koexistenz der primären Datenhaltung und deren abgeleiteten, sekundären Repräsentation, kann die IR-Komponente im Entwicklungswerkzeug verwendet werden. Durch die IR-Komponente werden keine Modifikationen an den primären Daten vorgenommen, was ebenfalls eine wichtige Anforderung aus der Praxis ist.

Leichtgewichtigkeit: Das semantische Netz muss nicht das komplette Datenmodell abbilden, sondern kann sich lediglich auf den für die IR-Komponente relevanten Teil beschränken. Irrelevante Strukturinformationen können ignoriert werden, so dass das semantische Netz nicht unnötig überladen wird. Das erlaubt die automatisierte Erstellung einer leichtgewichtigen Basisvariante semantischer Netze.

3.2.2 Aufbau des G&R-Basisnetzes

Beim Aufbau des semantischen Netzes erstellen wir aus dem Datensatz zunächst automatisiert durch direkte Transformationen ein einfaches semantisches Netz, welches im Folgenden auch als Basisnetz bezeichnet wird. Hierfür werden die in Abschnitt 2.2.3 vorgestellten Transformationen der Ontmalizer-Bibliothek verwendet. Dafür wird zuerst auf Basis des Datenmodells die Modellebene des semantischen Netzes durch *XSD2OWL*-Transformationen erstellt. Darauf aufbauend überführen wir die Daten mittels *XML2OWL*-Transformationen in das semantische Netz. Sie bilden die Instanzebene. Wir werden zudem beispielhaft zeigen, wie wir mit den vorgestellten Techniken das G&R-Basisnetz aufbauen. Durch den strukturierten Aufbau besitzen die Basisnetze spezielle Eigenschaften, die wir beleuchten werden. Diese Eigenschaften machen wir uns im späteren Retrievalverfahren zunutze.

3.2.2.1 Modellebene

In Tabelle 3.2 sind die Ergebnisse für die *XSD2OWL*-Transformationen der vier wichtigsten Konstrukte und Beziehungen am Beispiel eines Klassendiagrammausschnittes dargestellt, die wir im Folgenden zeilenweise beschreiben.

Jede *Klasse* wird zu einer Ressource vom Typ `owl:Class`, was im RDF-Graphen z.B. durch das Tripel `(<hara:Function> <rdf:type> <owl:Class>.)` abgebildet wird.

Generalisierung wird durch das Prädikat `rdfs:subclassOf`, im Beispiel durch das Tripel `(<hara:Malfunction> <rdfs:subclassOf> <hara:Function>.)`, etabliert.

Assoziationen entsprechen in OWL der abstrakten Property-Rolle zur Verknüpfung von Individuen und werden daher zu Ressourcen vom Typ `owl:ObjectProperty`. Ein Beispiel ist das Tripel `(<hara:hasMalfunctions> <rdf:type> <owl:ObjectProperty>.)`. Der Assoziationsbezug legt Definitions- und Wertebereich für die spätere Instanziierung fest: Das Prädikat `hara:hasMalfunctions` darf ausschließlich Instanzen des Typs `hara:Function` mit Instanzen des Typs `hara:Malfunction` verknüpfen. Solche Einschränkungen werden durch Restriktionen abgebildet, im Beispiel durch den anonymen Knoten `_:Restriction_1`. Da `hara:Malfunction` damit auch als Wertebereich agiert, wird zudem dessen Typbeziehung zu `rdfs:Datatype` etabliert. Sind beispielsweise Multiplizitäten im XML-Schema vorhanden, so werden diese in die passenden OWL-Konstrukte, z.B. `owl:maxCardinality`, `owl:minCardinality` oder `owl:someValuesFrom`, transformiert und durch weitere Restriktionen im RDF-Graphen sichtbar. Zugunsten eines leichtgewichtigen Basisnetzes nehmen wir Multiplizitäten nicht in die Modellebene auf, da wir von dieser Strukturinformation keinen Vorteil für die IR-Komponente erwarten.

Jedes *Attribut* einer Klasse entspricht der konkreten Property-Rolle zur Verknüpfung von Individuen mit konkreten Werten und wird durch eine Ressource von Typ `owl:DatatypeProperty` abgebildet, z.B. `(<hara:dtpName> <rdf:type> <owl:DatatypeProperty>.)`. Definitions- und Wertebereich werden ebenfalls über eine Restriktion, im Beispiel durch den anonymen Knoten `_:Restriction_2`, abgebildet. Nur Instanzen vom Typ `hara:Function` (und deren Subklassen) dürfen durch das Prädikat `hara:dtpName` und zudem ausschließlich mit Literalen des Typs `xsd:string` verknüpft werden.

3.2.2.2 Instanzebene

Abbildung 3.7 zeigt die Erweiterung des Beispiels aus Abschnitt 3.2.2.1 um die Instanziierung durch *XML2OWL*-Transformationen. Jede Instanz eines Typs wird als Individuum zu einer

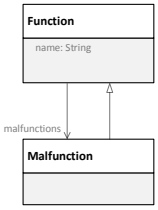
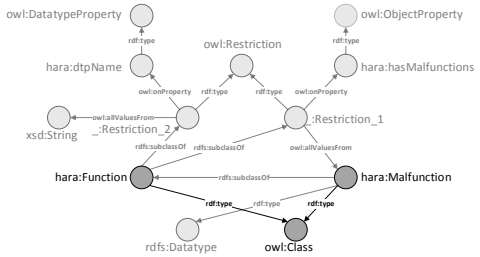
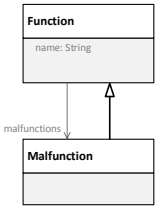
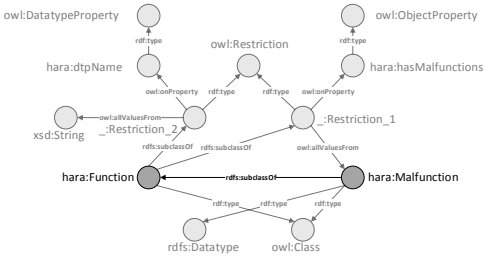
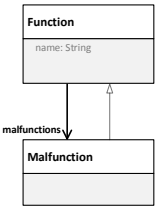
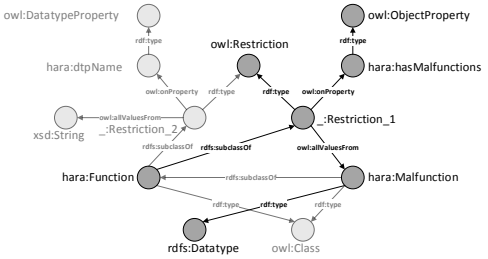
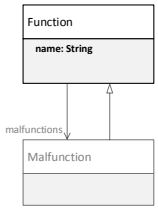
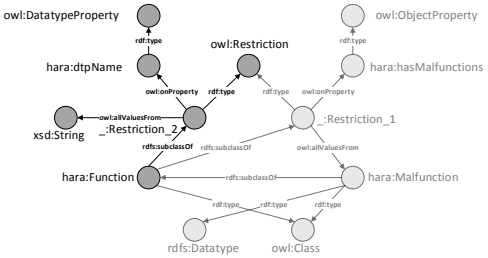
Konstrukte und Beziehungen	**RDF-Graph-Ausschnitt**
Klassen 	
Generalisierung 	
Assoziationen 	
Attribute 	

Tabelle 3.2: RDF-Graph nach *XSD2OWL*-Transformation am Beispiel selektierter Konstrukte und Beziehungen im Klassendiagramm

Ressource mit Typzugehörigkeit. Im Beispiel ist dies durch das Tripel (`<hara:Ins_Function_1> <rdf:type> <hara:Function>.`) abgebildet. Die konkreten Beziehungen zwischen den Instanzen werden passend zu den jeweiligen Restriktionen durch Prädikate des Typs `ObjectProperty` dargestellt. Im Beispiel erfüllt das Tripel (`<hara:Ins_Function_1> <hara:hasMalfunctions> <hara:Ins_Malfunction_3>.`) die oben beschriebene Restriktion. Ebenso werden die konkreten Attribute der Instanzen durch Verwendung der Prädikate des Typs `DatatypeProperty` dargestellt. Das Tripel (`<hara:Ins_Function_1> <hara:dtpName> <"Sitzheizung">.`) verknüpft beispielsweise die Ressource mit einem Literal vom Typ String.

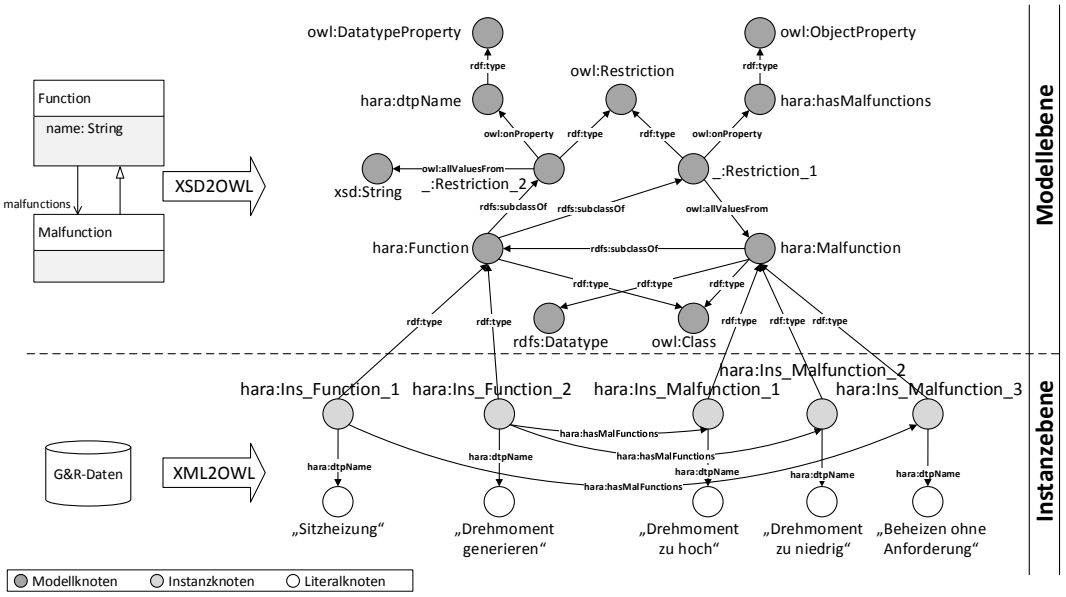


Abbildung 3.7: RDF-Graph nach *XSD2OWL*- und *XML2OWL*-Transformationen am Beispiel

3.2.2.3 Namenskonventionen

Die verwendeten Namenskonventionen für die Benennung der transformierten, domänenspezifischen Elemente entsprechen weitestgehend denen der verwendeten Ontmalizer-Bibliothek. Die Bezeichnung der Knoten und Kanten wird im Folgenden als *Label* bezeichnet. Ein Label (z.B. `hara:Ins_Function_1`) besteht in der Regel aus einem Namensraum-Präfix *ns* (z.B. `hara`), einem optionalen domänenspezifischen Präfix *dsPrefix* (z.B. `Ins`) sowie einem domänenspezifischen Suffix *dsSuffix* (z.B. `_Function_1`), das aus mindestens einem Element besteht. Der allgemeine Aufbau entspricht *<ns>:<dsPrefix><dsSuffix>*. Ausnahmen bilden die anonymen Knoten. Diese werden durch einen vorgesetzten Unterstrich und einen eindeutigen Namen in der Form *_:<blanknodeName>* bezeichnet. Zudem gilt für ein Literal, dass dessen Label seiner literalen Zeichenketten *literalName* entspricht. In Tabelle 3.3 sind die Namenskonventionen zur Benennung domänenspezifischer Elemente zusammengefasst.

	<ns>:	**<dsPrefix>**	**<dsSuffix>**
Klassenknoten	*<ns>*:		*<typeName>*
Instanzknoten	*<ns>*:	Ins	*<itemID>_<typeName>_<elementNumber>*
ObjectProperty	*<ns>*:	has	*<propertyName>*
DatatypeProperty	*<ns>*:	dtp	*<propertyName>*
Anonymer Knoten	_:		*<blanknodeName>*

Tabelle 3.3: Namenskonvention zur Benennung domänenspezifischer Elemente

Klassen werden gemäß ihres Klassen- beziehungsweise Typnamens *typeName* benannt, wobei das domänenspezifische Präfix entfällt. Der Name eines Properties *propertyName* entspricht dem jeweiligen Assoziations- oder Attributnamen. Die Art der Properties unterscheidet sich durch das domänenspezifische Präfix, wobei Prädikate des Typs `ObjectProperty` das Präfix `has` und Prädikate des Typs `DatatypeProperty` das Präfix `dtp` erhalten. Die differenziertere

Benennung von Instanzen ist einer der Hauptunterschiede zur Benennung der Ontmalizer-Bibliothek. Instanzen erhalten das domänenspezifische Präfix `Ins` und werden durch drei Elemente des domänenspezifischen Suffixes eindeutig spezifiziert. Der Typname *typeName* bezieht sich auf die Klasse, deren Instanz der jeweilige Knoten ist. Die *itemID* ist ein eindeutig generierter Identifizierer für jedes instanziierte Item, also für jede eingelesene G&R.[24] Dadurch kann jedes Element, egal welchen Typs, dem zugehörigen Item bereits anhand des Namens zugeordnet werden. Zudem hat dieses Vorgehen den Vorteil, dass die Elemente eines Typs lokal innerhalb eines Items eine erneute fortlaufende Nummerierung durch *elementNumber* erhalten können. Dies trägt vor allem zur Übersichtlichkeit bei.

3.2.2.4 Anpassungen und Erweiterungen der Ontmalizer-Bibliothek

Für die *XML2OWL*- und *XSD2OWL*-Transformationen verwenden wir die Ontmalizer-Bibliothek, welche grundsätzlich die in Abschnitt 2.2.3 vorgestellten Mappings implementiert. Zur Erstellung des Basisnetzes nehmen wir Anpassungen vor, beispielsweise um auch mehrere Schemata in das OWL-Modell zu integrieren oder das systemweise Einlesen der G&R-Daten zu ermöglichen. Außerdem erlaubt eine Reduzierung unnötiger anonymer Individuen und eine grundsätzliche Änderung in deren Namenskonvention, beispielsweise die in Abschnitt 3.2.2.3 beschriebene Benennung der Individuen, eine übersichtlichere Darstellung des Basisnetzes. Eine weitere Ergänzung stellt die Handhabung reflexiver Assoziationen dar. Beispielsweise können Funktionen in G&Rs über Systemgrenzen hinweg zu anderen Funktionen beitragen oder diese erfordern. Da die Daten systemweise sukzessive eingelesen und transformiert werden, können Referenzen auf Funktionen eines anderen Items nicht direkt, sondern erst nach deren Einlesen aufgelöst werden. Daher werden reflexive Assoziationen bei der Instanziierung erst abschließend in einem zusätzlichen Schritt aufgelöst.

3.2.2.5 Charakteristik des Basisnetzes

Durch die verwendeten Transformationen haben entstehende Basisnetze spezielle Eigenschaften. Daher beschreiben wir an dieser Stelle allgemeine Charakteristika von Basisnetzen und präsentieren anschließend die konkreten Eigenschaften des G&R-Basisnetzes nach den Transformation.

Schematische Struktur des Basisnetzes

Die Abbildung 3.8 stellt zur visuellen Unterstützung die schematische Struktur des Basisnetzes unterteilt in Modell- und Instanzebene dar. Auf Modellebene spielen die Informationen aus dem Datenmodell eine zentrale Rolle. Jeder *Klassenknoten* ist vom Typ `owl:Class` und daher mit diesem Knoten verbunden. Alle Klassenknoten, die gleichzeitig als Wertebereich für ein `ObjectProperty` agieren, sind zudem ein `rdfs:Datatype` und daher mit diesem Knoten verbunden. Jeder Klassenknoten, welcher den Definitionsbereich eines Properties darstellt, ist durch `rdfs:subclassOf`-Prädikate mit *Restriktionsknoten* verknüpft. Diese sind wiederum jeweils mit einem *Property-Knoten* verbunden. Ein Property-Knoten besitzt entweder eine Typbeziehung zu dem Knoten `owl:ObjectProperty` oder zu dem Knoten `owl:DatatypeProperty`. Jeder Restriktionsknoten für ein `ObjectProperty` ist wiederum durch ein `owl:allValuesFrom`-Prädikat mit einem Klassenknoten verbunden, während jeder Restriktionsknoten für ein

[24] In einigen Abbildungen, beispielsweise 3.7, wird zugunsten der Übersichtlichkeit eine verkürzte Benennung verwendet und auf die *itemID* verzichtet.

`DatatypeProperty` mit einem Knoten verbunden ist, der einen *primitiven Datentypen* repräsentiert, z.B. `xsd:string`.

Auf der Instanzebene stehen alle Individuen, die *Instanzknoten*, in einer Typbeziehung zum jeweiligen Klassenknoten. Zwischen den Instanzknoten existieren Verbindungen durch Prädikate des Typs `ObjectProperty`. Zudem existieren die Verbindungen zu den *Literalknoten* durch die Prädikate des Typs `DatatypeProperty`. Wegen des Grundprinzips der Objektidentität semantischer Netze ist jeder Knoten, auch jedes Literal, eindeutig identifizierbar. Das bedeutet, dass Literale mit gleichem Inhalt nur ein Mal vorkommen, weshalb Literale nicht zwangsläufig Blätter im semantischen Netz sind, sondern auch als Verbindungsknoten zwischen Instanzknoten agieren können.

semantischer Kern

Während auf Modellebene *domänenmodellspezifische Aussagen* abgebildet werden, bezeichnen wir die Instanzebene als den *semantischen Kern*, da hier die konkreten Informationen zu den Inhalten der G&Rs verortet sind. Die Klassenknoten stellen die Schnittstelle zwischen den Ebenen dar. Da das Wissen über den Typ eines Individuums, also ob es sich z.B. um eine Funktion, Fehlfunktion, Gefahr oder ASIL-Einstufung handelt, ebenfalls als konkrete Information gesehen werden kann, bezeichnen wir den Teil des Basisnetzes, welcher Instanzebene und domänenspezifische Klassenknoten umfasst, als *erweiterten semantischen Kern*.

erweiterter semantischer Kern

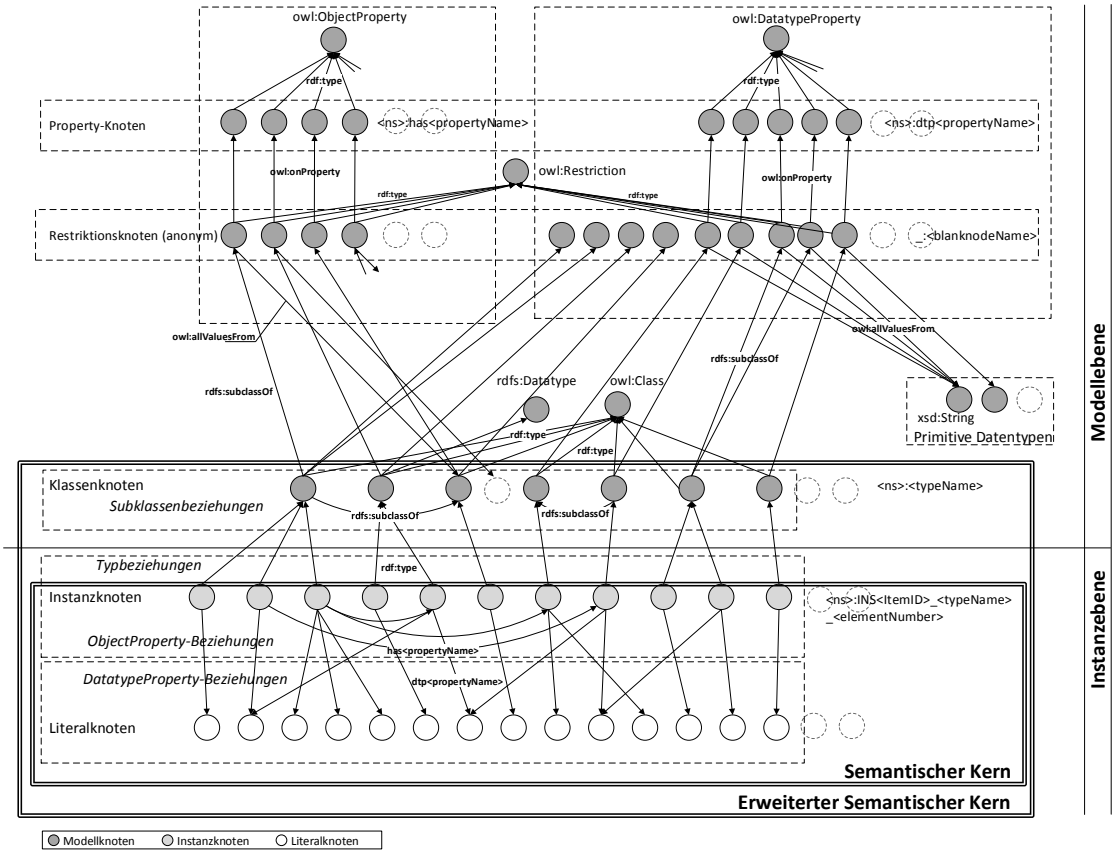


Abbildung 3.8: Schematische Struktur des Basisnetzes

Allgemeine Eigenschaften des Basisnetzes

Feste Modellebene: Die Modellebene besteht ausschließlich aus Ressourcen-Knoten und anonymen Knoten. Es existieren keine Literale oder domänenspezifische Instanzknoten. Das Basisnetz vermag durch Instanziierung wachsen, die Modellebene des Basisnetzes verändert sich dabei in der Regel nicht. Veränderungen auf Modellebene ergeben sich lediglich bei Änderungen des zugrundeliegenden Datenmodells, beispielsweise durch ein Software-Update.

Kontrolliert wachsende Instanzebene: Im Gegensatz zur Modellebene bewirkt das Hinzufügen neuer Daten auf Instanzebene das kontrollierte Wachstum gemäß der strukturellen Eigenheiten. Neue Instanzen werden demnach als Instanzknoten über Typbeziehungen zu den Klassenknoten etabliert und können konkrete Beziehungen zu anderen Instanzknoten sowie konkrete Beziehungen zu bereits vorhandenen oder neu zu erstellenden Literalen besitzen. Beim Wachstum des Netzes durch Instanziierung gelten besondere Gesetze, die denen skalenfreier[25] Netze ähneln. Charakteristisch ist beispielsweise der geringe Durchmesser[26], der bei Wachstum des Netzes durch Instanziierung nicht mitwächst. Stattdessen tragen die neuen Knoten zu einer stärkeren Verzweigung bestimmter Knoten bei.

Fokus auf Typbeziehungen: Das Wissen um die Zuordnung von Instanzen zu ihren Typen wird bei der Erstellung des Basisnetzes bewahrt. Da der in dieser Arbeit untersuchte Anwendungsfall (siehe Abbildung 3.3, Seite 57) auf die Suche nach Elementen bestimmter Typen fokussiert, eignet sich das so erstellte semantische Netz, da die Typzuordnung leicht extrahiert werden kann.

Verzweigung: Durch den starken Fokus auf Typzuordnung und Typbeziehungen tendieren Klassenknoten zu hohen Knotengraden und damit zu einer starken Verzweigung. Diese Eigenschaft kann in gleichem Maße auch auf Literalknoten mit häufig vorkommenden Zeichenketten zutreffen, welche aufgrund der Objektidentität mit vielen Instanzknoten verbunden sein können. Starke Verzweigungsknoten sind Knoten mit besonderer Bedeutung. Ihre Entfernung aus dem semantischen Netz sorgt für extremere Effekte als das Entfernen eines Knotens mit kleinem Knotengrad.

Hoher Informationsgehalt in Literalen: Durch die direkte Übernahme der Attributinhalte existieren Literalknoten mit sehr langen Zeichenketten. Dies resultiert insbesondere aus Attributen zu Beschreibungen oder Kommentaren. Literalknoten mit langen Zeichenketten sind semantisch sehr bedeutsam, da sie oftmals viele Informationen in nur einem Knoten bündeln. Dadurch sind Tripel zwar atomar, aber ein Literalknoten kann für einen komplexen Sachverhalt stehen.

[25] In skalenfreien Netzen unterliegt die Knotengradverteilung, anders als bei zufälligen Netzen, dem Potenzgesetz [BB04]. Das bedeutet, es existieren einige wenige Knoten mit hohem Knotengrad, während viele Knoten einen niedrigen Knotengrad besitzen. Diese Besonderheit wird in der Graph- und Netztheorie dafür verwendet, Knoten mit besonders hohem Einfluss zu identifizieren, da das Entfernen eines starken Verzweigers wesentlich größere Effekte auf ein Netz hat als die Wegnahme eines Knotens mit geringem Knotengrad. Beispiele für reale skalenfreie Netze sind das Internet, Strom- und Transportnetze [BB04]. Skalenfreie Netze weisen zudem meist das Kleine-Welt-Phänomen auf, welches die Eigenschaft beschreibt, dass jedes Knotenpaar über wenige Knoten direkt miteinander verbunden ist, also das Netz einen geringen Durchmesser besitzt [BB04]. Da RDF-Graphen im Gegensatz zu zufälligen Netzen kontrolliert und typspezifisch wachsen, teilen sie grundlegende Charakteristika skalenfreier Netze.

[26] Der Durchmesser ist der größtmögliche Abstand zweier Knoten in einem Graphen.

Konkrete Eigenschaften des G&R-Basisnetzes

Die Transformation der 155 G&Rs führt zu einem Basisnetz, welches aus 131.445 RDF-Tripeln und damit ebenso vielen Kanten besteht. Die Verteilung der Kanten und Knoten für das G&R-Basisnetz ist in Tabelle 3.4 aufgelistet. Mehr als die Hälfte der Kanten entfallen auf verwendete Datatype Properties, fast ein Viertel auf die Verwendung von Object Properties. Etwa ein Fünftel der Kanten werden zur Instanztypzuordnung verwendet. Damit beschreiben 79,72% der Kanten den semantischen Kern und 99,67% den erweiterten semantischen Kern. Die verbliebenen Kanten beziehen sich auf die restliche Modellebene.

Elemente	**Anzahl**
RDF-Tripel	131.445
Kanten, davon	131.445
- Datatype Properties	73.660
- Object Properties	31.132
- Instanztypbeziehungen	26.222
- andere Typbeziehungen	197
- Restriktionsbeziehungen	150
- Subklassenbeziehungen	84
Verwendete Kantentypen (Properties), davon	47
- domänenspezifisch, davon	43
- Datatype Properties	27
- Object Properties	16
- OWL-/RDF-spezifisch	4
Knoten, davon	49.525
- Instanzknoten	26.222
- Literalknoten	23.119
- Klassenknoten	23
- Property-Knoten, davon	54
- Datatype Properties	34
- Object Properties	20
- Restriktionsknoten	75
- Rest (z.B. OWL-/RDF-spezifische Knoten, anonyme Knoten)	32

Tabelle 3.4: Konkrete Eigenschaften des G&R-Basisnetzes

Die Kanten werden aus 47 verschiedenen Kantentypen gebildet, von denen lediglich vier OWL-, RDF- oder RDF-Schema-Konstrukte der Modellebene sind. Weitere 16 sind domänenspezifische Object Properties und 27 domänenspezifische Datatype Properties. Das Netz enthält 49.525 Knoten, von denen mehr als die Hälfte Instanzknoten und fast die Hälfte Literale abbilden. Die Ressourcentypen der Domäne werden durch 23 Klassenknoten repräsentiert. Damit beschreiben 99,63% der Knoten den semantischen Kern und 99,67% den erweiterten semantischen Kern. Die verbliebenen Knoten beziehen sich auf die restliche Modellebene. Beispielsweise existieren 20 Knoten für die Object Properties und 34 Knoten für die Datatype Properties.[27]

[27] Diese insgesamt 54 domänenspezifischen Property-Knoten sind auf der Modellebene abgebildet. Auf der Instanzebene müssen diese nicht zwangsläufig verwendet werden, was die Differenz zu den verwendeten 47 Kantentypen in der Tabelle erklärt.

3.2.3 Semantische Anreicherungen

Wir stellen im Folgenden Techniken zur Aufwertung der Basisnetze durch semantische Anreicherungen vor. Dafür analysieren wir zunächst die Komplexität der Literale im G&R-Basisnetz und zeigen, welche Teilartefakte und Prädikate besonders komplexe Textinhalte besitzen. Die erste Technik erlaubt eine Extraktion von Keywords aus Literalen mit komplexen Textinhalten. Die zweite Technik dient der Anreicherung durch den Umgang mit bekannten Synonymen und Akronymen. Beide ermöglichen den Ausbau semantischer Netze durch die differenzierte Betrachtung der Inhalte von Literalknoten und deren explizite Integration in das Netz. Durch semantische Anreicherungen werden zusätzliche Verbindungen zwischen kontextuell verbundenen Knoten geschaffen, was für das spätere IR von Bedeutung ist. Wir zeigen zudem, wie wir diese generischen Techniken für die Anreicherung der G&R-Basisnetze verwenden.

3.2.3.1 Analyse der Literale im G&R-Basisnetz

Literalknoten haben die Eigenschaft, dass sie in ihren literalen Inhalten eine Vielzahl von Aussagen bündeln können. Während auf Instanzebene verwendete Object Properties zu atomaren Aussagen über Instanzen führen, beispielsweise `<hara:Ins_Function_1><hara:hasMalfunctions> <hara:Ins_Malfunction_1>`, müssen verwendete Datatype Properties nicht zwangsläufig atomar sein, da ein Literal häufig nicht nur aus einem Wort besteht, sondern viele Aussagen enthalten kann. Als Beispiele können die Textfelder zur Systembeschreibung in Abbildung 3.1 (Seite 54) dienen. Hintergrund ist, dass im Entwicklungswerkzeug Freitextfelder nicht nur die Eingabe von Namen, sondern auch längerer Zeichenketten wie Beschreibungen oder Kommentaren erlauben. Das führt dazu, dass lange Ausführungen mit mehreren Aussagen an einen Knoten gebunden sind und nicht mehrere Tripel bilden, wie es idealerweise erwartet werden könnte. Literalknoten mit langen Zeichenketten sind aufgrund der potenziell vielen gebündelten Informationen in nur einem Knoten semantisch sehr bedeutsam. Daher ist es notwendig, eingebettete Semantik aus sehr ausführlichen Literalen zu extrahieren.

Analyse verwendeter Literale: In einer Analyse aller Tripel des Basisnetzes, die Literale verwenden, verschaffen wir einen Überblick über die Länge der Literale. Dafür unterteilen wir die Zeichenketten der Literale in einfache Token, indem diese an Stelle von Leerzeichen getrennt werden.[28] In Tabelle 3.5 sind die Ergebnisse unserer Analyse dargestellt. Die Gesamtanzahl der Literalverwendungen (73.660) entspricht der Anzahl aller Kanten, die Datatype Properties repräsentieren (siehe Tabelle 3.4, Seite 72), da wir die Literale jedes Tripels untersuchen. Zudem zeigt Abbildung 3.9 auf einer logarithmischen Skala, wie sich die durchschnittliche Tokenanzahl auf die verwendeten Prädikate (beziehungsweise Datatype Properties) verteilt und wie wir diese den Teilartefakten zuordnen können.

Auswertung: Das Literal mit leerer Zeichenkette existiert aufgrund der Objektidentität zwar nur ein Mal im Netz, wird aber 5.000 Mal verwendet. Mehr als die Hälfte aller Literalverwendungen beziehen sich auf Literale, die aus lediglich einem Token bestehen. Hingegen

[28] Ein Token entspricht in dieser Analyse nicht zwangsläufig einem Wort, sondern vielmehr einem durch Leerzeichen eingegrenzten Bereich. Da die Inhalte der Token an dieser Stelle nicht betrachtet werden, verzichten wir beispielsweise auf die Entfernung von Satzzeichen. Da diese Analyse darauf abzielt, einen Überblick über den Umfang und die Komplexität der Literalinhalte zu verschaffen, ist diese Herangehensweise hier zweckmäßig, da auch durch Leerzeichen separierte Ausdrücke wie beispielsweise „<3km/h“ gezählt werden, gleichfalls aber einfache Satzzeichen keine eigenen Token darstellen.

Tokenanzahl	Anzahl der Literalverwendungen	Anteil von Gesamtanzahl
0	5.000	6,79%
1	38.744	52,60%
2	3.892	5,28%
3	3.346	4,54%
4	3.307	4,49%
5	2.269	3,08%
6	1.702	2,31%
7	1.907	2,59%
8	1.650	2,24%
9	2.289	3,11%
10-19	6.885	9,35%
20-29	1.677	2,28%
20-39	452	0,61%
40-49	134	0,18%
50-99	238	0,32%
100-199	105	0,14%
200-299	34	0,05%
300-399	17	0,02%
400-499	5	0,01%
>500	7	0,01%
Gesamt	73.660	100%

Tabelle 3.5: Mengenmäßige und prozentuale Verteilung der Tokenanzahl aller verwendeten Literale

verwenden fast 13% aller Statements, die sich auf Literale beziehen, Zeichenketten aus mindestens zehn Token. Sehr lange Zeichenketten mit mehr als 100 Token werden in 0,23% der Statements beziehungsweise 168 Mal verwendet.

Die Zuordnung zu den Datatype Properties zeigt, dass sehr lange Zeichenketten insbesondere aus Attributen zu Beschreibungen oder Kommentaren resultieren, hier z.B. *dtpDescription* oder *dtpControllabilityComment* oder *dtpSeverityComment*. Auffällig ist auch das häufige Vorkommen langer Literale zur Beschreibung des Teilartefaktes Item, hier *dtpAssumptions* (∅ 59,5 Token) und *dtpBoundaries* (∅ 44,6 Token), welche wie die dortige *dtpDescription* (∅ 182,6 Token) aus der initialen Systembeschreibung stammen. Ebenso auffällig sind Prädikate wie z.B. *dtpAsil*, *dtpExposure*, *dtpSeverity* oder *dtpControllability*, die jeweils aus durchschnittlich einem Token bestehen, nämlich dem jeweiligen Parameter. Identifizierer bestehen ebenfalls meist aus einem Token.

Um den semantischen Gehalt expliziter abzubilden, stellen wir in den folgenden zwei Abschnitten zwei Techniken zur semantischen Anreicherung vor. Die erste Technik widmet sich der Integration von Keywords aus semantisch komplexen Literalen in das semantische Netz.

3.2.3.2 Integration von Keywords aus semantisch komplexen Literalen

Existieren in einem semantischen Netz Literale, welche aus mehreren Wörtern oder sogar mehreren Sätze bestehen, so ist der semantische Gehalt dieser Literalknoten möglicherweise sehr komplex. Repräsentiert wird jede dieser potenziell komplexen Einheiten dennoch lediglich durch einen Knoten, dem jeweiligen Literalknoten. Um den semantischen Gehalt expliziter abzubilden, stellen wir in diesem Abschnitt eine Technik vor, mit der wir Keywords zum einen

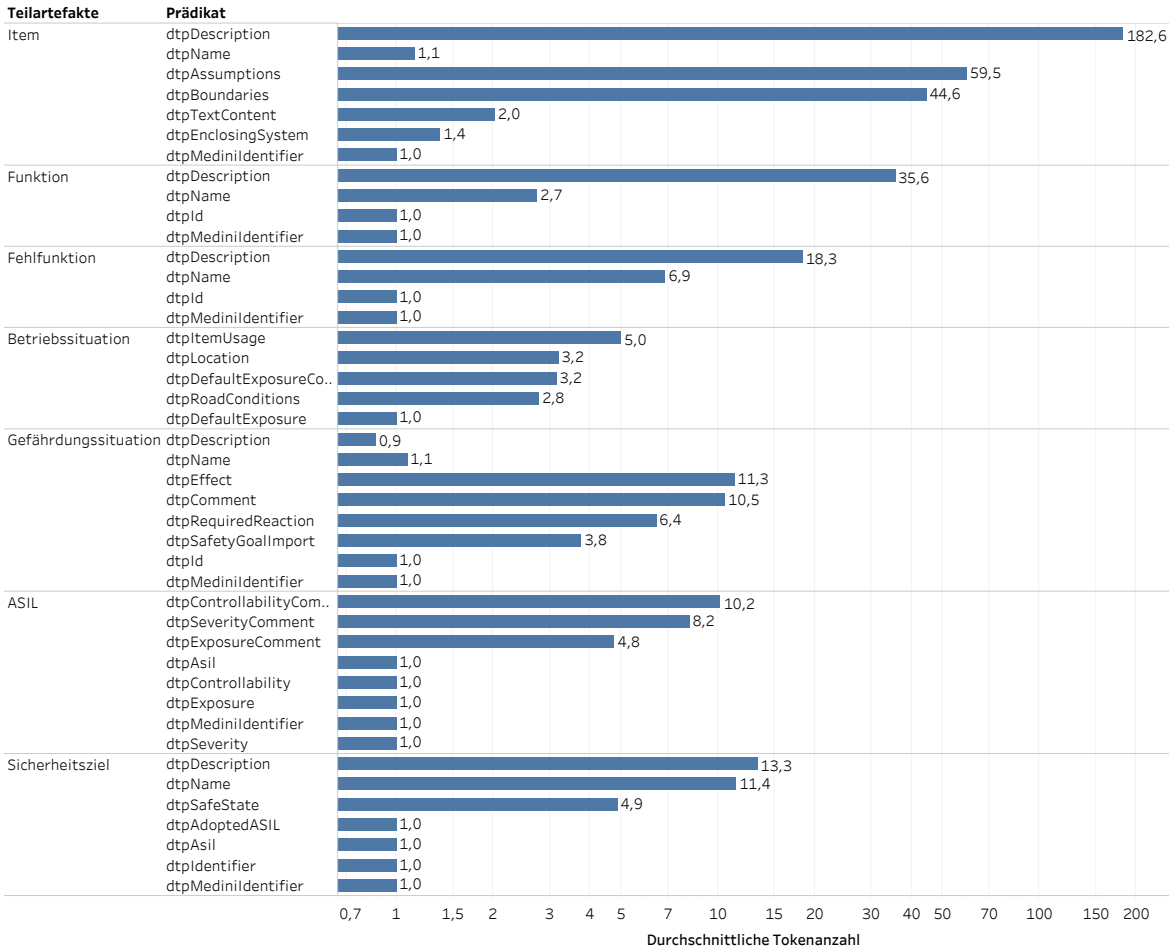


Abbildung 3.9: Durchschnittliche Tokenanzahl je Prädikat gruppiert nach zugeordneten Teilartefakten

aus semantisch komplexen Literalen extrahieren und zum anderen in das semantische Netz integrieren. Das Herzstück ist das Verfahren zur Extraktion der Keywords aus Literalen. Es ermöglicht mithilfe verschiedener Techniken des Natural Language Processing (NLP) (siehe Abschnitt 2.1.2.1) und einer Steuerung durch domänen- und anwendungsspezifische Filter die Identifikation relevanter Keywords. Durch Integration dieser automatisch extrahierten Keywords wird das semantische Netz dann im semantischen Kern angereichert.

Integration von Keywords auf Modell- und Instanzebene

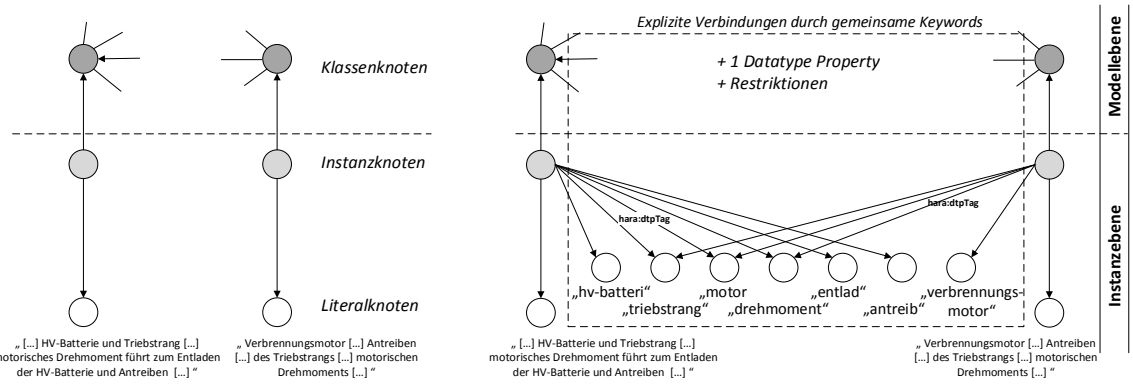
Erweiterung der Modellebene: Auf Modellebene wird ein neues Prädikat `hara:dtpTag` vom Typ `DatatypeProperty` mit den entsprechenden Restriktionen integriert. Dieses wird für alle Ressourcentypen angelegt, die potenzielle Kandidaten einer Keyword-Extraktion sind. Dadurch ist auf Instanzebene eine direkte Zuordnung identifizierter Keywords an den verbundenen Instanzknoten möglich. Das hat den Vorteil einer klaren und direkten Struktur und entspricht, wie bei der Erstellung des Basisnetzes, dem Grundgedanken eines auf Modellebene leichtgewichtigen Netzes.

Der Vorteil wird im Vergleich zur alternativen Integration einer eigenen Keyword-Klasse und der Etablierung von Instanzen für jedes konkrete Keyword deutlich. Diese sorgt insbesondere

auf Instanzebene für eine Vielzahl an zusätzlichen Knoten und Kanten, da jedes Keyword sowohl einen eigenen Instanzknoten als auch ein Literal mit dessen Textinhalt erhält. Zudem müssen in dem Fall zusätzliche Verbindungen zwischen den Keyword-Instanzen und den Instanzknoten, denen diese Keywords zugeordnet sind, geschaffen werden. Im Vergleich zur direkten Anbindung als Attribut hat dieses Vorgehen den Nachteil, dass aufgrund der indirekten Verbindung die Distanz, welche durch die Keywords erzeugt wird, größer ist als bei der direkten Anbindung durch ein Datatype Property.

Integration auf Instanzebene: Sei K die Menge extrahierter Keywords für ein Tripel (`<subject> <predicate> <object>.`), so wird auf Instanzebene für jedes k ∈ K ein Tripel (`<subject> <hara:dtpTag> <k>.`) erstellt und dem Netz hinzugefügt.

Als Beispiel sind in Abbildung 3.10a zwei Instanzknoten dargestellt, deren Literale einen langen Textinhalt haben, die aufgrund enthaltener Keywords eine thematische Nähe vermuten lassen. Trotzdem sind beide Instanzknoten auf Instanzebene voneinander isoliert. In Abbildung 3.10b hingegen sind die Keywords aus den Literalen extrahiert und explizit im Netz verankert.[29] Über gemeinsame Keywords entstehen dadurch Verbindungen zwischen den Instanzknoten.



(a) Ohne extrahierte Keywords (b) Mit extrahierten Keywords

Abbildung 3.10: Beispiel zweier Instanzknoten mit Literalen und gemeinsamen Keywords, a) isoliert trotz vermutlicher semantischer Verbindung und b) über explizite Keyword-Verbindungen miteinander verknüpft

Extraktion von Keywords

Die Abbildung 3.11 stellt unsere NLP-Pipeline zur Extraktion der Keywords dar, die für jedes Tripel durchlaufen wird, deren Objekt ein Literal ist. Die beiden Filterebenen ermöglichen eine domänenspezifische Anpassung und damit eine Steuerung der Keyword-Extraktion. Das kann eine Verfeinerung der Auswahl und eine Verbesserung der Keyword-Qualität bezogen auf die Anforderungen bewirken. Die Anpassungen der domänenspezifischen Filter obliegen dem Entwickler des IR-Systems, der im ersten Schritt der Methode zum Aufbau domänenspezifischer IR-Systeme eine Analyse der Domäne und Daten vornimmt und vorab Entscheidungen für die ansonsten automatisierte Extraktion der Keywords trifft. Der Verzicht

[29] Die hier dargestellten Keywords sind bereits auf ihre Stammform reduziert und unterscheiden sich daher leicht von ihren repräsentierten Wörtern, z.B. *entlad* für *entladen*.

führt in der Regel zu einer höheren Anzahl an Keywords und damit zu einer erhöhten Konnektivität von potenziell unspezifischen Literalinhalten. Die NLP-Schritte hingegen sind domänenunabhängig verwendbar. Die einzelnen Elemente der Pipeline werden im Folgenden kurz erläutert. Das Ergebnis der NLP-Pipeline sind eine Menge von Token, welche die Keywords für das Eingabetripel darstellen.

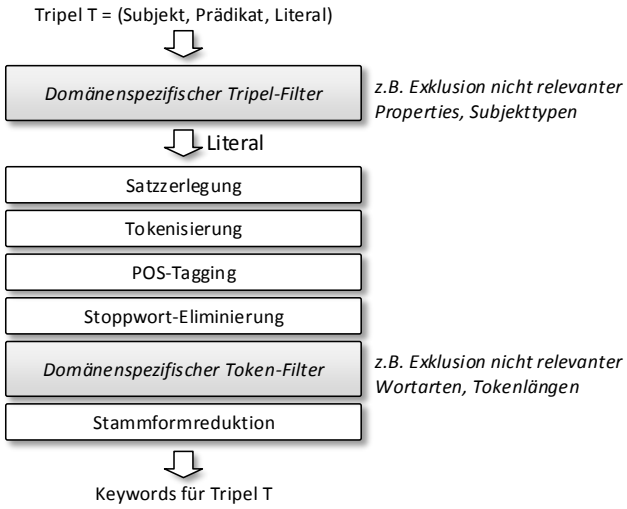


Abbildung 3.11: NLP-Pipeline zur Extraktion von Keywords

Domänenspezifischer Tripel-Filter: Jedes Tripel, welches als Objekt ein Literal enthält, ist ein potenzieller Kandidat für die Extraktion von Keywords. Der erste Filter ermöglicht eine Vorauswahl der Eingabetripel. Vorhandenes Domänenwissen und Kenntnisse über den Datensatz können zur Steuerung der Keyword-Extraktion beitragen. Damit ist eine Anpassung an die speziellen Anforderungen der Anwendung oder an bereits vorhandene Erkenntnisse zu relevanten und weniger relevanten Elementen möglich. Beispielsweise kann eine Exklusion von Tripeln mit nicht relevanten Properties sinnvoll sein. Das Gleiche gilt für Instanzen spezieller Ressourcentypen, für die Keywords keinen Mehrwert darstellen.

Satzzerlegung, Tokenisierung und PoS-Tagging: Die Literale der selektierten Tripel werden in einer ersten NLP-Phase in Sätze zerlegt und in eine Menge einzelner Token aufgeteilt. Den einzelnen Token wird durch einen PoS-Tagger die Wortart zugeordnet. Das ist erforderlich, um im späteren domänenspezifischen Token-Filter Keywords von Wortarten auszuklammern, die als nicht relevant erachtet werden.

Stoppwort-Eliminierung: In einer weiteren NLP-Phase eliminieren wir Stoppwörter. Als Grundlage dienen sprachabhängige Stoppwortlisten. Auch hier existiert die Möglichkeit zur domänenspezifischen Anpassung. Zum einen können Stoppwortlisten manuell erweitert werden, was den Nachteil hat, dass dies der subjektiven Einschätzung von Menschen zugrunde liegt. Zum anderen können Stoppwortlisten um genau jene Wörter angereichert werden, die im Text häufig vorkommen, allerdings nur geringe Domänenspezifizität besitzen. Dafür sind Analysen auf dem gesamten Textkorpus notwendig, die bewerten, wie charakteristisch ein Wort für diesen ist.

Domänenspezifischer Token-Filter: Entsprechend den Anforderungen an die Keywords lassen sich mit diesem Filter explizit bereits extrahierte Token selektieren oder exkludieren. Beispielsweise ermöglicht er das Filtern nach relevanten Wortarten. Oftmals gelten Substantive als besonders relevant, wohingegen Artikel zumeist als weniger relevant gelten. Ebenfalls kann entschieden werden, ob sich besonders kurze Token, die eine bestimmte Anzahl an Zeichen unterschreiten, überhaupt zur Keyword-Extraktion eignen.

Stammformreduktion: Anschließend reduzieren wir die verbleibenden Token auf ihre Stammform. Dadurch werden Token mit identischer Stammform zusammengefasst, was die Anzahl der Keywords mit semantisch gleicher Bedeutung reduziert. Dies erhöht zudem die indirekte Verbindung kontextuell ähnlicher Instanzknoten, was für das spätere IR von Bedeutung ist.

Extraktion und Integration von Keywords am Beispiel des G&R-Basisnetzes

G&R-spezifischer Tripel-Filter: Die Analyse der Literale in Abschnitt 3.2.3.1 hat gezeigt, dass die zumeist verwendeten Literale (52,6%) lediglich aus einem Token bestehen. Insbesondere Literale, die Identifizierer (z.B. *dtpID*) oder Parameter (z.B. *dtpSeverity*) beschreiben, bestehen zumeist lediglich aus einem Token (siehe Abbildung 3.9, Seite 75). Diese bestehen wiederum häufig maximal aus zwei Zeichen, z.B. *S1* oder *E3*. Identifizierer, Parameter und standardmäßig ausgefüllte Textfelder (z.B. *dtpDefaultExposureComment*) zeichnen sich durch sehr häufiges Auftreten wiederkehrender identischer Literale aus, da diese das Ergebnis standardmäßiger, vorbefüllter Eingaben oder zugeordneter Parameter sind. Sie differenzieren allerdings nicht sehr stark. Daher eignen sich diese drei Gruppen von Properties nicht für die Keyword-Extraktion im G&R-Basisnetz. Unter den Ressourcentypen existieren hingegen keine offensichtlichen Kandidaten für eine Exklusion durch den Tripel-Filter.

G&R-spezifischer Token-Filter: Selbst ein Literal mit lediglich einem Token besitzt trotz Kürze das Potenzial, ein Keyword zu liefern, welches auch aus anderen Literalen extrahiert wird. Es kann daher zu einer wertvollen Verbindung beitragen. Deshalb nehmen wir für den vorliegenden Datensatz keine Restriktion der Tokenanzahl vor. Sehr kurze Token hingegen enthalten oftmals Abkürzungen, Akronyme oder Parameter, welche wiederum aufgrund ihres sehr häufigen Auftretens wenig differenzieren. Wir legen daher fest, dass Keywords idealerweise aus mindestens zwei Zeichen bestehen und exkludieren kürzere extrahierte Token.

Für das G&R-Netz verwenden zudem zwei PoS-Filter, deren Auswirkungen wir in der späteren Evaluation vergleichen werden. Der *einfache* PoS-Filter ist streng und erlaubt ausschließlich Token der Wortart Substantiv (NN)[30] als Keywords. Der *erweiterte* PoS-Filter ist weniger streng und integriert zusätzlich die folgenden Wortarten: attributives Adjektiv (ADJA), adverbiales oder prädikatives Adjektiv (ADJD), Eigennamen (NE), Verben Infinitiv voll (VVINF) und finites Verb voll (VVFIN).

Tabelle 3.6 zeigt die Erweiterungen des Basisnetzes nach Keyword-Integration sowohl für die einfache als auch für die erweiterte Variante. Im Vergleich zu den 23.119 Literalknoten des Basisnetzes (siehe Tabelle 3.4, Seite 72) wird in der einfachen Keyword-Variante ein Zuwachs an Literalknoten von 32,4% erreicht. In der erweiterten Keyword-Variante vergrößert sich die Anzahl der Literalknoten sogar um 43,9%. Da die Keywords mehrfach im Netz verwendet

[30] In Klammern sind die PoS-Tags des STTS angegeben, welche üblicherweise von deutschsprachigen PoS-Taggern verwendet werden.

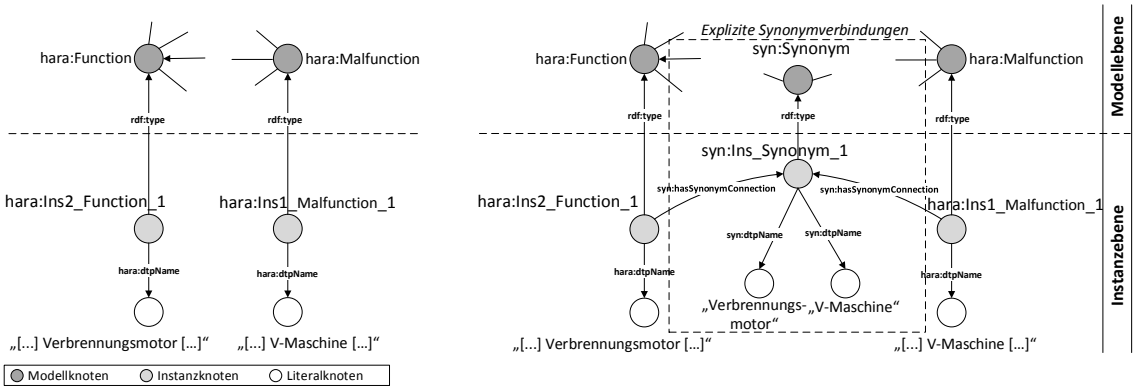
werden, steigt die Anzahl der Kanten wesentlich stärker. Schon in der einfachen Variante vergrößert sich die Anzahl der 131.445 Kanten des Basisnetzes um 63,1%, die der erweiterten Variante sogar um 102,4%, verdoppelt sich also etwa.

Elemente	**einfach**	**erweitert**
Erweiterung des Basisnetzes um Literalknoten	7.497 (32,4%)	10.156 (43,9%)
Erweiterung des Basisnetzes um Kanten durch Tripel mit Datatype Property `hara:dtpTag`	82.943 (63,1%)	134.605 (102,4%)

Tabelle 3.6: Keyword-spezifische Knoten und Tripel im erweiterten G&R-Basisnetz nach Integration der Keywords mit einfachem und erweiterten PoS-Filter

3.2.3.3 Integration von Synonymen und Akronymen

Nachdem wir im vorherigen Abschnitt die erste Technik zur semantischen Anreicherung vorgestellt haben, die eine Extraktion der Keywords aus Literalen ermöglicht, widmen wir uns in diesem Abschnitt der zweiten Technik. Im Folgenden präsentieren wir, wie vorhandenes Wissen über semantische Bedeutungsgleichheit von Wörtern, Wortgruppen oder Abkürzungen in das semantische Netz integriert und nutzbar gemacht werden kann. Ziel ist es, insbesondere im semantischen Kern des Netzes, also auf Instanzebene, semantische Verbindungen und Nähe zwischen jenen Instanzknoten herzustellen, die auch semantisch miteinander in Beziehung stehen. In Abbildung 3.12a sind beispielhaft zwei Instanzknoten dargestellt, deren Literale synonyme Wörter enthalten, hier *Verbrennungsmotor* und *V-Maschine*. Trotz der semantischen Nähe beider Instanzknoten sind sie auf Instanzebene voneinander isoliert. In Abbildung 3.12b hingegen sind die beiden Synonyme explizit im Netz verankert. Zudem sind auf Instanzebene jene Instanzknoten, die ebendiese bekannten Synonyme verwenden, miteinander verknüpft.



(a) Isolierte Instanzknoten (b) Über explizite Synonymverbindungen verknüpfte Instanzknoten

Abbildung 3.12: Beispiel zweier Instanzknoten mit Literalen und verwendeten Synonymen, a) isoliert trotz semantischer Verbindung und b) über explizite Synonymverbindungen miteinander verknüpft

Auf Instanzebene besteht unser Verfahren grundsätzlich aus zwei Integrationsschritten. Zunächst wird das aus externen Quellen verfügbare Wissen über Synonyme durch Instanziierung von Synonymgruppen explizit im semantischen Netz verankert. Mit dem zweiten Integrationsschritt wird zudem das implizite Auftreten synonymer Begriffe in den Literalen des Netzes explizit gemacht. Die konkrete Durchführung beider Schritte basiert auf konzeptuellen

Entscheidungen, die wir nach kurzer Begriffsklärung als generische Lösungen präsentieren. Diese konzeptuellen Entscheidungen zur Integration werden vom Entwickler des IR-Systems im ersten Schritt der Methode zum Aufbau domänenspezifischer IR-Systeme getroffen, so dass das semantische Netz mit seinen semantischen Anreicherungen automatisiert erstellt werden kann.

Begriffsklärung

Synonyme sind Wörter oder Wortgruppen, die bedeutungsgleich sind und daher ohne Bedeutungsänderung ausgetauscht[31] werden können. Wir verwenden den Begriff Synonym im Folgenden in einer offenen Interpretation, welche auch Akronyme[32] oder englischsprachige Formulierungen[33] einschließt, sofern diese die oben genannten Bedingungen nach Bedeutungsgleichheit und Austauschbarkeit erfüllen. Als *Synonymgruppe*[34] bezeichnen wir die Menge der synonymen Elemente, welche eine gemeinsame Bedeutung besitzen.

Integrationsschritt 1 - Instanziieren der Synonymgruppen

Die konkrete Durchführung der Schritte basiert auf konzeptuellen Entscheidungen, welche wir im Folgenden präsentieren und am Beispiel dreier Synonymgruppen veranschaulichen: *(V-Maschine, VM), (Verbrennungsmotor, V-Maschine)* und *(Momentengenerierung, Antreiben)*.

Singuläre und gruppierte Instanziierung: Jedes synonyme Element kann als individuelle Synonyminstanz in das Netz aufgenommen werden und über ein Datatype Property mit dem entsprechenden Textinhalt versehen werden. Diese bezeichnen wir als *singuläre Synonyminstanzen*. Gegenseitige Synonymie muss dann allerdings explizit markiert werden, beispielsweise durch ein Object Property, welches gegenseitig synonyme Elemente verbindet.

Eine Alternative ist die *Gruppierung gegenseitig synonymer Elemente*. Im Gegensatz zu den singulären Synonyminstanzen wird lediglich für jede Synonymgruppe eine Synonyminstanz erstellt. Der Vorteil ist, dass dadurch synonyme Elemente bereits durch die Zugehörigkeit zur gleichen Synonymgruppe auf Instanzebene direkt verbunden sind.

Additive und integrative Instanziierung: Wir unterscheiden zudem zwei Arten der Integration von gegenseitig synonymen Elementen in das semantische Netz. *Additive Instanziierung* bedeutet, dass wir gegenseitig synonyme Elemente ohne Abgleich mit bestehenden Synonymgruppen oder Synonyminstanzen in das Netz überführen. Jede Menge gegenseitig synonymer Elemente bleibt in sich isoliert. Dieser Ansatz geht davon aus, dass Synonymie

[31] Im Sprachgebrauch werden oftmals auch lediglich ähnliche, sinnverwandte Wörter als synonym bezeichnet. Hintergrund dieser strengen Interpretation des Synonymbegriffes ist, dass Bedeutungsgleichheit eine wichtige Voraussetzung für die Betrachtung von Synonymie als transitive Eigenschaft darstellt. Um unnötige Integration von Ungenauigkeit und Vagheit im semantischen Netz zu vermeiden, verzichten wir auf die exzessive Integration lediglich ähnlicher Begriffe weitestgehend.

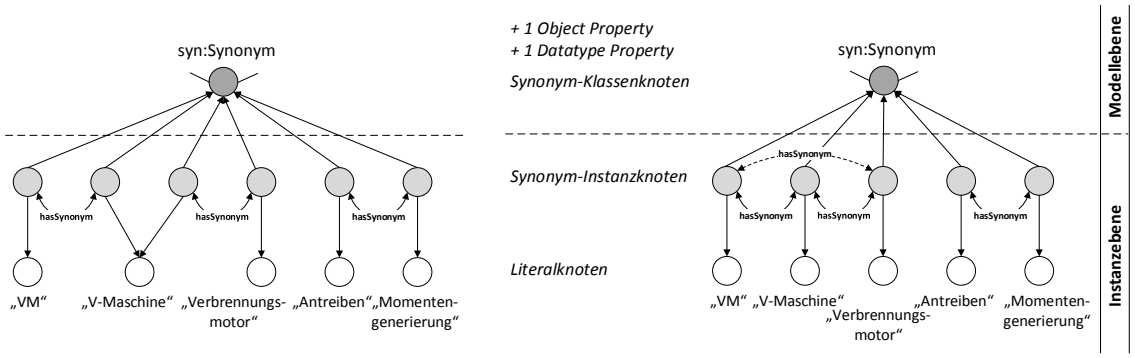
[32] Akronyme sind streng genommen keine Synonyme, da sie lediglich ein Initialwort, also eine verkürzte Variante eines Wortes oder einer Wortgruppe [CEE+10] darstellen. Mit Synonymen ist ihnen allerdings gemein, dass sie bedeutungsgleich sind und daher ohne Bedeutungsänderung ausgetauscht werden können. Daher behandeln wir sie wie Synonyme.

[33] Im Automobilbereich existieren viele Fachbegriffe, die sowohl in englischer als auch in deutscher Sprache gebräuchlich sind. Als Beispiel seien hier Motorsteuergerät, MSG, ECM und Engine Control Module genannt.

[34] Synonymgruppen entsprechen den SynSets in WordNet beziehungsweise GermaNet [HF97, HH10], welche Synonymmengen eines Konzeptes darstellen. Im Gegensatz zu SynSets werden keine zusätzlichen Beziehungen zwischen Synonymgruppen spezifiziert.

nicht zwangsläufig transitiv ist. Das bedeutet im konkreten Beispiel, dass im Netz nicht explizit sichtbar gemacht wird, dass *Verbrennungsmotor* und *V-Maschine* auch gegenseitig synonyme Elemente sind (oder sein könnten). Dadurch kann allerdings Redundanz entstehen, da auch für gleiche Synonymgruppen jeweils eine neue Instanz erstellt wird.

Integrative Instanziierung hingegen bedeutet, dass wir bei der Instanziierung gegenseitig synonymer Elemente einen Abgleich mit bereits bestehenden Synonymgruppen oder Synonyminstanzen vornehmen. Dadurch soll Redundanz vermieden werden. Daher werden bestehende singuläre Synonyminstanzen wiederverwendet und Synonymgruppen zunächst erweitert, falls diese bereits mindestens eines der Elemente einer neuen Gruppe enthalten. Lediglich im Falle tatsächlich neuer synonymer Elemente erstellen wir neue Synonyminstanzen und Synonymgruppen. Der Ansatz basiert auf der Annahme, dass Synonymie eine transitive Eigenschaft ist und diese auch explizit auf Instanzebene umgesetzt werden soll. Verstärkt wird dieser Grundgedanke durch die zusätzliche direkte Integration dieser transitiven Aussagen. Werden bei der Integration Verkettungen synonymer Elemente erkannt, so können diese optional durch Etablierung einer Synonymbeziehung zwischen diesen Elementen direkt sichtbar gemacht werden. Die Abbildungen 3.13 und 3.14 verdeutlichen die konzeptuellen Unterschiede zwischen singulärer und gruppierter, sowie additiver und integrativer Instanziierung.



(a) Additiv

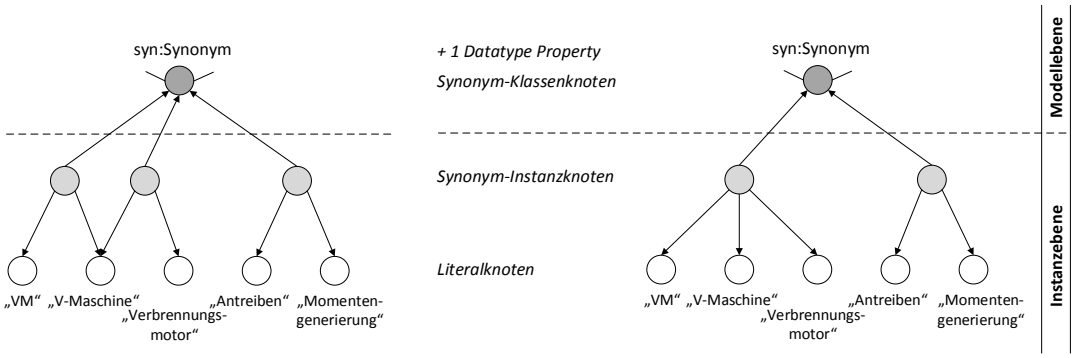
(b) Integrativ

Abbildung 3.13: Singuläre Synonyminstanzen mit a) additiver und b) integrativer Integration

In 3.13a sind für jedes synonyme Element einzelne Synonyminstanzen vorhanden, wobei die Instanzen der Synonympaare jeweils durch eine Beziehung *hasSynonym* verknüpft werden. Aufgrund der Objektidentität existieren zwei Instanzen mit identischem Literal *V-Maschine*. Auffällig ist hier der relativ große Abstand zwischen den Synonyminstanzknoten von *VM* und *Verbrennungsmotor* mit einer kürzesten Pfadlänge von vier Kanten auf Instanzebene. Der Abstand zwischen den Instanzknoten ist insofern von Bedeutung, da semantisch gleiche oder ähnliche Begriffe in semantischen Netzen idealerweise eine Nähe zueinander aufweisen. Darauf basieren auch die semantische Suchalgorithmen wie die in dieser Arbeit verwendete Aktivierungsausbreitung. In 3.13b hingegen wird auf die erneute Instanziierung des synonymen Elementes V-Maschine verzichtet, da dafür bereits ein Instanzknoten vorhanden ist. Dadurch reduziert sich die Pfadlänge zwischen *VM* und *Verbrennungsmotor* auf zwei Kanten. Wird während der Instanziierung direkte Integration der Synonymverkettung berücksichtigt, im Bild durch die gestrichelte Kante dargestellt, so reduziert sich der Abstand auf eine Kante. Integrative Instanziierung führt in der Regel zu weniger Synonyminstanzknoten mit gleichzeitig größerer Konnektivität.

In Abbildung 3.14 sind die gegebenen Synonympaare gruppiert als jeweils eine Synonyminstanz dargestellt. Das hat den Vorteil, dass aufgrund der direkten Zugehörigkeit zu einer

Synonyminstanz die explizite Beziehung durch ein Object Property entfallen kann. Während in Abbildung 3.14a die Synonympaare dennoch isoliert dargestellt werden, werden in Abbildung 3.14b vorhandene Gruppen erweitert. Dadurch reduziert sich die kleinste Pfadlänge zwischen den Instanzknoten von *VM* und *Verbrennungsmotor* bei additiver Integration auf zwei Kanten. Bei integrativer Integration beträgt diese null Kanten, da sich die beiden Literale auf denselben Instanzknoten beziehen.



(a) Additiv (b) Integrativ

Abbildung 3.14: Gruppierte Synonyminstanzen mit a) additiver und b) integrativer Integration

Integrationsschritt 2 - Verknüpfen verwendeter Synonyme

Im ersten Integrationsschritt machen wir die gegebenen Synonymgruppen im Netz explizit und verbinden die Literale der Synonyminstanzen direkt mit Instanzen anderer Typen. Aufgrund der Objektidentität passiert dies allerdings nur bei identischem Textinhalt automatisch (im Beispiel das Literal *V-Maschine*). Literale, die Elemente der Synonymgruppen lediglich als Teil der Zeichenkette enthalten, bleiben bei der direkten Integration von Synonymgruppen in das semantische Netz noch unberücksichtigt. Um auch diese Beziehungen explizit zu machen, wird in einem weiteren Schritt das Vorkommen synonymer Elemente in den bereits existierenden Literalen durch ein zusätzliches Object Property *hasSynonymConnection* markiert, welches Synonyminstanzen und Instanzen anderer Typen verbindet. Dadurch wird explizit gemacht, was nur implizit in den Literalen enthalten ist und indirekt durch die Synonymgruppen mitgemeint ist.

lokale Gültigkeit

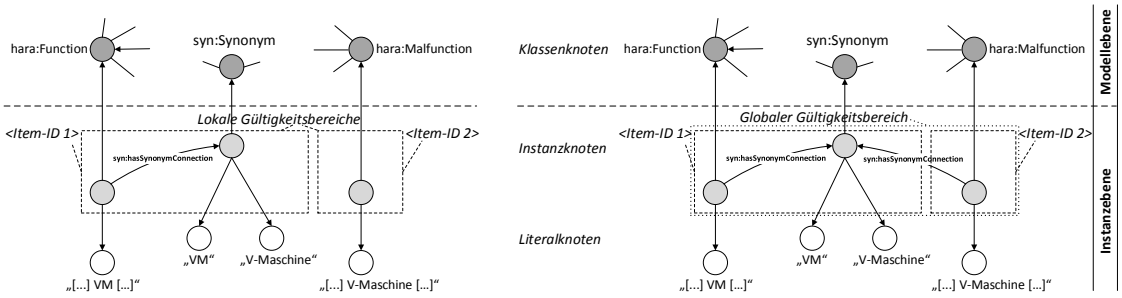
Lokale und globale Gültigkeit: Wir unterscheiden beim Verknüpfen den Gültigkeitsbereich der Synonyme. Mit *lokaler Gültigkeit* bezeichnen wir die Zuordnung von Synonymen zu einem konkreten Gültigkeitsbereich. So wird beispielsweise in der Informatik die Abkürzung VM weniger mit Verbrennungsmotoren als mit virtuellen Maschinen in Verbindung gebracht. In dem Fall ist eine Trennung des Gültigkeitsbereiches sinnvoll. Im Rahmen der G&R ist der Gültigkeitsbereich üblicherweise der Kontext eines Items. Aus praktischer Sicht entspricht dies der Integration von Synonymlisten, welche in der Praxis oftmals spezifisch für jedes G&R-Dokument angefertigt werden. Sind diese Synonyme ausschließlich lokal gültig, so wird dies sowohl bei der Instanziierung als auch beim Verknüpfen berücksichtigt. Die Einschränkung lokaler Gültigkeit hat bereits Auswirkungen auf die integrative Instanziierung, da dort beim Abgleich mit bereits enthaltenen Synonyminstanzen nur die Teilmenge der lokalen Instanzknoten, also alle Instanzknoten mit gleicher Item-ID, berücksichtigt werden. Syn-

onymverbindungen zu Instanzknoten werden dann ausschließlich in diesem Gültigkeitsbereich etabliert, also nur für Instanzknoten mit der gleichen Item-ID wie die des Synonyminstanzknotens. Dadurch ist nur die Integration in die Synonymgruppen eines Items möglich, was dazu führt, dass möglicherweise einige Synonyminstanzen nicht in vorhandene Synonymgruppen integriert, sondern stattdessen additiv hinzugefügt werden. Größeren Einfluss hat die Berücksichtigung lokaler Gültigkeit beim Verknüpfen der verwendeten Synonyme. Die Synonymverwendung wird dann ausschließlich innerhalb des Gültigkeitsbereiches eines Synonyms explizit gemacht, also nur zwischen Instanzknoten mit identischer Item-ID.

globale Gültigkeit

Globale Gültigkeit bedeutet, dass Synonyme nicht nur im Kontext eines Items gelten, sondern auch darüber hinaus nutzbar gemacht werden können. Unabhängig von potenziellen Zuordnungen von Synonymen zu konkreten Items wird auf Einschränkungen bei der Etablierung von Synonymverbindungen verzichtet. Dies entspricht einer Synonymliste, die für die gesamte Wissensbasis gelten soll. Selbstverständlich sind hier auch hybride Ansätze möglich, in denen sowohl global als auch lokal gültige Synonyme existieren.

Abbildung 3.15 verdeutlicht die Unterschiede. Das dargestellte Synonympaar ist in Abbildung 3.15a dem Item mit der *Item-ID 1* zugeordnet. Beim Verknüpfen der Synonymverwendung in den Literalen wird daher lediglich die Synonymverbindung zu dem Instanzknoten, welcher sich ebenfalls im selben Gültigkeitsbereich befindet, sichtbar gemacht. Die Synonymverwendung von *V-Maschine* wird nicht Item-übergreifend etabliert. In Abbildung 3.15b hingegen besitzt das Synonym globale Gültigkeit, unabhängig von einer potenziellen Zuweisung zu einer konkreten Item-ID. Alle Instanzknoten, welche synonyme Elemente verwenden, werden daher mit expliziten Synonymverbindungen zu den jeweiligen Synonyminstanzen verbunden.



(a) Lokal (b) Global

Abbildung 3.15: Verknüpfung verwendeter synonymer Elemente mit a) lokalen Gültigkeitsbereichen und b) globalem Gültigkeitsbereich

Die Entscheidung für globale oder lokale Gültigkeit ist abhängig von der Verfügbarkeit der Synonyme und der jeweiligen Anwendung. Sind Synonymgruppen global verfügbar, so stellt sich diese Frage nicht. Existieren allerdings lokale Synonymgruppen, so ist zu entscheiden, ob diese lokal oder global betrachtet werden sollen. Der Nachteil der Beschränkung auf lokale Gültigkeit ist der Verlust vieler potenzieller Verbindungen zugunsten der Sicherheit, dass die etablierten Synonymverbindungen gültig und wahr sind. Betrachtet man sie als global, so kann es sein, dass wesentlich mehr Synonymverbindungen gefunden und explizit gemacht werden können. Allerdings können sich unter diesen auch welche befinden, die im Kontext eines bestimmten G&R-Dokumentes ihre Gültigkeit verloren haben. Es ist eine Designentscheidung, die sorgfältige Abwägung erfordert. Bei geringer Domänenspezifizität besteht die Gefahr der Überflutung der Netze mit Elementen des Grundwortschatzes. Die explizite Integration nicht domänenspezifischer Wörter kann daher auch zur Ablenkung

von tatsächlich relevanten Verknüpfungen führen. Fehlende Domänenspezifizität ist der Hauptgrund für unseren Verzicht auf externe Synonym-Bibliotheken, beispielsweise den sogenannten *SynSets* aus WordNet, einem lexikalisch-semantischen Netz in englischer Sprache sowie dessen erweiterte deutschsprachige Variante GermaNet [HF97, HH10]. Wir vermuten hingegen, dass Synonyme und Akronyme bei hoher Domänenspezifizität innerhalb einer Domäne bei globaler Betrachtung oftmals mehr Nutzen als Schaden verursachen. Zudem gilt insbesondere die Verknüpfung zwischen unterschiedlichen G&Rs als vielversprechend für den untersuchten Anwendungsfall dieser Arbeit. Lokale Gültigkeit birgt die Gefahr, relevantes Wissen über Synonymverbindungen zugunsten der Bewahrung systemweiter Synonymlisten nicht zu integrieren und Potenziale, welche durch die Verbindungen entstehen, nicht zu nutzen.

Integration von Synonymen und Akronymen in das G&R-Basisnetz

Wir haben verschiedene Ansätze zur Integration von Synonymen und Akronymen in semantische Netze vorgestellt und diskutiert. Für die explizite Integration in das G&R-Basisnetz treffen wir folgende Designentscheidungen, die wir kurz begründen.

1. Gruppierte und integrative Instanziierung
2. Globale Gültigkeit der Synonymgruppen

Wir entscheiden uns für die gruppierte und integrative Instanziierung, um eine möglichst kompakte Integration der Synonyme zu erreichen. Insbesondere die gruppierte integrative Instanziierung bietet das größte Potenzial zur Zusammenfassung gleichbedeutender Begriffe unter denselben Instanzknoten. Davon profitieren wir bei der geplanten Anwendung der Aktivierungsausbreitung als semantischen Suchalgorithmus auf dem semantischen Netz.

In der Praxis werden Akronyme und Synonyme bei der Durchführung einer G&R oftmals für jedes System einzeln gesammelt, sind also bei der Spezifikation lokal gemeint. Diese Listen sind anfällig für Unvollständigkeit oder werden für einzelne Projekte gar nicht erst angelegt. Eine Auswertung der verfügbaren Listen und ein Abgleich mit FuSi-Experten zeigten allerdings starke Überschneidungen und eine sehr hohe Domänenspezifizität. Daher betrachten wir diese Listen auch über die Grenzen einzelner G&R-Projekte hinweg als übertragbar und interpretieren die eigentlich lokalen Synonymgruppen als global gültig.

Insgesamt erweitern 270 Synonymgruppen mit 652 enthaltenen synonymen Elementen das Basisnetz. Deren Verwendung haben wir über die Beziehung `hasSynonymConnection` mit 9.760 Instanzknoten explizit gemacht. Die Anzahl der zusätzlichen Tripel des durch Synonyme angereicherten Basisnetzes ist in Tabelle 3.7 aufgelistet. Im Vergleich zu den 131.445 Kanten beziehungsweise Tripeln des Basisnetzes (siehe Tabelle 3.4, Seite 72) wird durch die synonymspezifischen Tripel im erweiterten G&R-Basisnetz ein Zuwachs von 8,13% erreicht.

Elemente	**Anzahl (%)**
Erweiterung des Basisnetzes um Tripel, davon	10.689 (8,13%)
- Datatype Property `syn:dtpName`	652
- Object Property `syn:hasSynonymConnection`	9.760
- Instanztypbeziehungen	270
- Beziehungen auf Modellebene (z.B. Subklassenbeziehungen, Restriktionen)	7

Tabelle 3.7: Synonymspezifische Tripel im erweiterten G&R-Basisnetz nach Integration der Synonyme

3.2.4 Semantische Reduktionen

Die semantischen Anreicherungen setzen darauf, die semantische Komplexität im Netz explizit zu machen. Mit der Reduktion des Netzes streben wir die Konzentration auf semantische Kernaussagen an, indem wir Tripel entfernen, deren domänenspezifischer und semantischer Beitrag gering ist. Hintergrund für potenzielle Reduktionen ist die später folgende Anwendung der graphbasierten Suchalgorithmen im IR-Schritt. Diese nutzen assoziative Verbindungen im Netz, um die gegenseitige Relevanz von Knoten zu bewerten. Die Grundidee basiert darauf, dass Verbindungen zwischen Knoten semantische Nähe repräsentieren. Um diese Grundidee beim Aufbau der Netze zu unterstützen, präsentieren wir Techniken zur Reduktion der Netze sowohl auf Modellebene als auch auf Instanzebene. Dies hat das Ziel, den Fokus auf die semantisch relevantesten Zonen im Netz zu lenken und semantisch irrelevante Umwege auf den Pfaden im Netz zu verringern.

3.2.4.1 Potenziale zur Reduktion - Identifikation von Tripeln mit geringer domänenspezifischer Bedeutsamkeit

Das domänenspezifische Wissen ist vorwiegend auf Instanzebene verortet, während die Modellebene das terminologische Wissen und daher Aussagen über die Struktur des Netzes enthält (siehe Abschnitt 2.2.2). Wir beschreiben im Folgenden für beide Ebenen Reduktionspotenziale, die auf eine Aufwertung der Netze abzielen, indem der Fokus auf die semantisch relevantesten Zonen im Netz gelenkt wird.

Modellspezifische Aussagen auf Modellebene

In dieser Arbeit verwenden wir semantische Suchalgorithmen, welche auf Aktivierungsausbreitung basieren. Da diese nicht auf die logische Auswertung der Aussagen auf Modellebene setzen, sondern stattdessen auf die assoziativen Verbindungen, können Aussagen auf Modellebene als semantisch weniger relevant als domänenspezifische Aussagen auf Instanzebene betrachtet werden. In Abbildung 3.8 (Seite 70) haben wir bereits eine Differenzierung bezüglich semantischer Relevanz vorgenommen und zwei Zonen im Netz als semantisch besonders relevant markiert. Der semantische Kern umfasst demnach die Instanzebene, da hier das domänenspezifische Wissen instanziiert ist. Der erweiterte semantische Kern enthält zudem die Aussagen zu Typbeziehungen zwischen domänenspezifischen Klassen und ihren Instanzen sowie die Subklassenbeziehungen der domänenspezifischen Klassen. Diese Zonen sind Grundlage für potenzielle Reduktionen auf Modellebene.

Bei der Reduktion auf den semantischen Kern schließen wir alle Aussagen auf Modellebene aus. Damit erzeugen wir ein streng domänenspezifisches Netz, welches stark auf die Instanzknoten und deren Beziehungen zu anderen Instanzknoten und zu literalen Knoten fokussiert. Im Gegensatz dazu verbleiben im semantischen Netz bei der Reduktion auf den erweiterten semantischen Kern alle Tripel zu Typbeziehungen zwischen domänenspezifischen Klassen und ihren Instanzen sowie die Subklassenbeziehungen der domänenspezifischen Klassen. Das kann den Vorteil haben, dass die Klassenknoten als Verbindungsknoten zu allen Instanzen der jeweiligen Klasse agieren. Dadurch kann beispielsweise der Wirkungskreis bei der Aktivierungsausbreitung schneller ausgeweitet werden.

Domänenspezifische Aussagen auf Instanzebene

Auf Instanzebene ist das Ziel die Entfernung aller Tripel, die zwar in der domänenspezifischen Zone der Instanzebene verortet sind, aber dennoch keinen semantisch wertvollen Beitrag leisten. Ohne zusätzliches Domänenwissen exkludieren wir zunächst alle Tripel, deren Objekt das Literal mit leerer Zeichenkette ist. Im Basisnetz existiert aufgrund der Objektidentität genau ein Literal mit leerer Zeichenkette, welches allerdings als Objekt von 5.000 Tripeln (siehe Tabelle 3.5, Seite 74) verwendet wird. Die häufige Verwendung suggeriert eine große Bedeutsamkeit, tatsächlich ist sie allerdings gering. Weitere Reduktionen bedürfen einer domänenspezifischen Analyse und werden vom Entwickler des IR-Systems festgelegt, der im ersten Schritt der Methode zum Aufbau des domänenspezifischen IR-Systems eine Vorab-Analyse der Domäne und Daten vornimmt. Die domänenspezifischen Reduktionen können dann automatisiert umgesetzt werden.

3.2.4.2 Reduktionen im G&R-Basisnetz

Für die Reduktion des G&R-Basisnetzes verwenden wir zwei Reduktionsstufen. Die Qualität des unreduzierten Netzes und der beiden reduzierten Netze werden wir in der Evaluation dieser Arbeit miteinander vergleichen. Die beiden Reduktionsstufen unterscheiden sich in ihren Reduktionen auf Modellebene, verwenden allerdings die gleichen Reduktionen auf Instanzebene.

Reduktionen auf Instanzebene: Die Reduktionen auf Instanzebene entstammen einer Untersuchung des G&R-Datensatzes, bei der wir domänenspezifische Tripel identifiziert haben, die wir als semantisch nicht bedeutsam einstufen. In dem G&R-Basisnetz existieren beispielsweise einige Properties, deren Verwendung keinen semantischen Mehrwert für das semantische Netz ergibt. Dazu gehören vor allem die Properties verschiedener Identifizierer (`dtpId, dtpIdentifier, dtpMediniIdentifier`) sowie das Property `dtpSafetyGoalImport`, welches ebenfalls lediglich Abkürzungen für spezielle Identifizierer für Sicherheitsziele enthält. Wir beschränken uns auf die Exklusion von Tripeln, über deren geringe semantische Bedeutsamkeit im Netz keine Zweifel bestehen. Wir verzichten auf die Exklusion von Tripeln, deren semantische Bedeutsamkeit fraglich ist, da zu wenig über die Auswirkungen der Modifikationen bekannt ist.

Starke Reduktion auf den semantischen Kern: Bei der Reduktion auf den semantischen Kern werden zunächst alle Tripel der Modellebene entfernt, so dass ausschließlich Tripel der Instanzebene, dem semantischen Kern, im Netz verbleiben. Zudem entfernen wir alle domänenspezifischen Tripel, die wir als semantisch nicht bedeutsam einstuften. Diese Ausprägung wird im Folgenden auch als *starke Reduktion* bezeichnet.

Einfache Reduktion auf den erweiterten semantischen Kern: Bei der Reduktion auf den erweiterten semantischen Kern verbleiben im semantischen Netz, im Gegensatz zur starken Reduktion, alle Tripel zu Typbeziehungen zwischen domänenspezifischen Klassen und ihren Instanzen sowie die Subklassenbeziehungen der domänenspezifischen Klassen. Zudem entfernen wir ebenfalls alle domänenspezifischen Tripel, die wir als semantisch nicht bedeutsam einstuften. Diese Ausprägung wird im Folgenden auch als *einfache Reduktion* bezeichnet.

3.2.5 Anfragerepräsentation

In Abschnitt 3.1.2 und insbesondere in Abbildung 3.3 auf Seite 57 wird der in dieser Arbeit untersuchte Anwendungsfall für das IR-System vorgestellt, welcher das Informationsbedürfnis der folgenden Form betrachtet:

Welche Elemente vom Typ <X> stehen in semantischem Zusammenhang zu Element <Y>?

Diese Frage enthält zwei grundlegende Parameter. Das Element <Y> stellt den Ausgangspunkt der semantischen Suche dar und kann in einer allgemeinerer Form aus mehreren Elementen bestehen. Der Typ <X> beschreibt das Suchziel und ist in einer allgemeineren Form nicht auf einen Typ reduziert, sondern kann aus mehreren Typen bestehen. Beide Parameter charakterisieren das Informationsbedürfnis und werden im Folgenden kurz beschrieben.

Ausgangspunkt: Das Element <Y> ist der Ausgangspunkt für die Relevanzbewertung und entspricht in der Anfrage einem oder mehreren Knoten, den sogenannten *Startknoten*, welche diesen Ausgangspunkt repräsentieren. Der *direkte Ausgangspunkt* wird im einfachsten Fall durch genau einen Knoten repräsentiert, nämlich jenen Instanzknoten, für welchen das Informationsbedürfnis formuliert wird. Wir können allerdings auch den Kontext, in dem sich der Ausgangspunkt befindet, hinzuziehen. Dadurch kann sich die Repräsentation des Ausgangspunktes auf Knoten ausweiten, welche im Kontext des Informationsbedürfnisses eine gewisse Bedeutung besitzen. Am Beispiel des IR-Systems für G&Rs kann z.B. auch der Instanzknoten, welcher das dazugehörige Item repräsentiert, den Ausgangspunkt erweitern. Ebenso ist denkbar, dass sich beispielsweise die Navigationshistorie im Entwicklungswerkzeug oder Keywords in der Repräsentation des Ausgangspunktes widerspiegeln. Die Entscheidung für eine automatisierte, kontextbezogene Ausweitung der Ausgangspunkte wird vom Entwickler des IR-Systems nach Analyse der Domäne und Daten getroffen. Die Nutzer erwarten vom IR-System automatisierte Verfahren zur Befriedigung ihrer Informationsbedürfnisse, die während der Durchführung von G&Rs entstehen. Ein Mehraufwand für die manuelle Formulierung komplexer Anfragen ist nicht akzeptabel.

Als Beispiel dient erneut das folgende Informationsbedürfnis. *Welche <Fehlfunktionen> (X) stehen in semantischem Zusammenhang zur <Funktion Momentengenerierung des Systems Neues Antriebssystem> (Y)?* Der direkte Ausgangspunkt ist hier der Instanzknoten, welcher die *Funktion Momentengenerierung des Systems Neues Antriebssystem* repräsentiert. Erweitert man den Ausgangspunkt um den Kontext, so kann beispielsweise der Instanzknoten, der das entsprechende System *Neues Antriebssystem* repräsentiert, zur Repräsentation des Ausgangspunktes hinzugezogen werden.

Suchziel: Der Typ <X> charakterisiert, wonach gesucht werden soll, hier beschrieben durch einen oder mehrere Ressourcentypen der gesuchten Elemente. Wir nennen diesen Parameter daher auch das *Suchziel*. Mit dem Suchziel lässt sich die Ergebnisliste genauer spezifizieren und damit je nach Bedürfnis eingrenzen. Diese Eingrenzung ist zwar prinzipiell nicht notwendig, da die gegenseitige Relevanz zwischen jedem beliebigen Knotenpaar, unabhängig von Ressourcentypen, bestimmt werden kann. Eine konkrete Fragestellung unterstützt allerdings auch eine möglichst konkrete Antwort des IR-Systems. Im Beispiel ist als Suchziel der Typ Fehlfunktion angegeben. Erweitern wir das Suchziel beispielsweise auf alle verfügbaren Ressourcentypen, so wird die Retrieval-Antwort nicht auf Fehlfunktionen beschränkt sein und stattdessen sämtliche Zusammenhänge im Netz verfügbar machen. Für den Nutzer erfolgt die Auswahl des Suchziels idealerweise direkt im Entwicklungswerkzeug, beispielsweise über das Kontextmenü, wie in Abbildung 3.3 (Seite 57) des Anwendungsfalls dargestellt.

Definition Anfrage: In der späteren Aufbereitung der Ergebnisse des IR-Systems verwenden wir eine suchzielspezifische Filterung, welche die definierten Suchziele verwendet, um die Antworten des IR-Systems an das Informationsbedürfnis des Nutzers anzupassen. Daher folgt eine formale Definition der Anfrage, die in Abschnitt 3.4.1 verwendet wird.

Sei T die nichtleere Menge aller domänenspezifischen Ressourcentypen und N_I die nichtleere Menge von Instanzknoten eines semantischen Netzes, dann ist eine Anfrage Q ein 2-Tupel (N_Q, T_Q), wobei N_Q den Ausgangspunkt darstellt und T_Q das Suchziel. Ferner gilt, dass $N_Q \neq \emptyset, T_Q \neq \emptyset, N_Q \subseteq N_I, T_Q \subseteq T$.

3.2.6 Zusammenfassung und Beitrag

Der Abschnitt 3.2 widmete sich der Komponente zur Repräsentation von Wissen und Anfragen im IR-System. Wir haben gezeigt, wie wir domänenspezifische semantische Netze automatisiert aufbauen können, indem wir uns am Datenmodell des Entwicklungswerkzeugs orientieren, aus dem das Wissen stammt. Dadurch konnten wir den Aufwand beim Aufbau der Netze erheblich reduzieren und überwinden damit eine der großen Hürden bei der Verwendung semantischer Netze. Wir haben zusätzliche Techniken zur semantischen Anreicherung der Netze vorgestellt, die eine Extraktion von Keywords aus Literalen mit komplexen Textinhalten erlauben und den Umgang mit bekannten Synonymen und Akronymen ermöglichen. Dadurch erreichen wir eine Aufwertung der Netze. Wir stellten außerdem Techniken zur Reduktion der Netze vor. Diese ermöglichen es, den Fokus auf die semantisch relevantesten Zonen im Netz zu lenken, indem sie zum einen auf den (erweiterten) semantischen Kern der Netze reduzieren und zum anderen Möglichkeiten zur domänenspezifischen Exklusion nicht relevanter Informationen im Netz bieten. Wir haben die Anwendung der Techniken gezeigt, indem wir das Basisnetz am Beispiel eines industriellen Datensatzes mit 155 G&Rs automatisiert aufgebaut haben. Außerdem entstanden Varianten semantischer Netze unter Anwendung aller vorgestellten Techniken zur Anreicherung und Reduktion, deren Charakteristika wir untersuchten und darstellten. Zusätzlich zur graphbasierten Repräsentation des Wissens, präsentierten wir ein Konzept zur graphbasierten Repräsentation strukturierter Anfragen. Diese enthalten den Ausgangspunkt der semantischen Suche und das Suchziel. Die strukturierte Formulierung des Suchziels unterstützt uns in der späteren Ausgabe der Ergebnisse bei der typbasierten Eingrenzung der Ergebnisliste.

Mit dem Abschnitt 3.2 unterstützen wir den zweiten Schritt der in Abschnitt 3.1.2.4 vorgestellten Methode zum Aufbau domänenspezifischer IR-Systeme. In diesem zweiten Schritt erstellt der Entwickler eines IR-Systems die semantischen Netze anhand der domänenspezifischen Daten und erweitert beziehungsweise reduziert diese bei Bedarf. Mit den vorgestellten Techniken stehen dem Entwickler eines IR-Systems generische Lösungen zum automatisierten, datenmodellbasierten Aufbau der semantischen Netze zur Verfügung. Seine Aufgabe besteht vorrangig darin, sich für das Ausmaß an semantischer Anreicherung und Reduktion zu entscheiden. Diese Entscheidungen werden idealerweise an die Domäne und Anwendung angepasst. In der Evaluation unseres IR-Systems für G&Rs in Kapitel 4 untersuchen wir 18 Netzvarianten, die sich aus den hier vorgestellten semantischen Anreicherungen und Reduktionen des Basisnetzes ergeben. Wir werden beispielsweise zeigen, dass die semantischen Anreicherungen und Reduktionen der Netze essenziell für qualitativ gute Ergebnisse des IR-Systems für G&Rs sind.

Im nächsten Abschnitt zeigen wir, wie wir das Retrievalverfahren Aktivierungsausbreitung auf den semantischen Netzen anwenden. Damit unterstützen wir den dritten Schritt der Methode zum Aufbau domänenspezifischer IR-Systeme.

3.3 Aktivierungsausbreitung in semantischen Netzen

In diesem Abschnitt widmen wir uns dem assoziativen Retrievalverfahren Aktivierungsausbreitung, welches wir im IR-System für G&Rs verwenden. In Abschnitt 2.3 haben wir die Grundprinzipien der Aktivierungsausbreitung im IR vorgestellt. Algorithmen, die auf Aktivierungsausbreitung basieren, folgen diesen Grundprinzipien, variieren dennoch häufig im Detail, zumeist abhängig von der konkreten Anwendung. Um eine reproduzierbare und konfigurierbare Verwendung der Aktivierungsausbreitung zu gewährleisten, konzeptualisieren und formalisieren wir in diesem Abschnitt etablierte Prinzipien und neuartige Techniken. Dafür werden wir zunächst die Aktivierungsverteilungsfunktionen beschreiben und verschiedene Modi für Beschränkungen etablieren, die eine Konfigurierbarkeit gewährleisten. Anschließend stellen wir den Algorithmus für die Aktivierungsausbreitung vor, der das Zusammenspiel der Aktivierungsverteilungsfunktionen formal beschreibt. Wir erweitern die Aktivierungsausbreitung zudem um sogenannte Ausbreitungsmodi, welche eine zusätzliche Steuerung des Ausbreitungsverhaltens im Graphen erlauben. Diese sind grundsätzlich generisch und bilden die Grundlage für die Verwendung der Aktivierungsausbreitung im IR-System für G&Rs.

3.3.1 Aktivierungsverteilungsfunktionen

Aktivierungsausbreitung wird gesteuert von Aktivierungsverteilungsfunktionen, die wir in diesem Abschnitt für eine reproduzierbare und konfigurierbare Verwendung formal beschreiben. Dafür definieren wir zunächst das semantische Netz und führen zur Unterstützung des Verständnisses einige Vorbetrachtungen durch. Anschließend folgt die formale Definition der Aktivierungsverteilungsfunktionen.

Definition semantisches Netz: Sei L eine Menge von Labels, dann ist ein semantisches Netz G als zusammenhängender, gerichteter und gelabelter Multigraph[35] durch ein 6-Tupel $(\mathrm{N}, \mathrm{E}, \mathrm{s}, \mathrm{t}, \mathrm{l}, \omega)$ beschrieben. Hierbei sind N eine nichtleere Menge von Knoten, E eine nichtleere Menge von Kanten, $\mathrm{s} : \mathrm{E} \to \mathrm{N}$ eine Quellknoten-Abbildung für Kanten, $\mathrm{t} : \mathrm{E} \to \mathrm{N}$ eine Zielknoten-Abbildung für Kanten, $\mathrm{l} : \mathrm{N} \cup \mathrm{E} \to \mathrm{L}$ die Labelfunktion und $\omega : \mathrm{E} \to \mathbb{R}$ eine Kantengewichtsfunktion.

Funktionsdeklarationen: Das semantische Netz ist Grundlage für die Aktivierungsverteilungsfunktionen, die in Pulsen $\mathrm{p} \in \mathbb{N}$ auf dem Netz ausgeführt werden. Puls $\mathrm{p} = 0$ beschreibt den initialen Aktivierungszustand aller Knoten und damit den Ausgangspunkt des Algorithmus. Für $\mathrm{n} \in \mathrm{N}$, $\mathrm{e} \in \mathrm{E}$ und Puls $\mathrm{p} > 0$ beschreiben wir folgende Aktivierungsverteilungsfunktionen:

- die Output-Funktion $\mathrm{out} : \mathrm{N} \times \mathrm{E} \times \mathbb{R} \times \mathbb{N}_{>0} \to \mathbb{R}$ bestimmt die Output-Aktivierung $\mathrm{o}_{\mathrm{n,e}}^{(\mathrm{p})}$ für Knoten n über Kante e im Puls p,
- die Input-Funktion $\mathrm{in} : \mathrm{N} \times \mathrm{N} \times \mathrm{E} \times \mathbb{R} \to \mathbb{R}$ bestimmt die Input-Aktivierung $\mathrm{i}_{\mathrm{n,e}}^{(\mathrm{p})}$ für Knoten n über Kante e im Puls p,
- die Input-Funktion $\mathrm{in} : \mathrm{N} \times \mathbb{N}_{>0} \to \mathbb{R}$ bestimmt die Input-Aktivierung $\mathrm{i}_{\mathrm{n}}^{(\mathrm{p})}$ für Knoten n über alle Kanten im Puls p und
- die Aktivierungsfunktion $\mathrm{act} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ bestimmt den Aktivierungszustand $\mathrm{a}_{\mathrm{n}}^{(\mathrm{p})}$ für Knoten n im Puls p.

[35] Die formalen Annahmen sind angelehnt an [BBK+09]. Zusätzlich berücksichtigen wir Multigraphen.

Die Funktionsdeklarationen enthalten die vollständige Liste aller Parameter einer Funktion. Da wir die Aktivierungsverteilungsfunktionen teilweise redefinieren, verwendet nicht jede Definition einer Funktion alle in der Funktionsdeklaration gelisteten Parameter. Zur besseren Lesbarkeit werden im Folgenden nicht verwendete Parameter in den Funktionsdefinitionen weggelassen.

Hilfsdefinitionen:

Die Knoten $n, m \in N$ und die sie verbindende Kante $e \in E$ bilden ein *Tripel*.

$$\text{trip}(n, m, e) \Leftrightarrow (s(e) = n \wedge t(e) = m) \vee (s(e) = m \wedge t(e) = n) \tag{3.1}$$

Ein Knoten $n \in N$ und eine Kante $e \in E$ sind *inzident*, wenn n und e miteinander verbunden sind.

$$\text{inz}(n, e) \Leftrightarrow s(e) = n \vee t(e) = n \tag{3.2}$$

Die Knoten $n, m \in N$ sind *adjazent*, wenn sie durch eine Kante $e \in E$ miteinander verbunden sind. Adjazente Knoten werden auch als *Nachbarknoten* bezeichnet.

$$\text{adj}(n, m) \Leftrightarrow \exists e \in E.\text{trip}(n, m, e) \tag{3.3}$$

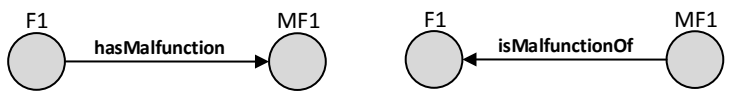
In dieser Arbeit bezeichnen wir einen Knoten als *aktiviert*, wenn dieser positive Aktivierungswerte besitzt. Daher gilt für $n \in N$, $p \geq 0$, dass

$$\text{akt}(n, p) \Leftrightarrow a_n^{(p)} > 0. \tag{3.4}$$

Nachdem ein Knoten aktiviert wurde, kann die Stärke der Aktivierung in Folgepulsen zu- oder abnehmen. Daher kann ein Knoten im Rahmen der hier vorgestellten Methode auch deaktiviert werden.

Ausbreitungsrichtungen: Semantische Netze sind gerichtete Graphen, deren Kanten eine gerichtete Bedeutung besitzen. Ob die Kantenrichtung bei der Verteilung von Aktivierung zwischen Knoten berücksichtigt wird, also Aktivierung nur in Kantenrichtung weiterverteilt wird, sollte anwendungsspezifisch überlegt werden.

Das Umkehren einer Kante verändert zwar die Leserichtung der Aussage eines Tripels und damit die semantische Bedeutung, aber nicht zwangsläufig die semantische Bedeutsamkeit für die sie verbindenden Knoten. Als Beispiel ist in Abbildung 3.16 ein Tripel abgebildet. Dessen Aussage `F1 hasMalfunction MF1` kann durch Umkehren der Kante auch als `MF1 isMalfunctionOf F1` gelesen werden. Dabei wendet sich zwar die Aussage des Tripels, die gegenseitige Bedeutsamkeit beider Knoten entsteht allerdings bereits durch die Existenz der sie verbindenden Kante und bleibt daher trotz Umkehrung bestehen.



(a) `F1` hat Fehlfunktion `MF1` (b) `MF1` ist Fehlfunktion von `F1`

Abbildung 3.16: Beispieltripel mit semantisch äquivalenten Aussagen trotz unterschiedlicher Leserichtung

Spielt die Leserichtung der kantenbeschreibenden Eigenschaft also eine untergeordnete Rolle, sollten im Falle einer möglichen Aktivierungsverteilung zwischen den Knoten beide Richtungen gleichberechtigt behandelt werden. Spielt hingegen die sequentielle Abfolge von Knoten in Form gerichteter Pfade benachbarter Knoten eine besondere Rolle im Anwendungsfall, kann eine Aktivierungsausbreitung entlang gerichteter Kanten sinnvoll sein.

In dieser Arbeit gehen wir davon aus, dass die Bedeutsamkeit einer Kante für die Knoten im Aktivierungsprozess nicht von der Kantenrichtung sondern lediglich von ihrer Existenz abhängt. Aus diesem Grund werden alle Aktivierungsverteilungsfunktionen ohne Berücksichtigung der Kantenrichtung vorgestellt.

Output-Aktivierung

Die Output-Aktivierung $o_{n,e}^{(p)}$ für Knoten $n \in N$ über Kante $e \in E$ wird in jedem Puls $p > 0$ bestimmt durch die Output-Funktion

$$o_{n,e}^{(p)} = \mathrm{out}(n, e, a_n^{(p-1)}), \tag{3.5}$$

wobei

$$\mathrm{out}(n, e, a) := \begin{cases} a & \text{if } \mathrm{inz}(n, e), \\ 0 & \text{else.} \end{cases} \tag{3.6}$$

Input-Aktivierung

Die Input-Aktivierung $i_{n,e}^{(p)}$ für Knoten $n, m \in N$ über Kante $e \in E$ mit Kantengewicht $\omega(e)$ wird in jedem Puls $p > 0$ bestimmt durch die Input-Funktion

$$i_{n,e}^{(p)} = \mathrm{in}(n, m, e, o_{m,e}^{(p)}), \tag{3.7}$$

wobei

$$\mathrm{in}(n, m, e, o) := \begin{cases} o \cdot \omega(e) & \text{if } \mathrm{trip}(n, m, e), \\ 0 & \text{else.} \end{cases} \tag{3.8}$$

Die Input-Aktivierung für Knoten n, gewonnen über alle Kanten, wird konsolidiert durch die Input-Funktion

$$i_n^{(p)} = \mathrm{in}(n, p), \tag{3.9}$$

wobei

$$\mathrm{in}(n, p) := \sum_{e \in E} i_{n,e}^{(p)}. \tag{3.10}$$

Aktivierungsfunktion

Der Aktivierungszustand $a_n^{(p)}$ beschreibt den Aktivierungswert für Knoten $n \in N$ in Puls $p \geq 0$, wobei $a_n^{(0)}$ den initialen Aktivierungszustand von Knoten n darstellt. Für $p > 0$ wird der Aktivierungszustand bestimmt durch die Aktivierungsfunktion

$$a_n^{(p)} = \mathrm{act}(i_n^{(p)}, a_n^{(p-1)}), \tag{3.11}$$

welche, abhängig von einer anwendungsspezifisch gewählten Transferfunktion $\zeta : \mathbb{R} \to \mathbb{R}$, einen neuen Aktivierungswert aus der Summe des Aktivierungswertes des vorherigen Pulses und der neuen Input-Aktivierung berechnet, so dass

$$\mathrm{act}(\mathrm{i}, \mathrm{a}) := \zeta(\mathrm{i} + \mathrm{a}). \tag{3.12}$$

Knoten können bereits nach wenigen Pulsen sehr hohe Aktivierungswerte besitzen, wenn diese beispielsweise aufgrund hoher Knotengrade stark verzweigt sind und wiederholt ohne Dämpfung über viele Kanten Aktivierung beziehen. Über die Transferfunktionen lässt sich der Wertebereich der Aktivierungswerte steuern. In dieser Arbeit untersuchen wir die Auswirkung unterschiedlicher Transferfunktionen. Dafür werden verschiedene *Aktivierungsmodi* definiert, die in Tabelle 3.8 mit Funktionsgraphen dargestellt sind.

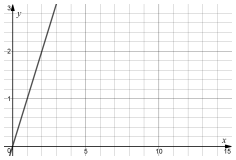
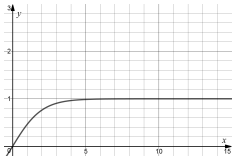
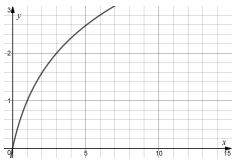
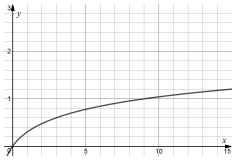
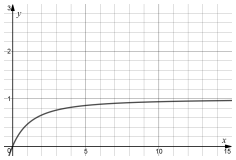
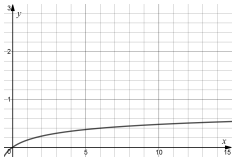
Aktivierungsmodus	**Transferfunktion** $\zeta(\mathrm{x})$		**Funktionsgraph**
IDENTITY	$\zeta(\mathrm{x}) := \mathrm{x}$	(3.13)	
SIG	$\zeta(\mathrm{x}) := 2 \cdot (\frac{1}{1+\mathrm{e}^{-\mathrm{x}}} - 0.5)$	(3.14)	
LOG2	$\zeta(\mathrm{x}) := \log_2(\mathrm{x}+1)$	(3.15)	
LOG10	$\zeta(\mathrm{x}) := \log_{10}(\mathrm{x}+1)$	(3.16)	
LOG2SIG	$\zeta(\mathrm{x}) := 2 \cdot (\frac{1}{1+\mathrm{e}^{-(\log_2(\mathrm{x}+1))}} - 0.5)$	(3.17)	
LOG10SIG	$\zeta(\mathrm{x}) := 2 \cdot (\frac{1}{1+\mathrm{e}^{-(\log_{10}(\mathrm{x}+1))}} - 0.5)$	(3.18)	

Tabelle 3.8: Aktivierungsmodi mit Transferfunktionen und Funktionsgraphen

Die Identitätsfunktion verändert den Eingabewert nicht und unterstützt daher den potenziell schnellen, ungebremsten Anstieg der Aktivierungswerte. Der Aktivierungsmodus *SIG* verwendet eine auf den Wertebereich [0,1] normierte Sigmoidfunktion, mit der Aktivierungswerte beschränkt werden, die allerdings nur für niedrige Werte des Definitionsbereiches differenziert

und die darüber hinaus sehr schnell gegen 1 konvergiert. Die Logarithmusfunktionen der Aktivierungsmodi *LOG2* und *LOG10* reduzieren die Wertebereiche unterschiedlich stark, beschränken sie allerdings nicht. Mit den Aktivierungsmodi *LOG2SIG* und *LOG10SIG* wird durch Kombination von Logarithmus- und nachgesetzter Sigmoidfunktion eine Beschränkung auf den Wertebereich [0,1] mit stärkerer Differenzierung im unteren Definitionsbereich erreicht.

3.3.2 Beschränkungen

Wie in Abschnitt 2.3.3 vorgestellt, ermöglichen Beschränkungen eine zusätzliche Konfiguration der Aktivierungsverteilungsfunktionen und eine gezielte Steuerung der Aktivierungsausbreitung. Im Folgenden werden Aktivierungsbeschränkung, pulsweise Dämpfung und Verzweigungsbeschränkung in die Aktivierungsverteilungsfunktionen integriert. Diese drei für diese Arbeit relevanten Beschränkungen beeinflussen die Stärke der Output-Aktivierung und erfordern daher Anpassungen in der Output-Funktion.

Mit Aktivierungsschwellwert $\tau \in \mathbb{R}$, Dämpfungsfaktor $\mathrm{d} \in \mathbb{R}$, $0 \leq \mathrm{d} \leq 1$, der Dämpfungsfunktion $\delta : \mathbb{R} \times \mathbb{N}_{>0} \to \mathbb{R}$ und der Verzweigungsfunktion $\beta : \mathrm{N} \to \mathbb{R}$ definieren wir die Output-Funktion

$$\mathrm{out}(\mathrm{n}, \mathrm{e}, \mathrm{a}, \mathrm{p}) := \begin{cases} \delta(\mathrm{d}, \mathrm{p}) \cdot \beta(\mathrm{n}) \cdot \mathrm{a} & \text{if } \mathrm{inz}(\mathrm{n}, \mathrm{e}), \mathrm{a} \geq \tau, \\ 0 & \text{else.} \end{cases} \tag{3.19}$$

Aktivierungsbeschränkung wird gesteuert durch den Aktivierungsschwellwert τ. Wird Aktivierungsbeschränkung nicht angewandt, so ist der Schwellwert entsprechend durch $\tau = 0$ festgelegt.

Dämpfung wird gesteuert durch den Dämpfungsfaktor d und die Dämpfungsfunktion δ. Wir definieren verschiedene *Dämpfungsmodi*, welche sich in der Wahl der Dämpfungsfunktion unterscheiden und in Tabelle 3.9 dargestellt sind. Die abgebildeten Funktionsgraphen haben wir beispielhaft für den Dämpfungsfaktor 0,9 erstellt. Wird Dämpfung nicht angewandt, so entspricht das dem Dämpfungsmodus *IGNORE*. Der Dämpfungsmodus *FIXED* verwendet eine konstante Dämpfung, während der Dämpfungsmodus *INCREASING* mit steigender Pulsanzahl eine steigende Dämpfung ermöglicht.

Verzweigungsbeschränkung wird gesteuert durch die Verzweigungsfunktion β. Wir definieren verschiedene Verzweigungsmodi, welche sich in der Wahl der Verzweigungsfunktion unterscheiden und in Tabelle 3.10 dargestellt sind. Der Verzweigungsmodus *DEGREE* verteilt die Aktivierung gleichmäßig auf alle inzidenten Kanten eines Knotens.[36] Je stärker die Verzweigung, desto stärker ist auch die Diskriminierung bei der Berechnung der Output-Aktivierung. Der Verzweigungsmodus *FANOUT* hingegen setzt die Anzahl der adjazenten Knoten ins Verhältnis zur Gesamtanzahl der Knoten eines Netzes. Die Diskriminierung ist besonders hoch, wenn ein Knoten nicht nur lokal, sondern auch global betrachtet sehr viele Nachbarknoten besitzt. Existiert nur ein Nachbarknoten, so erhält dieser allerdings fast die

[36] Da aufgrund der in Abschnitt 3.2.4 beschriebenen Reduktionsstufen der Zusammenhang des semantischen Netzes nicht gewährleistet ist, ist die Existenz von isolierten Knoten mit Knotengrad 0 möglich. Eine Division durch 0 ist dennoch aufgrund der erforderlichen Inzidenz zu Kante e ausgeschlossen.

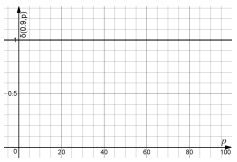
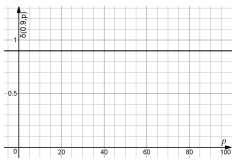
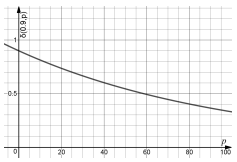
Dämpfungsmodus	**Dämpfungsfunktion** $\delta(\mathrm{d},\mathrm{p})$		**Funktionsgraph**
IGNORE	$\delta(\mathrm{d},\mathrm{p}) := 1$	(3.20)	
FIXED	$\delta(\mathrm{d},\mathrm{p}) := \mathrm{d}$	(3.21)	
INCREASING	$\delta(\mathrm{d},\mathrm{p}) := 0{,}99^{\mathrm{p}} \cdot \mathrm{d}$	(3.22)	

Tabelle 3.9: Dämpfungsmodi mit Dämpfungsfunktionen und Funktionsgraphen für $\delta(0{,}9,\mathrm{p})$

volle Aktivierung.[37] Im folgenden Abschnitt 3.3.4.2 wird im Rahmen der Einführung von Ausbreitungsmodi diese Liste um einen weiteren Verzweigungsmodus erweitert.

Verzweigungsmodus	**Verzweigungsfunktion** $\beta(\mathrm{n})$	
NONE	$\beta(\mathrm{n}) := 1$	(3.23)
DEGREE	$\beta(\mathrm{n}) := \frac{1}{\lvert\{\mathrm{e} \in \mathrm{E} : \mathrm{inz}(\mathrm{n},\mathrm{e})\}\rvert}$	(3.24)
FANOUT	$\beta(\mathrm{n}) := 1 - \frac{\lvert\{\mathrm{m} \in \mathrm{N} : \mathrm{adj}(\mathrm{n},\mathrm{m})\}\rvert}{\lvert\mathrm{N}\rvert}$	(3.25)

Tabelle 3.10: Verzweigungsmodi mit Verzweigungsfunktionen

3.3.3 Algorithmus

Die Aktivierungsverteilungsfunktionen werden in einer spezifizierten Reihenfolge ausgeführt, die den Beschreibungen in Abschnitt 2.3 entspricht. Ein neuer Puls startet mit der Berech-

[37] Als Alternative ist die Subtraktion der Zahl 1 von der Anzahl der adjazenten Knoten möglich. Dies bewirkt, dass im Falle, dass ein Knoten den Knotengrad 1 besitzt, dieser die volle Aktivierung verteilen kann. Subtrahiert man die 1 nicht, wie in Formel 3.25, so wird eine Dämpfung vorgenommen, die aufgrund der üblicherweise hohen Anzahl an Knoten allerdings sehr leicht ausfällt.

nung der Output-Funktionen. Diese basieren auf dem Aktivierungszustand des Graphen im vorherigen Puls. Anschließend werden die Input-Aktivierungswerte jedes Knotens mittels Input-Funktionen bestimmt. Diese werden darauffolgend in den Aktivierungsfunktionen für die Berechnung der neuen Aktivierungszustände der Knoten genutzt. Der Ablauf ist im folgenden Algorithmus 1 dargestellt.

Algorithmus 1 : Aktivierungsausbreitung

Data : $G = (N, E, s, t, l, \omega)$, $p_{max} \in \mathbb{N}_{>0}$, initiale Aktivierung $a_n^{(0)}$ for $n \in N$
Result : $a_n^{(p_{max})}$ for $n \in N$

1 *Initialisierung*
2 $p \leftarrow 0$
3 **for** $n \in N, e \in E$ **do**
4 $\quad i_{n,e}^{(0)} \leftarrow 0$
5 $\quad i_n^{(0)} \leftarrow 0$
6 *Aktivierungsausbreitung*
7 **while** $p < p_{max}$ **do**
8 $\quad p \leftarrow p + 1$
9 $\quad$ *Output-Aktivierung*
10 $\quad$ **for** $n \in N, e \in E$ **do**
11 $\quad\quad o_{n,e}^{(p)} \leftarrow out(n, e, a_n^{(p-1)}, p)$;
12 $\quad$ *Input-Aktivierung*
13 $\quad$ **for** $n \in N$ **do**
14 $\quad\quad$ **for** $m \in N, e \in E$ **do**
15 $\quad\quad\quad i_{n,e}^{(p)} \leftarrow in(n, e, o_{m,e}^{(p)})$;
16 $\quad\quad i_n^{(p)} \leftarrow in(n, p)$;
17 $\quad\quad$ *Aktualisieren der Gesamtaktivierung*
18 $\quad\quad a_n^{(p)} \leftarrow act(i_n^{(p)}, a_n^{(p-1)})$;

Der vorgestellte Algorithmus fokussiert auf Verständlichkeit und weist daher offensichtliches Optimierungspotenzial auf. Der implementierte Algorithmus ist bezüglich Laufzeit optimiert, aber dennoch ergebnisgleich. Der Charakter der Optimierung liegt vor allem in der Auslassung unnötiger Berechnungen. So ist es beispielsweise nicht notwendig, die Zuweisung in Zeile 5 des Algorithmus 1 für jede Kante durchzuführen. Zudem werden unnötige Berechnungen vermieden, indem bereits vor der Berechnung der Output-Aktivierung in Zeile 11 die Knoten-Kanten-Paare ausgeschlossen werden, die keine Sendeerlaubnis besitzen oder die den Aktivierungsschwellwert ohnehin nicht überschreiten.

Wir haben in diesem Abschnitt eine reproduzierbare Formalisierung etablierter Prinzipien der Aktivierungsverteilung vorgenommen und eine erweiterbare Liste verschiedener Modi zur Konfiguration der Beschränkungen bestehend aus Aktivierungsmodi, Dämpfungsmodi, Verzweigungsmodi vorgestellt. Diese Beschränkungen steuern die Berechnungen der Aktivierungsverteilungsfunktionen und wirken sich lediglich indirekt, beispielsweise über Kantengewichte oder Schwellwerte, auf das Ausbreitungsverhalten im Graphen aus. Daher erweitern wir die Aktivierungsausbreitung im folgenden Abschnitt um ein neuartiges Konzept zur Steuerung des Ausbreitungsverhaltens, das eine differenzierte Behandlung der Aktivierungsausbreitung ermöglicht.

3.3.4 Steuerung des Ausbreitungsverhaltens durch Ausbreitungsmodi

Bei der puren, unbeschränkten Aktivierungsausbreitung dürfen in jedem Puls potenziell alle aktivierten Knoten Aktivierung an Nachbarknoten verteilen und die Nachbarknoten diese auch empfangen. Die zuvor vorgestellten Beschränkungen steuern die Berechnungen der Aktivierungsverteilungsfunktionen und wirken sich lediglich indirekt auf das Ausbreitungsverhalten im Graphen aus. Kantengewichte oder Schwellwerte der Aktivierungsbeschränkung können zwar die Ausbreitung über bestimmte Kanten verhindern. Allerdings geschieht dies abhängig von den Berechnungen der Aktivierungsverteilungsfunktionen und dementsprechend erst in der Aktivierungsausbreitungs- oder der Nachbearbeitungsphase. Wir ergänzen die Aktivierungsausbreitung um das neuartige Konzept der Sendemodi und Ausbreitungsmodi, mit denen bereits in der Vorverarbeitungsphase eine Teilmenge der Knoten-Kanten-Paare bestimmt werden kann, für welche die Aktivierungsverteilungsfunktionen in der Aktivierungsausbreitungsphase ausgeführt werden. Im Folgenden erläutern wir zunächst unsere Ziele bei der Steuerung des Ausbreitungsverhaltens. Wir werden dann verschiedene Sendemodi sowohl formal als auch am Beispiel beschreiben und anschließend zu sinnvollen Ausbreitungsmodi kombinieren.

3.3.4.1 Ausbreitungsverhalten

Ausbreitungsverhalten

Das *Ausbreitungsverhalten* bezieht sich auf die Pfade, auf welchen die Aktivierung in jedem Puls durch den Graphen transportiert wird, und nicht auf die Stärke der Aktivierung. Es beschreibt also, ob Aktivierung von einem Knoten über eine Kante zu einem Knoten übertragen wird. Durch die Steuerung des Ausbreitungsverhaltens wird entschieden, ob in einem Tripel mit $n, m \in N, e \in E$ Aktivierung von einem Knoten zum anderen verteilt werden darf. Dieser Sachverhalt kann aus drei Perspektiven betrachtet werden, die in Abbildung 3.17 dargestellt sind, die Sende-, die Empfangs- und die Transportperspektive. Bei der Ausbreitung bezieht sich die Sendeperspektive darauf, ob Knoten n Aktivierung über Kante e verteilen darf. Die Erlaubnis bezieht sich intuitiv auf die Output-Funktion von Knoten n. Aus Empfangsperspektive stellt sich die Frage, ob Knoten m Aktivierung über Kante e empfangen darf. Die Erlaubnis zum Empfang bezieht sich intuitiv auf die Input-Funktion von Knoten m. Betrachtet man die Ausbreitung allerdings aus Transportperspektive, so ist bedeutend, ob Kante e Aktivierung zwischen den Knoten n und m transportieren darf. Aus dieser Sicht bezieht sich die Erlaubnis sowohl auf die Output-Funktion von Knoten n als auch auf die Input-Funktion von Knoten m.

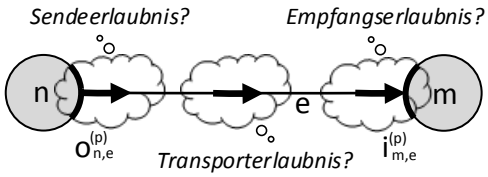


Abbildung 3.17: Sende-, Empfangs- und Transportperspektive

Die unterschiedlichen Sichten sind bei Überlegungen zu einer sinnvollen Steuerung der Aktivierungsausbreitung hilfreich. Einschränkungen zum Transport oder Empfang von Aktivierung lassen sich allerdings immer derart formulieren, dass sie sich auf das Senden, und damit auf

die Output-Funktion, beziehen.[38] Die Anwendung auf die Output-Funktion hat den Vorteil, dass durch das frühzeitige Ausschließen von Knoten-Kanten-Paaren unnötige Berechnungen von Aktivierungsverteilungsfunktionen verhindert beziehungsweise verringert werden können, was effizientere Berechnungen ermöglicht. In dieser Arbeit werden daher Maßnahmen zur Steuerung ausschließlich auf das Sendeverhalten bezogen und derart formuliert, dass sie sich auf die Output-Funktion anwenden lassen, selbst wenn sich diese direkt oder indirekt auf das Empfangsverhalten beziehen.

Ziele der Steuerung des Ausbreitungsverhaltens

Mit der aktiven Steuerung des Ausbreitungsverhaltens verfolgen wir folgende Ziele.

Gezielte statt ungebremste Flutung des Netzes: Da die Aktivierungsausbreitung grundsätzlich die wiederholte Aktivierungsverteilung aller aktivierten Knoten sowie den Aktivierungsempfang aller Knoten, und damit die schnelle gegenseitige Rückverteilung von Aktivierung zwischen benachbarten Knoten erlaubt, tendiert sie zu einer Favorisierung der direkten Nachbarschaft. Die Aktivierungswerte eines Knotens sind das Maß für dessen Relevanz im Graphen, welche sich durch Empfang von Aktivierung erhöht und deren Ausbleiben nachlassendes oder nicht vorhandenes Interesse symbolisiert. Die aktive Steuerung zielt darauf ab, ein differenzierteres Verhalten bei der Wahl der Pfade im Verlauf der Aktivierungsausbreitung zu erreichen. Sie ist zudem eine Maßnahme, um beispielsweise eine extrem oszillierende Aktivierungsverteilung über Pulse hinweg zu umgehen.

Begünstigung und Bestrafung struktureller Besonderheiten im Graphen: Sind strukturelle Besonderheiten sowie deren Bedeutung in domänenspezifisch aufgebauten Netzen bekannt, so bieten diese Potenzial zur gezielten Steuerung zu erwartender Effekte der Aktivierungsausbreitung. Beispielsweise zeichnet sich das Basisnetz der G&R-Dokumente durch starke Verzweigung zwischen Klassenknoten und deren Instanzknoten aus. Die sich pulsweise wiederholende gegenseitige Aktivierungsverteilung kann den eher negativen Effekt der undifferenzierten Aktivierung aller Instanzknoten bewirken. Ebenso ist der geringe Durchmesser des Netzes durch die Ableitung aus dem Datenmodell charakteristisch, was darauf hindeutet, dass wenige Pulse bereits zur Aktivierung aller Knoten des Netzes ausreichen könnten. Die aktive Steuerung der Aktivierungsausbreitung zielt darauf ab, das vorhandene Wissen über die strukturellen Besonderheiten zur Begünstigung oder Bestrafung zu nutzen, indem erwünschte und unerwünschte Effekte bei der Aktivierungsverteilung verstärkt oder abgeschwächt werden.

Reduktion der notwendigen Berechnungen: Aufgrund der oftmals hohen Anzahl an Knoten-Kanten-Paaren sind entsprechend viele Berechnungen für die Aktivierungsverteilungsfunktionen erforderlich. Die gezielte Auswahl der Knoten-Kanten-Paare für eine potenzielle Ausbreitung kann eine Reduzierung des Aufwands bewirken.

[38] Jedes Tripel ist einmalig im RDF-Graphen vorhanden. Nur wenn sowohl Senden, Transport und Empfangen erlaubt sind, darf Aktivierung über eine Kante verteilt werden. Soll der Empfang oder der Transport unterbunden werden, lässt sich dies auch durch den Entzug der Sendeerlaubnis realisieren.

3.3.4.2 Sendemodi

Sendemodus

Ein *Sendemodus* beschreibt ein Ausbreitungsverhalten von Aktivierung im Graphen. Er bestimmt in jedem Puls vor Berechnung der Aktivierungswerte, zwischen welchen Knoten und auf welchen Kanten Aktivierung transportiert werden darf. Dafür wird, im Gegensatz zu den Aktivierungsverteilungsfunktionen, zusätzlich zum aktuellen Aktivierungsstatus auch die Aktivierungshistorie vergangener Pulse betrachtet. Ein *Ausbreitungsmodus* kann sich aus verschiedenen Sendemodi zusammensetzen und somit deren unterschiedliche Sendeverhalten kombinieren. Daher beschreibt ein Ausbreitungsmodus das gesamte Ausbreitungsverhalten von Aktivierung im Graphen. Sei SM eine nichtleere Menge von Sendemodi, dann ist AM die Menge aller Ausbreitungsmodi.

Ausbreitungsmodus

$$\mathrm{AM} := \mathcal{P}(\mathrm{SM}) \setminus \emptyset \tag{3.26}$$

Mithilfe der Sende- und Ausbreitungsmodi lässt sich für jeden Knoten die Menge aller Knoten-Kanten-Paare mit Sendeerlaubnis ermitteln. $\mathrm{SP}^{(\mathrm{p})}_{\mathrm{n,sm}} \subseteq \mathrm{N} \times \mathrm{E}$ ist die Menge aller Knoten-Kanten-Paare mit Sendeerlaubnis für Knoten $\mathrm{n} \in \mathrm{N}$ in Puls $\mathrm{p} > 0$ und im Sendemodus $\mathrm{sm} \in \mathrm{SM}$. Zudem stellt $\mathrm{SP}^{(\mathrm{p})}_{\mathrm{n,am}} \subseteq \mathrm{N} \times \mathrm{E}$ die Menge aller Knoten-Kanten-Paare mit Sendeerlaubnis für Knoten $\mathrm{n} \in \mathrm{N}$ in Puls $\mathrm{p} > 0$ und im Ausbreitungsmodus $\mathrm{am} \in \mathrm{AM}$ dar. Dabei gilt, dass sich die Menge aller Knoten-Kanten-Paare mit Sendeerlaubnis bei Kombination unterschiedlicher Sendeverhalten durch verschiedene Sendemodi aus deren Schnittmenge ergibt, so dass

$$\mathrm{SP}^{(\mathrm{p})}_{\mathrm{n,am}} := \bigcap_{\mathrm{mode} \in \mathrm{am}} \mathrm{SP}^{(\mathrm{p})}_{\mathrm{n,mode}}. \tag{3.27}$$

Das bedeutet, ein Knoten-Kanten-Paar erhält in einem Ausbreitungsmodus genau dann eine Sendeerlaubnis, wenn es für jeden im Ausbreitungsmodus enthaltenen Sendemodus eine Sendeerlaubnis besitzt. Formal definieren wir die Sendeerlaubnis über die Erlaubnisfunktion $\varphi : \mathrm{N} \times \mathrm{E} \times \mathbb{N}_{>0} \times \mathrm{AM} \to \{0, 1\}$. Sie bestimmt, ob ein Knoten $\mathrm{n} \in \mathrm{N}$ über eine Kante $\mathrm{e} \in \mathrm{E}$ in Puls $\mathrm{p} > 0$ und im Ausbreitungsmodus $\mathrm{am} \in \mathrm{AM}$ Aktivierung verteilen darf und ist definiert als

Erlaubnisfunktion

$$\varphi(\mathrm{n}, \mathrm{e}, \mathrm{p}, \mathrm{am}) := \begin{cases} 1 & \text{if } (\mathrm{n}, \mathrm{e}) \in \mathrm{SP}^{(\mathrm{p})}_{\mathrm{n,am}}, \\ 0 & \text{else.} \end{cases} \tag{3.28}$$

Die Output-Funktion unter Berücksichtigung der Ausbreitungsmodi wird für $\mathrm{n} \in \mathrm{N}, \mathrm{e} \in \mathrm{E}$, Puls $\mathrm{p} > 0$ und Ausbreitungsmodus $\mathrm{am} \in \mathrm{AM}$ wie folgt definiert.

$$\mathrm{out}_{\mathrm{am}}(\mathrm{n}, \mathrm{e}, \mathrm{a}, \mathrm{p}) := \varphi(\mathrm{n}, \mathrm{e}, \mathrm{p}, \mathrm{am}) \cdot \mathrm{out}(\mathrm{n}, \mathrm{e}, \mathrm{a}, \mathrm{p}) \tag{3.29}$$

Mit der Definition der Sendeerlaubnis erweitern wir in Tabelle 3.11 die Verzweigungsmodi aus Tabelle 3.10 um eine weitere Verzweigungsfunktion, die bei der Diskriminierung der Verzweigerknoten nur die Knoten-Kanten-Paare berücksichtigt, über die Aktivierung gemäß Sendemodus verteilt werden darf.

Im Folgenden werden ausgewählte Sendemodi vorgestellt. Die Liste ist generell erweiterbar. So ließe sich beispielsweise auch die kantengerichtete Aktivierungsausbreitung in einem Sendemodus beschreiben, in dem jeder Knoten lediglich über ausgehende Kanten und nicht über eingehende Kanten Sendeerlaubnis erhält. Da dieser Modus in dieser Arbeit nicht verwendet wird, verzichten wir auf dessen Definition.

Verzweigungsmodus	**Verzweigungsfunktion** $\beta(\mathrm{n})$	
BRANCH	$\beta(\mathrm{n}) := \frac{1}{\max(1, \lvert\{\mathrm{e} \in \mathrm{E} : \varphi(\mathrm{n}, \mathrm{e}, \mathrm{p}, \mathrm{am}) = 1, \mathrm{p} > 0, \mathrm{am} \in \mathrm{AM}\}\rvert)}$	(3.30)

Tabelle 3.11: Erweiterung der Verzweigungsmodi

Sendemodus Basic

Der Sendemodus *Basic (B)* erlaubt jedem Knoten das Senden von Aktivierung über jede inzidente Kante und stellt den Standard-Sendemodus dar. Die Menge der Knoten-Kanten-Paare mit Sendeerlaubnis für Puls $\mathrm{p} > 0$ ist definiert durch

$$\mathrm{SP}^{(\mathrm{p})}_{\mathrm{n,B}} := \{(\mathrm{n}, \mathrm{e}) \in \mathrm{N} \times \mathrm{E} \mid \mathrm{inz}(\mathrm{n}, \mathrm{e})\}. \tag{3.31}$$

Bedeutung: Dieser Modus bewirkt, dass grundsätzlich jeder Knoten Aktivierung über inzidente Kanten an benachbarte Knoten verteilen darf. Das Netz wird geflutet. Abbildung 3.18 stellt das Ausbreitungsverhalten ausgehend vom Startknoten A für die Pulse 1 bis 4 am Beispiel dar. Knoten, die aktiviert sind oder im Laufe der Aktivierungsausbreitung aktiviert werden, sind grau markiert.[39] Zudem ist der Bereich der Aktivierung verteilenden Knoten umkreist. Der Sendemodus B zeichnet sich durch ungebremste Ausbreitung und einen wachsenden Anteil der verteilenden Knoten aus.

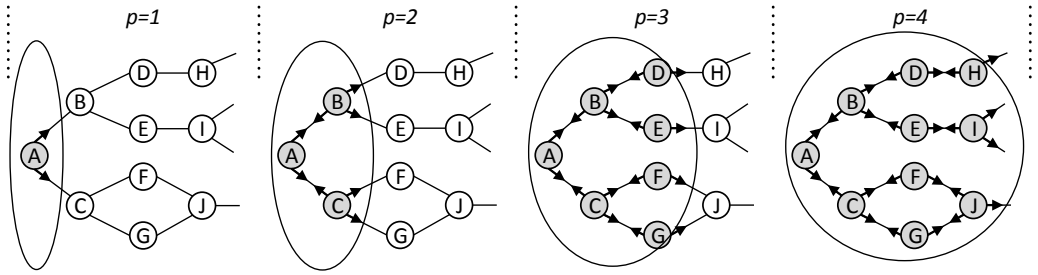


Abbildung 3.18: Ausbreitungsverhalten im Sendemodus *B*

Sendemodus RecentReceiver

Der Sendemodus *RecentReceiver (RR)* erlaubt in einem Puls lediglich denjenigen Knoten das Verteilen von Aktivierung, welche im vorherigen Puls Empfänger von Aktivierung waren. Die Menge der Knoten-Kanten-Paare mit Sendeerlaubnis für Puls $\mathrm{p} > 0$ ist definiert durch

$$\mathrm{SP}^{(1)}_{\mathrm{n,RR}} := \mathrm{SP}^{(1)}_{\mathrm{n,B}}, \tag{3.32}$$

[39] In diesem Beispiel wird der Fall einer möglichen Deaktivierung von Knoten nicht betrachtet. Stattdessen werden alle Knoten, welche theoretisch aufgrund des Ausbreitungsverhalten Aktivierung erhalten können, auch als aktiviert betrachtet und sind daher im Beispiel grau markiert.

sowie

$$\mathrm{SP}^{(\mathrm{p})}_{\mathrm{n,RR}} := \{(\mathrm{n},\mathrm{e}) \in \mathrm{N} \times \mathrm{E} \mid \mathrm{inz}(\mathrm{n},\mathrm{e}), \mathrm{i}^{(\mathrm{p}-1)}_{\mathrm{n}} > 0\} \text{ für } \mathrm{p} > 1. \tag{3.33}$$

Bedeutung: Knoten, welche im vorherigen Puls Aktivierung empfangen haben, nennen wir Recent Receiver. Sie markieren in jedem Puls einen Bereich im Graphen, der das momentane Zentrum der Relevanzbewertung darstellt. Dieser Bereich kann sich im Laufe aufeinanderfolgender Pulse vergrößern, verkleinern oder über den Graphen verschieben. Im Sendemodus RecentReceiver werden Knoten, die außerhalb dieses Bereiches liegen, temporär von der aktiven Teilnahme an der Aktivierungsverteilung ausgeschlossen. Dieser Modus verhindert damit die wiederholte Verteilung von Aktivierung durch Knoten, welche im vorherigen Puls keine Aktivierung erhielten und dadurch aktuell nicht im Zentrum des Interesses liegen. Damit bleibt die stetige Nachaktivierung auf bereits bekannten Pfaden aus. Stattdessen rückt der aktuelle Kontext in den Vordergrund. Der Ausschluss wird mit erneutem Aktivierungsempfang aufgehoben. Oftmals alterniert die Sendeerlaubnis pulsweise zwischen Knoten. Abbildung 3.19 zeigt das Ausbreitungsverhalten am Beispiel. Hier ist deutlich die Veränderung des Bereiches der Aktivierung verteilenden Knoten sichtbar, der sich über die Pulse hinweg teilweise alternierend verschiebt.

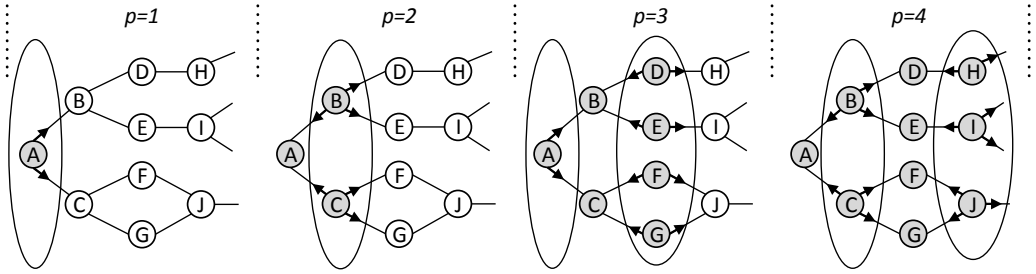


Abbildung 3.19: Ausbreitungsverhalten im Sendemodus *RR*

Sendemodus Forward

Der Sendemodus *Forward (FW)* verhindert, dass ein Knoten direkt über jene Kante Aktivierung zurückverteilt, über die er im vorherigen Puls Aktivierung erhalten hat. Die Menge der Knoten-Kanten-Paare mit Sendeerlaubnis für Puls $\mathrm{p} > 0$ ist definiert durch

$$\mathrm{SP}^{(1)}_{\mathrm{n,FW}} := \mathrm{SP}^{(1)}_{\mathrm{n,B}}, \tag{3.34}$$

sowie

$$\mathrm{SP}^{(\mathrm{p})}_{\mathrm{n,FW}} := \{(\mathrm{n},\mathrm{e}) \in \mathrm{N} \times \mathrm{E} \mid \mathrm{inz}(\mathrm{n},\mathrm{e}), \mathrm{i}^{(\mathrm{p}-1)}_{\mathrm{n,e}} = 0\} \text{ für } \mathrm{p} > 1. \tag{3.35}$$

Bedeutung: Mit diesem Modus wird die ständige gegenseitige Verteilung von Aktivierung zwischen zwei Knoten eines Tripels unterbunden. Durch den zwischen den Pulsen alternierenden oder gar gleichzeitigen gegenseitigen Aktivierungsempfang werden Knoten in der Nähe von Startknoten oftmals stark bevorzugt. In diesem Modus wird stattdessen eine Vorwärtsbewegung von Aktivierung im Graphen bewirkt. Abbildung 3.20 stellt das Ausbreitungsverhalten am Beispiel dar.

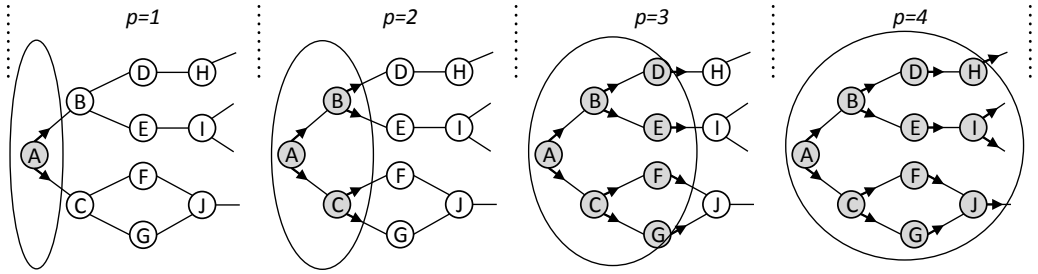


Abbildung 3.20: Ausbreitungsverhalten im Sendemodus *FW*

Sendemodus ForwardLoop

Der Sendemodus *ForwardLoop (FWL)* ist eine weniger restriktive Variante des Sendemodus Forward. Einem Knoten wird das direkte Zurückverteilen von Aktivierung über jene Kanten, über die er im vorherigen Puls Aktivierung erhalten hat, nicht grundsätzlich untersagt. Eine Ausnahme wird dann gemacht, wenn Aktivierung zusätzlich über andere Kanten empfangen wurde. Die Menge der Knoten-Kanten-Paare mit Sendeerlaubnis für Puls $p > 0$ ist definiert durch

$$SP^{(1)}_{n,FWL} := SP^{(1)}_{n,B}, \tag{3.36}$$

sowie

$$\begin{aligned} SP^{(p)}_{n,FWL} := \{(n,e) \in N \times E \mid inz(n,e), \\ (i^{(p-1)}_{n,e} = 0) \vee (\exists e_1 \in E, e \neq e_1, i^{(p-1)}_{n,e_1} > 0)\} \text{ für } p > 1. \end{aligned} \tag{3.37}$$

Bedeutung: Diesem Modus liegt die Idee zugrunde, dass die gegenseitige Aktivierungsverteilung zwischen zwei Knoten nur dann unterbunden werden soll, wenn ausschließlich die zuletzt übertragene Aktivierung hin und her transportiert wird. Wird das Zurückverteilen bei Aktivierungsempfang über verschiedene Kanten erlaubt, so entstehen neue Aktivierungspfade, was insbesondere die Aktivierung von Kreisstrukturen stärkt und dadurch deren besondere semantische Bedeutung im Graphen hervorheben kann. Abbildung 3.21 stellt das Ausbreitungsverhalten am Beispiel dar. Im Vergleich zu Sendemodus *FW* unterscheidet sich dieser Sendemodus durch die Behandlung der Kreisstruktur (*C-F-J-G*). Da Knoten *J* in Puls 3 aus mehreren Richtungen Aktivierung erhält, kann dieser in Puls 4 Aktivierung gegen die absolute Vorwärtsbewegung zurückverteilen. Dadurch sind die Aktivierungspfade *F-J-G* und *G-J-F* möglich, welche auf der Kreisstruktur dennoch als Vorwärtsbewegung gesehen werden kann.

3.3.4.3 Ausbreitungsmodi

Ausbreitungsmodi bestehen aus Kombinationen von Sendemodi (siehe Formel 3.26, Seite 99). Da der Sendemodus Basic die Verteilung über alle inzidenten Knoten-Kanten-Paare erlaubt, gilt $SP^{(p)}_{n,B} \cap SP^{(p)}_{n,sm} = SP^{(p)}_{n,sm}, \forall sm \in SM$. Da $SP^{(p)}_{n,FW} \subseteq SP^{(p)}_{n,FWL}$, gilt zudem, dass $SP^{(p)}_{n,FW} \cap SP^{(p)}_{n,FWL} = SP^{(p)}_{n,FW}$. Durch Subsumption entfallen Kombinationen, so dass sich aus den vier vorgestellten Sendemodi die sechs Ausbreitungsmodi in Tabelle 3.12 ergeben.

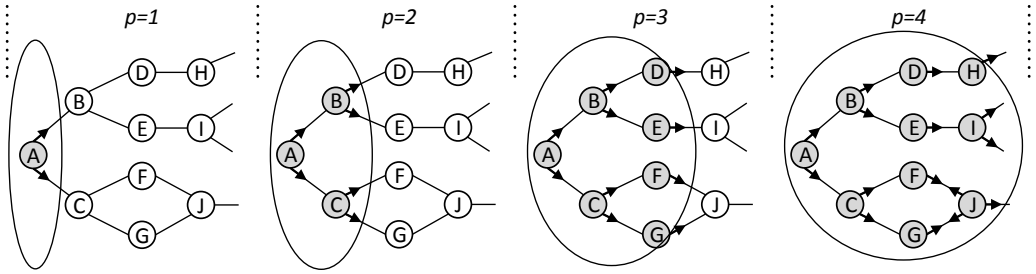


Abbildung 3.21: Ausbreitungsverhalten im Sendemodus *FWL*

Ausbreitungsmodus	**am =**
B	{Basic}
RR	{RecentReceiver}
FW	{Forward}
FW-RR	{Forward, RecentReceiver}
FWL	{ForwardLoop}
FWL-RR	{ForwardLoop, RecentReceiver}

Tabelle 3.12: Resultierende Ausbreitungsmodi aus beteiligten Sendemodi

Zusätzlich zu den reinen Sendemodi ergeben sich zwei kombinierte Ausbreitungsmodi: *FW-RR* und *FWL-RR*.

Bedeutung: Der Ausbreitungsmodus *FW-RR* favorisiert eine Vorwärtsbewegung der Knoten, die im vorherigen Puls Empfänger von Aktivierung waren. Dadurch wird, wie in Abbildung 3.22 dargestellt, der Bereich der verteilenden Knoten eingeschränkt und zudem über die Pulse im Graphen verschoben. Bei Graphen, die viele und insbesondere kleine Kreisstrukturen enthalten, besteht allerdings die Gefahr einer schnellen Stagnation. Diesem Umstand wird im Ausbreitungsmodus *FWL-RR* durch die erlaubte Rückwärtsbewegung auf Kreispfaden Rechnung getragen. Wie in Abbildung 3.23 dargestellt, wird zwar eine klare Vorwärtsbewegung präferiert, allerdings in Kreisen weniger restriktiv behandelt.

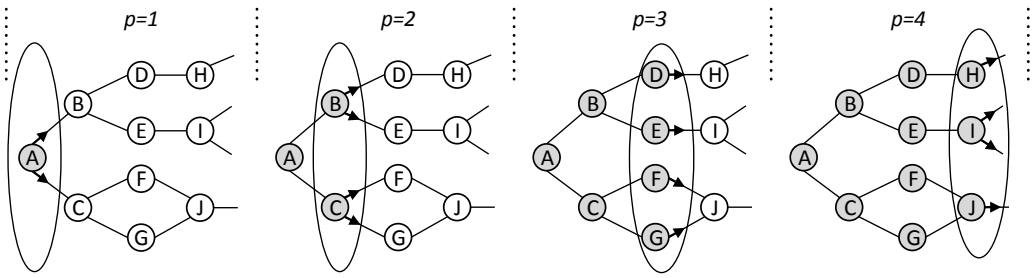


Abbildung 3.22: Ausbreitungsverhalten im Ausbreitungsmodus *FW-RR*

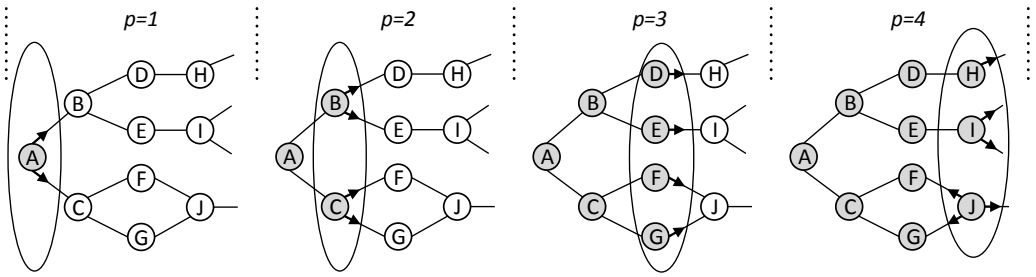


Abbildung 3.23: Ausbreitungsverhalten im Ausbreitungsmodus *FWL-RR*

3.3.5 Zusammenfassung und Beitrag

In diesem Abschnitt 3.3 haben wir etablierte Prinzipien und neuartige Techniken der Aktivierungsausbreitung derart aufbereitet, dass sie zum einen für die Anwendung im IR-System für G&Rs verwendbar und zum anderen auf andere Domänen übertragbar sind. Dafür haben wir die Aktivierungsverteilung durch bekannte Aktivierungsverteilungsfunktionen reproduzierbar formalisiert und Beschränkungen konzeptualisiert. Entstanden ist eine erweiterbare Liste verschiedener Modi zur Konfiguration der Beschränkungen bestehend aus Aktivierungsmodi, Dämpfungsmodi, Verzweigungsmodi. Wir haben zudem die Aktivierungsausbreitung um ein neuartiges Konzept zur Steuerung des Ausbreitungsverhaltens erweitert. Dieses unterscheidet sich von den üblicherweise zur Steuerung bemühten Aktivierungsverteilungsfunktionen in der grundsätzlichen Idee, nicht die Berechnung von Aktivierungswerten, sondern das Ausbreitungsverhalten im Graphen zu steuern und damit strukturelle Eigenheiten verschiedener semantischer Netzwerke zu berücksichtigen. Die Ausbreitungsmodi folgen Ausbreitungsregeln und steuern somit die Pfade, über die im Aktivierungsprozess über die Pulse Aktivierung verteilt wird. Das Ergebnis ist eine erweiterbare Liste verschiedener Sendemodi und Ausbreitungsmodi zur aktiven Steuerung des Ausbreitungsverhaltens im Verlauf der Aktivierungsausbreitung.

Mit dem Abschnitt 3.3 unterstützen wir den dritten Schritt der in Abschnitt 3.1.2.4 vorgestellten Methode zum Aufbau domänenspezifischer IR-Systeme. In diesem dritten Schritt konfiguriert der Entwickler des IR-Systems die Aktivierungsausbreitung. Die Ergebnisse der Konzeptualisierungen stellen generische Lösungen dar, welche die Retrieval-Komponente des IR-Systems anreichern. Dem Entwickler eines IR-Systems steht damit eine Bandbreite an generischen Konfigurationsmöglichkeiten zur Verfügung. Die Auswahl geeigneter Modi ist schwierig, da die Effekte abhängig sind von der Gesamtheit der Konfigurationsentscheidungen und den komplexen Eigenschaften der verwendeten Graphen. Im Rahmen der Evaluation des IR-Systems für G&Rs in Kapitel 4 werden wir unter anderem den Einfluss der vorgestellten Modi für die Aktivierungsverteilungsfunktionen, hier Aktivierungsmodi, Dämpfungsmodi und Verzweigungsmodi, sowie der präsentierten Ausbreitungsmodi auf die Qualität der Ergebnisse untersuchen.

Im nächsten Abschnitt zeigen wir, wie die Ergebnisse des IR-Systems aufbereitet und erklärt werden können. Damit unterstützen wir den vierten und damit letzten Schritt der Methode zum Aufbau domänenspezifischer IR-Systeme.

3.4 Automatische Aufbereitung der Ergebnisse

In diesem Abschnitt präsentieren wir, wie wir die Ergebnisse der Aktivierungsausbreitung aufbereiten und die Anfrage an das IR-System beantworten. Zusätzlich stellen wir ein Verfahren vor, welches automatisiert textbasierte Erklärungen für die Ergebnisse des IR-Systems generiert, so dass der Nutzer die Ergebnisse nachvollziehen kann.

3.4.1 Suchzielspezifische Filterung

Das Ergebnis der Aktivierungsausbreitung ist ein Netz, in dem jeder Knoten einen Aktivierungswert besitzt, dessen Höhe das Maß für die Relevanz zu den Startknoten darstellt, welche die Anfrage repräsentiert. Zur Ausgabe von Empfehlungen wird das Ergebnis daher nachbearbeitet. Zuerst werden durch eine suchzielspezifische Filterung jene Knoten identifiziert, welche dem Informationsbedürfnis des Nutzers entsprechen. Diese Knoten werden nach Relevanz, hier entsprechend nach der Höhe ihrer Aktivierungswerte, sortiert. Diese Liste wird optional auf eine zu spezifizierende Anzahl an Elementen gekürzt, so dass dem Nutzer lediglich die Top-Elemente als Empfehlungen zur Verfügung gestellt werden können. Dieses Vorgehen stellen wir im Folgenden vor.

Hilfsdefinitionen

Sei ein Label L gemäß der Namenskonvention aus Abschnitt 3.2.2.3 aufgebaut aus verschiedenen optionalen Parametern und sei NS die Menge aller domänenspezifischen Namensraum-Präfixe, DP die Menge aller domänenspezifischen Präfixe, TN die Menge aller Ressourcentypen und stehe $e \notin NS \cup DP \cup TN$ für das undefinierte Element, so sind folgende Funktionen definiert:

- $dp : L \to DP \cup \{e\}$ eine Abbildung domänenspezifischer Präfixe,
- $ns : L \to NS \cup \{e\}$ eine Abbildung domänenspezifischer Namensräume und
- $typeName : L \to TN \cup \{e\}$ eine Abbildung domänenspezifischer Ressourcentypen.

Des Weiteren beschreibt die Relation $isInstance \subseteq L$, ob ein Label $l \in L$ einem domänenspezifischen Instanzknoten zugeordnet ist.

$$isInstance(l) \iff ns(l) \neq e \wedge dp(l) = INS \tag{3.38}$$

Mit der Relation $ofType \subseteq L \times TN$ überprüfen wir, ob ein Knoten einem speziellen Ressourcentypen entspricht.

$$ofType(l, t) \iff isInstance(l) \wedge typeName(l) = t \tag{3.39}$$

Filtern

Die Menge der Knoten, die als Empfehlung in Frage kommen, enthält alle domänenspezifischen Instanzknoten, welche dem in der Anfrage $Q = (N_Q, T_Q)$ spezifizierten Suchziel T_Q entsprechen. Daher beschreibt R_T die entsprechend gefilterte Menge aller Knoten eines Netzes $G = (N, E, s, t, l, \omega)$.

$$R_T := \{n \in N \setminus N_Q \mid isInstance(l(n)), \exists t \in T_Q \mid ofType(l(n), t)\} \tag{3.40}$$

Sortieren

Sei die Relation $\leq_L \subseteq L \times L$ eine lexikographische Totalordnung auf der Grundmenge der Labels L, so ist die nach Höhe der Aktivierungswerte geordnete Folge $r : \mathbb{N} \to R_T$ wie folgt definiert, wobei aktivierungsgleiche Knoten lexikographisch sortiert werden.

$$r := (r_i)_{i=1,\cdots,|R_T|}, \text{ so dass } r_i \in R_T, a(r_i) > a(r_{i+1}) \vee (a(r_i) = a(r_{i+1}) \wedge l(r_i) \leq_L l(r_{i+1})) \tag{3.41}$$

Kürzen

Als Empfehlungen verwenden wir die Teilfolge r_{top} von r, wobei mit dem Parameter $top \in \mathbb{N}, top \leq |R_T|$ die gewünschte Anzahl der Ergebnisse der Folge spezifiziert wird.

$$r_{top} := (r_{i_k})_{k=1,\cdots,top} \tag{3.42}$$

3.4.2 Generierung von Erklärungen für Empfehlungen

In diesem Abschnitt stellen wir eine zusätzliche Komponente des IR-Systems vor, welche automatisiert Erklärungen für die vom IR-System bereitgestellten Empfehlungen generiert. Analog zum Informationsbedürfnis befriedigen sie ein Erklärungsbedürfnis des Nutzers. Erklärungskomponenten haben üblicherweise die Aufgabe, die Nutzerakzeptanz und das Vertrauen in das jeweilige System zu erhöhen. Daher werden mittels Generierung von Erklärungen die Ergebnisse des IR-Systems für den Nutzer auf Anfrage derart aufbereitet, dass dieser das Ergebnis als Empfehlung für sein Informationsbedürfnis nachvollziehen kann. Die dadurch erzeugte Transparenz ermöglicht die Prüfbarkeit, was wiederum die Beurteilung von Güte und Relevanz der Ergebnisse unterstützt.

Die Methode zur Erklärung der Resultate des in der vorliegenden Arbeit entwickelten IR-Systems für G&Rs ist im Rahmen einer Masterarbeit [Mic16] entstanden und zum Teil in einem Konferenzartikel [MH16] veröffentlicht. Die Inhalte dieses Abschnittes stammen daher teilweise aus diesen Arbeiten oder sind an dortige Darstellungen und Beschreibungen angelehnt. Die Methode ist in Abbildung 3.24 dargestellt. Als Eingabe wird das semantische Netz als Ausgangsgraph sowie ein Ausbreitungsgraph verwendet, den wir aufgrund des rekonstruktiven Charakters aus der Aktivierungshistorie aufbauen. Zudem wird spezifiziert, für welchen Knoten ein Erklärungsbedürfnis besteht. Diesen Knoten nennen wir Zielknoten. Als Ergebnis der dreischrittigen Methode wird eine textbasierte Erklärung ausgegeben. Der Ausbreitungsgraph und die drei Schritte der Methode werden im Folgenden konzeptuell beschrieben. Eine detaillierte Beschreibung sowie formale Hintergründe und ein umfangreicheres praxisnahes Beispiel findet sich bei Michalke [Mic16].

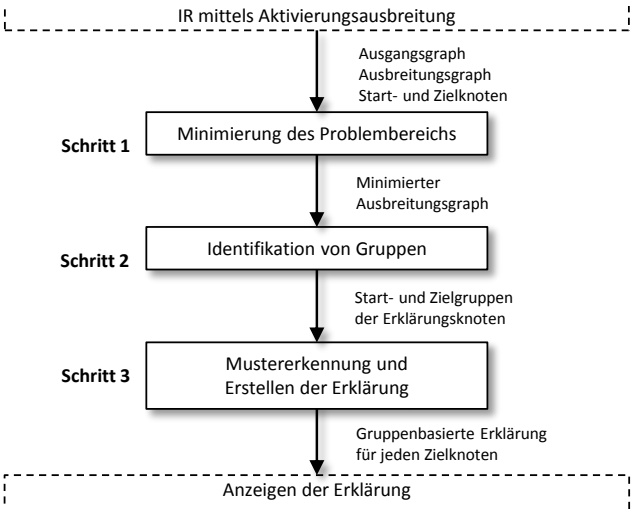


Abbildung 3.24: Methode zur automatischen Erklärung der generierten Empfehlungen für G&Rs nach [Mic16]

3.4.2.1 Ausbreitungsgraph zur Darstellung der Aktivierungshistorie

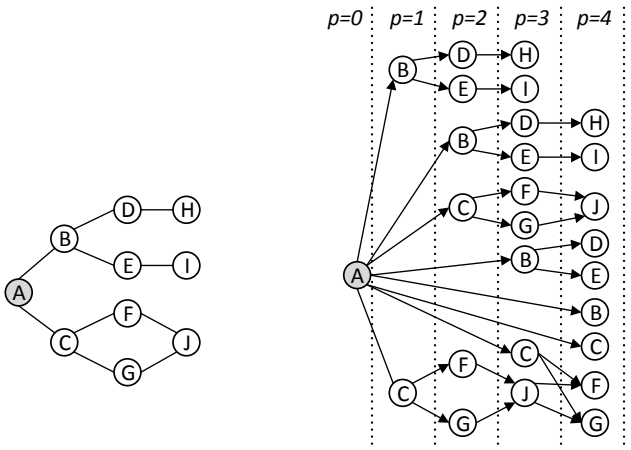
Ausbreitungsgraph

Zur Darstellung der Aktivierungshistorie, welche Aktivierungswerte und Ausbreitungsverhalten aller Pulse einer Aktivierungsausbreitung zu einer Anfrage dokumentiert, erstellen wir einen sogenannten Ausbreitungsgraphen. Ein *Ausbreitungsgraph* ist ein gerichteter Graph, dessen Quellknoten den oder die Startknoten der Anfrage darstellen und dessen Kanten und

Knoten mit denen des Ausgangsgraphen korrespondieren. Jeder Knoten im Ausbreitungsgraphen repräsentiert den Aktivierungszustand des korrespondierenden Knotens im Ausgangsgraphen in einem bestimmten Puls, während jede Kante für die Aktivierungsverteilung der korrespondierenden Kante des jeweiligen Pulses steht. Da Knoten mehrfach Aktivierung erhalten können, treten deren korrespondierende Knoten und Kanten im Ausbreitungsgraphen möglicherweise mehrfach auf. Sie sind dennoch eindeutig identifizierbar, da Annotationen an Kanten und Knoten zusätzliche Informationen tragen. Beispielsweise erhalten Knoten einen eindeutigen Identifizierer, eine Bezeichnung, den Typ, die Input-Aktivierung und den Aktivierungswert, Kanten hingegen den Typ, den Puls und die transportierte Aktivierung. In Abbildung 3.25 ist der Ausbreitungsgraph für das in Abbildung 3.21 (Seite 103) vorgestellte Beispielnetz im Sendemodus *FWL* dargestellt. Genau genommen entsprechen die Knoten, welche in jeder Pulsspalte dargestellt sind, den zuvor vorgestellten RecentReceiver-Knoten. Da im Sendemodus *FWL* nicht nur *RR*-Knoten Aktivierung verteilen, sind in der Abbildung einige Knoten auch über die Pulsspalten hinweg verbunden, z.B. verteilt Knoten A in jedem Puls Aktivierung an die Knoten B und C.

Das IR-System identifiziert im aktivierten Netz durch suchzielspezifische Filterung, wie in Abschnitt 3.4.1 beschrieben, die Knoten mit der stärksten Relevanz für die gestellte Anfrage, so dass diese dem Nutzer als Antwort beziehungsweise Empfehlung ausgegeben werden können. Die folgenden Schritte zur Erklärungsgenerierung für Empfehlungen widmen sich der Frage, warum ein bestimmter Knoten, den wir im Folgenden auch *Zielknoten* nennen, relevant für den oder die Startknoten ist. Dieses Vorgehen ist generell für jeden Knoten im Netz anwendbar und nicht beschränkt auf die Knoten, die als Empfehlung ausgewählt werden. Die Erklärung beschreibt, was den Zusammenhang zwischen Startknoten und Zielknoten ausmacht.

Zielknoten



(a) Ausgangsgraph (b) Ausbreitungsgraph

Abbildung 3.25: Ausgangsgraph zu Ausbreitungsgraph für das Beispiel in Abbildung 3.21

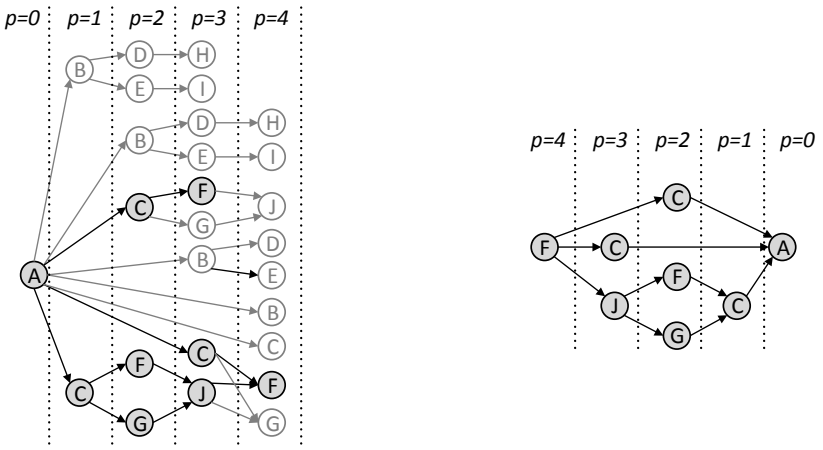
3.4.2.2 Minimierung des Problembereiches

Der Ausbreitungsgraph enthält die gesamte Aktivierungshistorie und damit auch Pfade, die weder direkt noch indirekt einen Beitrag zur Aktivierung eines Zielknotens leisten. Sei im vorliegenden Beispiel der Knoten F der Zielknoten, also jener Knoten, für den eine Erklärung generiert werden soll, so sind in Abbildung 3.26a jene Knoten grau markiert,

die während der Aktivierungsausbreitung einen Beitrag zu dessen Aktivierung leisten. Um die relevanten Pfade zu identifizieren, wird ein Backtracking-Algorithmus angewandt, der ausgehend vom Zielknoten alle Knoten und Kanten mit Aktivierungsbeteiligung identifiziert. Dabei muss berücksichtigt werden, dass es für jeden Knoten im Ausgangsgraphen mehrere korrespondierende Knoten im Ausbreitungsgraphen geben kann. Daher werden während des Backtrackings für jeden selektierten Knoten auch die weiteren Vorkommen des Knotens in vorherigen Pulsen selektiert, weil diese einen indirekten Beitrag leisten.

minimierter Ausbreitungsgraph

Die so zurückverfolgten, potenziell relevanten Pfade werden in einen minimierten Ausbreitungsgraphen überführt, der für das Beispiel in Abbildung 3.26b dargestellt ist. Der *minimierte Ausbreitungsgraph* ist ebenfalls ein gerichteter Graph, dessen Kantenrichtungen beim Backtracking derart umgekehrt werden, dass alle Pfade vom Zielknoten zu dem oder den Startknoten führen. Er ergibt sich aus ebenjenen Knoten und Kanten, die zur Aktivierung von Knoten F beitragen.



(a) Backtracking im Ausbreitungsgraphen

(b) minimierter Ausbreitungsgraph

Abbildung 3.26: Der zur Abbildung 3.25 gehörende a) Ausbreitungsgraph, in dem die durch Backtracking identifizierten Pfade, die zur Aktivierung von Knoten F beigetragen haben, markiert sind, sowie b) dessen minimierter Ausbreitungsgraph

3.4.2.3 Identifikation von Gruppen

Da die Größe der Netze und die Anzahl an Pulsen üblicherweise die des Beispiels übertreffen, sind selbst minimierte Ausbreitungsgraphen in der Regel sehr groß. Um einer Erklärung die notwendige Kürze zu verleihen und zudem eine besonders relevante und prägnante Erklärung zu generieren, werden im minimierten Ausbreitungsgraphen besonders markante Bereiche

Gruppen

gesucht, deren Beitrag zur Aktivierung des Zielknotens besonders groß ist. Diese Bereiche nennen wir im Folgenden Gruppen. Dafür identifizieren wir in einem ersten Teilschritt

Erklärungsknoten

die relevantesten Knoten, die wir Erklärungsknoten nennen. In einem zweiten Teilschritt bestimmen wir die relevantesten Pfade und Gruppen um diese Erklärungsknoten.

Auswahl der Erklärungsknoten

Zur Reduktion der Komplexität wird zunächst mittels eines Parameters $n_{max} \in \mathbb{N}_{>0}$ die Anzahl der gesuchten Erklärungsknoten festgelegt. Damit lässt sich Genauigkeit und Länge der später generierten Erklärung beeinflussen. Da n_{max} bei wenigen Pulsen nicht zu klein und bei vielen Pulsen nicht zu groß werden sollte, wird n_{max} durch eine nichtlineare Funktion in Abhängigkeit der Pulsanzahl bestimmt, hier die Wurzelfunktion der Pulsanzahl.

Kandidaten

Zudem wird die Menge der potenziellen Erklärungsknoten, die wir auch *Kandidaten* nennen, auf den semantischen Kern des Netzes beschränkt, da die Knoten und Kanten der Modellebene als wenig hilfreich für domänenspezifische Erklärungen angenommen werden. Durch eine domänenspezifische Voranalyse sind weitere Einschränkungen denkbar, beispielsweise die Exklusion bestimmter Ressourcentypen.

Für jeden Kandidaten, der im minimierten Ausbreitungsgraphen verbleibt, berechnen wir einen Relevanzwert, der dessen individuelle Relevanz für die Erklärung des Zielknotens bestimmt. Damit werden alle Kandidaten miteinander vergleichbar. Drei Faktoren, die sich aus der lokalen Netzstruktur des minimierten Ausbreitungsgraphen um einen Kandidaten ableiten, fließen in die Berechnung seines Relevanzwertes ein. Zum einen wird die Input-Aktivierung des Kandidaten berücksichtigt, wobei eine Abschwächung für Input-Aktivierung, die von Knoten der Modellebene eingeht, erfolgt. Zum anderen wird das zahlenmäßige Verhältnis eingehender Kanten von Knoten der Modellebene zu Knoten der Instanzebene berücksichtigt. Damit wird die Relevanz der Kandidaten gestärkt, die ihre Aktivierung größtenteils von Knoten aus dem semantischen Kern erhalten, unter der Annahme, dass sich domänenspezifische Knoten des semantischen Kerns besser für domänenspezifische Erklärungen eignen. Hiermit wird demnach fehlende Domänenspezifizität bestraft. Viele eingehende Kanten im minimierten Ausbreitungsgraphen bedeuten, dass der Kandidat im Ausbreitungsgraphen Aktivierung an viele Nachbarknoten verteilt hat. Um hier einen dämpfenden Effekt für starke Verteilerknoten zu bewirken, verwenden wie als dritten Faktor ein Maß für die Vernetzung des Kandidaten, ähnlich dem im Rahmen des Aktivierungsausbreitung beschriebenen Verzweigungsmodus *DEGREE* (siehe Formel 3.24, Seite 95).

Die Erklärungsknoten sind die n_{max} Kandidaten mit den höchsten individuellen Relevanzwerten. Da es für jeden Knoten des Ausgangsgraphen mehrere korrespondierende Knoten im Ausbreitungsgraphen geben kann, achten wir bei der Selektion darauf, dass maximal ein korrespondierender Knoten ausgewählt wird, in dem Fall der korrespondierende Knoten mit dem größten Relevanzwert.

Auswahl der relevantesten Pfade und Gruppen

Die identifizierten Erklärungsknoten befinden sich im minimierten Ausbreitungsgraphen zwischen Start- und Zielknoten. Die Menge aller potenziell relevanten Pfade ergibt sich aus allen Pfaden, die Start- und Zielknoten über einen relevantesten Knoten verbinden. Zudem definieren wir Teilpfade, die beidseitig eines Erklärungsknoten zum einen Zielknoten und Erklärungsknoten und zum anderen Erklärungsknoten und Startknoten verbinden. Die Aufteilung in diese Teilpfade unterstützt die Erstellung der Erklärung vor allem im sprachlichen Aufbau. So wird die Nähe zu Ziel- oder Startknoten bei der Rekonstruktion der Aussagen auf den Pfaden verwendet, insbesondere zur Auswahl sinnvoller Subjekte und Objekte im Satzbau. Für eine Erklärung eignen sich unter der potenziell hohen Anzahl an Teilpfaden nur die relevantesten. Die Wichtigkeit der Pfade bestimmen wir über die transportierte Aktivierung auf ihren Kanten. Analog zum Parameter n_{max}, mit dem wir die Länge der

Erklärung beschränken, begrenzen wir mittels eines weiteren Parameters m die Tiefe der Erklärung in Form der maximalen Anzahl an Pfaden, die je Erklärungsknoten und Teilpfad gewählt wird. Dieser Parameter wird in Abhängigkeit von der Pulsanzahl des jeweiligen Teilpfades berechnet. Für jeden Erklärungsknoten ergeben sich beidseitig aus den relevantesten Pfaden zwei Teilgraphen, die bereits vorgestellten Gruppen. Diese Gruppen unterteilen wir entsprechend der Lage zu ihrem Erklärungsknoten in *Startgruppe* und *Zielgruppe*. Seien G und J identifizierte Erklärungsknoten, so sind in Abbildung 3.27 deren Ziel- und Startgruppen dargestellt.

Startgruppe
Zielgruppe



(a) für Erklärungsknoten G (b) für Erklärungsknoten J

Abbildung 3.27: Die zum minimierten Ausbreitungsgraphen gehörenden Ziel- und Startgruppen für a) Erklärungsknoten G, sowie b) Erklärungsknoten J

3.4.2.4 Mustererkennung und Erstellen der Erklärung

Die Anordnung der Knoten und Kanten der Gruppen weisen Muster auf, die bei der sprachlichen Gestaltung der Erklärung die Sinnhaftigkeit und Korrektheit unterstützen können und vor allem zur Kürze und Ausdrucksstärke beitragen und somit die Verständlichkeit erhöhen. Abbildung 3.28 zeigt vier Muster sowie die Formation, die auf kein spezielles Muster schließen lässt. Die Muster wurden zudem basierend auf der Stärke ihrer Verbindungskraft und ihrem potenziellen Nutzen priorisiert.

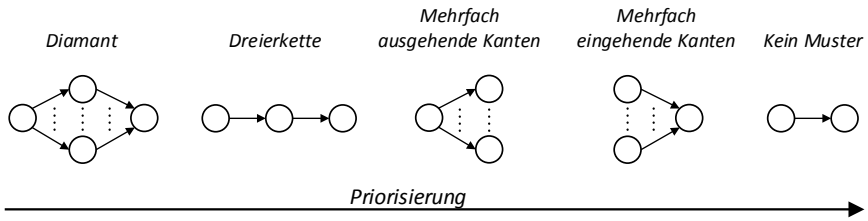


Abbildung 3.28: Identifizierbare Muster und deren Priorisierung

Für jedes Muster legen speziell angepasste Grammatiken den Aufbau der Sätze fest. Dabei basieren alle Erklärungen auf dem grundsätzlichen Aufbau von Aussagen im RDF-Graphen (Subjekt-Prädikat-Objekt) des verwendeten semantischen Netzes. In die Grammatiken fließt zudem Wissen über die Bedeutung der unterschiedlichen Properties und Knotentypen sowie deren Zugehörigkeit zur Modell- oder Instanzebene. Zudem wird sprachliche Kompaktheit beispielsweise durch sinnvolles Zusammenfassen von Knoten gleichen Typs erreicht. Die textuellen Erklärungen erstellen wir basierend auf den Mustern für jede Ziel- und Startgruppe. Sie bestehen je nach gewählten Parametern für Breite und Tiefe der Erklärung aus mehreren Sätzen.

3.4.2.5 Evaluation der Methode

Die Erklärungskomponente wurde im Rahmen der Masterarbeit [Mic16] einer Evaluation unterzogen, die zwei unterschiedliche Ansätze verfolgt. Aufgrund des beschränkten Rahmens einer Masterarbeit sind diese zwar einfach gehalten, vermitteln aber dennoch einen Eindruck über die Nützlichkeit der vorgestellten Methode.

Zum einen verwendet Michalke einen erklärungszielabhängigen Evaluationsansatz nach Tintarev und Masthoff [TM11] und argumentiert, dass die Ziele Transparenz und Prüfbarkeit sowie Erhöhung von Effektivität und Effizienz mit der vorgestellten Erklärungskomponente erfüllt werden. Zum anderen führt sie eine Bewertung der Güte durch. Dafür werden drei Anfragen auf jeweils vier semantischen G&R-Netzen ausgeführt. Parallel wird für jeden der zwölf Durchläufe manuell das erwartete Ergebnis festgelegt. Erwartete und generierte Erklärungen werden dann bezüglich der gemeinsamen Knoten und Kanten verglichen und durch die Qualitätsmaße Genauigkeit und Trefferquote bewertet. Eine Genauigkeit von 100% bedeutet, dass die generierte Erklärung keine unerwarteten Knoten beziehungsweise Kanten enthält. Eine Trefferquote von 100% bedeutet, alle erwarteten Knoten beziehungsweise Kanten sind Teil des generierten Ergebnisses. Für die verschiedenen Durchläufe liegt die Genauigkeit der Knoten und Kanten zwischen 33 und 100%, deren Trefferquote zwischen 60 und 100%.[40] Da die Qualität einer Erklärung mehr von einem hohen Anteil erwarteter Elemente in der Erklärung profitiert als sie durch die zusätzlichen Knoten und Kanten verliert, ist die Trefferquote im untersuchten Feld wichtiger als die Genauigkeit. Mithilfe des gewichteten F_2-Maßes zwischen 62 und 100% zeigt Michalke, dass für die meisten Durchläufe ein akzeptables Maß an Güte erreicht wird.

Der Hauptgrund für Durchläufe mit geringer Genauigkeit und Trefferquote ist, dass es zumeist nicht nur eine mögliche Erklärung gibt, sondern stattdessen unterschiedliche Erklärungen ein Erklärungsbedürfnis befriedigen können. Die potenzielle Existenz mehrerer Knoten und Kanten mit gleichem individuellem Relevanzwert bei gleichzeitiger Restriktion von Länge und Tiefe der Erklärungen führt dazu, dass nicht immer alle Knoten und Kanten in einer generierten Erklärung verwendet werden und die erwartete Erklärung demnach ebenfalls schwierig manuell zu erstellen ist.

[40] Das Erreichen optimaler Genauigkeitswerte von 100% ist zwar wünschenswert, in komplexen Evaluationen allerdings meist unüblich. Dass in der hier vorgestellten Evaluation 100% erreicht werden, liegt an dem vereinfachten Evaluationsaufbau. Eine komplexere Betrachtung, in der die Genauigkeit nicht ausschließlich auf dem Enthaltensein bestimmter Knoten und Kanten basiert, sondern auch die Bewertung der Sinnhaftigkeit der Erklärungen durch Nutzer einschließt, wird seltener zu einer Genauigkeit von 100% führen.

3.4.3 Zusammenfassung und Beitrag

In diesem Abschnitt 3.4 haben wir beschrieben, wie wir die Ergebnisse des Retrievalverfahrens für den Nutzer aufbereiten, so dass dieser Empfehlungen bezogen auf sein ursprüngliches Informationsbedürfnis erhält. Während der Aktivierungsausbreitung hat jeder Knoten im semantischen Netz ein Maß an Aktivierung erhalten, welches Aufschluss über dessen Relevanz zur gestellten Anfrage liefert. Zur Auswertung dieser Ergebnisse nutzen wir eine suchzielspezifische Filterung und erstellen eine nach Relevanz sortierte und gekürzte Ergebnisliste, um die Anfrage bestmöglich zu beantworten. Die suchzielspezifische Filterung wird insbesondere durch den struktur- und datenmodellbasierten Ansatz möglich.

Zudem haben wir ein Verfahren entwickelt, welches für die Ergebnisse des IR-Systems automatisiert textbasierte Erklärungen generiert, so dass der Nutzer das Ergebnis als Empfehlung für sein Informationsbedürfnis nachvollziehen kann. Der vorgestellte Ansatz macht sich die Aktivierungshistorie zunutze und rekonstruiert graphbasiert, wie ein Ergebnis zustande kam. Erklärungen erhöhen üblicherweise die Nutzerakzeptanz und das Vertrauen in das jeweilige System, da es Transparenz schafft und zur Prüfbarkeit beiträgt. Dadurch wird unser domänenspezifisches IR-System für G&Rs aufgewertet.

Mit dem Abschnitt 3.4 unterstützen wir den vierten und letzten Schritt der in Abschnitt 3.1.2.4 vorgestellten Methode zum Aufbau domänenspezifischer IR-Systeme. In diesem vierten Schritt konfiguriert der Entwickler des IR-Systems die Ausgabe der Ergebnisse und Erklärungen des IR-Systems. Wir haben generische Lösungen vorgestellt, die vom Entwickler des IR-Systems angepasst werden können. Beispielsweise kann dieser den Umfang der auszugebenden Ergebnislisten festlegen oder die beschriebene Länge und Tiefe der Erklärungen anwendungsspezifisch konfigurieren.

Wir haben die in diesem Kapitel vorgestellten Techniken und Konzepte als prototypisches IR-System implementiert, unter anderem um damit eine Evaluation der Konzepte zu ermöglichen. Im folgenden Abschnitt beschreiben wir diese prototypische Umsetzung.

3.5 Prototypische Umsetzung

Die in dieser Arbeit vorgestellten Konzepte und Techniken haben wir in einem eigenständigen Prototyp umgesetzt, den wir Hazard Analysis and Risk assessment dEcision Support Tool (HARvESTer) nennen. Der Prototyp enthält Funktionalitäten, die zwei Nutzergruppen adressieren. Zum einen haben wir alle im generischen Rahmenwerk verfügbaren Konzepte und Techniken implementiert, mit deren Hilfe Entwickler domänenspezifische IR-Systeme erstellen können. Zum anderen enthält der Prototyp ein domänenspezifisches IR-System für G&Rs, welches deren Nutzer bei der Durchführung von G&Rs unterstützt. Dieser Abschnitt widmet sich zunächst den implementierten Funktionalitäten für beide Nutzergruppen und anschließend der Beschreibung des Prototypaufbaus.

3.5.1 Nutzersichten

Im Folgenden stellen wir die Funktionalitäten aus Sicht der beiden Nutzergruppen detaillierter vor.

Nutzer des IR-Systems: Das IR-System für G&Rs richtet sich an Durchführende von G&Rs, die mithilfe einer grafischen Benutzeroberfläche die wesentlichen Funktionalitäten des domänenspezifischen IR-Systems nutzen können. Diese Funktionalitäten umfassen unter anderem:

- Laden eines Datensatzes aus G&Rs
- Browsen im G&R-Datensatz
- Absetzen von Anfragen für Empfehlungen über das Kontextmenü, wobei die Suchziele abhängig vom jeweiligen Ausgangspunkt sinnvoll eingeschränkt werden
- Absetzen von Anfragen für Erklärungen von Ergebnissen über das Kontextmenü
- Browsen der Ergebnislisten
- Konfiguration der Aktivierungsausbreitung und der Ausgabe (z.B. Länge der Ergebnisliste, Erklärungen)

Dieser Teil der Implementierung verfolgt das Ziel, die entwickelten Konzepte und Techniken sowie das geplante Vorgehen für potenzielle Anwender anschaulich zu gestalten und mithilfe der grafischen Benutzeroberfläche zeigen, ausprobieren und diskutieren zu können. Er dient zudem als Gesprächsgrundlage, um die Nutzerakzeptanz einzuschätzen. Da im untersuchten industriellen Umfeld (siehe Abschnitt 3.1.1.2) das Entwicklungswerkzeug für G&Rs ANSYS medini analyze verwendet wird, haben wir Aussehen und Handhabung (*Look and Feel*) von ANSYS medini analyze in unserem Prototyp nachgebildet. Das Entwicklungswerkzeug ANSYS medini analyze ist eine kommerzielle Eclipse Rich Client Platform (RCP) basierte Applikation, die in Java umgesetzt ist.[41] Daher haben wir das IR-System für G&Rs ebenfalls als Eclipse RCP Applikation in Java konzipiert.

Wir haben die Retrieval-Funktionalität so in das Werkzeug integriert, dass die Generierung von Empfehlungen und Erklärungen direkt über das Kontextmenü angesteuert werden kann. Dadurch erfordert die Nutzung keinen manuellen Zusatzaufwand und keine relevanten Vorkenntnisse. Diese Umsetzung ist im praktischen Anwendungsfeld auf gute Nutzerakzeptanz getroffen, da die zusätzliche Integration des IR-Systems das etablierte Vorgehen nicht stört, sondern lediglich zusätzliche Unterstützung anbietet. In der Nutzeranwendung erlauben wir

[41] Das bezieht sich auf den Zeitpunkt der Entstehung dieses Prototyps in dieser Arbeit.

die Konfiguration der Aktivierungsausbreitung und der Ergebnis- und Erklärungsausgabe. Im realen Einsatz stellt das für den Nutzer des IR-Systems keine notwendige Funktionalität dar. In unserer prototypischen Umsetzung diente dies allerdings der Veranschaulichung der verschiedenen Konzepte.

Entwickler des IR-Systems: Im Prototyp sind zusätzliche Funktionalitäten enthalten, die den Entwickler domänenspezifischer IR-Systeme adressieren. Im Gegensatz zum IR-System für G&Rs, welches über eine grafische Benutzeroberfläche gesteuert wird, haben wir für die Aufgaben des Entwicklers auf eine grafische Benutzerschnittstelle verzichtet. Stattdessen können die verschiedenen Entwickleranwendungen über die Kommandozeile aufgerufen werden. Die Funktionalitäten umfassen unter anderem:

- Erstellen der semantischen Basisnetze sowie deren semantische Anreicherungen und Reduktionen
- Erstellen der semantischen Netzwerkskelette (siehe Abschnitt 5.2)
- Simulation der Aktivierungsausbreitung auf semantischen Netzwerkskeletten (siehe Abschnitt 5.4)
- Vorbereitung, Durchführung und Auswertung der Evaluation
- Optimierung von Konfigurationen unter Verwendung evolutionärer Algorithmen

3.5.2 Aufbau

In Abbildung 3.29 ist die Architektur unseres Prototyps dargestellt. Im Folgenden stellen wir seinen strukturellen Aufbau mit den wichtigsten Komponenten und Besonderheiten kurz vor.

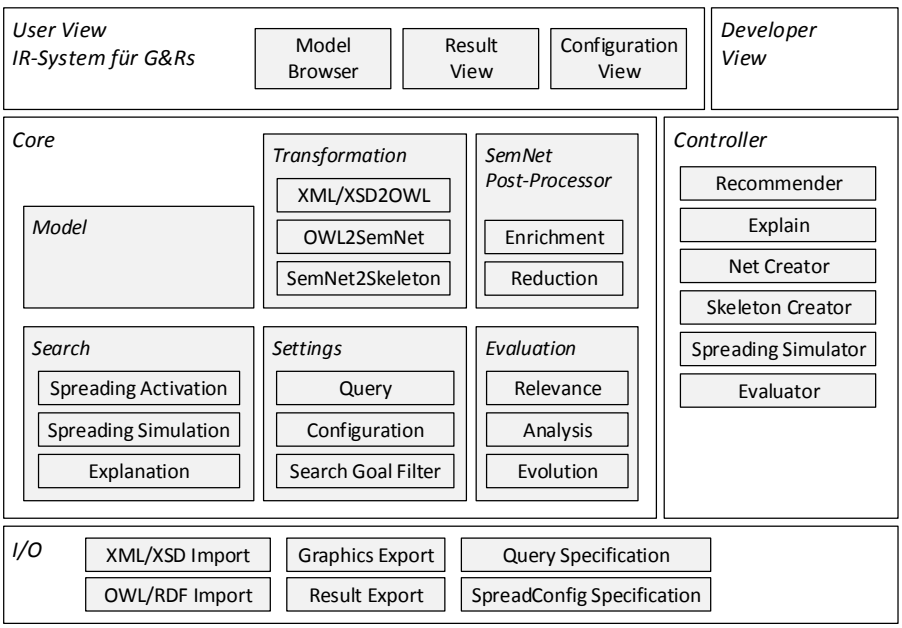


Abbildung 3.29: Architektur des entwickelten Prototyps

View-Komponenten

User View: Für das IR-System für G&Rs stellen wir eine grafische Benutzerschnittstelle zur Verfügung. Ein *Model Browser* ermöglicht das Browsen in den G&Rs des Datensatzes. Ein Tree-View zeigt die hierarchische Strukturierung der G&Rs, in einem Table-View sind deren Inhalte tabellenartig abgebildet. Das Erscheinungsbild ist angelehnt an die Benutzeroberfläche von ANSYS medini analyze. Im *Result View* erscheint die Ausgabe der Ergebnisse von Anfragen für Empfehlungen und die Ausgabe von Erklärungen für die Ergebnisse. Der *Configuration View* ermöglicht Anpassungen in den Konfigurationen der Aktivierungsausbreitung und verschiedene Einstellungen für die Ergebnisausgabe.

Developer View: Wir stellen eine eigene Entwicklerschnittstelle zur Verfügung, mit der die oben genannten Funktionalitäten aufgerufen werden können.

Core-Komponente

Model: Diese Komponente beschreibt die interne Repräsentation der semantischen Netze. Für die Knoten und Kanten des Netzes verwenden wir das Open Source Java Framework Apache Jena [Jena], welches Programmierschnittstellen und umfassende Unterstützung im Umgang mit RDF und OWL bietet. Jena ermöglicht unter anderem das Einlesen, Speichern und Modifizieren von RDF- und OWL-Daten, sowie die Anwendung von Abfragesprachen (SPARQL-API). In Wrapper-Klassen fügen wir den Jena-Knoten und -Kanten zusätzliche Eigenschaften hinzu. Für das Netz verwenden wir die Java-Bibliothek JGraphT [JGraphT], deren Graphstrukturen mit den Wrapper-Klassen für Knoten und Kanten umgehen können und zudem nützliche Methoden für die Graphverarbeitung zur Verfügung stellen.

Transformation: Diese Komponente enthält Unterkomponenten, die verschiedene Transformationen ermöglichen. *XML/XSD2OWL* erlaubt die Transformation von XML-Schema zu OWL-Model und XML zu OWL-Instanzen gemäß den Ausführungen in Abschnitt 2.2.3. Sie basiert auf der dort beschriebenen Java-Bibliothek *Ontmalizer* und unseren in Abschnitt 3.2.2.4 beschriebenen Anpassungen und Erweiterungen. Der Umgang mit RDF- und OWL-Daten, insbesondere das Modifizieren und Speichern, wird auch im *Ontmalizer* mit Apache Jena realisiert. *OWL2SemNet* ermöglicht das Einlesen von RDF- und OWL-Daten und deren Abbildung in der internen Repräsentation für semantische Netze. *SemNet2Skeleton* erzeugt aus semantischen Netzen deren semantische Netzwerkskelette, eine Abstraktion semantischer Netze, die wir in Kapitel 5 präsentieren.

SemNet Post-Processor: Diese Komponente ermöglicht die Nachbearbeitung semantischer Netze und erlaubt unterschiedliche Anreicherungs- und Reduktionsstufen der Netze. Die Unterkomponente *Enrichment* ermöglicht die Erweiterung der Netze um Keywords und Synonyme gemäß den Beschreibungen in Abschnitt 3.2.3. Dabei kommen Techniken des NLP zum Einsatz, für die wir das Stanford CoreNLP Natural Language Processing Toolkit [MSB+14] einsetzen. Die Stammformreduktion nehmen wir mit dem deutschsprachigen SnowballStemmer der Bibliothek Apache OpenNLP [OpenNLP] vor. Die Unterkomponente *Reduction* entfernt Knoten und Kanten des semantischen Netzes gemäß den Beschreibungen in Abschnitt 3.2.4.

Search: Diese Komponente enthält Unterkomponenten für die verwendeten Suchalgorithmen. *Spreading Activation* stellt den Algorithmus der Aktivierungsausbreitung mit den in Abschnitt 3.3 formalisierten Aktivierungsverteilungsfunktionen bereit sowie alle in dieser Arbeit vorgestellten Techniken und Konzepte zur Aktivierungsausbreitung, mit denen der Algorithmus parametrisiert werden kann. Die Unterkomponente *Spreading Simulation* enthält überarbeitete Aktivierungsverteilungsfunktionen gemäß dem in Abschnitt 5.4 vorgestellten Näherungsverfahren. Die Unterkomponente *Explanation* ermöglicht die Erklärungsgenerierung. Sie wurde im Rahmen der Masterarbeit von Michalke [Mic16] in die Eclipse RCP Applikation eingebunden.

Settings: Diese Komponente ermöglicht den Umgang mit den für die semantische Suche notwendigen Informationen und Eingaben. Die Unterkomponenten *Query* und *Configuration* verarbeiten Anfragen und Konfigurationen für die Aktivierungsausbreitung. *Search Goal Filter* ermöglicht die suchzielspezifische Filterung.

Evaluation: Diese Komponente unterstützt die Vorbereitung, Durchführung und Auswertung der Evaluation. *Relevance* enthält Funktionalitäten zum Umgang mit objektiven Relevanzen und Systemrelevanzen sowie den Distributionen, die beide vereinen. *Analysis* ermöglicht die Auswertung der Distributionen anhand der verwendeten Bewertungsmaße EP und NDCG (siehe Abschnitt 2.1.3.3). *Evolution* ermöglicht die Optimierung der Konfigurationen für die Aktivierungsausbreitung mittels evolutionärer Algorithmen.

Controller-Komponente

Diese Komponente erwartet die Eingaben aus der Benutzeroberfläche des IR-Systems für G&Rs, nutzt zur Ausführung die Core-Komponenten und zur Ausgabe die grafische Benutzeroberfläche oder die I/O-Komponente. *Recommender* und *Explain* gehören zum IR-System für G&Rs. Über die Entwicklerschnittstelle werden der *Net Creator* zum Erstellen der gewünschten Netzvariante, der *Skeleton Creator* zum Aufbau semantischer Netzwerkskelette, der *Spreading Simulator* zur Simulation der Aktivierungsausbreitung auf dem Netzwerkskelett und der *Evaluator* zur Durchführung von Evaluationen angesprochen.

I/O-Komponente

Diese Komponente ermöglicht die Ein- und Ausgabe von Daten verschiedener Formate. *XML/XSD Import* verarbeitet XML/XML-Schema-Dateien des industriellen Datensatzes. *OWL/RDF Import* ermöglicht das Einlesen von OWL-Dateien zum Aufbau semantischer Netze. *Query Specification* und *SpreadConfig Specification* erlauben das Einlesen von Anfragen und Konfigurationen für die Aktivierungsausbreitung aus Spezifikationsdateien zur Abarbeitung während der Evaluation. *Results Export* und *Graphics Export* ermöglichen die Ausgabe von Ergebnissen. Dazu gehören Suchergebnisse der Aktivierungsausbreitung, Evaluationsergebnisse und visuelle Darstellungen der graphbasierten Daten.

Im Prototyp sind die Konzepte und Techniken, die wir in Kapitel 3 vorgestellt haben, umgesetzt. Damit ermöglicht der Prototyp die Durchführung einer Evaluation, in der wir zum einen die qualitative Bewertung des IR-Systems für G&Rs vornehmen werden und zum anderen untersuchen können, welchen Einfluss die neuen Techniken auf die Qualität der Ergebnisse haben. Dieser Evaluation widmen wir uns im folgenden Kapitel.

4 Evaluation

In diesem Kapitel präsentieren wir die Ergebnisse unserer Evaluation, die vor allem die qualitative Bewertung des IR-Systems für G&Rs und die Analyse der Einflussfaktoren in den Fokus nimmt. Dafür präsentieren wir zunächst unseren Evaluationsplan. Insbesondere stellen wir die Evaluationsfragen vor, denen sich die anschließende Evaluation widmet. Im ersten Teil der Evaluation analysieren wir die Urteilerübereinstimmung. Wir werden zeigen, dass ein ausreichendes Maß an Einigkeit zwischen den Domänenexperten bezüglich der beurteilten Szenarien erreicht wird. Das ist die Basis zur Verwendung ihrer objektiven Relevanzbewertungen in der darauffolgenden qualitativen Evaluation der Ergebnisse des IR-Systems. Diese qualitative Bewertung beginnen wir mit einer explorativen Analyse zufällig generierter Konfigurationen. Sie erlaubt einen ersten Überblick und zeigt deutlich die Notwendigkeit einer geeigneten Parametrisierung der Konfigurationen und Auswahl der Netze. Mittels evolutionärer Algorithmen optimieren wir anschließend Konfigurationen und analysieren, unter welchen Bedingungen gute Ergebnisse erreicht werden konnten. Wir werden beispielsweise zeigen, dass die von uns eingeführten Ausbreitungsmodi zur zusätzlichen Steuerung der Aktivierungsausbreitung sowie die semantischen Anreicherungen und Reduktionen der Netze essenziell für qualitativ gute Ergebnisse des IR-Systems sind. Die optimierten Konfigurationen verwenden wir zur Bewertung der Qualität unseres IR-Systems für G&Rs. Wir werden zeigen, dass es mit geeigneten Konfigurationen relevante und nützliche Ergebnisse liefert.

4.1 Evaluationsplan

Im Folgenden stellen wir unseren komparativen Evaluationsaufbau vor, der den Vergleich der Ergebnisse des IR-Systems bezüglich der folgenden drei Dimensionen erlaubt: Netzvarianten, Anfragen und Konfigurationsparameter. Ferner erläutern wir unser Vorgehen bei der objektiven Relevanzbestimmung, da wir die Urteile mehrerer Domänenexperten in die Evaluation einbeziehen. Abschließend präsentieren wir in Abschnitt 4.1.4 eine detaillierte Aufschlüsselung der Evaluationsfragen und der Strategien zu ihrer Beantwortung.

4.1.1 Komparative Evaluationsdurchführung

Für die Evaluation verfolgen wir einen komparativen Evaluationsansatz gemäß den Beschreibungen in Abschnitt 2.1.3.2. Abbildung 2.2 (Seite 16) zeigt die grundsätzliche Evaluationsdurchführung für ein Informationsbedürfnis, welches mit dem IR-System auf einem semantischen Netz durchgeführt wird. Jedes Informationsbedürfnis wird in eine Anfrage an das IR-System umgewandelt. Jede Anfrage wird vom IR-System bearbeitet, als Ergebnis wird eine Liste mit Relevanzwerten ausgegeben. In absteigender Reihenfolge stellen diese die Empfehlungen des IR-Systems dar. Die gleichen Informationsbedürfnisse werden zur Relevanzbewertung an die Beurteiler gegeben. Diese entscheiden, welche Elemente sie subjektiv für relevant halten. Zusätzlich zum grundsätzlichen Aufbau in Abbildung 2.2 (Seite 16), führen wir eine notwendige Vor- und Nachbearbeitung im Rahmen der Erstellung einer objektiven Relevanz durch. Zum einen werden Pooling und Maßnahmen zur Reduktion der Elementlisten vorgenommen. Das ist notwendig, um eine Angemessenheit des Aufwands für die Rater zu gewährleisten. Zum anderen müssen in der Nachbearbeitung die zuvor reduzierten Elemente wieder in die Elementlisten aufgenommen werden und die unterschiedlichen Ratermeinungen zu einer gemeinsamen objektiven Relevanz je Anfrage zusammengefügt werden. Diese Vor- und Nachbearbeitung wird in Abschnitt 4.1.2 detaillierter beleuchtet. Nach dem Erstellen

der objektiven Relevanz kann durch Gegenüberstellung mit der Systemrelevanz ein Durchlauf des IR-Systems bewertet werden, also ein Informationsbedürfnis auf einem Netz unter Verwendung einer einzigen Konfiguration.

In unserer Evaluation werden allerdings mehrere Informationsbedürfnisse untersucht, die jeweils mit verschiedenen Konfigurationen des IR-Systems auf jeweils unterschiedlichen Varianten der semantischer Netze durchgeführt werden. Abbildung 4.1 stellt die in dieser Evaluation untersuchten Dimensionen dar. Die Konfigurationen des IR-Systems beziehen sich auf die Parametrisierung der Aktivierungsausbreitung. Im Vergleich zum zweidimensionalen komparativen Evaluationsaufbau in Abbildung 2.3 (Seite 16), bei dem die Ergebnisse verschiedener IR-Systeme für verschiedene Informationsbedürfnisse miteinander verglichen werden können, stellt dieser Aufbau eine Erweiterung um die Dimension der Netzvarianten dar. Somit ist ein Vergleich der verschiedenen Ausbaustufen der semantischen Netze möglich. Das ist notwendig, um die Frage zu beantworten, welche Ausbaustufe die Qualität der Ergebnisse begünstigt und welche einen gegenteiligen Effekt auf die Ergebnisse hat.

Ermöglicht wird dieser Vergleich dadurch, dass die potenzielle Ergebnismenge für alle Netzvarianten, trotz Anreicherungen und Reduktionen, übereinstimmt. In keiner Netzvariante werden Instanzknoten des semantischen Kerns entfernt oder hinzugefügt. Anreicherungen und Reduktionen werden innerhalb des domänenspezifischen semantischen Kerns ausschließlich durch Hinzufügen von Kanten und Literalknoten vorgenommen. Da außerhalb des semantischen Kerns befindliche Knoten ohnehin nicht Teil der Ergebnismenge sein können, beeinträchtigen Änderungen auf Modellebene nicht die Vergleichbarkeit der Ergebnisse.

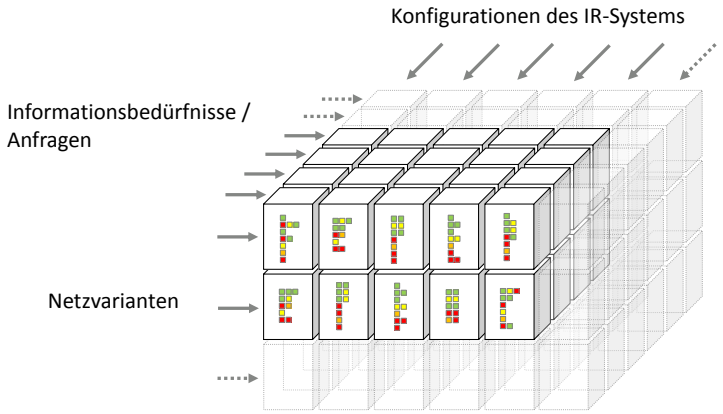


Abbildung 4.1: Dimensionen der komparativen Evaluation in dieser Arbeit

Datensatz

Für die Evaluation verwenden wir 155 abgeschlossene G&Rs der Firma Daimler AG. Eine Beschreibung des Datensatzes befindet sich in Abschnitt 3.1.1.2.

Netzvarianten

Die Tabelle 4.1 gibt einen Überblick darüber, mit welchen semantischen Anreicherungen und Reduktionen die in der Evaluation verwendeten Netzvarianten erstellt wurden. Wir untersuchen damit sechs Kombinationen semantischer Anreicherungen und drei Ausprägungen von Reduktionen, die wir im Folgenden Anreicherungs- und Reduktionsstufen nennen. Anreicherungsstufe 1 und Reduktionsstufe 1 stehen für die nicht angereicherte beziehungsweise

nicht reduzierte Basisvariante. Die Benennung der Netzvarianten enthält die Information zur jeweiligen Anreicherungs- und Reduktionsstufe, z.B. steht *N5-2* für die Kombination von Anreicherungsstufe 5 und Reduktionsstufe 2. In Tabelle 4.2 werden zudem einige Eigenschaften bezüglich ihrer Größe zusammengefasst.[42]

Netzvariante	**Anreicherungen**			**Reduktionen**	
	Synonyme	**Keywords**			
		einfach	**erweitert**	**einfach**	**erweitert**
N1-1	-	-	-	-	-
N1-2	-	-	-	x	-
N1-3	-	-	-	-	x
N2-1	x	-	-	-	-
N2-2	x	-	-	x	-
N2-3	x	-	-	-	x
N3-1	-	x	-	-	-
N3-2	-	x	-	x	-
N3-3	-	x	-	-	x
N4-1	-	-	x	-	-
N4-2	-	-	x	x	-
N4-3	-	-	x	-	x
N5-1	x	x	-	-	-
N5-2	x	x	-	x	-
N5-3	x	x	-	-	x
N6-1	x	-	x	-	-
N6-2	x	-	x	x	-
N6-3	x	-	x	-	x

Tabelle 4.1: Netzvarianten für die Evaluation

Netzvariante	**# Tripel/Kanten**	**# Knoten**	**# Literalknoten**
N1-1	131.505	49.543	23.119
N1-2	106.128	37.523	11.283
N1-3	79.906	37.483	11.283
N2-1	142.187	50.363	23.669
N2-2	116.810	38.344	11.833
N2-3	90.318	38.303	11.833
N3-1	214.448	57.040	30.616
N3-2	189.071	45.020	18.780
N3-3	162.849	44.980	18.780
N4-1	266.110	59.699	33.275
N4-2	240.733	47.679	21.439
N4-3	214.511	47.639	21.439
N5-1	225.130	57.860	31.166
N5-2	199.753	45.841	19.330
N5-3	173.261	45.800	19.330
N6-1	276.792	60.518	33.824
N6-2	251.415	48.499	21.988
N6-3	224.923	48.458	21.988

Tabelle 4.2: Eigenschaften der evaluierten Netzvarianten

[42] Es treten kleine Abweichungen zur Knoten- und Kantenanzahl in den Abschnitten 3.2.2 bis 3.2.3 auf, da im Rahmen der Evaluation bereits für das Basisnetz N1-1 alle Erweiterungen auf Modellebene vorgenommen wurden, die für die potenziellen semantischen Anreicherungen erforderlich sind.

Informationsbedürfnisse

Wir untersuchen acht Informationsbedürfnisse, die das IR-System als Anfragen verarbeitet. Diese Informationsbedürfnisse haben wir so erstellt, dass sie sich auf verschiedene Systeme des verwendeten Datensatzes beziehen und unterschiedliche Typen von Informationsbedürfnissen abdecken. Sie sind realistische Informationsbedürfnisse, die beim Erstellen von G&Rs auftreten können, was wir im Gespräch mit einem Entwickler überprüft haben. Eine detaillierte Beschreibung der einzelnen Informationsbedürfnisse und deren Kontexte ist aus Gründen der Geheimhaltung an dieser Stelle leider nicht möglich. Die Tabelle 4.3 soll dennoch einen Eindruck vermitteln, aus welchen Systemen und welchen Domänen die Informationsbedürfnisse stammen, sowie welche Anfragetypen die Informationsbedürfnisse abdecken.

Anfrage	System	Teildomäne	Ausgangspunkt	Suchziel	Kontext
Q23	Elektrischer Fensterheber	Innenraum	Funktion	Funktion	Systemname, Systembeschreibung, Funktion
Q35	Außenlicht	Innenraum	Funktion	Fehlfunktion	Systemname, Systembeschreibung, Funktion
Q46	Elektrisches Antriebssystem	Powertrain	Fehlfunktion	Sicherheitsziel	Systemname, Systembeschreibung, Funktion, Fehlfunktion
Q54	Parksperre	Fahrdynamik	Gefährdungssituation, Fehlfunktion, Betriebssituation	Gefährdungssituation (erwartete Reaktion)	Systemname, Systembeschreibung, Funktion, Fehlfunktion, Betriebssituation, Effekt
Q61	Schiebehebedach	Innenraum	Gefährdungssituation, Fehlfunktion, Betriebssituation	Gefährdungssituation (Effekt)	Systemname, Systembeschreibung, Funktion, Fehlfunktion, Betriebssituation
Q64	Wischer	Innenraum	Gefährdungssituation, Fehlfunktion, Betriebssituation	Gefährdungssituation (Effekt)	Systemname, Systembeschreibung, Funktion, Fehlfunktion, Betriebssituation
Q81	Sitzheizung/-lüftung	Innenraum	Gefährdungssituation, Fehlfunktion, Betriebssituation	IsoAsil (S+C, inkl. Kommentar)	Systemname, Systembeschreibung, Funktion, Fehlfunktion, Betriebssituation, Effekt
Q85	Heckscheibenrollo	Innenraum	Gefährdungssituation, Fehlfunktion, Betriebssituation	IsoAsil (S+C, inkl. Kommentar)	Systemname, Systembeschreibung, Funktion, Fehlfunktion, Betriebssituation, Effekt

Tabelle 4.3: Evaluierte Informationsbedürfnisse

Die untersuchten Informationsbedürfnisse entstammen Systemen der Teildomänen Innenraum, Fahrdynamik und Powertrain. Die Mehrzahl der Informationsbedürfnisse wird der Teildomäne Innenraum zugeordnet, da diese den größten Anteil im Datensatz ausmachen (siehe Tabelle 3.1, Seite 55). Die Zuordnung zu den Teildomänen ist lediglich informativ, da die semantische Suche ohnehin nicht nur innerhalb der Domänengrenzen stattfindet. Stattdessen werden Empfehlungen system- und domänenübergreifend generiert. Beispielsweise werden Gefährdungen durch Brand im Innenraum bei Heizungssystemen wie der Stand- oder Sitzheizung betrachtet. Sie spielen aber auch im Powertrain beim Batteriemanagement oder in induktiven Ladesystemen eine Rolle. Ähnliche domänenübergreifende Betrachtungen existieren beispiels-

weise für die Gefährdung durch Einklemmen. Jedes Informationsbedürfnis der Tabelle 4.3 wird als Anfrage vom IR-System bearbeitet. Die aufgelisteten Ausgangspunkte spiegeln die Ressourcentypen der Startknoten für den Algorithmus wider. Das Suchziel entspricht dem gesuchten Ressourcentyp. Liegt der Fokus des Suchziels auf einem speziellen Prädikat, so wird dies in Klammern vermerkt. Als Beispiel werden in Q54 Zusammenhänge zu erwarteten Reaktionen von Gefährdungssituationen gesucht. Die Spalte Kontext dokumentiert, welche zusätzlichen Informationen den Ratern in den Bewertungsszenarien zur Verfügung gestellt werden. Der Aufbau der Bewertungsszenarien wird im folgenden Abschnitt 4.1.2.2 detailliert beschrieben.

4.1.2 Bestimmung der objektiven Relevanz

Abbildung 4.2 zeigt den Ablauf beim Erstellen der objektiven Relevanzbewertung. Für deren Gewinnung erstellen wir für die vorgestellten Informationsbedürfnisse sogenannte Bewertungsszenarien, um diese von Experten der Domäne beurteilen zu lassen. Aufgrund der großen Datenmengen verwenden wir Pooling zur Reduktion der Anzahl der zu bewertenden Elemente. In einem zweiten manuellen Reduktionsschritt entfernen wir zum einen eindeutig nicht relevante Elemente aus dem Pool und zum anderen fassen wir eindeutig gleichbedeutende Elemente derart zusammen, dass von diesen nur jeweils ein Element bewertet werden muss. Das führt zu Bewertungsszenarien mit reduzierten Elementlisten. Die verschiedenen Bewertungen fassen wir abschließend zu einer gemeinsamen objektiven Relevanzbewertung zusammen. In diesem Abschnitt stellen wir die einzelnen Schritte detailliert vor.

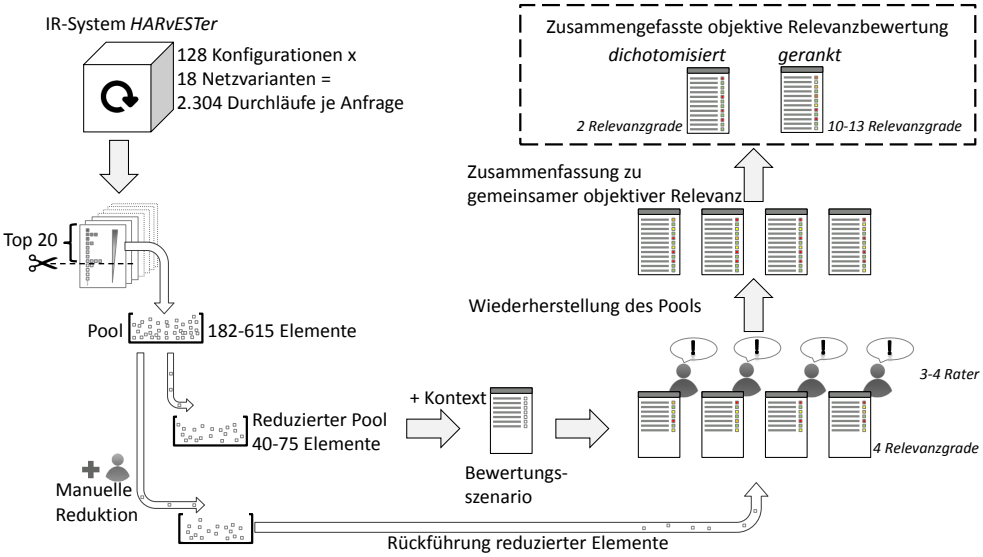


Abbildung 4.2: Erstellen objektiver Relevanzbewertungen mittels Pooling, Raterbewertungen und deren Zusammenfassung

4.1.2.1 Pooling

Zur Erstellung eines Pools zu bewertender Elemente führen wir die vorgestellten Anfragen vom IR-System (*HARvESTer*) mit verschiedenen Konfigurationen und auf verschiedenen Varianten semantischer Netze aus. Tabelle 4.4 listet die ausgewählten Konfigurationseinstellungen auf, aus deren Kombinationen sich 128 Konfigurationen ergeben. Jede Anfrage wird auf jedem der

18 semantischen Netze und mit jeder der 128 Konfigurationen untersucht. Daraus ergeben sich 2.304 Ausführungen je Anfrage und Pool. Für jede dieser Ausführungen selektieren wir die 20 besten Elemente der Ergebnislisten als Beitrag für den Pool. Für das gesamte Pooling ergeben sich 18.432 Ausführungen.

Eigenschaft	**Verwendete Konfigurationsparameter**
Anzahl der Pulse	15
Aktivierungsmodus	*IDENTITY*
Schwellwert	ohne ($\tau = 0$)
	mit ($\tau = 0{,}001$)
Dämpfungsmodus	*IGNORE* (d = 1,0)
	FIXED (d = 0,5)
	FIXED (d = 0,8)
	INCREASING
Verzweigungsmodus	*NONE*
	FANOUT
Kantengewichtung	ohne (ω(e) = 1 für alle Kanten)
	mit (ω(e) kantentypspezifisch angepasst)
Ausbreitungsmodi	*B*
	RR
	FW
	FW-RR

Tabelle 4.4: Basisset verwendeter Konfigurationseinstellungen

Reduktion der Pools

Die Pools enthalten für jede Anfrage zwischen 182 und 615 Elemente, abhängig von der Überschneidung der besten Ergebnisse für die einzelnen Konfigurationen. Aufgrund der hohen Anzahl an Elementen in den Pools und wegen beschränkter Ressourcen bei der Bewertung durch die Rater, werden die Elementlisten in einem zusätzlichen Reduktionsschritt weiter reduziert, so dass für jede Anfrage lediglich 40 bis 75 Elemente im Pool verbleiben. Diese Reduktion wird manuell im Vier-Augen-Prinzip[43] vorgenommen, wobei Reduktionsentscheidungen nur dann getroffen werden dürfen, wenn kein Zweifel an der Eindeutigkeit besteht.

Dafür identifizieren wir manuell zuerst jene Elemente, welche sicher als nicht relevant erkannt werden. Diese werden aus dem Pool entfernt und bereits vorab mit dem Relevanzgrad 0 bewertet. Dies betrifft hauptsächlich Elemente, welche aufgrund ihrer Unvollständigkeit als irrelevant betrachtet werden und Elemente, die offensichtlich in keinem nennenswerten Zusammenhang zur Anfrage stehen. In einem weiteren Schritt identifizieren wir im Pool Gruppen von Elementen, welche gleichbedeutende Elemente enthalten. Dies betrifft zum einen syntaktisch sehr ähnliche Elemente, die sich unter anderem lediglich durch verschiedene Schreibweisen unterscheiden oder semantisch ähnliche Elemente (z.B. *nicht notwendig* und *unnötig*). Aus jeder dieser Gruppen verbleibt lediglich ein Element im Pool, da von einer gleichen Bewertung der anderen Elemente der Gruppe ausgegangen wird. Die verbleibenden Elemente werden in das Bewertungsszenario für die Rater zur Beurteilung gegeben. Tabelle 4.5 zeigt die Anzahl der Bewertungselemente in den Pools jeder Anfrage sowohl vor als auch

[43] Die Entscheidungen zur Reduktion erfolgten nach Absprache mit einem Entwickler aus dem industriellen Anwendungsfeld.

nach der Reduktion. Zudem ist die Anzahl der Elemente dokumentiert, die aufgrund von Irrelevanz oder Gruppierung von der Bewertung ausgeschlossen sind.

Anfrage	# Bewertungs-elemente vor Reduktion	# irrelevant	# Gruppen	# reduzierte gruppierte Elemente	# Bewertungs-elemente nach Reduktion
Q23	240	181	18	19	40
Q35	360	263	30	52	45
Q46	235	133	10	27	75
Q54	187	25	21	87	75
Q61	195	116	1	5	74
Q64	182	128	1	4	50
Q81	565	321	68	199	45
Q85	615	336	88	229	50

Tabelle 4.5: Anzahl der Bewertungselemente für die einzelnen Stufen des Poolings

Wiederherstellung der Pools

Nach Bewertung durch die Rater machen wir die Reduktionsschritte wieder rückgängig. Das Ergebnis ist ein beurteilter Top-20-Pool für jede Anfrage. Dafür rollen wir zuerst die gruppierten Elemente aus und fügen sie dem Pool wieder hinzu. Jedes der gruppierten Elemente erhält dabei den Relevanzgrad des Elements, von welchem es im reduzierten Pool vertreten wurde. Anschließend geben wir alle aus dem Pool entfernten irrelevanten Elemente mit ihrem zugeordneten Relevanzgrad 0 in den Pool zurück.

4.1.2.2 Bewertungsszenarien für die Beurteilung durch Rater

Jedes Informationsbedürfnis überführen wir in ein Bewertungsszenario, welches dem groben Aufbau aus Abbildung 4.3 folgt und für die Rater als digitaler Fragebogen konzipiert ist. Die grauen Felder enthalten gegebene Inhalte, das weiße Feld ist editierbar und für die Eingaben der Rater gedacht.

Der *Kontext* enthält Hintergrundinformationen, die zum Zeitpunkt der Anfrage entsprechend dem Ausgangspunkt und dem G&R-Schritt vorhanden sind und zeigt daher auf, worauf sich die Anfrage bezieht. Mit der Bereitstellung des Kontextes wird simuliert, dass ein auftretendes Informationsbedürfnis in der Regel in einen Vorgang bei der Erstellung einer G&R eingebettet ist. Dadurch wird das Verständnis der Rater bezüglich der zu bewertenden Szenarien erhöht. Der Kontext enthält mindestens Name und Beschreibung des Systems, aus dem die Anfrage stammt, sowie optional weitere Informationen, die in Tabelle 4.3 aufgelistet sind.

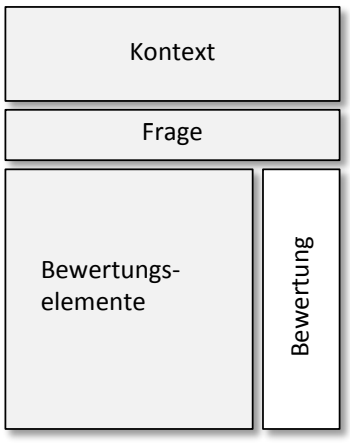


Abbildung 4.3: Bewertungsszenario

Die *Frage* entspricht der in Abschnitt 3.2.5 vorgestellten allgemeinen Form eines Informationsbedürfnisses, z.B. *Welche Fehlfunktionen stehen in einem semantischen Zusammenhang zur Funktion Momentengenerierung?*, das sich auf den bereitgestellten Kontext bezieht.

Im Feld *Bewertungselemente* sind die nach manueller Reduktion im Pool verbliebenen Bewertungselemente in zufälliger Reihenfolge abgebildet. Für die *Bewertung* der Relevanz jedes Bewertungselementes werden Relevanzgrade verwendet, welche über eine Drop-down-Liste ausgewählt werden. Folgende vier ordinalskalierte Relevanzgrade von 0 bis 3 stehen zur Verfügung.

Relevanzgrad 0: nicht relevant, 1: eher nicht relevant, 2: eher relevant, 3: sehr relevant

Die Entscheidung, welcher Relevanzgrad den einzelnen Antwortmöglichkeiten zugewiesen werden sollte, ist nicht immer einfach und eindeutig. Daher haben wir den Ratern Anhaltspunkte anhand eines Beispiels aus ihrer Domäne zur Verfügung gestellt. Eine Antwort sollte als sehr relevant (Grad 3) gelten, wenn sie eine korrekte Antwort auf die Frage bezüglich des Kontextes darstellt, also so auch übernommen werden könnte. Nicht relevante Antworten (Grad 0) haben keinerlei Bezug zur Anfrage. Mit den mittleren Relevanzgraden wird eine Tendenz in die Richtungen relevant oder nicht relevant angezeigt. Eine Antwort gilt als eher relevant (Grad 2), wenn sie nicht komplett passt, aber dennoch einen starken Bezug zur Anfrage hat. Als eher nicht relevant (Grad 1) können Antworten markiert werden, die zwar keinen starken Bezug zur Anfrage besitzt, aber auch nicht ganz irrelevant sind.

4.1.2.3 Rater

An der Evaluation nahmen vier Experten aus dem Bereich Funktionssicherheit der Firma Daimler AG teil, unter anderem der Teamleiter und ein Reviewer von G&Rs. Alle Rater sind mit G&Rs vertraut, da sie operativ an der Erstellung von G&Rs beteiligt sind oder waren. Im Rahmen der Evaluation nahmen sie eine Selbsteinschätzung ihres Experten- und Konfidenzlevels vor und beantworteten die folgenden Fragen mit Antwortmöglichkeiten einer fünfstufigen verbalen Ratingskala. Abbildung 4.4 zeigt die Ergebnisse dieser Selbsteinschätzung.

1. Wie erfahren sind Sie in der Durchführung von G&Rs?

(Level 1: sehr unerfahren, 2: eher unerfahren, 3: mittel, 4: eher erfahren, 5: sehr erfahren)

2. Wie sicher waren Sie in Ihren Relevanzbewertungen?

(Level 1 - sehr unsicher, 2 - eher unsicher, 3 - mittel, 4 - eher sicher, 5 - sehr sicher)

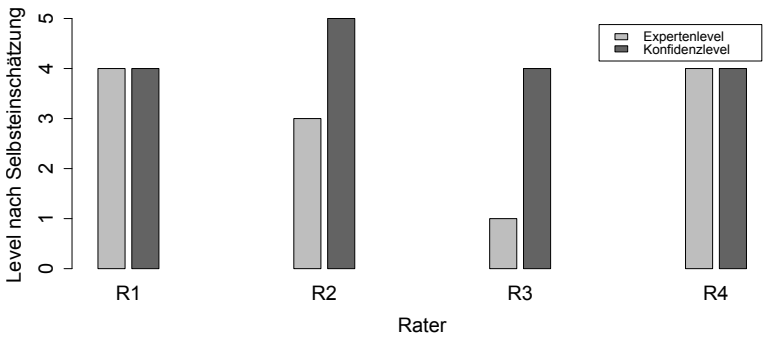


Abbildung 4.4: Selbsteinschätzung der Rater bezüglich ihres Experten- und Konfidenzlevels

Die Rater R1, R2 und R4 halten sich für mittel bis eher erfahren. Auffällig ist die zurückhaltende Selbsteinschätzung von Rater R3 als eher unerfahren bei der Durchführung von G&Rs. Bezüglich der Relevanzbewertungen ist dennoch eine hohe Konfidenz vorhanden. Alle Rater sind sich eher sicher oder sehr sicher bei der Beurteilung der einzelnen Bewertungsszenarien.

4.1.2.4 Zusammenfassen der Relevanzbewertungen zu einer gemeinsamen objektiven Relevanz

Ziel ist es, für jedes Bewertungsszenario eine Relevanzbewertung zu erhalten, welche als Goldstandard in der Evaluation verwendet wird. Bei der Berechnung der Genauigkeit können die Relevanzgrade nicht berücksichtigt werden, sondern lediglich, ob ein Element relevant ist oder nicht. Daher werden die Relevanzbewertungen für die Evaluation der Genauigkeit dichotomisiert, für die Evaluation der Nützlichkeit werden Relevanzgrade beibehalten. Da kein übliches Vorgehen und keine allgemeingültige Empfehlung für die Art der Zusammenführung der verschiedenen Relevanzbewertungen existiert, werden im Folgenden verschiedene Varianten und eigene Überlegungen vorgestellt.

Zusammenfassen zu einer dichotomisierten objektiven Relevanz

Dichotomisierung und Auflösen der Ordinalskala: Für die Berechnung der Genauigkeit ist eine Dichotomisierung der Relevanzbewertungen notwendig. Dafür werden die Relevanzgrade aufgelöst, so dass alle Bewertungselemente, welchen durch den Relevanzgrad 0 keine Relevanz zugeordnet wurde, diesen Relevanzgrad behalten. Allen anderen Bewertungselementen mit einem Mindestmaß an Relevanz, also mit Relevanzgraden von 1 bis 3, wird der Wert 1 zugeordnet. Die Bewertung ist nunmehr eine Boolesche Bewertung, in der jedes Bewertungselement entweder als relevant (0) oder als nicht relevant (1) markiert ist.

Erstellen der objektiven Relevanzbewertung: Für die Zusammenfassung der dichotomisierten objektiven Relevanz betrachten wir zwei Varianten. Jedes Element, welches mindestens von einem Rater einen Relevanzgrad größer als 0 zugewiesen bekommt, wird in der ersten Variante der dichotomisierten objektiven Relevanzbewertung ORD1 als relevant markiert. Dadurch werden eventuell zu viele Elemente als relevant markiert, also auch solche, die nur von wenigen Nutzern als relevant empfunden werden. In einer zweiten Variante werden, im Gegensatz zu Variante 1, nur die Elemente als relevant markiert, die von einer Mindestanzahl an Ratern oder einem vorher definierten Anteil der Raterstimmen als relevant bewertet wurden. Wir definieren für die dichotomisierte Variante ORD2, dass mindestens die Hälfte aller Rater das jeweilige Element als relevant empfinden muss. Elemente, die beispielsweise von einer Person als relevant eingestuft werden, von zwei anderen aber nicht, werden demnach als objektiv nicht relevant markiert. Dadurch werden eventuell zu wenig Elemente als relevant markiert. Tabelle 4.6 zeigt für jede Anfrage die Anzahl der als relevant markierten Bewertungselemente zum einen durch die Rater und zum anderen für die beiden Varianten der objektiven Relevanz.

Zusammenfassen zu einer gerankten objektiven Relevanz

Jedes Bewertungselement erhält einen neuen Relevanzgrad, welcher sich aus den einzelnen Bewertungen der Rater zusammensetzt. Dafür werden die durch sie vergebenen Relevanzgrade addiert. Ordinalskalen eignen sich üblicherweise nicht für die Anwendung von Rechenoperationen, da nicht zwangsläufig Äquidistanz zwischen den Graden vorausgesetzt werden kann. In diesem Fall ist die Aufsummierung der Bewertungen eine Möglichkeit zur Zusammenfassung der verschiedenen Bewertungen, da sich so die mehrfache Bewertung in einer erwarteten Form auf das gemeinsame objektive Ranking auswirkt.

Abbildung 4.5 stellt die anteilige Verteilung der Bewertungen für jede Anfrage durch jeden

Anfrage	R1	R2	R3	R4	ORD1	ORD2
Q23	30	18	14	31	35	28
Q35	48	-	4	50	55	44
Q46	32	-	13	47	47	34
Q54	20	40	-	69	83	30
Q61	24	17	13	22	32	28
Q64	8	12	17	20	25	17
Q81	6	-	2	4	8	2
Q85	9	-	11	47	51	9

Tabelle 4.6: Anzahl der als relevant markierten Bewertungselemente der verschiedenen Rater und in den beiden Varianten der objektiven Relevanz

Rater sowie die verschiedenen zusammengefassten objektiven Relevanzen dar. Während die Bewertungen der Rater auf der vierstufigen Ordinalskala stattfinden, ist die Ordinalskala der dichotomisierten Daten reduziert auf zwei Stufen und die Ordinalskala der zusammengefassten gerankten Daten erweitert auf zehn oder 13 Stufen, abhängig von der Anzahl der Rater. Es ist erkennbar, dass die Ratings nicht gleichmäßig verteilt sind. Der Anteil der Elemente mit Relevanzgrad 0 ist beispielsweise wesentlich höher als der Anteil anderer Relevanzgrade. Zudem entstehen durch die relativ geringe Stufenanzahl der Ordinalskala Überlappungen auf den Rängen, die Rangbindung. Diese Besonderheiten haben wir bei der Auswahl der Bewertungsmaße berücksichtigt (siehe Abschnitt 2.1.3.4).

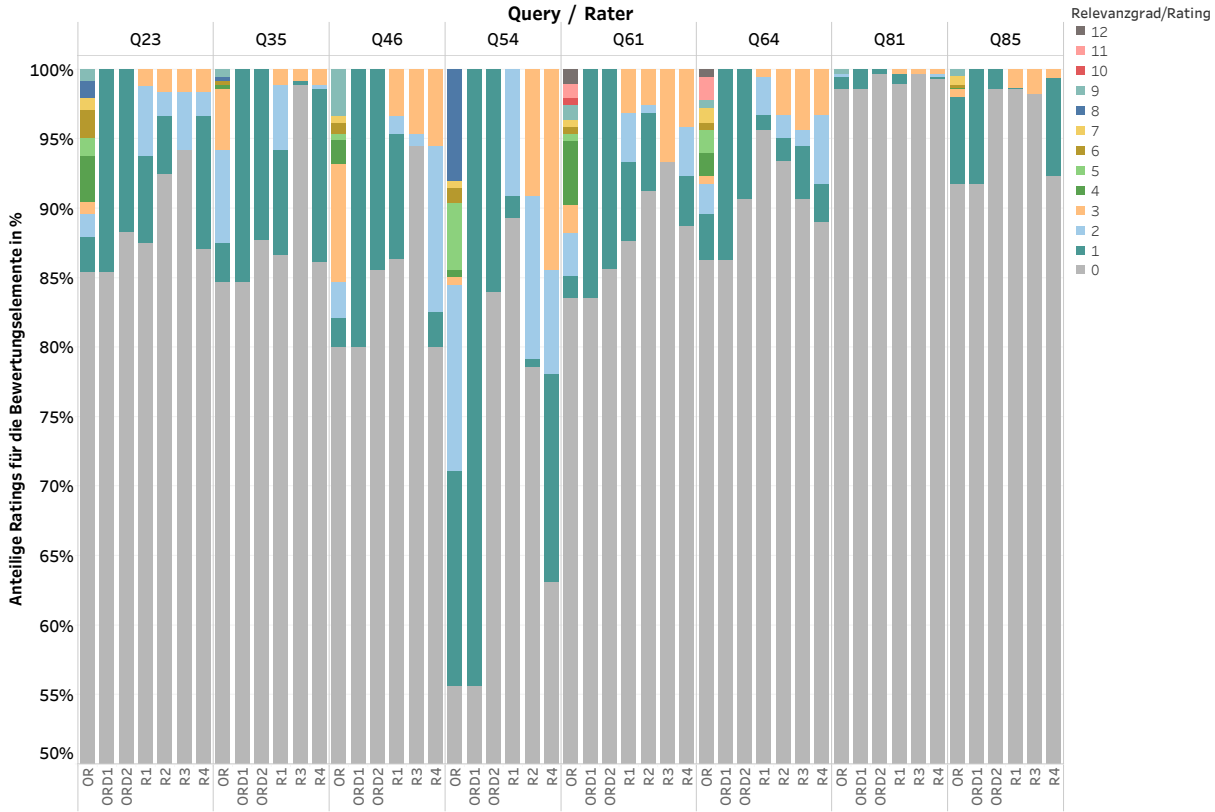


Abbildung 4.5: Anteilige Bewertungen für die Anfragen (Queries) je Rater und je zusammengefasster objektiver Relevanz: ORD1 und ORD2 entsprechen der dichotomisierten Zusammenfassung und OR der gerankten Zusammenfassung

4.1.3 Berechnung der Systemrelevanz und Distributionen

Das Ergebnis des IR-Systems ergibt für jede Kombination aus Konfiguration, Anfrage und Netzvariante eine Systemrelevanz, die in Kombination mit den beiden zusammengefassten objektiven Relevanzbeurteilungen zwei Distributionen ergeben, eine dichotomisierte und eine gerankte. Die Distributionen sind die Basis zur Bewertung jeder Ausführung und damit zur Beantwortung der Evaluationsfragen.

4.1.4 Evaluationsfragen

Im Folgenden formulieren wir Evaluationsfragen, deren Beantwortung die Bewertung der Effektivität des in dieser Arbeit entwickelten IR-Systems ermöglicht. Die Fragen dienen der Einschätzung, unter welchen Umständen das IR-System besonders gute Ergebnisse liefert. Tabelle 4.7 listet die Evaluationsfragen auf. Zudem wird für jede Evaluationsfrage die Strategie zu ihrer Beantwortung kurz aufgeführt sowie der Abschnitt ihrer Beantwortung.

EQ	Evaluationsfrage	Evaluationsstrategie	Abschnitt
1	Ist die Übereinstimmung der Rater ausreichend?		4.2
1a	In welchem Maß stimmen die Bewertungen der Rater überein?	Diese Frage wird mit dem Brennan-Prediger-Koeffizienten und dem Gwet-AC_1-Koeffizienten beantwortet.	4.2.1 (S. 132)
1b	In welchem Maß stimmt die Abstufung der Bewertungen in den Rangreihen überein?	Diese Frage wird mit Kendalls Konkordanzkoeffizient W beantwortet.	4.2.2 (S. 134)
1c	Ist die Übereinstimmung nach Zusammenfassung zu einer objektiven Relevanz noch ausreichend?	Diese Frage wird für die dichotomisierte objektive Relevanz mit dem Brennan-Prediger-Koeffizienten und für die gerankte objektive Relevanz mit Kendalls Konkordanzkoeffizient W beantwortet. Dafür werden alle Rater-OR-Paare untersucht.	4.2.3 (S. 136)
2	Wie stark ist der Einfluss der Netzvarianten, Konfigurationsparameter und Anfragen auf die Qualität der Ergebnisse? Welche Bedeutung hat die Parametrisierung?	Diese Frage wird durch die Auswertung der Ergebnisse für zufällig generierte Konfigurationen beantwortet.	4.3 (S. 144)
3	Welche Faktoren wirken sich günstig auf die Qualität der Ergebnisse aus?	Zur Bewertung der Faktoren werden zunächst Konfigurationen mittels evolutionärer Algorithmen optimiert. Anschließend analysieren wir die Top-Konfigurationen der Durchläufe.	4.4
3a	Ist der Einsatz der semantischen Anreicherungen und der Reduktionen der Netze sinnvoll? Welchen Einfluss haben sie auf die Qualität der Ergebnisse?		4.4.3.5 (S. 168)
3b	Ist der Einsatz der Ausbreitungsmodi sinnvoll? Welchen Einfluss haben sie auf die Qualität der Ergebnisse?		4.4.3.5 (S. 169)
3c	Welchen Einfluss haben die Konfigurationsparameter und deren Kombinationen auf die Qualität der Ergebnisse?		4.4.3.5 (S. 169)
4	Sind die Ergebnisse des IR-Systems relevant und nützlich?	Die Qualität des IR-Systems wird durch Berechnung von Bewertungsparametern für die verschiedenen Anfragen mit optimierten Konfigurationen und Netzen untersucht.	4.4.4 (S. 172)

Tabelle 4.7: Evaluationsfragen

4.2 Urteilerübereinstimmung

In diesem Abschnitt stellen wir die Ergebnisse zur Urteilerübereinstimmung gemäß den Beschreibungen aus Abschnitt 2.1.3.4 vor.

4.2.1 Brennan-Prediger- und Gwet-AC_1-Koeffizient

Die Ergebnisse der Übereinstimmungsanalyse sind für den Brennan-Prediger-Koeffizienten κ_{BP} in Tabelle 4.8 und für den Gwet-AC_1-Koeffizient in Tabelle 4.9 dargestellt. Des Weiteren sind für jede Anfrage die tatsächlich erreichte Übereinstimmung pa, der berechnete Standardfehler se, das daraus folgende 95%-Konfidenzintervall mit der unteren Grenze CI^- und der oberen Grenze CI^+, der Prüfwert t sowie die Signifikanzgrenze $t_{1-\alpha,N-1}$ aufgelistet. Die Beschreibung und Berechnung dieser zufallskorrigierten Übereinstimmungskoeffizienten sowie die Interpretation der Ergebnisse haben wir in Abschnitt 2.1.3.4 auf den Seiten 22–23 dargelegt.

Anfrage	**pa**	κ_{BP}	**se**	CI^-	CI^+	**t**	$t_{1-\alpha,N-1}$
Q23	0,8986	0,8648	0,0218	0,8218	0,9078	39,6283	1,6513
Q35	0,8861	0,8481	0,0197	0,8093	0,887	42,9639	1,6491
Q46	0,8596	0,8128	0,0281	0,7575	0,8681	28,9586	1,6514
Q54	0,6863	0,5817	0,0351	0,5125	0,6509	16,595	1,6531
Q61	0,8957	0,861	0,024	0,8136	0,9083	35,8565	1,6527
Q64	0,913	0,884	0,0229	0,8388	0,9292	38,6254	1,6533
Q81	0,9929	0,9906	0,0038	0,983	0,9981	258,1903	1,6476
Q85	0,9474	0,9211	0,011	0,8996	0,9427	83,9843	1,6473

Tabelle 4.8: Ergebnisse der Brennan-Prediger-Koeffizient-Analyse

Anfrage	**pa**	**Gwet AC_1**	**se**	CI^-	CI^+	**t**	$t_{1-\alpha,N-1}$
Q23	0,8986	0,8921	0,0184	0,8558	0,9284	51,2237	1,6513
Q35	0,8861	0,8791	0,0165	0,8465	0,9116	55,909	1,6491
Q46	0,8596	0,8475	0,0241	0,7999	0,895	37,0624	1,6514
Q54	0,6863	0,6396	0,0339	0,5727	0,7065	21,1762	1,6531
Q61	0,8957	0,889	0,0201	0,8493	0,9287	46,3711	1,6527
Q64	0,913	0,9085	0,0189	0,8712	0,9457	50,308	1,6533
Q81	0,9929	0,9929	0,0029	0,9872	0,9986	343,442	1,6476
Q85	0,9474	0,9455	0,0078	0,9302	0,9608	124,7448	1,6473

Tabelle 4.9: Ergebnisse der Gwet-AC_1

Die vorliegenden Urteile der Rater werden als nominalskalierte Daten interpretiert. Das bedeutet, das implizite Ranking der Ordinalskala bleibt bei der Bewertung der Übereinstimmung unberücksichtigt. Ähnlichkeit beziehungsweise Nähe im Ranking wird nicht belohnt, wie in gewichteten Analysen für ordinalskalierte Daten üblich.[44] Dieser Verzicht führt zu einer strengeren Betrachtung der Urteile, da eine Übereinstimmung nur bei perfekter nominaler Über-

[44] Berechnungen mit linearer und quadratischer Gewichtung für ordinalskalierte Daten führten sogar zu besseren Ergebnissen. Diese sind in Anhang C.1 abgebildet.

einstimmung vorhanden ist. Die Abbildungen 4.6 und 4.7 zeigen die 95%-Konfidenzintervalle für jede Anfrage für beide Koeffizienten.

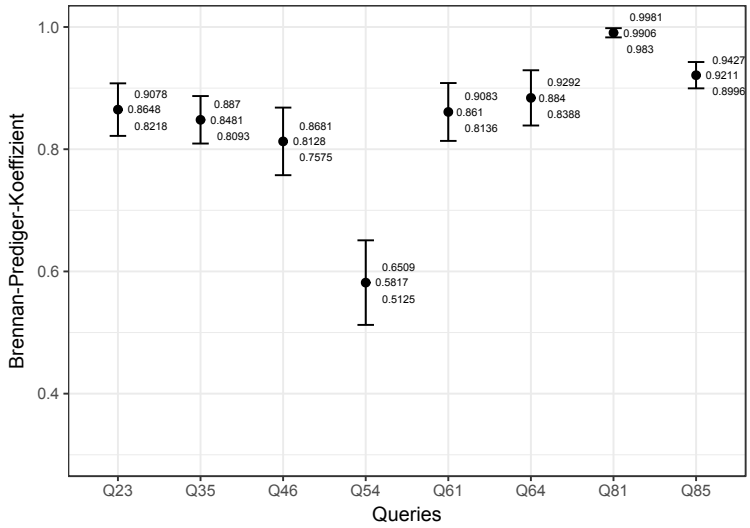


Abbildung 4.6: 95%-Konfidenzintervalle des κ_{BP}-Koeffizienten

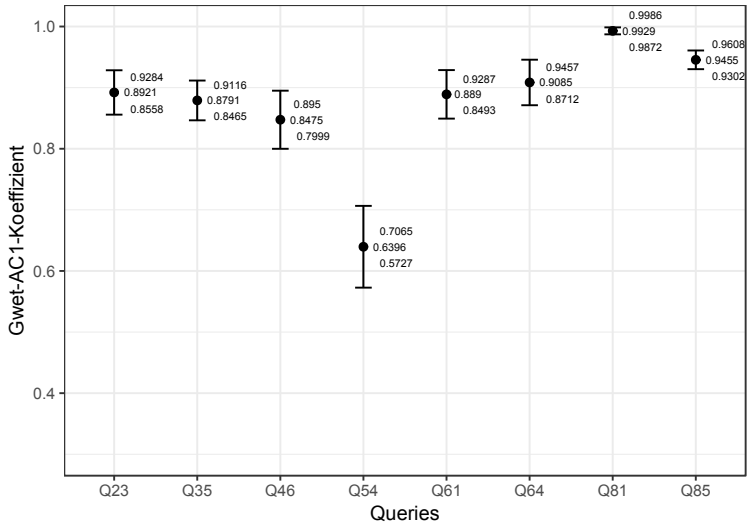


Abbildung 4.7: 95%-Konfidenzintervalle des Gwet-AC_1-Koeffizienten

Hypothese H_0**:** Es wird vermutet, dass die Übereinstimmung nicht besser als die zufällig erwartete Übereinstimmung ist. Als Signifikanzniveau und damit als obere Grenze für die Irrtumswahrscheinlichkeit wird $\alpha = 5\%$ gewählt.

Interpretation: Die berechneten Brennan-Prediger-Koeffizienten liegen für alle Anfragen im Bereich $0{,}5817 \leq \kappa_{\mathrm{BP}} \leq 0{,}9906$. Die Anfrage Q54 kann gemäß Bewertungsskala von Landis et al. (siehe Tabelle 2.1, Seite 23) lediglich eine mittelhohe (moderate) Übereinstimmung erreichen, die anderen Anfragen jedoch eine hohe bis sehr hohe (substantial, almost perfect) Übereinstimmung. Die berechneten Gwet-AC_1-Koeffizienten liegen für ebendiese Anfragen

im Bereich $0{,}6396 \leq AC_1 \leq 0{,}9929$. Hier erreicht die Anfrage Q54 wieder lediglich einen mittelhohen Wert, alle anderen Anfragen liegen allerdings in einem hohen bis sehr hohen Bereich. Sowohl für den Brennan-Prediger-Koeffizient als auch für den AC_1-Koeffizient gilt nach rechtsseitigem Einstichproben-t-Test, dass für jede Anfrage der Prüfwert t weitaus größer ist als die Signifikanzgrenze, also $t > t_{1-\alpha,N-1}$. Damit kann die Nullhypothese abgelehnt werden. Die berechnete Übereinstimmung ist für alle Anfragen signifikant größer als die zufällig erwartete Übereinstimmung.

Das Konfidenzintervall gibt den Wertebereich an, der den wahren Koeffizienten in der Population mit Wahrscheinlichkeit von 95% enthält. Der sehr kleine Standardfehler und die daraus mit beidseitigem Einstichproben-t-Test berechneten engen Konfidenzintervalle erlauben eine positive Aussage über die Güte der Koeffizientenschätzung. So enthält beispielsweise für die Anfrage Q23 der Wertebereich $[0{,}8218, 0{,}9078]$ das wahre κ_{BP} mit einer Wahrscheinlichkeit von 95%. Die unteren Konfidenzgrenzen liegen für Q54 oberhalb von 0,5, für alle anderen Anfragen sogar oberhalb von 0,75.

Anmerkung: Die große Stichprobenzahl (Anzahl der Bewertungselemente) hat positiven Einfluss auf die Konfidenzintervalle und damit auf die Güte der Schätzung. Alternativ kann die Berechnung der Übereinstimmungsmaße auf den bewerteten reduzierten Pools durchgeführt werden, welche eine wesentlich geringere Stichprobenzahl aufweisen. Für die Bestimmung einer objektiven Relevanz steht allerdings die Einigkeit der Rater bezüglich der reduzierten Selektion von Bewertungselementen weniger im Vordergrund als die Frage, wie stark die Übereinstimmung der bewerteten Pools mit allen Bewertungselementen ist. Die Ergebnisse der Übereinstimmungsanalyse für die reduzierten Pools sind im Anhang C.1.2 aufgelistet. Die berechnete Übereinstimmung fällt im Vergleich zur nicht reduzierten Poolgröße zwar niedriger aus, ist aber dennoch für jede Anfrage signifikant größer als die zufällig erwartete Übereinstimmung.

Entscheidung

***EQ1a**: In welchem Maß stimmen die Bewertungen der Rater überein?*

Das Gesamtbild zeigt eine mittelhohe bis sehr hohe und signifikant überzufällige Übereinstimmung für die berechneten Koeffizienten und für alle Anfragen. Für beide Übereinstimmungsmaße kann mit hoher Konfidenz auf eine gute Abschätzung geschlossen werden. Die Anfrage Q54 hebt sich zwar deutlich von anderen Anfragen ab, ist aber dennoch in einem akzeptablen Bereich.

4.2.2 Konkordanzkoeffizient nach Kendall

Tabelle 4.10 listet die Ergebnisse von Kendalls Konkordanzanalyse zur Quantifizierung der Rangreihenübereinstimmung auf. Die Beschreibung und Berechnung von Kendalls Konkordanzkoeffizienten W sowie das Vorgehen bei der Interpretation der Ergebnisse haben wir in Abschnitt 2.1.3.4 auf den Seiten 23–24 dargelegt.

Anfrage	**# Rater m**	**# Bewertungs-elemente N**	**W**	χ^2_r	$\chi^2_{1-\alpha,N-1}$	**p-Wert**
Q23	4	240	0,7469	713,9999	276,0624	8,21e-49
Q35	3	360	0,6576	708,2381	404,1821	4,10e-25
Q46	3	235	0,7419	520,8256	270,6839	6,98e-24
Q54	3	187	0,6706	374,1713	218,8205	9,38e-15
Q61	4	195	0,6612	513,0843	227,4964	1,18e-30
Q64	4	182	0,7041	509,7613	213,3906	4,64e-33
Q81	3	565	0,6675	1129,337	620,357	4,42e-40
Q85	3	615	0,5632	1037,4622	672,7548	3,13e-24

Tabelle 4.10: Ergebnisse der Konkordanzanalyse nach Kendall

Hypothese H_0: Es wird vermutet, dass keine Konkordanz zwischen den Rangreihen besteht. Als Signifikanzniveau und damit als obere Grenze für die Irrtumswahrscheinlichkeit wird $\alpha = 5\%$ gewählt.

Interpretation: Die ermittelten Konkordanzkoeffizienten für die Anfragen liegen im Bereich $0{,}5632 \leq W \leq 0{,}7469$ bei einem Durchschnittswert von 0,6766. Sie zeigen, dass eine mittelhohe Übereinstimmung der Rangreihen besteht, Abweichungen aber dennoch existieren. Der Friedman-Test ergibt für jede Anfrage, dass $\chi^2_r > \chi^2_{0.95,N-1}$. Da die Prüfgröße größer als die kritische Schranke ist, kann für jede Anfrage die Nullhypothese verworfen werden. Das wird durch den sehr geringen p-Wert unterstützt, mit dem die Irrtumswahrscheinlichkeit abgeschätzt werden kann. Die Rangreihen können somit als konkordant gelten.

Paarweise Rangreihenübereinstimmung

Die Prüfgröße χ^2_r ist linear abhängig von der Anzahl der Rater und der Bewertungselemente (siehe Formel 2.18, Seite 24). Sie steigt mit zunehmender Anzahl von Ratern bei gleichbleibender Anzahl an Bewertungselementen. Eine zusätzliche Berechnung aller paarweisen Konkordanzkoeffizienten ergibt allerdings ein ähnlich eindeutiges Bild. Tabelle 4.11 zeigt die Ergebnisse von Kendalls Konkordanzanalyse für alle paarweisen Kombinationen der vier Rater R1, R2, R3 und R4. Für die durch drei Rater bewerteten Bewertungsszenarien sind die Ergebnisse der resultierenden drei paarweisen Kombinationen abgebildet. Für die durch vier Rater bewerteten Bewertungsszenarien sind die Ergebnisse der resultierenden sechs paarweisen Kombinationen abgebildet.

Hypothese H_0: Es wird vermutet, dass keine Konkordanz zwischen den Rangreihen der einzelnen Raterpaare besteht. Als Signifikanzniveau wird ebenfalls $\alpha = 5\%$ gewählt.

Anfrage	Raterpaar	W	χ^2_r	$\chi^2_{1-\alpha,Fg}$	p-Wert
Q23	R1 R2	0,854	408,1981	276,0624	5,49e-11
	R1 R3	0,7906	377,9008		2,50e-08
	R1 R4	0,939	448,8221		5,65e-15
	R2 R3	0,6933	331,4146		6,98e-05
	R2 R4	0,8594	410,7721		3,17e-11
	R3 R4	0,7756	370,7413		9,59e-08
Q35	R1 R3	0,5608	402,6696	404,1821	0,0556
	R1 R4	0,9314	668,7293		5,86e-21
	R3 R4	0,581	417,1598		0,0184
Q46	R1 R3	0,751	351,4581	270,6839	1,02e-06
	R1 R4	0,898	420,2704		9,16e-13
	R3 R4	0,7308	342,0176		5,12e-06
Q54	R1 R2	0,7931	295,051	218,8205	6,08e-07
	R1 R4	0,7931	264,5875		1,34e-04
	R2 R4	0,7931	283,2693		5,53e-06
Q61	R1 R2	0,7312	283,6864	227,4964	2,81e-05
	R1 R3	0,6756	262,1499		8,04e-04
	R1 R4	0,953	369,7578		4,43e-13
	R2 R3	0,8575	332,7278		2,19e-09
	R2 R4	0,7172	278,2565		6,89e-05
	R3 R4	0,6878	266,8579		4,05e-04
Q64	R1 R2	0,7996	289,4635	213,3906	5,33e-07
	R1 R3	0,7782	281,7047		2,36e-06
	R1 R4	0,6791	245,8231		9,59e-04
	R2 R3	0,8409	304,423		2,55e-08
	R2 R4	0,844	305,5385		2,01e-08
	R3 R4	0,8496	307,5382		1,32e-08
Q81	R1 R3	0,7513	847,5053	620,3570	9,52e-14
	R1 R4	0,6997	789,2786		1,00e-09
	R3 R4	0,8341	940,889		2,41e-21
Q85	R1 R3	0,8984	1103,2464	672,7548	2,22e-30
	R1 R4	0,6449	791,9676		1,46e-06
	R3 R4	0,617	757,6275		6,27e-05

Tabelle 4.11: Konkordanzkoeffizient nach Kendall für jedes Raterpaar

Interpretation: Die ermittelten Konkordanzkoeffizienten für die Anfragen liegen im Bereich $0{,}5608 \leq W \leq 0{,}953$ bei einem Durchschnittswert von 0,7683. Auch hier kann von einer mittleren bis hohen Übereinstimmung der Rangreihen ausgegangen werden. Lediglich bei Anfrage Q35 kann ein einziges Raterpaar (R1 R3) keine signifikante Konkordanz erreichen. Hier wird nur eine Irrtumswahrscheinlichkeit von 5,56% erreicht. Das bedeutet, für dieses Raterpaar kann nicht ausgeschlossen werden, dass der mittelhohe Konkordanzkoeffizient von $W = 0{,}5608$ nur auf zufälliger Rangreihenübereinstimmung basiert. Die Rangreihen der verbleibenden Raterpaare von Q35, sowie aller anderen bewerteten Szenarien sind bei paarweiser Kombination signifikant konkordant.

Entscheidung

***EQ1b**: In welchem Maß stimmt die Abstufung der Bewertungen in den Rangreihen überein?*

Die Kendallsche Konkordanzanalyse zeigt eine signifikante Konkordanz der von den Ratern in Rangreihe gebrachten Elemente der Bewertungsszenarien. Die zusätzlich durchgeführte Konkordanzanalyse für alle Raterpaarkombinationen unterstützt das Ergebnis. Bis auf einen Fall konnte auch hier signifikante Konkordanz gezeigt werden. In der Gesamtbetrachtung aller Rater der Anfrage Q35 ist der mittelhohe Konkordanzkoeffizient $W = 0{,}6576$ signifikant, so dass auch die Bewertungen dieses Szenarios für die Evaluation verwendbar sind.

4.2.3 Übereinstimmung der zusammengefassten objektiven Relevanz

Eine zusätzliche Analyse beurteilt die Güte der zusammengefassten objektiven Relevanz. Dafür wird die Übereinstimmung mittels Brennan-Prediger-Koeffizienten κ_{BP} für die zusammengefasste dichotomisierte objektive Relevanz sowohl für alle Rater als auch für jedes Rater-OR-Paar[45] bestimmt. Die zusammengefasste gerankte objektive Relevanz wird hingegen mittels Kendalls Konkordanzanalyse auf Rangreihenübereinstimmung zwischen jedem Rater-OR-Paar überprüft.

4.2.3.1 Übereinstimmung der zusammengefassten dichotomisierten objektiven Relevanz

Analog zu der Analyse in Abschnitt 4.2.1 wird die Hypothese H_0 überprüft, die besagt, dass die Übereinstimmung der dichotomisierten Beurteilungen nicht besser als die zufällig erwartete Übereinstimmung ist. Als Signifikanzniveau wird $\alpha = 5\%$ gewählt. In Tabelle 4.12 sind die Ergebnisse aufgelistet.

Anfrage	**pa**	$\kappa_{\mathbf{BP}}$	**se**	$\mathbf{CI^-}$	$\mathbf{CI^+}$	**t**	$\mathbf{t_{1-\alpha,N-1}}$
Q23	0,9382	0,8764	0,0222	0,8327	0,9201	39,5102	1,6513
Q35	0,9037	0,8074	0,0247	0,7588	0,8561	32,6382	1,6491
Q46	0,8979	0,7957	0,0314	0,7339	0,8576	25,3474	1,6514
Q54	0,7611	0,5223	0,0469	0,4298	0,6148	11,1413	1,6531
Q61	0,9162	0,8325	0,0303	0,7727	0,8923	27,4672	1,6527
Q64	0,9386	0,8773	0,0263	0,8255	0,9291	33,417	1,6533
Q81	0,9929	0,9858	0,0058	0,9745	0,9971	171,3068	1,6476
Q85	0,9523	0,9046	0,0139	0,8774	0,9318	65,2283	1,6473

Tabelle 4.12: Ergebnisse der Brennan-Prediger-Koeffizient-Analyse für die dichotomisierten Beurteilungen aller Rater

Interpretation: Die tatsächliche Übereinstimmung pa ist für die dichotomisierten Daten höher als für die nicht dichotomisierten Daten in Tabelle 4.8. Der Übereinstimmungskoeffizient κ_{BP} fällt abgesehen von der Anfrage Q23 geringfügig niedriger aus. Die Brennan-Prediger-Koeffizienten der Beurteilungen aller Rater liegen für alle Anfragen im Bereich $0{,}5223 \leq \kappa_{\mathrm{BP}} \leq 0{,}9858$.[46] Wie bei der Übereinstimmungsanalyse für nicht dichotomisierte Daten erreicht die Anfrage Q54 lediglich eine mittelhohe Übereinstimmung, während die anderen Anfragen eine hohe bis sehr hohe Übereinstimmung aufweisen. Die berechnete Übereinstimmung ist für alle Anfragen signifikant größer als die zufällig erwartete Übereinstimmung. Die Ergebnisse der paarweisen Übereinstimmung zwischen den einzelnen Ratern und den zusammengefassten dichotomisierten objektiven Relevanzvarianten (ORD1 und ORD2) sind in Anhang C.2 in Tabelle C.13 dokumentiert. Die Ergebnisse zeigen, dass die berechnete Übereinstimmung auch hier für alle Anfragen signifikant größer ist als die zufällig erwartete Übereinstimmung. In einigen Fällen konnte sogar eine totale Übereinstimmung gezeigt werden, z.B. ORD1 und R4. Insbesondere in der Anfrage Q54, die ohnehin die geringste Übereinstimmung zwischen

[45] Hierbei wird jede Variante (ORD1 und ORD2) mit jeweils jedem Rater untersucht.

[46] Bei der Dichotomisierung der Beurteilungen verändert sich die erwartete zufällige Übereinstimmung p_e von 0,25 für die vier Relevanzgrade auf 0,5 für die verbliebenen zwei Relevanzgrade. Deswegen muss sich der berechnete Übereinstimmungskoeffizient κ_{BP} für die dichotomisierten Daten nicht zwangsläufig verbessern.

allen Ratern besitzt, ist erkennbar, wie schwierig das Zusammenfassen zu einer objektiven Relevanz ist. Bei Variante ORD1 fällt die Übereinstimmung mit R1 und R2 gering aus, mit der Variante ORD2 hingegen die Übereinstimmung zu R4. Dies passt auch zu dem Bild, welches Tabelle 4.6 (Seite 128) und Abbildung 4.5 (Seite 128) vermitteln. Offensichtlich ist die Anzahl der 187 Bewertungselemente, die mit einem Relevanzgrad > 0 beurteilt wurden, sehr unterschiedlich, hier 20, 40 und 69. Zusammengefasst ergeben sich mit Variante ORD1 83 und ORD2 30 Elemente.

4.2.3.2 Übereinstimmung der zusammengefassten gerankten objektiven Relevanz

Analog zu der Analyse in Abschnitt 4.2.2 wird die Hypothese H_0 überprüft, die besagt, dass keine Konkordanz zwischen den Rangreihen der einzelnen Rater-OR-Paare besteht. Als Signifikanzniveau wird ebenfalls $\alpha = 5\%$ gewählt. Die Ergebnisse dieser Analyse sind in Anhang C.2 in Tabelle C.12 dokumentiert.

Interpretation: Die ermittelten Konkordanzkoeffizienten für die Anfragen liegen im Bereich $0{,}5745 \leq W \leq 0{,}9983$ bei einem Durchschnittswert von 0,8523. Auch hier kann von einer mittleren bis hohen Übereinstimmung der Rangreihen ausgegangen werden. Die Rangreihen der einzelnen Rater und der zusammengefassten gerankten objektiven Relevanz sind signifikant konkordant.

4.2.3.3 Entscheidung zur Verwendung der zusammengefassten objektiven Relevanz

***EQ1c**: Ist die Übereinstimmung nach Zusammenfassung zu einer objektiven Relevanz noch ausreichend?*

Das Gesamtbild zeigt eine mittelhohe bis sehr hohe und signifikant überzufällige Übereinstimmung für die berechneten Koeffizienten und für alle Anfragen. Sowohl für die zusammengefasste dichotomisierte objektive Relevanz als auch für die zusammengefasste gerankte objektive Relevanz kann mit hoher Konfidenz auf eine gute Abschätzung geschlossen werden. Damit ist die Verwendung der zusammengefassten objektiven Relevanzbeurteilungen für die Bewertung der Güte des IR-Systems möglich. Ferner wird für die dichotomisierte objektive Relevanz die Variante ORD1 gewählt. Das hat den Vorteil, dass dadurch, wie auch bei der gerankten OR, alle Relevanzbeurteilungen der Rater in die zusammengefasste OR einfließen. Im Gegensatz dazu entfielen bei Auswahl der Variante ORD2 einzelne Bewertungen, die keine Mehrheit erhalten. Auch wenn die berechneten Übereinstimmungskoeffizienten häufig höher in der Variante ORD2 ausfallen, sind sie bei Verwendung von ORD1 dennoch in einem akzeptablen mittelhohen bis hohen Bereich und zudem für alle Anfragen signifikant größer als die zufällig erwartete Übereinstimmung.

4.3 Explorative Evaluation des IR-Systems für zufällige Konfigurationen

In diesem Abschnitt untersuchen wir die Qualität zufällig generierter Konfigurationen. Wir stellen zunächst unser Vorgehen bei der Durchführung vor und präsentieren anschließend die Auswertung der generierten Daten. Wir werden einen Überblick geben, wie sich die Qualität der Konfigurationen auf die verwendeten Netzvarianten, Konfigurationsparameter und Anfragen verteilt. Dieser Teil der Evaluation wird deutlich machen, wie wichtig sowohl die geeignete Parametrisierung der Konfigurationen als auch die Auswahl der Netze ist.

4.3.1 Durchführung

Für jede der 18 Netzvarianten aus Tabelle 4.1 (Seite 121) werden 500 Konfigurationen zufällig generiert und für jede der acht Anfragen ausgeführt. Jede dieser Ausführungen nennen wir im Folgenden auch Durchlauf. Es ergeben sich 9.000 Konfigurationen, für die wir insgesamt 72.000 Durchläufen durchführen. Die Wertebereiche der veränderlichen Konfigurationsparameter sind in Tabelle 4.13 aufgelistet. Die Einschränkungen von Dämpfungsfaktor und Aktivierungsschwellwert sind das Ergebnis vorheriger Experimente, die bereits zeigten, dass eine zu starke Dämpfung oftmals eine zu schwache Trennung zwischen relevanten und nicht relevanten Elementen bewirkt und zu hohe Schwellwerte oftmals unerreichbar sind und daher zum frühzeitigen Ende der Aktivierungsausbreitung führen.

Konfigurationsparameter	**Wertebereich**
Ausbreitungsmodus	{*B, RR, FW, FW-RR, FWL, FWL-RR*}
Aktivierungsmodus	{*IDENTITY, SIG, LOG2, LOG10, LOG2SIG, LOG10SIG*}
Dämpfungsmodus	{*IGNORE, FIXED, INCREASING*}
Verzweigungsmodus	{*NONE, DEGREE, FANOUT, BRANCH*}
Kantengewichtung	{*true, false*}
Dämpfungsfaktor d	[0,8..1]
Aktivierungsschwellwert τ	[0..0,15]
Pulsanzahl	[10..50]

Tabelle 4.13: Wertebereiche der Konfigurationsparameter für den Zufallsgenerator

Zur Bewertung unterschiedlicher Netzvarianten und Konfigurationsparameter verwenden wir die in Abschnitt 2.1.3.3 vorgestellten Bewertungsmaße für Genauigkeit (EP) und Nützlichkeit (NDCG). Für die Gesamtbetrachtung der Güte berechnen wir zusätzlich ein kombiniertes Bewertungsmaß, welches im Folgenden Fitness genannt wird, und dessen Berechnung für einen Durchlauf, bestehend aus der Konfiguration conf und der Anfrage $q \in Q$, in Formel 4.1 dargestellt ist. Dafür kombinieren wir EP und $NDCG_{Min}$ gleichberechtigt, da sowohl die Genauigkeit als auch die Nützlichkeit zur Güte beitragen. Die Überlegungen, die zu dieser Entscheidung führten, sind in Anhang C.3.1 dargelegt. *Fitness*

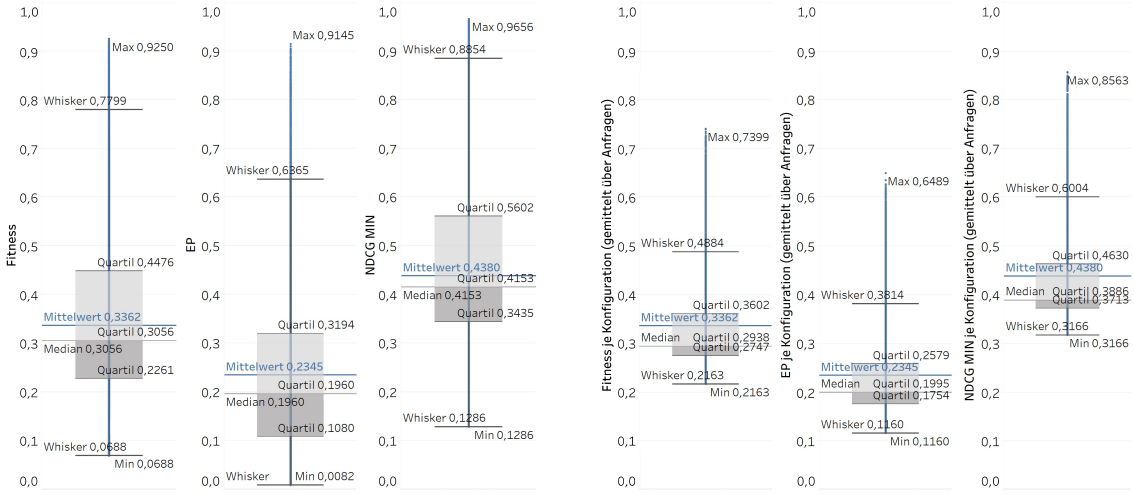
$$\text{Fitness}(conf, q) := \frac{\text{EP}(conf, q) + \text{NDCG}_{\text{Min}}(conf, q)}{2} \tag{4.1}$$

Zur Bewertung, wie gut eine Konfiguration für verschiedene Anfragen performt, wird die Fitness der Durchläufe der jeweiligen Anfragen gemittelt. Die Fitness je Konfiguration, gemittelt über alle Anfragen $q \in Q$, ist in Formel 4.2 definiert.

$$\text{Fitness}(conf) := \frac{1}{|Q|} \cdot \sum_{q \in Q} \text{Fitness}(conf, q) \tag{4.2}$$

4.3.2 Auswertung

Abbildung 4.8 zeigt die Verteilung der Werte für Fitness, EP und $NDCG_{Min}$ für die untersuchten Konfigurationen, zum einen für die einzelnen Durchläufe (Abbildung 4.8a) und zum anderen gemittelt über die Anfragen (Abbildung 4.8b).



(a) (b)

Abbildung 4.8: Verteilung der berechneten Werte für Fitness, EP und $NDCG_{Min}$ auf a) alle Durchläufe und b) alle Konfigurationen gemittelt über alle Anfragen

Für einzelne Durchläufe werden hohe EP-Werte bis 0,9145 und $NDCG_{Min}$-Werte bis 0,9656 erreicht. Die Streubreite für die Bewertungsmaße deckt allerdings fast den gesamten Wertebereich ab. Der Mittelwert[47] der Fitness für alle Durchläufe von 0,3362 liegt unsymmetrisch zwischen dem kleinsten Wert 0,0688 und dem größten Wert 0,925. Der Boxplot zeigt, dass sich die mittleren 50% der Werte, der sogenannte Interquartilbereich, auf niedrigem Niveau zwischen 0,2261 und 0,4476 befinden. Zudem ist eine Rechtsschiefe erkennbar, da das untere Quartil der Box, abgegrenzt durch den Medianwert 0,3056, einen kleineren Wertebereich markiert. Insgesamt betrachtet dominieren niedrige Fitnesswerte. Dennoch existieren Durchläufe mit Fitnesswerten im oberen Wertebereich, häufig im Ausreißerbereich außerhalb der oberen Whisker[48]. Es ist zudem auffällig, dass sich der Interquartilbereich der $NDCG_{Min}$-Werte oberhalb des Interquartilbereichs der EP-Werte befindet, also in mehr Durchläufen hohe $NDCG_{Min}$-Werte als hohe EP-Werte erreicht werden.

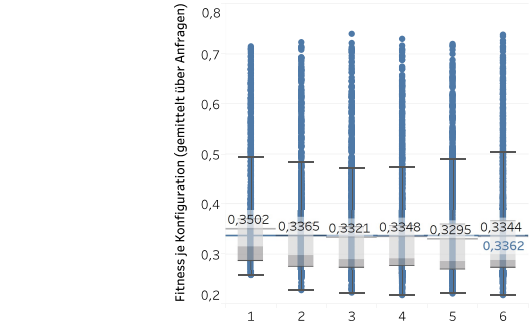
Diese Beobachtungen lassen sich auf die Boxplotdiagramme in Abbildung 4.8b übertragen, in denen die Datenwerte gemittelt über die Anfragen, also je Konfiguration, dargestellt sind. Hier ist erkennbar, dass sich die Streubreite für alle Bewertungsmaße deutlich verringert hat. Diese Stauchung deutet darauf hin, dass die Qualität der Ergebnisse bei gleicher Konfiguration für die unterschiedlichen Anfragen oftmals unterschiedlich ausfällt.

Zwischenfazit: Die Konfigurationen haben starken Einfluss auf die Qualität der Ergebnisse. Eine zufällige Auswahl von Konfigurationen ist aufgrund der großen Streubreite und der Dominanz niedriger Fitnesswerte nicht sinnvoll.

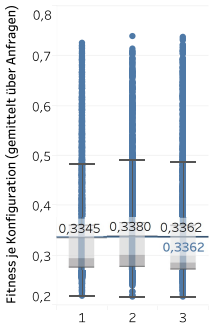
[47] Die Mittelwerte werden in den Abbildungen der Evaluation durch blaue Linien und Beschriftungen dargestellt.
[48] Die Whisker der Boxplotdiagramme in dieser Arbeit kennzeichnen den 1,5-fachen Interquartilabstand.

4.3.2.1 Einfluss der Netzvarianten

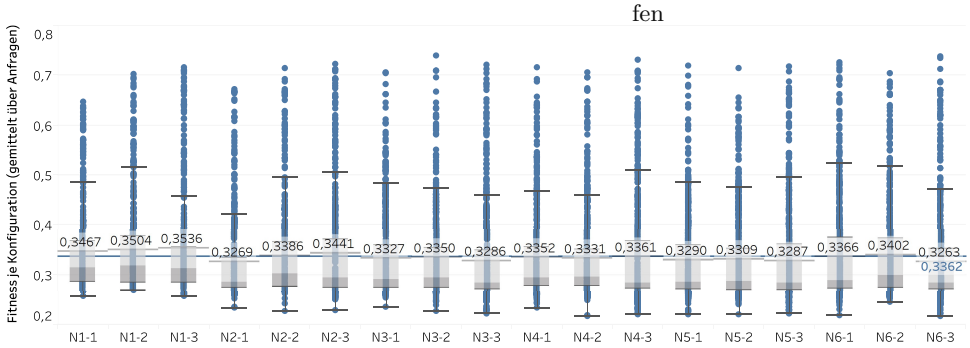
In den Abbildungen 4.9a–4.9c sind die Datenwerte aggregiert über die Anreicherungs- und Reduktionsstufen sowie über alle Netzvarianten dargestellt. Die Mittelwerte der verschiedenen Netzvarianten heben sich, im Vergleich zum Mittelwert aller Fitnesswerte der untersuchten Konfigurationen von 0,3362, nicht sehr stark ab. Die Abweichungen der aggregierten Mittelwerte je Netzvariante zum Gesamtmittelwert reichen von circa -3,0% für das Netz *N6-3* bis zu +5,2% für *N1-3*. Ein ähnliches Bild ergibt sich bei Betrachtung der Verteilung auf die Anreicherungs- und Reduktionsstufen, die aufgrund der Aggregation keine Tendenz für die Favorisierung einer speziellen Netzvariante erkennen lässt.



(a) Anreicherungsstufen



(b) Reduktionsstufen

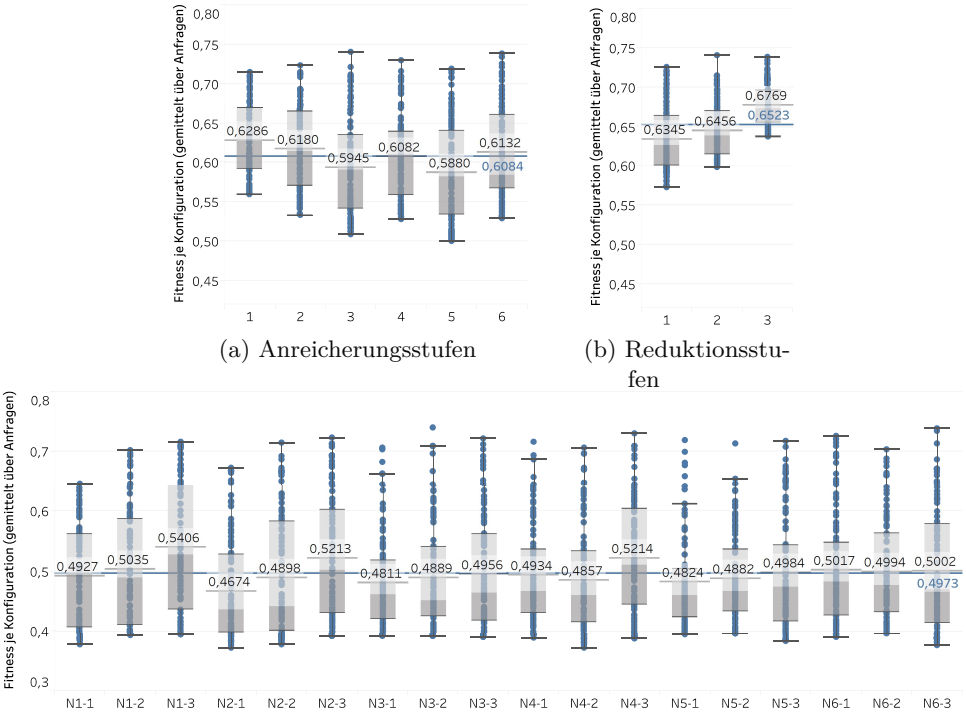


(c) Netzvarianten

Abbildung 4.9: Verteilung der Fitnesswerte je Konfiguration auf a) die Anreicherungsstufen der Netzvarianten, b) die Reduktionsstufen der Netzvarianten und c) die Netzvarianten

Für den Erfolg des untersuchten IR-Systems ist es nicht erforderlich, dass alle Konfigurationen auf allen Netzvarianten sehr gute Fitnesswerte garantieren. Genau genommen suchen wir genau die Konfiguration und die Netzvariante, die sich am besten für alle Anfragen eignen. Daher untersuchen wir zusätzlich Tendenzen im Bereich der besten Konfigurationen. Zu diesem Bereich zählen wir im Folgenden die *Top100* Konfigurationen jedes aggregierten Merkmals, also jeder Spalte.[49] In Abbildung 4.10 ist die Verteilung für die *Top100* Datenwerte jeder Netzvariante dargestellt. Unter den Top-Werten sind leichte Tendenzen erkennbar. Die *Top100* Fitnesswerte der Reduktionsstufe 3 sind im Mittel höher als die der Stufe 2, welche wiederum die Stufe 1 überragen. Trotz der höheren Mittelwerte von Reduktionsstufe 3 existiert der

[49] Am Beispiel bedeutet das, dass in Abbildung 4.10a jeweils die 100 besten Konfigurationen jeder Anreicherungsstufe oder in Abbildung 4.10c jeweils die 100 besten Konfigurationen jeder Netzvariante betrachtet werden. Die Menge aller *Top100* Werte beider Abbildungen muss dabei nicht identisch sein. Daher unterscheiden sich auch die dargestellten Mittelwerte.



(a) Anreicherungsstufen (b) Reduktionsstufen

(c) Netzvarianten

Abbildung 4.10: Verteilung der *Top100* Fitnesswerte je Konfiguration auf a) die Anreicherungsstufen der Netzvarianten, b) die Reduktionsstufen der Netzvarianten und c) die Netzvarianten

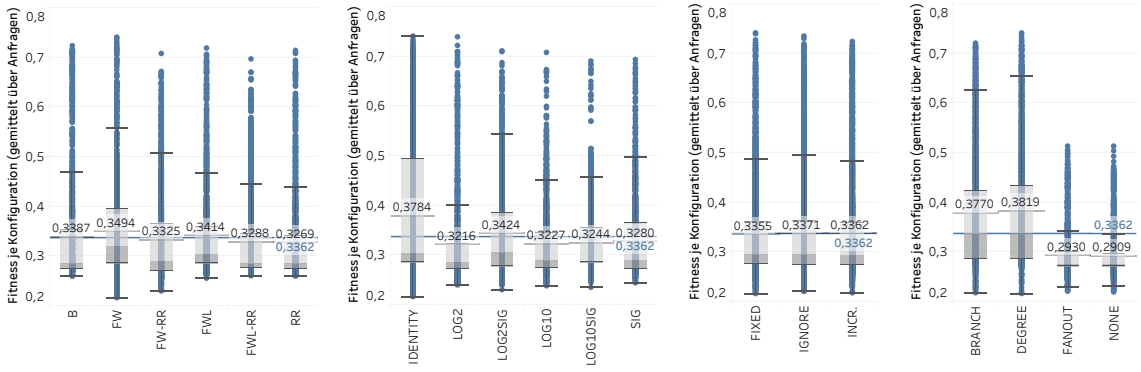
höchste Fitnesswert für einen Durchlauf bei Reduktionsstufe 2. Die Anreicherungsstufen zeigen leichte Abweichungen vom Mittelwert 0,6084. Den niedrigsten Mittelwert besitzt Anreicherungsstufe 5, gefolgt von der Stufe 3, welche allerdings den größten berechneten Fitnesswert besitzt. Die Anreicherungsstufen 1 und 2 haben hingegen die höchsten Mittelwerte, wobei Stufe 1 den höchsten Interquartilbereich aufweist. Die Netzvarianten N3-2 und N6-3 heben sich durch die höchsten Datenwerte ab, während die Netzvarianten N1-3, N2-3 und N4-3 stabile hohe Mittelwerte aufweisen. Die Netzvarianten N2-1, N3-1 und N5-1 fallen hingegen aufgrund ihrer niedrigen Mittel- und Medianwerte sowie der niedrigen Interquartilbereiche negativ auf.

Zwischenfazit: Im Top-Bereich der zufällig generierten Konfigurationen gibt es leichte Tendenzen für den Einsatz semantischer Anreicherungen und Reduktionen von Netzen.

4.3.2.2 Einfluss der Konfigurationsparameter

In Abbildung 4.11 ist die Verteilung der Fitnesswerte je Konfiguration, aggregiert über die verschiedenen Modi zur Konfiguration, dargestellt. Unter den Ausbreitungsmodi erreicht der Modus *FORWARD* sowohl den höchsten Mittelwert als auch den größten berechneten Fitnesswert, weist allerdings aufgrund der großen Streubreite eine geringe Stabilität auf. Eine schwache Tendenz zu niedrigen Fitnesswerten haben die Modi mit RecentReceiver-Beteiligung *RR*, *FW-RR* und *FWL-RR*, die durch geringe Mittel- und Medianwerte auffallen. In den Aktivierungsmodi hebt sich der Modus *IDENTITY* durch den höchsten Mittelwert und die

höchsten Fitnesswerte innerhalb der Whisker, also nicht im Ausreißerbereich, hervor. Der Aktivierungsmodus *LOG2* zeigt den geringsten Mittelwert und die stärkste Verdichtung im unteren Bereich, da mehr als 75% der Werte unterhalb des Gesamtmittelwerts liegen. Dennoch existieren einige Konfigurationen mit sehr hohen Fitnesswerten, unter anderem mit der höchsten berechneten Fitness. Die Dämpfungsmodi weisen keine Tendenz auf. Die Verzweigungsmodi hingegen zeigen klar, dass die Modi *BRANCH* und *DEGREE* nicht nur im Mittel besser als die Modi *FANOUT* und *NONE* sind, sondern insbesondere im oberen Fitnessbereich deutlich mehr Potenzial besitzen.



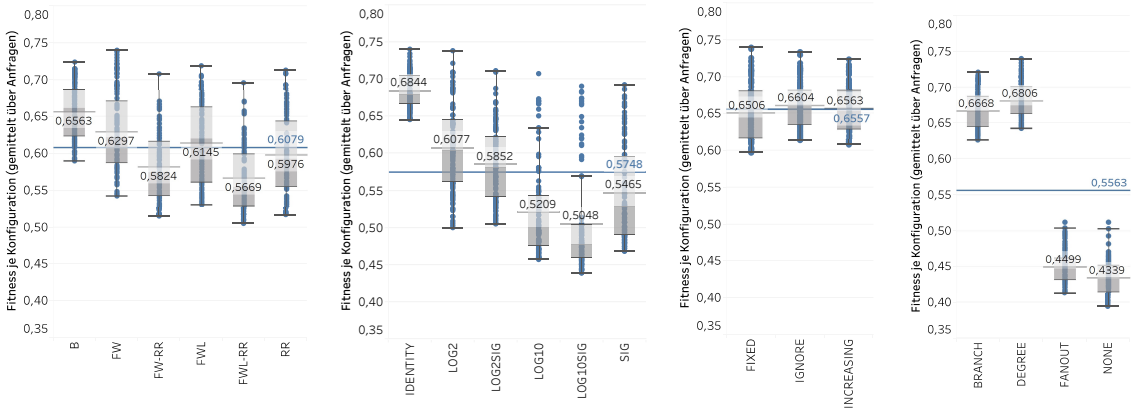
(a) Ausbreitungsmodus (b) Aktivierungsmodus (c) Dämpfungsmodus (d) Verzweigungsmodus

Abbildung 4.11: Verteilung der Fitnesswerte je Konfiguration auf die a) Ausbreitungsmodi, b) Aktivierungsmodi, c) Dämpfungsmodi und d) Verzweigungsmodi

Betrachtet man in Abbildung 4.12 die *Top100* Datenwerte je Konfiguration, so ist deutlicher erkennbar, dass die Verzweigungsmodi *FANOUT* und *NONE* keinen positiven Effekt auf die Fitnesswerte haben. Unter den Aktivierungsmodi besitzt *IDENTITY* die stabilsten und höchsten Fitnesswerte, gefolgt von Modus *LOG2*, für den dennoch lediglich das oberste Quartil die gesamte Streubreite von *IDENTITY* abdeckt. Das bedeutet, 75% aller *Top100* Werte von *LOG2* sind kleiner als jeder *Top100* Fitnesswert von *IDENTITY*. Die geringste Fitness zeigen die Modi *LOG10* und *LOG10SIG*, die jeweils mehr als 75% ihrer Top Datenwerte unterhalb des Mittelwerts verorten. Unter den Ausbreitungsmodi zeigt das Diagramm insbesondere im Modus *B* eine höhere Verdichtung im oberen Fitnessbereich, der Modus *FW* hingegen kann höhere Einzelwerte aufweisen. Wie bereits in Abbildung 4.11a existiert eine schwache Tendenz zu niedrigen Fitnesswerten unter den Modi mit *RR*-Beteiligung.

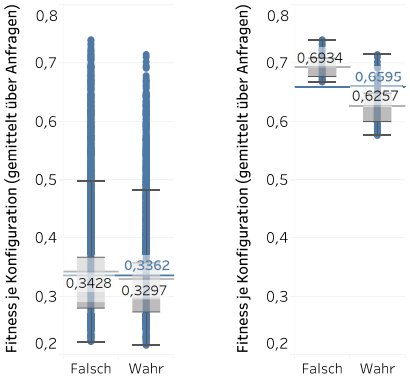
In Abbildung 4.13 sind die Fitnesswerte für die Verwendung der Kantengewichtung dargestellt. Während bei Betrachtung aller Durchläufe lediglich eine leichte Tendenz gegen die Verwendung der Kantengewichtung erkennbar ist, zeigt sich unter den *Top100* Fitnesswerten beider Optionen eine deutlich klarere Tendenz. Es existieren im dargestellten Bereich deutlich mehr sehr gute Konfigurationen ohne gewichtete Kanten als mit Kantengewichtung. Alle ungewichteten *Top100* Konfigurationen besitzen bessere Fitnesswerte als 75% der gewichteten.

In Abbildung 4.14 sind Streudiagramme zur Verteilung der Fitnesswerte auf den Dämpfungsfaktor und den Aktivierungsschwellwert dargestellt. Es ist eine deutliche Verdichtung im unteren Fitnessbereich über die gesamten Wertebereiche beider Faktoren erkennbar. Dämpfungsfaktor 1 entspricht dem Dämpfungsmodus *IGNORE* und enthält daher mehr Datenwerte als der übrige Bereich. Die Diagramme zeigen eine schwache Tendenz kleiner Schwellwerte und hoher Dämpfungsfaktoren zu höheren Fitnesswerten.



(a) Ausbreitungsmodus (b) Aktivierungsmodus (c) Dämpfungsmodus (d) Verzweigungsmodus

Abbildung 4.12: Verteilung der *Top100* Fitnesswerte je Konfiguration auf die a) Ausbreitungsmodi, b) Aktivierungsmodi, c) Dämpfungsmodi und d) Verzweigungsmodi



(a) Je Durchlauf (b) Top100

Abbildung 4.13: Verteilung der Fitnesswerte auf die Kantengewichtung für a) alle Konfigurationen und b) die *Top100* Konfigurationen

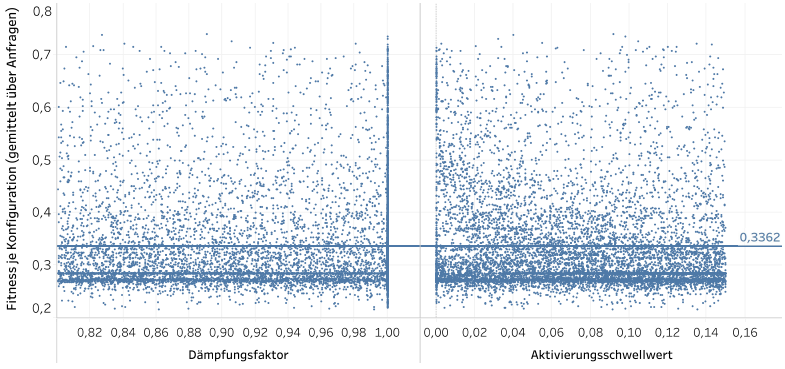


Abbildung 4.14: Verteilung der Fitnesswerte je Konfiguration auf den Dämpfungsfaktor und den Aktivierungsschwellwert

Abbildung 4.15 zeigt die Verteilung der Fitnesswerte aller Konfigurationen auf die Anzahl der Pulse ohne deutliche Tendenz.

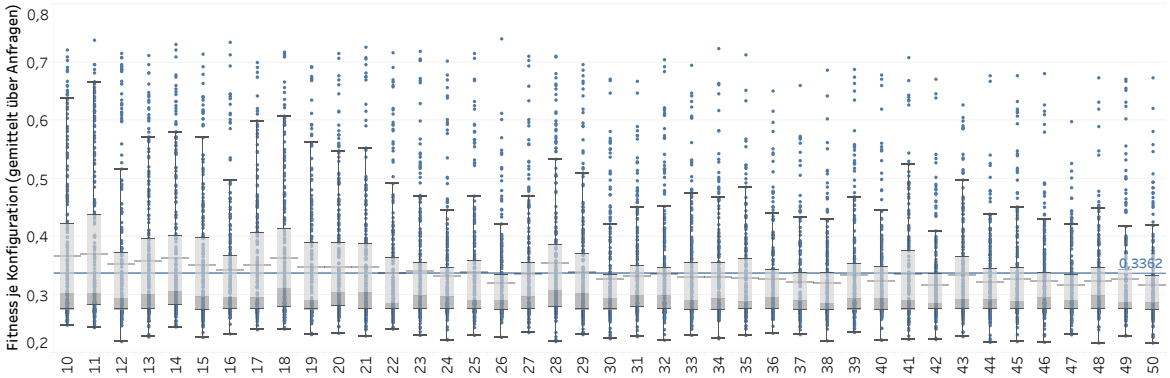


Abbildung 4.15: Verteilung der Fitnesswerte je Konfiguration auf die Anzahl der Pulse

Zwischenfazit: Die Verzweigungsmodi *FANOUT* und *NONE* haben keinen positiven Effekt auf die Qualität der Ergebnisse. Für den sinnvollen Einsatz der Ausbreitungsmodi, Aktivierungsmodi, Dämpfungsmodi, Dämpfungsfaktoren und Aktivierungsschwellwerte und für die Verwendung der Kantengewichtung sind in den zufällig generierten Konfigurationen lediglich schwache Tendenzen erkennbar.

4.3.2.3 Einfluss der Anfragen

Abbildung 4.16 stellt die Verteilung der Fitnesswerte aller Konfigurationen je Anfrage dar.

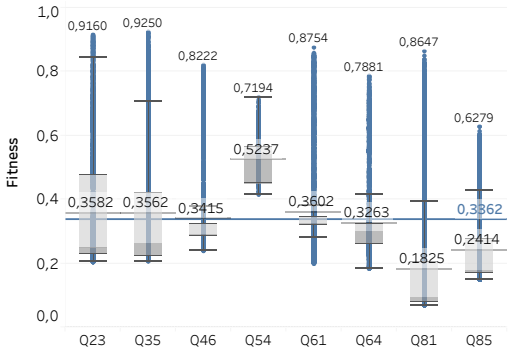


Abbildung 4.16: Verteilung der Fitnesswerte je Konfiguration auf die Anfragen

Das Diagramm zeigt deutliche Unterschiede in der Streubreite der Fitnesswerte und ihrer Höhe. Während die Anfrage Q54 eine geringe Streubreite bei höchstem aggregierten Mittelwert aufweist, besitzen beispielsweise die Anfragen Q23, Q35 und Q81 eine wesentlich höhere Streubreite bei niedrigerem aggregierten Mittelwert, aber dennoch für viele Konfigurationen höhere individuelle Fitnesswerte. Die Anfrage Q54 zeigt sich im Vergleich zu den anderen Anfragen relativ unabhängig von den Konfigurationen.

Zwischenfazit: Die Anfragen haben starken Einfluss auf die Qualität der Ergebnisse.

4.3.3 Fazit

***EQ2**: Wie stark ist der Einfluss der Netzvarianten, Konfigurationsparameter und Anfragen auf die Qualität der Ergebnisse? Welche Bedeutung hat die Parametrisierung?*

Die Auswahl der Netzvarianten und die Auswahl der Konfigurationsparameter haben großen Einfluss auf die Qualität der Ergebnisse. Die untersuchten zufälligen Konfigurationen sind aufgrund der großen Streubreite der Fitnesswerte und deren rechtsschiefer Lage mit starker Verdichtung im unteren Fitnessbereich häufiger ungeeignet als geeignet. Die Parametrisierung hat daher eine große Bedeutung, denn eine gezielte Auswahl geeigneter Konfigurationen kann zu guten Ergebnissen des IR-Systems führen. Die Auswahl einer geeigneten Konfiguration ist allerdings sehr schwer. Die Datenauswertung in Abschnitt 4.3.2 zeigt oftmals lediglich schwache Tendenzen. So lässt sich beispielsweise keine klare Tendenz für die Favorisierung einer speziellen Netzvariante, dennoch aber eine schwache Tendenz für zunehmende Reduktionsstufen erkennen. Auch bei den Modi zur Konfiguration sind lediglich schwache Tendenzen zu erkennen, beispielsweise für die Ausbreitungsmodi *B* und *FW* und die Aktivierungsmodi *IDENTITY* und *LOG2*. Eine starke Tendenz zu sehr geringen Fitnesswerten zeigen die Verzweigungsmodi *FANOUT* und *NONE*. Weitere schwache Tendenzen zu guten Fitnesswerten zeigen zudem die Konfigurationen mit ungewichteten Kanten sowie niedrige Aktivierungsschwellwerte und hohe Dämpfungsfaktoren. Zudem konnten wir zeigen, dass die Qualität der Ergebnisse bei gleicher Konfiguration für die unterschiedlichen Anfragen sehr unterschiedlich ausfällt. Das Gesamtbild der anfrageabhängigen Datenanalyse führt zu der Vermutung, dass die Anfragen starken Einfluss auf die Qualität der Ergebnisse haben.

Wir vermuten, dass komplexe Abhängigkeiten zwischen den Konfigurationsparametern und eventuell auch zu den Netzvarianten bestehen. Die zufällig generierten Konfigurationen eignen sich nicht für deren Analyse. Aufgrund der hohen Anzahl von Konfigurationen mit geringen Fitnesswerten kann das Spektrum geeigneter Konfigurationen nicht optimal ausgewertet werden. Im folgenden Abschnitt werden daher mithilfe evolutionärer Algorithmen optimierte Konfigurationen gesucht und untersucht. Die Optimierungsläufe enthalten, im Gegensatz zu den zufällig generierten Konfigurationen, üblicherweise eine höhere Anzahl an Konfigurationen mit hohen Fitnesswerten. Daher kann eine anschließende Datenanalyse aussagekräftigere Ergebnisse zur Beantwortung der Evaluationsfrage EQ3 liefern, die sich mit der Bewertung der Effekte der einzelnen Faktoren (Netzvariante, Konfigurationsparameter) befasst.

4.4 Konfigurationsoptimierung mit evolutionären Algorithmen

Während sich Abschnitt 4.3 den Ergebnissen der Suchläufe im IR-System mit zufälligen Konfigurationen widmet, untersuchen wir in diesem Abschnitt, wie weit die Güte der Ergebnisse durch Optimierung der Konfigurationen verbessert werden kann. Dafür verwenden wir evolutionäre Algorithmen, um optimale Konfigurationen zunächst für die Gesamtheit der Anfragen und anschließend für die einzelnen Anfragen zu identifizieren. Die Top-Konfigurationen und ihre Parametrisierungen verwenden wir zur Auswertung, welche Konfigurationsparameter und Netzvarianten sich besonders gut eignen. Die Erkenntnisse aus den Analysen beider Optimierungen verwenden wir in einer abschließenden Auseinandersetzung dazu, welches Vorgehen bei der Auswahl geeigneter Konfigurationen sinnvoll ist.

4.4.1 Durchführung

Für jede der 18 Netzvarianten führen wir einen Optimierungslauf mittels evolutionärer Algorithmen durch. Dafür erstellen wir je Generation eine Population von 50 Individuen. Jedes Individuum entspricht einer Konfiguration, für die das IR-System für jede der acht Anfragen die Fitness berechnet. Während die erste Generation vollständig zufällig erzeugt wird, besteht jede folgende Generation zu definierten Anteilen aus überlebenden Konfigurationen, aus Konfigurationen, welche durch Mutation verändert wurden, und aus erneut zufällig erzeugten Konfigurationen. Als Abbruchkriterium legen wir zum einen das Erreichen der größtmöglichen Fitness mit dem Wert 1 fest. Zum anderen beenden wir einen Optimierungslauf bei Stagnation der Fitness über 15 Generationen oder nach der Berechnung von 200 Generationen. Im Folgenden erläutern wir, wie wir die evolutionären Operationen Selektion und Mutation in dieser Arbeit verwenden.

Optimierungsziele: Wir untersuchen zwei Optimierungsziele, das anfrageübergreifende OZ1 und das anfragespezifische OZ2.

OZ1: Wir suchen für jede Netzvariante die Konfigurationen, die für die Gesamtheit der untersuchten Anfragen bestmögliche Ergebnisse liefern. Daher optimieren wir die über die Menge aller Anfragen gemittelte Fitness je Netzvariante (siehe Formel 4.2, Seite 137). Optimale Konfigurationen sind demnach universell für alle untersuchten Anfragen einsetzbar.

OZ2: Wir suchen für jede Netzvariante nach Konfigurationen, die für jede Anfrage bestmögliche Ergebnisse liefern. Optimiert wird daher die Fitness jeder einzelnen Anfrage je Netzvariante (siehe Formel 4.1, Seite 137). Optimale Konfigurationen sind dann spezifisch für jede Anfrage einsetzbar.

Zufällige Erzeugung von Konfigurationen: Die Wertebereiche der Konfigurationsparameter entsprechen denen aus Tabelle 4.13 (Seite 137), mit der Ausnahme, dass die Verzweigungsmodi *NONE* und *FANOUT* aufgrund der Erkenntnisse in Abschnitt 4.3 bereits exkludiert werden.

Selektion von Konfigurationen: Werden für eine neue Generation lediglich die besten Individuen der vorherigen Generation ausgewählt, so kann eine Häufung von Individuen der gleichen Abstammung im oberen Fitnessbereich zu geringer Diversität und einem zu schnellen Ausschluss von Individuen mit gutem Fitnesspotenzial führen. Daher fassen wir alle Individuen zu Familien zusammen. Eine Familie umfasst alle Individuen, die ursprünglich

vom gleichen Individuum abstammen. In der nach Fitnesswerten sortierten Liste verbleiben lediglich die vier besten Individuen jeder Familie. Die besten Individuen dieser Liste bilden ohne Veränderung 5% der neuen Generation. Für weitere 65% der neuen Generation werden ebenfalls die besten Individuen der Liste gewählt. Diese werden, durch Mutation verändert, als Nachfolger in die neue Generation aufgenommen. Die verbleibenden 30% werden zufällig erzeugt.

Mutation von Konfigurationen: Für jede zu mutierende Konfiguration wird zuerst zufällig ein Konfigurationsparameter bestimmt. Dieser wird so verändert, dass eine neue Konfiguration entsteht. Ist der zu mutierende Konfigurationsparameter ein Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus, Verzweigungsmodus oder die Kantengewichtung, so wird zufällig eines der verbleibenden Elemente des jeweiligen Wertebereichs gewählt, der nicht in der Ausgangskonfiguration verwendet wurde. Für Dämpfungsfaktor und Aktivierungsschwellwert wird zufällig ein neuer Wert innerhalb des Wertebereichs bestimmt, der jeweils einen maximalen Abstand zum Ausgangswert von 0,02 haben darf. Dadurch wird die Nähe zum Ausgangswert garantiert. Zudem wird sichergestellt, dass der Aktivierungsschwellwert mit einer Wahrscheinlichkeit von 5% den Wert 0 annehmen kann, um auch den Verzicht auf Aktivierungsschwellwerte untersuchen zu können. Bei Mutation der Pulsanzahl wird zufällig ein ganzzahliger Wert bestimmt, der einen maximalen Abstand von 2 zum Ausgangswert besitzt.

Es folgt die Analyse des ersten Optimierungsziels zur Identifikation anfrageübergreifend optimaler Konfigurationen.

4.4.2 Auswertung der anfrageübergreifenden Optimierung

In Abbildung 4.17 ist die Entwicklung des maximalen Fitnesswertes über die Generationen der Optimierungsläufe für die einzelnen Netzvarianten dargestellt. Die Abbildung zeigt, dass die Fitness in allen Optimierungsläufen steigt.

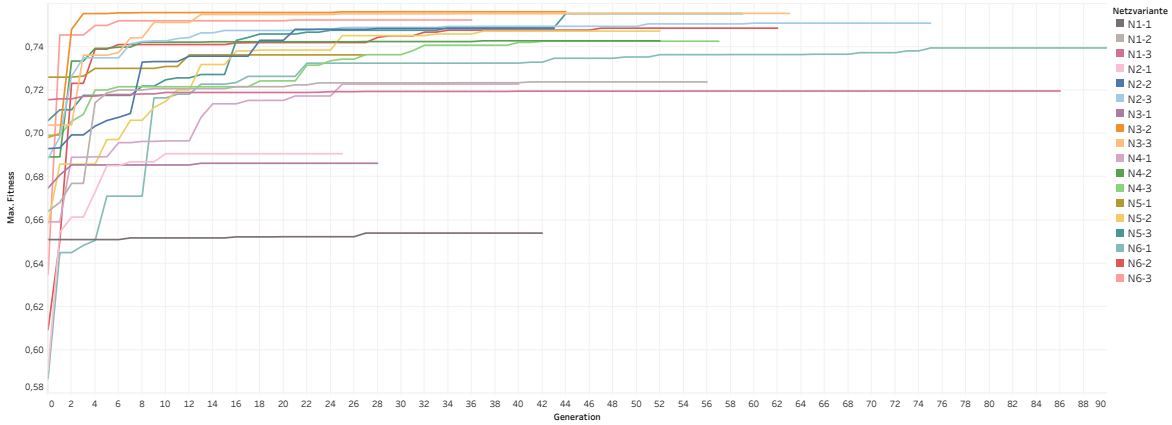
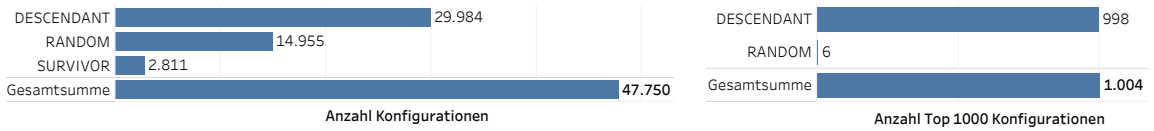


Abbildung 4.17: Entwicklung des maximalen Fitnesswertes über die Generationen der Optimierungsläufe je Netzvariante

Abbildung 4.18a dokumentiert den Umfang der Optimierungsläufe für *OZ1*. Sie stellt die Anzahl aller untersuchten Konfigurationen sowie die zahlenmäßigen Anteile der zufällig gewählten Konfigurationen (*RANDOM*), der mutierten Konfigurationen (*DESCENDANT*) und der überlebenden Konfigurationen (*SURVIVOR*) dar. In den Optimierungsläufen für das *OZ1*

wurden insgesamt 47.750 Konfigurationen für jede der acht Anfragen in dementsprechend 382.000 Durchläufen des IR-Systems untersucht. Davon sind 29.984 Konfigurationen durch Mutation erzeugte Nachfolger der besten Konfigurationen und 14.995 zufällig erzeugte Konfigurationen. Die 2.811 überlebenden Konfigurationen sind ausgewählte Kopien zufälliger oder bereits mutierter Konfigurationen, die aufgrund ihrer Güte der Population einer neuen Generation unverändert beigefügt wurden. Diese doppelten Konfigurationen werden in folgenden Analysen ausgeschlossen.



(a) (b)

Abbildung 4.18: Anteilige Anzahl der Evolutionstypen von a) allen untersuchten Konfigurationen und b) den *Top1000* Konfigurationen

Die beste anfrageübergreifende Konfiguration wurde für das Netz *N3-2* gefunden und besitzt eine Fitness von 0,7563, die sich aus einem EP-Wert von 0,6645 und einem $NDCG_{Min}$-Wert von 0,8480 zusammensetzt. Die Parameter der Konfiguration sind in Tabelle 4.14 aufgelistet. Diese Konfiguration werden wir in Abschnitt 4.4.4 verwenden, um die Relevanz und Nützlichkeit der Ergebnisse des IR-Systems zu bewerten und die Evaluationsfrage EQ4 zu beantworten.

Optimierte Konfiguration	$\mathbf{K_Q}$
Ausbreitungsmodus	*FW*
Aktivierungsmodus	*IDENTITY*
Dämpfungsmodus	*INCREASING*
Verzweigungsmodus	*DEGREE*
Kantengewichtung	*false*
Dämpfungsfaktor d	0,8971
Aktivierungsschwellwert τ	0,1090
Pulsanzahl	26
Netzvariante	*N3-2*
Fitness	0,7563
EP	0,6645
$NDCG_{Min}$	0,8480

Tabelle 4.14: Parametrisierung der besten anfrageübergreifenden Konfiguration

In Abbildung 4.19 wird die Verteilung der berechneten Werte für Fitness, EP und $NDCG_{Min}$ sowohl für alle Durchläufe der Konfigurationen für alle Anfragen dargestellt, als auch für die über alle Anfragen gemittelten Konfigurationen. Die Werte für die höchste Fitness von 0,9333 und die höchste über die Anfragen gemittelte Fitness von 0,7563 liegen oberhalb der entsprechenden Werte für die ausschließlich zufällig generierten Konfigurationen in Abbildung 4.8 (Seite 138). Im Gegensatz zu den zufällig generierten Konfigurationen, die eine Rechtsschiefe und viele Ausreißer aufweisen, zeigen die optimierten Konfigurationen deutlich mehr gute Fitnesswerte und eine ausgewogenere Verteilung.

Die Optimierungsläufe enthalten, im Gegensatz zu den zufällig generierten Konfigurationen, eine höhere Anzahl an Konfigurationen mit hohen Fitnesswerten. Wir konnten zwar eine

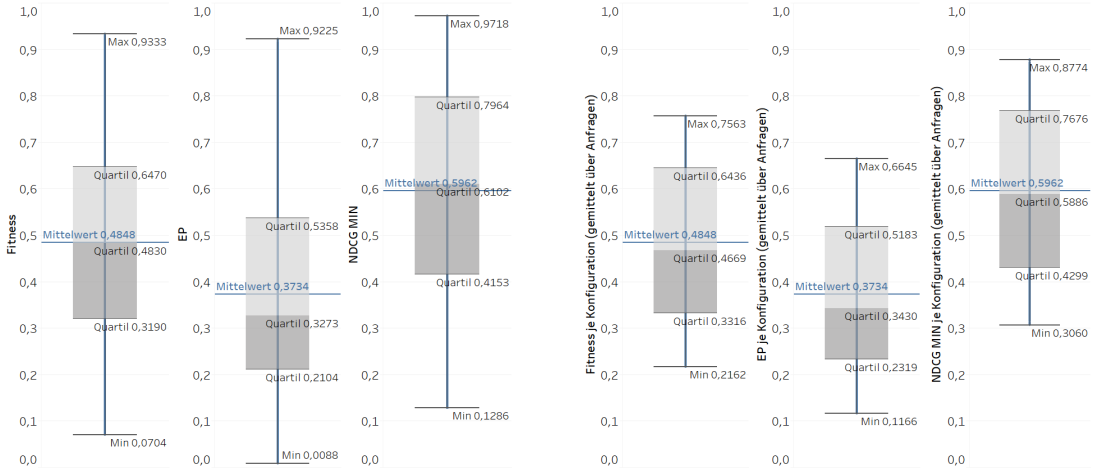


Abbildung 4.19: Verteilung der berechneten Werte für Fitness, EP und $NDCG_{Min}$ auf a) alle Durchläufe und b) alle Konfigurationen gemittelt über alle Anfragen

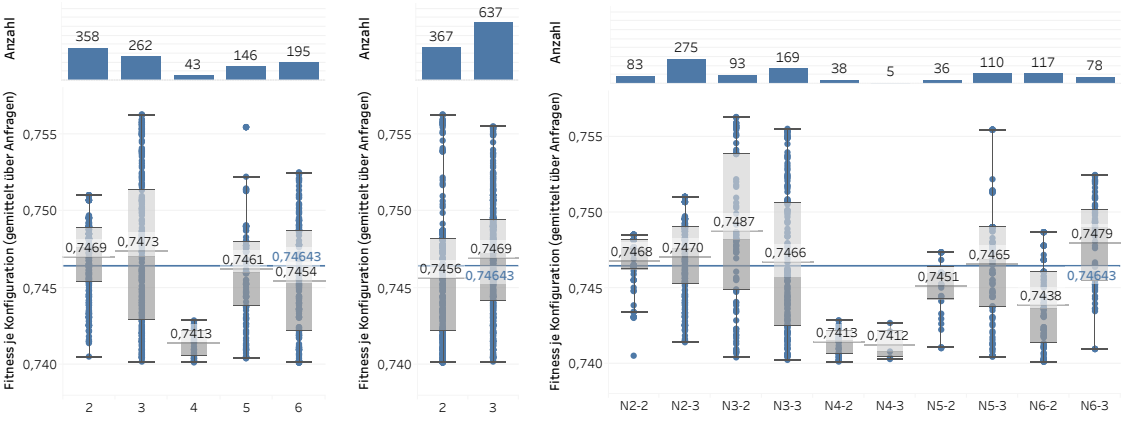
beste Konfiguration identifizieren, die Ergebnisse der Top-Konfigurationen liegen allerdings nah beieinander. Für eine differenzierte Beantwortung der Evaluationsfrage EQ3 untersuchen wir daher in den folgenden Analysen nicht nur die beste Konfiguration, sondern die Menge der besten Konfigurationen. Diese legen wir auf die *Top1000* fest. Da mehrere Konfigurationen den gleichen Fitnesswert wie die tausendste Konfiguration besitzen, umfasst die Menge der *Top1000* genau genommen 1.004 Konfigurationen mit Fitnesswerten zwischen 0,7401 und 0,7563. Wir bezeichnen sie im Folgenden dennoch als *Top1000*.

4.4.2.1 Auswirkungen der Netzvarianten

In Abbildung 4.20 sind im unteren Bereich die Verteilungen der Fitnesswerte auf die Anreicherungsstufen und Reduktionsstufen der Netzvarianten sowie auf die einzelnen Netzvarianten (Tabelle 4.1, Seite 121) dargestellt. Im oberen Bereich sehen wir die zahlenmäßige Verteilung der Konfigurationen. Es existieren acht Netzvarianten, für die keine Konfigurationen gefunden werden konnten, deren Fitness das Mindestmaß erreicht. Dazu gehören alle sechs Netze der Reduktionsstufe 1, also alle unreduzierten Netze. Des Weiteren sind alle Netze der Anreicherungsstufe 1, also die Netze ohne spezielle semantische Anreicherungen in allen drei Reduktionsstufen, nicht in der Auswahl der Netze mit besten Fitnesswerten enthalten. Daraus folgern wir, dass sowohl semantische Anreicherungen als auch Reduktionen beim Aufbau semantischer Netze sinnvoll sind und einen positiven Einfluss auf die Qualität der Ergebnisse haben.

> ***Zwischenfazit:*** Der Einsatz der semantischen Anreicherungen und der Reduktionen der Netze ist sinnvoll. Sie tragen essenziell zur Qualität der Ergebnisse bei.

Die Anreicherungsstufe, die für die semantische Anreicherung mit einfachen Keywords (3) steht, zeigt den höchsten gemittelten Fitnesswert, gefolgt von der Anreicherung durch ausschließlich Synonyme (2), durch einfache Keywords und Synonymbeziehungen (5) und durch erweiterte Keywords und Synonymbeziehungen (6). Die Anreicherungsstufe 4, die ausschließlich erweiterte Keywords enthält, besitzt den niedrigsten Mittelwert und zudem mit 43 Konfigurationen



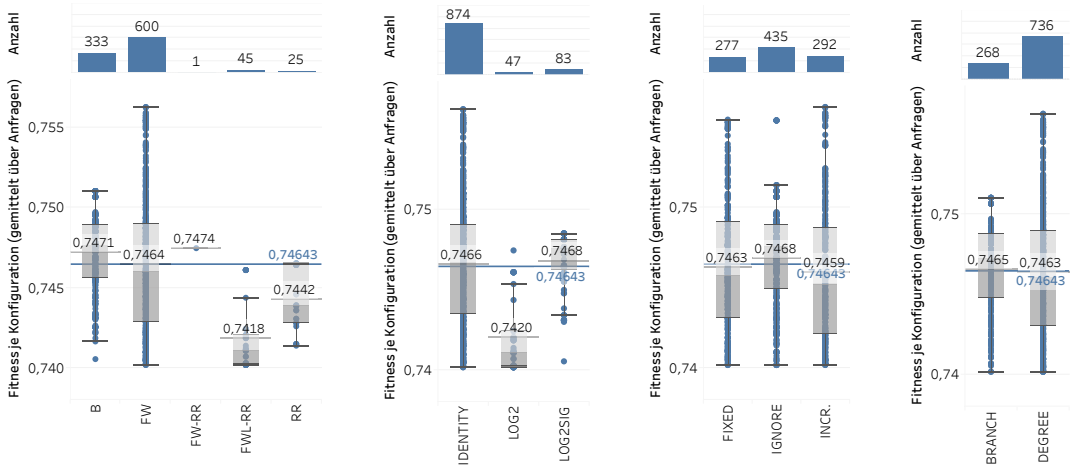
(a) Anreicherungsstufen (b) Reduktionsstufen (c) Netzvarianten

Abbildung 4.20: Anzahl und Verteilung der *Top1000* Fitnesswerte auf die a) Anreicherungsstufen der Netzvarianten, b) Reduktionsstufen der Netzvarianten und c) Netzvarianten

den geringsten Anteil in den *Top1000*. Insbesondere die Anreicherungsstufen 2 und 3 zeigen sowohl hohe als auch stabile Werte in den oberen 50% der Datenpunkte und besitzen zudem mit 358 beziehungsweise 262 Konfigurationen die größten Anteile an den *Top1000*. Die Anreicherungsstufen 5 und 6 zeigen zum einen weniger Konfigurationen (146 beziehungsweise 195), die zum anderen eine geringfügig geringere Qualität aufweisen. Unter den einzelnen Netzvarianten heben sich die Netze *N3-2*, *N3-3*, *N5-3* und *N6-3*, gefolgt von *N2-3* durch sehr hohe Fitnesswerte ab.

4.4.2.2 Auswirkungen der Konfigurationsparameter

Für die *Top1000* Konfigurationen sind in Abbildung 4.21 die Verteilung der Anzahl und der Fitnesswerte für die Ausbreitungsmodi, Aktivierungsmodi, Dämpfungsmodi und Verzweigungsmodi dargestellt.
Der Ausbreitungsmodus *FW* besitzt die meisten und höchsten Fitnesswerte, gefolgt von *B* mit weniger, niedrigeren, aber dennoch stabilen Fitnesswerten. Die Modi *FWL-RR* und *RR* hingegen erreichen deutlich weniger und niedrigere Fitnesswerte im oberen Bereich, der Ausbreitungsmodus *FW-RR* besitzt lediglich eine Konfiguration unter den *Top1000*. Für die Aktivierungsmodi *LOG10*, *LOG10SIG* und *SIG* konnte keine Konfiguration die *Top1000* erreichen. Unter den verbleibenden Aktivierungsmodi ist die Verwendung der Identitätsfunktion *IDENTITY* am erfolgreichsten, gefolgt von sowohl der Logarithmusfunktion mit Basis 2 und nachgesetzter Sigmoidfunktion *LOG2SIG* als auch der alleinigen Logarithmusfunktion *LOG2*. Sowohl die Verwendung der alleinigen Logarithmusfunktion mit Basis 10 oder mit deren nachgesetzter Sigmoidfunktion als auch die alleinige Verwendung der Sigmoidfunktion scheinen die Werte so stark zu diskriminieren und zu wenig zu differenzieren (siehe Funktionsgraphen in Tabelle 3.8, Seite 93), dass eine ungünstige Wirkung auf die Fitnesswerte erkennbar wird. Die Dämpfungsmodi für steigende Dämpfung *INCREASING* und konstante Dämpfung *FIXED* besitzen einige höhere Fitnesswerte als der Dämpfungsmodus für den Verzicht auf Dämpfung *IGNORE*. Dieser besitzt wiederum mehr Konfigurationen mit überdurchschnittlicher Fitness. Bei den Verzweigungsmodi dominiert der Modus *DEGREE* deutlich nach Anzahl und leicht nach Höhe der Fitnesswerte.



(a) Ausbreitungsmodus (b) Aktivierungsmodus (c) Dämpfungsmodus (d) Verzweigungsmodus

Abbildung 4.21: Anzahl und Verteilung der *Top1000* Fitnesswerte auf die a) Ausbreitungsmodi, b) Aktivierungsmodi, c) Dämpfungsmodi und d) Verzweigungsmodi

Zwischenfazit: Der Einsatz von Ausbreitungsmodi ist sinnvoll. Unter den *Top1000* anfrageübergreifenden Konfigurationen existieren mehr und höhere Fitnesswerte für den Ausbreitungsmodus *FW* als für die Basisvariante *B*, die für die ungebremste Flutung des Netzes ohne spezifisches Ausbreitungsverhalten steht.

Zwischenfazit: Der Aktivierungsmodus *IDENTITY* dominiert, während Aktivierungsmodi mit logarithmischer oder sigmoider Abschwächung entfallen oder eine untergeordnete Rolle im Bereich der Top-Konfigurationen spielen. Der Verzweigungsmodus *DEGREE* zeigt eine bessere Qualität als der Verzweigungsmodus *BRANCH*.

Kombination der Konfigurationsparameter: Abbildung 4.22 stellt die Verteilung der Fitnesswerte auf die verschiedenen Konfigurationsparameter in einem Diagramm mit farblich hervorgehoben Netzvarianten dar. Die Größe der Kreise symbolisiert die Pulsanzahl der Konfigurationen. Zunächst erkennen wir in diesem Diagramm, dass eine Vielzahl der 216 möglichen Kombinationen aus den sechs Ausbreitungsmodi, sechs Aktivierungsmodi, zwei Dämpfungsmodi und drei Verzweigungsmodi in den *Top1000* nicht mehr vertreten ist. Wenngleich die verbliebenen 29 enthaltenen Kombinationen zunächst komplex wirken, stellen sie bereits einen stark eingegrenzten Parameterraum für die Auswahl qualitativ guter Konfigurationen dar. Diese Reduktion geeigneter Kombinationen ist gut, da sie zum einen die Eignung bestimmter Kombinationen betont und zum anderen zeigt, dass die Auswahl nicht beliebig ist.

Das Diagramm in Abbildung 4.22 zeigt zudem, dass im dargestellten Fitnessbereich der *Top1000* Konfigurationen einige Konfigurationskombinationen und Netzvarianten besonders gut miteinander harmonieren. Darüber hinaus zeigen einige Konfigurationskombinationen sogar eine exklusive Eignung für bestimmte Netzvarianten und umgekehrt. Beispielsweise wird deutlich, dass für die Kombinationen *B* + *IDENTITY* und *RR* + *IDENTITY* fast alle Datenpunkte auf das Netz *N2-3* entfallen, welches zudem kaum positive Fitnesswerte in anderen Konfigurationskombinationen erreichen kann. Die Konfigurationen mit Beteiligung

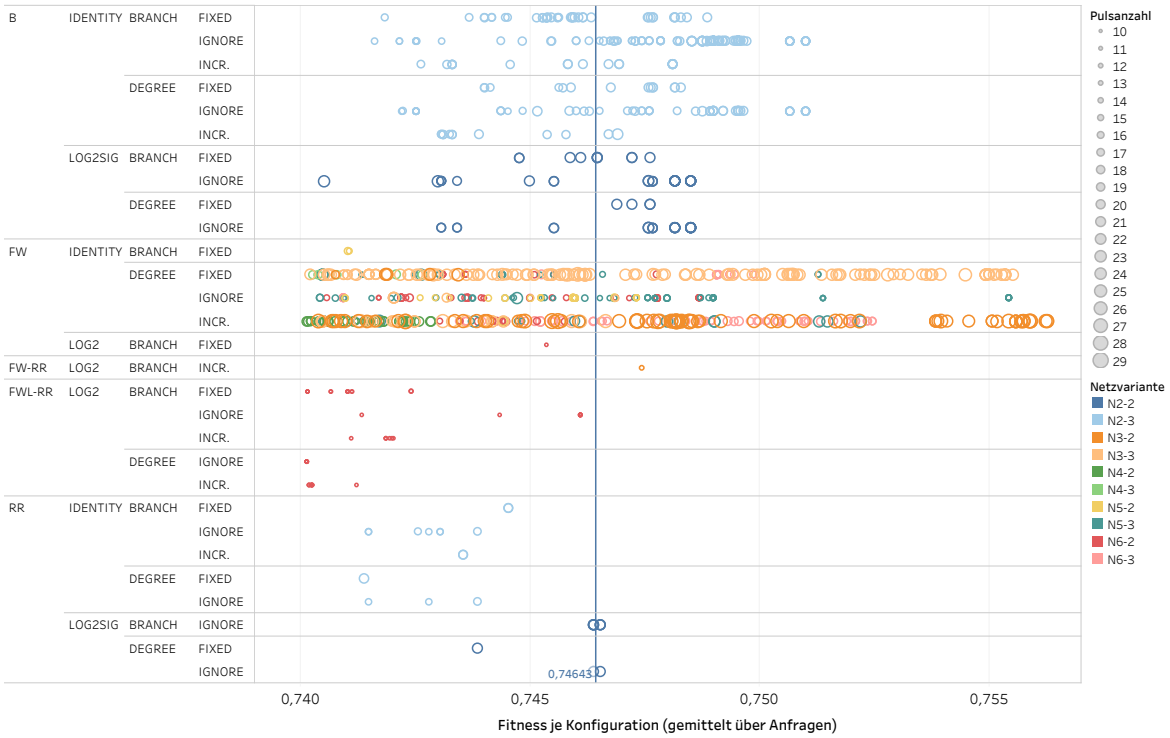


Abbildung 4.22: Verteilung der *Top1000* Fitnesswerte auf die Netzvarianten und die Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus

von *LOG2SIG* verteilen sich zum größten Teil auf das Netz *N2-2*, während *LOG2* zum größten Teil gute Fitnesswerte für das Netz *N6-2* zeigt. Die höchsten Fitnesswerte werden in der Kombination *FW* + *IDENTITY* + *DEGREE* erreicht, welche Datenpunkte für verschiedene Netze besitzt, insbesondere die der Anreicherungsstufen 3, 4 und 5. Fixe Dämpfungsfaktoren zeigen dabei bessere Ergebnisse für das Netz *N3-3* und steigende Dämpfungsfaktoren bessere Fitnesswerte für das Netz *N3-2*. Die Harmonie zwischen Netzvarianten und bestimmten Kombinationen bedeutet auch, dass die Anzahl geeigneter Kombinationen sogar weiter reduziert werden kann, wenn wir uns auf eine Netzvariante beschränken.

In Abbildung 4.23 ist dargestellt, welche Kombinationen von Konfigurationsparametern in den *Top1000* besonders häufig gemeinsam auftreten. Die Ausbreitungsmodi sind farblich markiert und die Größe der Kreise skaliert mit der Anzahl an Knoten, die sie repräsentieren. Damit sind Aussagen zur Stabilität von Kombinationen aus Konfigurationsparameter möglich. Die Stabilität einer Kombination zeigt sich durch die möglichst hohe Anzahl an Konfigurationen, die sie in den *Top1000* enthalten. Die Kombination *FW* + *IDENTITY* + *DEGREE* ist beispielsweise sehr stabil. Sie dominiert auch zahlenmäßig sehr stark, wobei mehr Konfigurationen (241) steigende als konstante (190) Dämpfungsfaktoren besitzen, gefolgt von keiner Dämpfung (166). Die Kombination *B* + *IDENTITY* hat ebenfalls einen großen Anteil in den *Top1000*, insbesondere bei Verzicht auf Dämpfung in beiden Verzweigungsmodi. Diese Beobachtung lässt sich ebenfalls auf die Kombination des Ausbreitungsmodus *B* mit dem Aktivierungsmodus *LOG2SIG* übertragen. Der Ausbreitungsmodus *FWL-RR* tritt seltener auf und ist ausschließlich in Kombination mit dem Aktivierungsmodus *LOG2* vertreten.

Für die acht untersuchten Anfragen lässt sich eindeutig die beste Konfiguration identifizieren. Soll die Analyse allerdings zur Auswahl einer geeigneten Konfiguration für ein IR-System

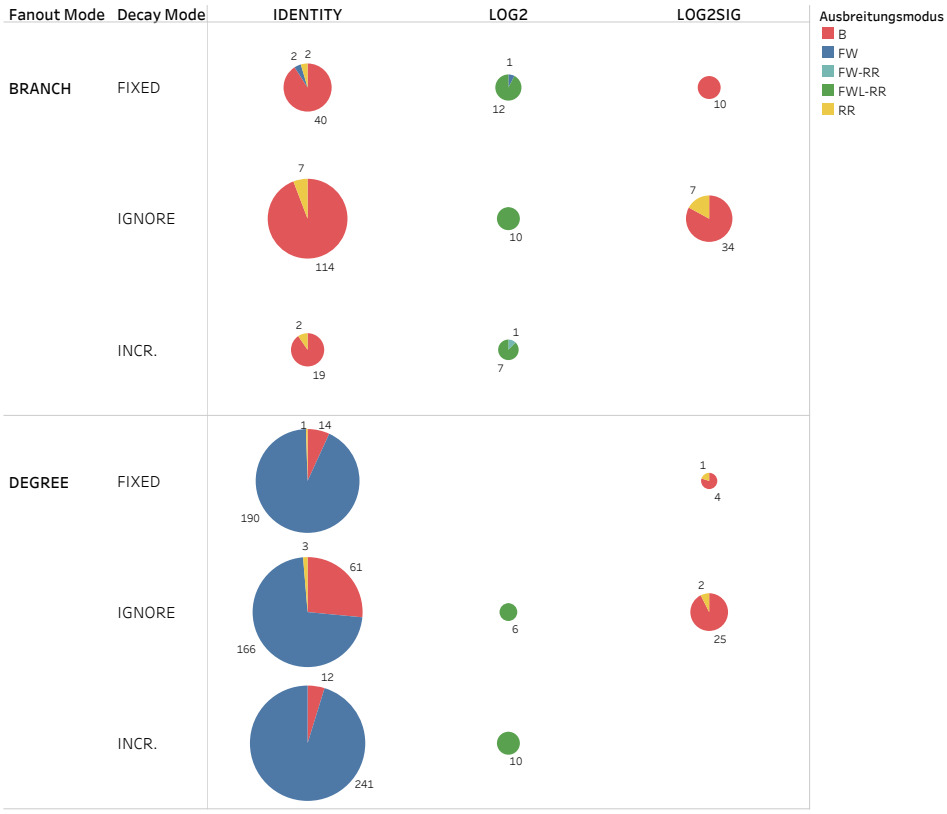


Abbildung 4.23: Anzahl der Kombinationen für die Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus in den *Top1000*

genutzt werden, welches auch für andere Anfragen gute Ergebnisse liefert, so dienen die Abbildungen 4.22 und 4.23 als Hilfsmittel zur Auswahl einer geeigneten Konfiguration. Ist die Konfiguration mit höchstem Fitnesswert beispielsweise ein Ausreißer, so ist sie für die allgemeine Verwendung für andere Anfragen vermutlich nicht geeignet. Daher berücksichtigen wir bei der Bewertung von Kombinationen neben der hohen Fitness und einer passenden Netzvariante auch die Stabilität der Kombination.

Zwischenfazit: Kombinationen von Konfigurationsparametern zeigen eine exklusive Eignung für spezifische Netzvarianten. Die stark reduzierte Anzahl geeigneter Kombinationen ermöglicht eine spezifische Auswahl geeigneter Konfigurationsparameter.

Pulsanzahl: In Abbildung 4.24 ist die Anzahl und Verteilung der *Top1000* Konfigurationen bezüglich ihrer Pulsanzahl dargestellt. Es konnten keine Konfigurationen mit einer Pulsanzahl ≥ 30 in den *Top1000* gefunden werden. Die beste Qualität erreichen Konfigurationen mit einer Pulsanzahl zwischen 22 und 27 und mit der Pulsanzahl 14. Der Bereich zwischen 13 und 27 sticht zudem durch die große Anzahl an Konfigurationen hervor. In Abbildung 4.22 ist durch die Größe der Kreise bereits eine Zuordnung der Pulsanzahl zu den Konfigurationen bezüglich der verschiedenen Konfigurationsparameter ersichtlich. Hier ist deutlich erkennbar, dass insbesondere die Konfigurationen mit *LOG2SIG* beziehungsweise die Kombinationen *FW* + *IDENTITY* + *DEGREE* mit steigender und konstanter Dämpfung hohe Pulszahlen besitzen. Die Kombinationen *B* + *IDENTITY* und *RR* + *IDENTITY* besitzen zumeist

mittlere Pulsanzahlen, die dargestellten Konfigurationen mit *LOG2* hingegen lediglich kleine Pulsanzahlen.

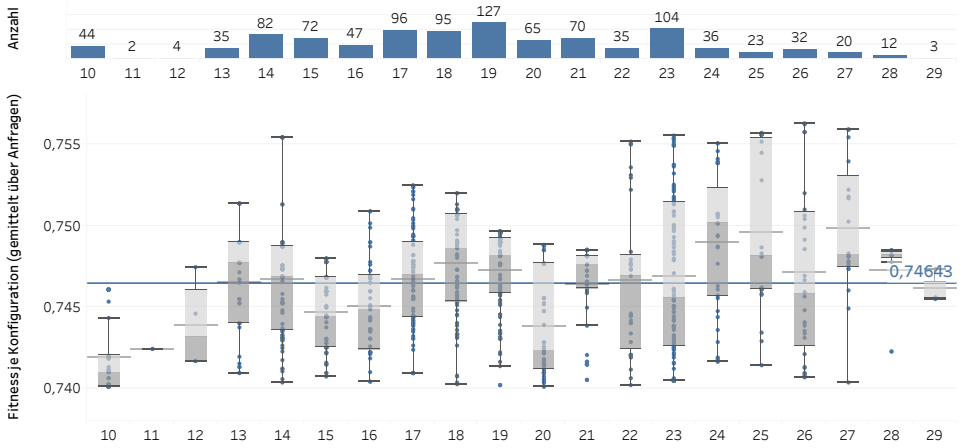


Abbildung 4.24: Anzahl und Verteilung der *Top1000* Fitnesswerte auf die Pulsanzahl

Kantengewichtung: Für die Kantengewichtung konnte in den Optimierungsläufen keine Konfiguration gefunden werden, die eine über die Anfragen gemittelte Fitness $\geq 0{,}7401$ im Bereich der *Top1000* erreicht. Daraus schließen wir, dass die Verwendung der festgelegten Kantengewichtung in anfrageübergreifenden Konfigurationen keinen positiven Effekt auf die Qualität der Ergebnisse des IR-Systems hat. Das Potenzial, das die Kantengewichtung dennoch besitzt, kann in zukünftigen Arbeiten durch weitere Experimente identifiziert werden. Dazu können evolutionäre Algorithmen eingesetzt werden, die beispielsweise die Gewichtung der verschiedenen Kantentypen optimieren.

Zwischenfazit: Die Verwendung der festgelegten Kantengewichtung hat in anfrageübergreifenden Konfigurationen keinen positiven Effekt auf die Qualität der Ergebnisse.

4.4.2.3 Auswirkungen der Anfragen und Anfragetypen

In Abbildung 4.25 wird die Verteilung der Fitnesswerte für jede Anfrage, aufgeteilt nach den verschiedenen Konfigurationsparametern, dargestellt.

Auch hier sind lediglich die *Top1000* Konfigurationen im Diagramm vertreten, allerdings sind alle Durchläufe der acht Anfragen einzeln und nicht gemittelt dargestellt. Daher weicht der Fitnessbereich von den gemittelten Fitnesswerten ab. Durch die farbliche Markierung der untersuchten Anfragen wird deutlich, dass die verschiedenen Anfragen einen unterschiedlichen Beitrag zu den gemittelten Fitnesswerten leisten. So sind die Fitnesswerte für die Anfrage Q85 und Q54 für alle Konfigurationskombinationen deutlich niedriger als für die Anfragen Q23 und Q35, die die besten Fitnesswerte erreichen. Das unterstützt die Vermutung aus Abschnitt 4.3.2.3, dass die Anfragen Einfluss auf die Qualität der Ergebnisse haben. Das Optimierungsziel *OZ1* besteht darin, bestmögliche Konfigurationen für die Gesamtheit der untersuchten Anfragen zu identifizieren. Durch die Mittelwertbildung über die Ergebnisse der einzelnen Anfragen können jene Anfragen, für die es weniger sehr gute Konfigurationen gibt, in den Optimierungsläufen benachteiligt werden. Wir vermuten, dass wir bei Optimierung jeder einzelnen Anfrage bessere Fitnesswerte erreichen können. Im folgenden Abschnitt untersuchen

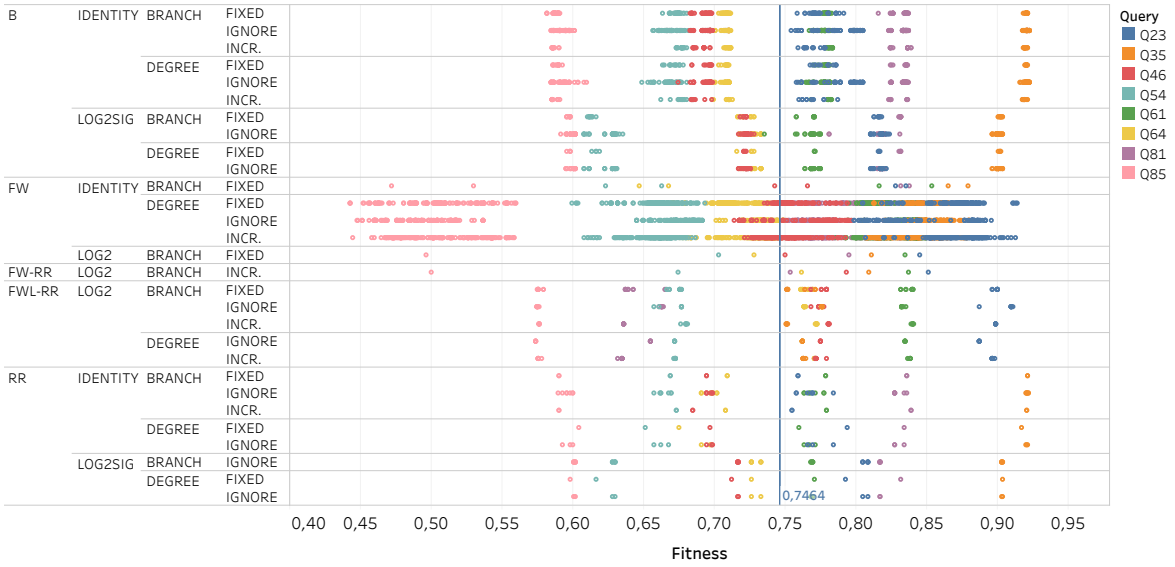


Abbildung 4.25: Verteilung der *Top1000* Fitnesswerte je Anfrage auf Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus

wir daher mit dem *OZ2* zusätzliche Optimierungsläufe, die wir für jede einzelne Anfrage durchführen und deren Ergebnisse wir anschließend analysieren.

4.4.2.4 Fazit

In der vorgestellten Analyse von anfrageübergreifenden Konfigurationen wurde für jede Netzvariante ein Optimierungslauf mit evolutionären Algorithmen durchgeführt. Das IR-System berechnete für jede der 47.500 untersuchten Konfigurationen die Fitnesswerte für jede der acht Anfragen und optimierte den über die Anfragen gemittelten Fitnesswert.

Wir konnten zeigen, dass die semantischen Anreicherungen und die Reduktionen, die wir in den Abschnitten 3.2.3 und 3.2.4 vorstellten, essenziell zur Qualität der Ergebnisse beitragen. In den *Top1000* Konfigurationen waren keine Netze beteiligt, die auf semantische Anreicherung oder Reduktion verzichten.

Außerdem konnten wir in der Analyse zeigen, dass der Einsatz der Ausbreitungsmodi, die wir in Abschnitt 3.3.4 als zusätzliches Werkzeug zur Steuerung der Aktivierungsausbreitung einführten, sinnvoll ist. Unter den *Top1000* der anfrageübergreifenden Konfigurationen dominierte der Ausbreitungsmodus *FW*, der eine Vorwärtsbewegung der Aktivierung im Netz präferiert, vor dem Ausbreitungsmodus *B*, der für die ungebremste Flutung des Netzes steht.

Des Weiteren ermöglichte die Analyse Aussagen darüber, welche der übrigen Parameter in den *Top1000* Konfigurationen im allgemeinen dominieren. Beispielsweise zeigte der Aktivierungsmodus *IDENTITY* meist bessere Ergebnisse als die Aktivierungsmodi mit logarithmischer oder sigmoider Abschwächung, der Verzweigungsmodus *DEGREE* zeigte eine bessere Qualität als der Verzweigungsmodus *BRANCH*. Die Analysen zeigten zudem, dass die Verwendung der vorher festgelegten Kantengewichtung bei der anfrageübergreifenden Optimierung keinen positiven Effekt auf die Qualität der Ergebnisse hat. Wir vermuten, dass eine eigene Optimierung der Kantengewichte bezüglich der verschiedenen Kantentypen, eine Verbesserung der Qualität bewirken kann.

Die Analyse zeigt auch deutlich, dass die Parameter nicht einzeln, sondern in ihrer Kombinati-

on betrachtet werden sollten. Bei Betrachtung der *Top1000* Konfigurationen reduzierten sich die 216 Kombinationen, die sich allein aus den Konfigurationsparametern Ausbreitungsmodi, Aktivierungsmodi, Dämpfungsmodi und Verzweigungsmodi ergeben, auf nur noch 29 Kombinationen. Da zudem die Kombinationen von Konfigurationsparametern eine exklusive Eignung für spezifische Netzvarianten zeigen und umgekehrt, konnten wir den Parameterraum weiter einschränken. Insbesondere die Abbildung 4.22 (Seite 151) visualisiert gute Kombinationen der Konfigurationsparameter und die zugehörige Netzvariante. Zusammen mit der Abbildung 4.23 (Seite 152), die eine zusätzliche Betrachtung der Stabilität der Kombinationen ermöglicht, dient sie uns als Guideline für die Auswahl geeigneter Konfigurationen für das IR-System auf Basis der untersuchten Anfragen.

4.4.3 Auswertung der anfragespezifischen Optimierung

In diesem Abschnitt stellen wir die Ergebnisse der Optimierungsläufe vor, deren Ziel die Optimierung der Fitness für jede Anfrage und jede Netzvariante ist.

4.4.3.1 Allgemeine Auswertung der Optimierungsläufe

Die besten anfragespezifischen Konfigurationen werden in den Tabellen 4.15 und 4.16 aufgelistet. Diese Konfigurationen und die anfrageübergreifende beste Konfiguration aus Tabelle 4.14 (Seite 147) werden wir in Abschnitt 4.4.4 verwenden, um die Relevanz und Nützlichkeit der Ergebnisse des IR-Systems zu bewerten und die Evaluationsfrage EQ4 zu beantworten.

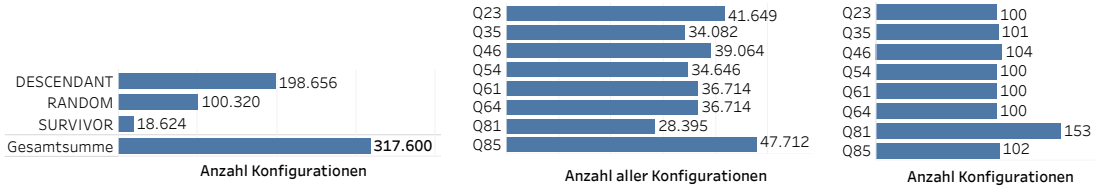
Optimierte Konfiguration	$\mathbf{K}_{Q23}$	$\mathbf{K}_{Q35}$	$\mathbf{K}_{Q46}$	$\mathbf{K}_{Q54}$
Anfrage	Q23	Q35	Q46	Q54
Ausbreitungsmodus	*FWL*	*FW*	*B*	*FWL*
Aktivierungsmodus	*IDENTITY*	*LOG10*	*IDENTITY*	*LOG10SIG*
Dämpfungsmodus	*FIXED*	*IGNORE*	*INCREASING*	*INCREASING*
Verzweigungsmodus	*DEGREE*	*BRANCH*	*DEGREE*	*BRANCH*
Kantengewichtung	*false*	*false*	*true*	*false*
Dämpfungsfaktor d	0,9480	1	0,8317	0,8104
Aktivierungsschwellwert τ	0,0627	0,0	0,1340	0,0
Pulsanzahl	18	18	10	46
Netzvariante	*N6-1*	*N2-3*	*N6-2*	*N4-1*
Fitness	0,9262	0,9319	0,8477	0,7367
EP	0,8944	0,9230	0,7444	0,6340
$NDCG_{Min}$	0,9579	0,9409	0,9510	0,8393

Tabelle 4.15: Parametrisierung der besten anfragespezifischen Konfigurationen für Q23-Q54

Optimierte Konfiguration	$\mathbf{K}_{Q61}$	$\mathbf{K}_{Q64}$	$\mathbf{K}_{Q81}$	$\mathbf{K}_{Q85}$
Anfrage	Q61	Q64	Q81	Q85
Ausbreitungsmodus	*FW-RR*	*FWL-RR*	*FW-RR*	*FWL*
Aktivierungsmodus	*IDENTITY*	*IDENTITY*	*SIG*	*LOG2*
Dämpfungsmodus	*INCREASING*	*IGNORE*	*FIXED*	*INCREASING*
Verzweigungsmodus	*BRANCH*	*BRANCH*	*BRANCH*	*BRANCH*
Kantengewichtung	*false*	*false*	*true*	*false*
Dämpfungsfaktor d	0,9451	1	0,9837	0,8034
Aktivierungsschwellwert τ	0,1256	0,0371	0,0062	0,0024
Pulsanzahl	32	14	10	15
Netzvariante	*N4-2*	*N6-2*	*N4-3*	*N2-3*
Fitness	0,9084	0,8111	0,9453	0,6446
EP	0,8840	0,6867	0,9139	0,3732
$NDCG_{Min}$	0,9328	0,9355	0,9768	0,9159

Tabelle 4.16: Parametrisierung der besten anfragespezifischen Konfigurationen für Q61-Q85

Abbildung 4.26 gibt einen Überblick über den Umfang dieser Optimierungsläufe.[50] Das



(a) (b) (c)

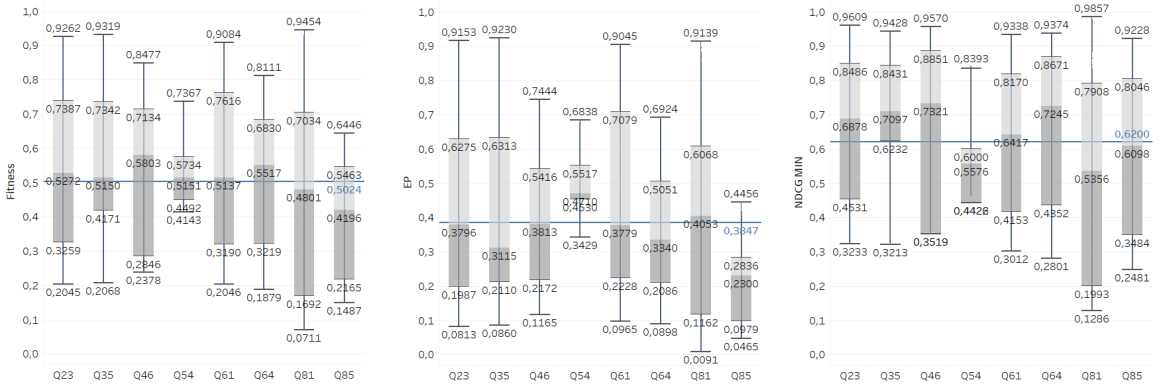
Abbildung 4.26: Anzahl der Konfigurationen a) aufgeteilt nach Evolutionstypen, b) aufgeteilt nach Anfragen und c) *Top100* aufgeteilt nach Anfragen

IR-System berechnete für insgesamt 317.600 Konfigurationen Fitnesswerte (Abbildung 4.26a). Wie bereits in den Durchläufen für *OZ1*, werden auch hier die Knoten des Evolutionstyps *SURVIVOR* (18.264 Konfigurationen) aufgrund ihrer mehrfachen Existenz in der folgenden Analyse zur Qualität der Ergebnisse ausgeschlossen. Für die unterschiedlichen Anfragen untersuchten wir jeweils zwischen 28.395 und 47.712 Konfigurationen (Abbildung 4.26b), von denen wir jeweils die Konfigurationen mit den einhundert besten Fitnesswerten, die *Top100*, für die folgenden Analyse betrachten. In Abbildung 4.26c ist erkennbar, dass die Menge der *Top100* teilweise mehr als einhundert Konfigurationen enthält. Das ist möglich, da mehrere Konfigurationen den gleichen Fitnesswert wie das hundertste Element besitzen.[51]

Abbildung 4.27 zeigt die Verteilung der berechneten Werte für Fitness, EP und NDCG_{Min}. Wir konnten für alle Anfragen Konfigurationen mit besseren Fitnesswerten finden als in den zufälligen (Abbildung 4.16, Seite 143) und den anfrageübergreifenden (Abbildung 4.25, Seite 154) Durchläufen. Die maximalen Fitnesswerte je Anfrage unterscheiden sich stark. Insbesondere die Anfrage Q85 kann lediglich den Fitnesswert 0,6446 erreichen. Diese Unterschiede führen wir nach der Vielzahl an Durchläufen nicht auf eine ungünstige Wahl der Konfigurationsparameter oder Netzvariante zurück. Während wir sehr hohe Werte für die im Fitnesswert enthaltene Nützlichkeit NDCG_{Min} erreichen, ist die Genauigkeit EP für jede Anfrage geringer. Das bedeutet zwar, dass das IR-System nicht alle Elemente, die von den Ratern als relevant eingestuft wurden, weit oben in der Ergebnisliste ansiedelt. Der dennoch hohe Nützlichkeitswert ist allerdings ein Zeichen dafür, dass die Elemente mit hohem Relevanzgrad auch in der Ergebnisliste des IR-Systems eine hohe Position erhalten. Das Ergebnis unterstützt unsere Vermutung, dass die Anfragen Einfluss auf die Qualität der Ergebnisse haben. Dies gilt insbesondere für die Berechnung der EP-Werte. Ein Grund hierfür kann die unterschiedliche Bewertung durch die Rater sein. In Q85 wurden beispielsweise von den Ratern R1 und R3 wenige (9 und 11) Elemente als relevant markiert, diese dafür aber meist mit höchstem Relevanzgrad (siehe Abbildung 4.5, Seite 128). Rater R4 hat insgesamt 47 Elemente markiert, davon ebenfalls wenige sehr relevante Elemente, aber zusätzlich sehr viele Elemente mit dem geringen Relevanzgrad 1 markiert. Bei der Berechnung von EP werden

[50] Die maximalen Fitnesswerte steigen im Verlauf der Generationen der durchgeführten Optimierungsläufe je Netzvariante und Anfrage. Auf eine Darstellung der Entwicklung, analog zu Abbildung 4.17 auf Seite 146, verzichten wir an dieser Stelle.

[51] Eine mögliche Erklärung für die hohe Anzahl ranggleicher Konfigurationen für die Anfrage Q81 ist eine längere Suche im Parameterraum, die weder eine Verbesserung noch eine Verschlechterung bewirkt hat. Das kann bei schlechten Konfigurationen beispielsweise dann passieren, wenn sich Konfigurationsparameter gegenseitig stark diskriminieren, so dass die Aktivierungsausbreitung bereits nach wenigen Pulsen abbricht. Mutationen einzelner Parameter zeigen dann oftmals keine Wirkung. Bei guten Konfigurationen, wie im Fall der Q81 kann dieser Effekt auftreten, da die sehr guten Konfigurationen über mehrere Generationen hinweg wiederholt selektiert und mutiert werden.



(a) Fitness (b) EP (c) $NDCG_{Min}$

Abbildung 4.27: Verteilung der berechneten Werte a) Fitness, b) EP und c) $NDCG_{Min}$ für alle Durchläufe aufgeteilt nach Anfragen

alle relevanten Elemente unabhängig des Relevanzgrades gleichbehandelt. In die Bewertung des $NDCG_{Min}$-Wertes fließen die Relevanzgrade ein. Eine Diskrepanz wie in Abbildung 4.27 ist möglich, wenn beispielsweise die wichtigsten Elemente zwar als solche identifiziert werden, einige Elemente mit geringer Relevanz aber nicht im obersten Bereich der Ergebnisliste erscheinen. Diese Beobachtung unterstützt uns in der Kombination der beiden Bewertungsmaße für Genauigkeit und Nützlichkeit, da wir beide Eigenschaften als wichtig für die Güte der Ergebnisse erachten.

Zwischenfazit: Die anfragespezifischen Konfigurationen erreichen eine bessere Qualität als die anfrageübergreifenden und die zufällig generierten Konfigurationen.

4.4.3.2 Darstellung der Ergebnisse

Auf den folgenden Seiten sind die Ergebnisse der anfragespezifischen Optimierungsläufe in verschiedenen Diagrammen zusammengefasst, die wir anschließend zur Analyse der Auswirkungen von Netzvarianten und Konfigurationsparametern heranziehen. Wir stellen zunächst diese Diagramme vor. Sie beleuchten unterschiedliche Aspekte beziehungsweise aggregieren die Daten teilweise unterschiedlich, um die komplexen Zusammenhänge zu erfassen.

In den Abbildungen 4.28a–4.28h ist die Verteilung der *Top100* Fitnesswerte auf die Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus dargestellt. Die Netzvarianten sind farblich markiert und die Pulsanzahl wird durch die Größe der Kreise symbolisiert. Konfigurationen mit Kantengewichtung sind zusätzlich als Stern markiert. Diese Abbildungen zeigen für jede Anfrage, wie sich ihre *Top100* Konfigurationen zusammensetzen, für welche Netzvariante sie gefunden wurden und welche Fitness sie erreichen. In der Abbildung sind die Aktivierungsschwellwerte und Dämpfungsfaktoren nicht explizit dargestellt. Daher können beispielsweise zwei Konfigurationen, die lediglich in den beiden Parametern variieren, in der Abbildung nicht unterschieden werden. Sie liegen entsprechend übereinander.

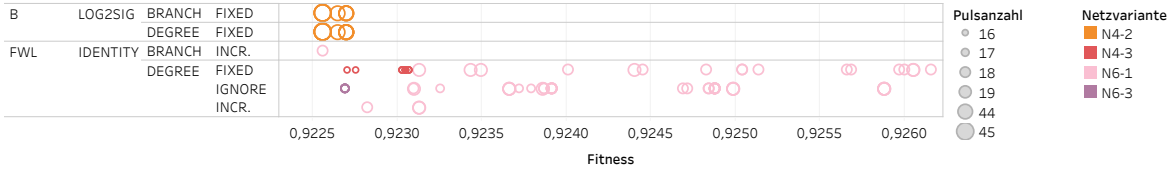
In Abbildung 4.29 stellen wir die anzahlmäßige Verteilung der Kombinationen für die Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus in den *Top100* je Netzvariante dar. Hier aggregieren wir über die Anfragen. Da die betrachteten Konfigurationen der Menge der *Top100* entstammen und dadurch bereits

eine hohe Qualität aufweisen, bleiben die Fitnesswerte hier unberücksichtigt. Das Balkendiagramm erlaubt eine Aussage dazu, welche Kombinationen der Konfigurationsparameter für die einzelnen Netzvarianten besonders häufig vorkommen und sich daher besonders eignen.

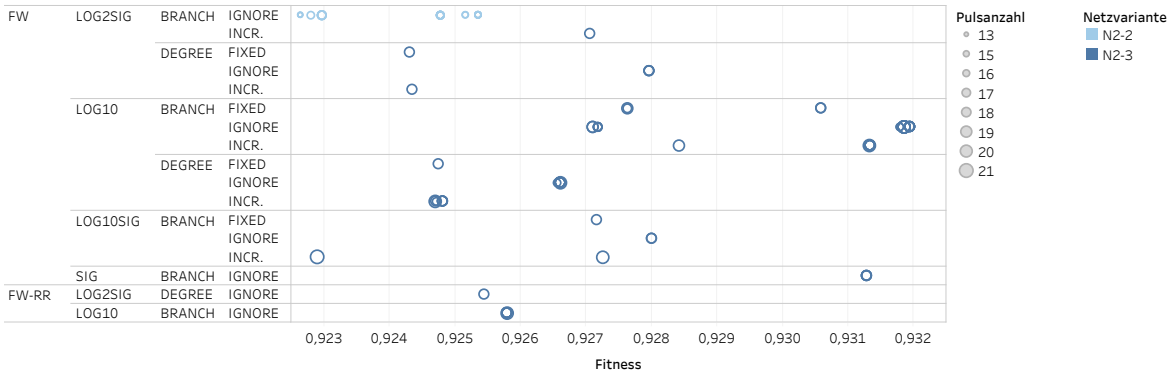
Die Abbildung 4.30 zeigt die Anzahl der Kombinationen für die Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus in den *Top100* für jede Anfrage unabhängig von den Netzvarianten. Diese Abbildung gibt Aufschluss darüber, welche Kombinationen für jede Anfrage besonders oft gut harmonieren. Die Darstellung als Kreisdiagramme mit skalierter Größe verdeutlicht visuell die Präferenz spezieller Kombinationen.

In Abbildung 4.31 werden die Werte aus Abbildung 4.30 über die Anfragen aggregiert, so dass die Anzahl der Kombinationen für die Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus in den *Top100* und der spaltenweisen Gesamtsumme dargestellt sind. Sie zeigt ein Gesamtbild, wie gut verschiedene Konfigurationsparameter der untersuchten Konfigurationen miteinander harmonieren, unabhängig von der Art der Anfrage.

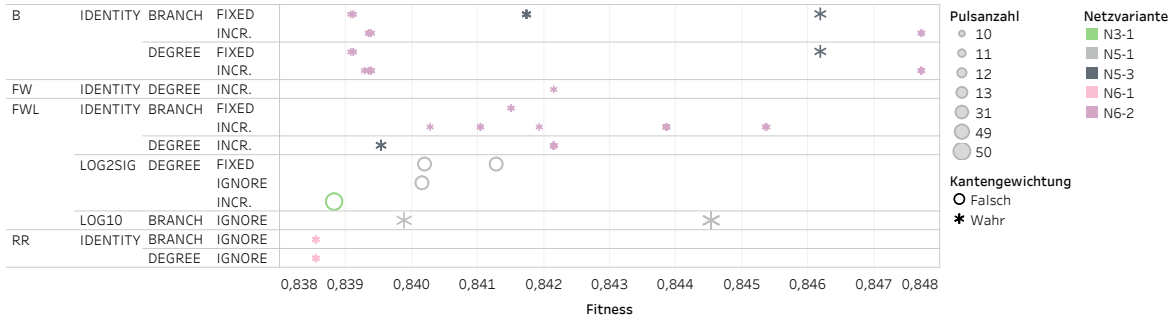
Diese Abbildungen bilden die Grundlage der folgenden Auswertungen, die im Detail durch kleinere Diagramme angereichert werden. Im Anschluss an die Abbildung der Diagramme widmen wir uns den Auswertungen der Ergebnisse der anfragespezifischen Optimierung.



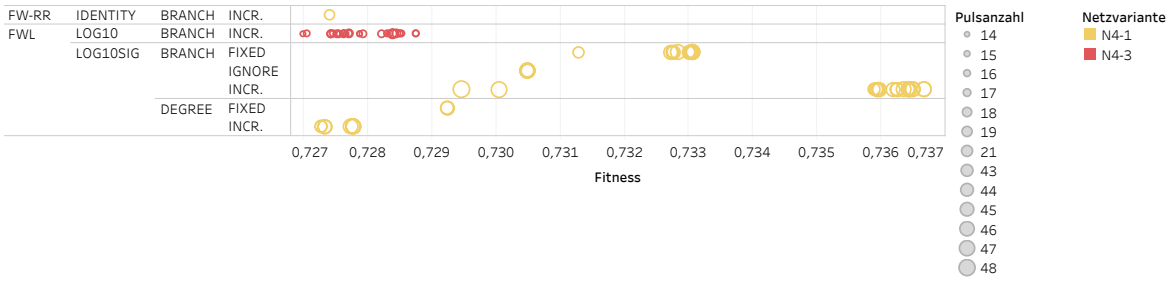
(a) Q23



(b) Q35

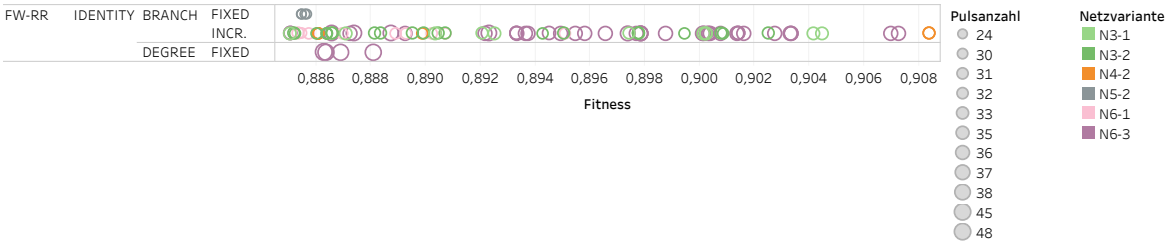


(c) Q46

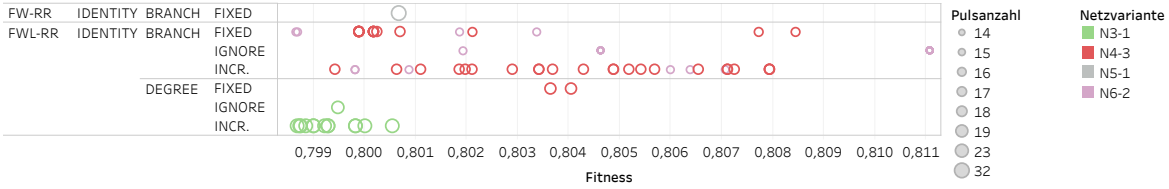


(d) Q54

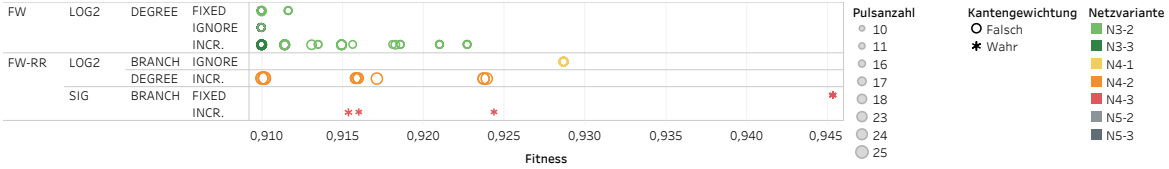
Abbildung 4.28: Verteilung der *Top100* Fitnesswerte auf die Netzvarianten und die Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus, Kantengewichtung und Pulsanzahl (1/2)



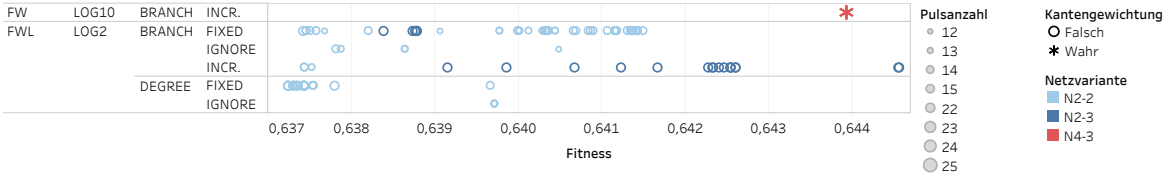
(e) Q61



(f) Q64



(g) Q81



(h) Q85

Abbildung 4.28: Verteilung der *Top100* Fitnesswerte auf die Netzvarianten und die Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus, Kantengewichtung und Pulsanzahl (2/2)

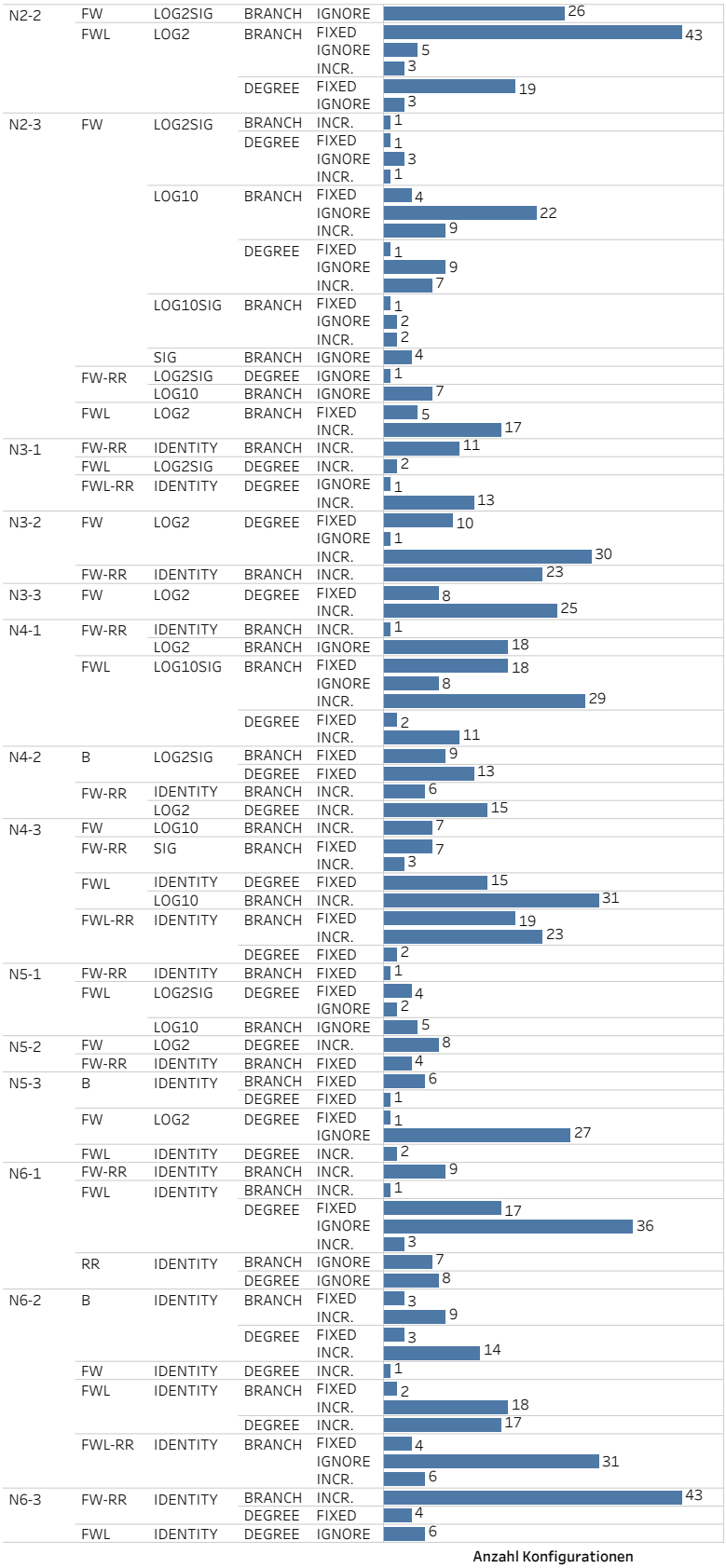
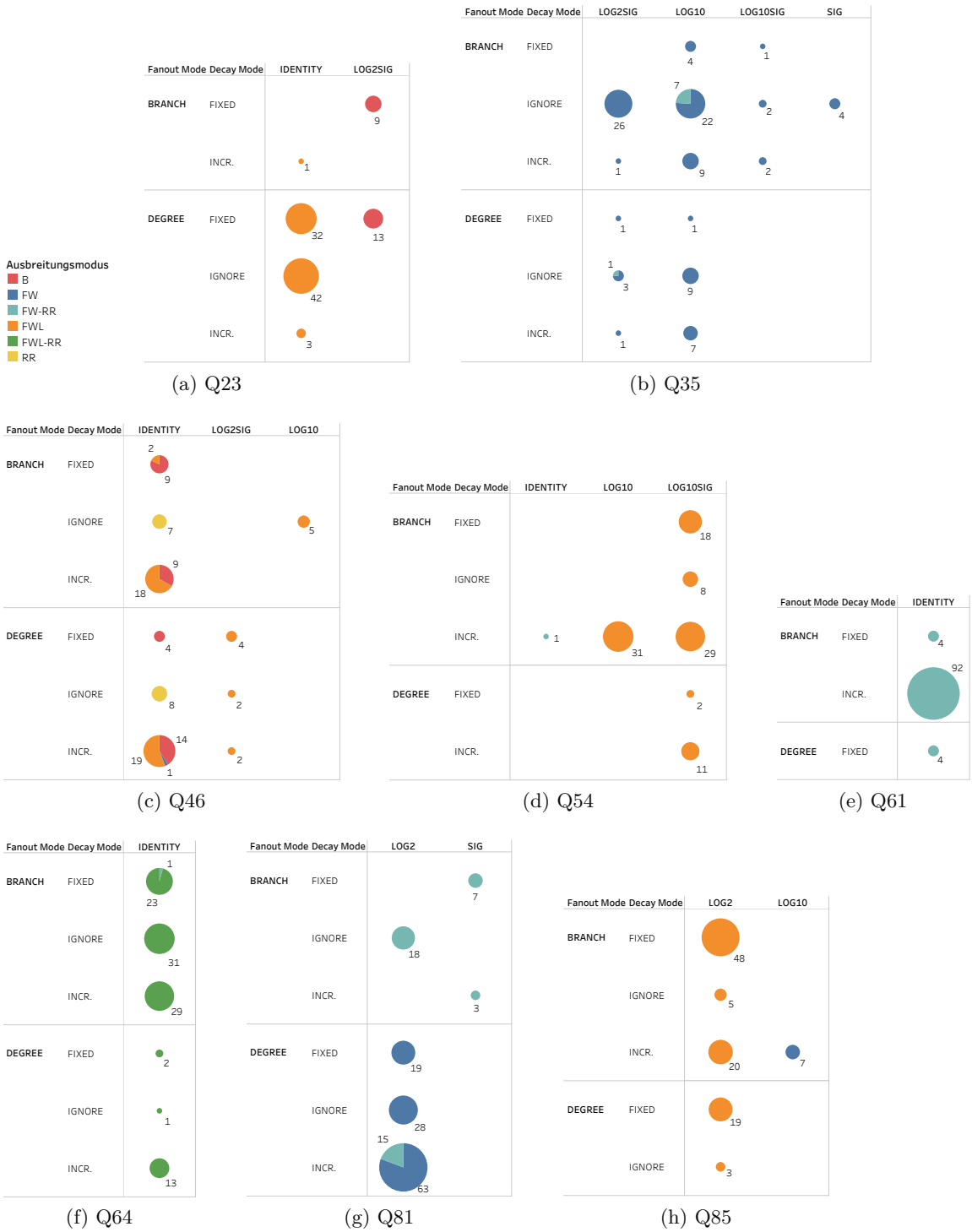


Abbildung 4.29: Anzahl der Kombinationen für die Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus in den *Top100* je Netzvariante



(a) Q23 (b) Q35

(c) Q46 (d) Q54 (e) Q61

(f) Q64 (g) Q81 (h) Q85

Abbildung 4.30: Anzahl der Kombinationen für die Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus in den *Top100* je Anfrage

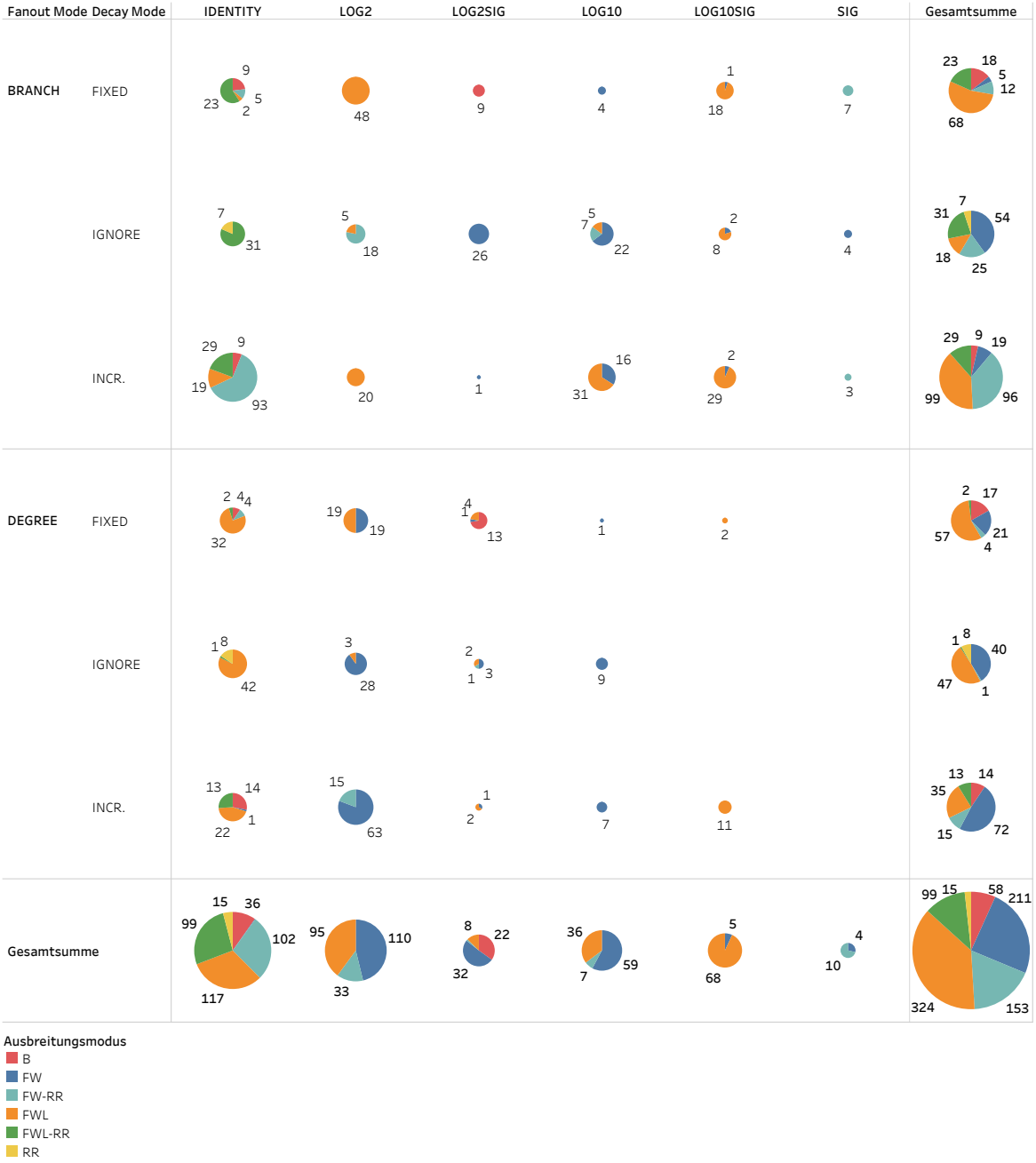


Abbildung 4.31: Anzahl der Kombinationen für die Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus in den *Top100*

4.4.3.3 Einfluss der Netzvarianten

Wie bereits bei der anfrageübergreifenden Optimierung, erreichen auch bei der anfragespezifischen Optimierung nicht alle Netzvarianten Fitnesswerte in den *Top100* Konfigurationen. Die genaue Verteilung aller anfragespezifischen *Top100* Konfigurationen auf die Netzvarianten ist in Abbildung 4.32 dargestellt. Alle Netze ohne semantische Anreicherung (*N1-1*, *N1-2*

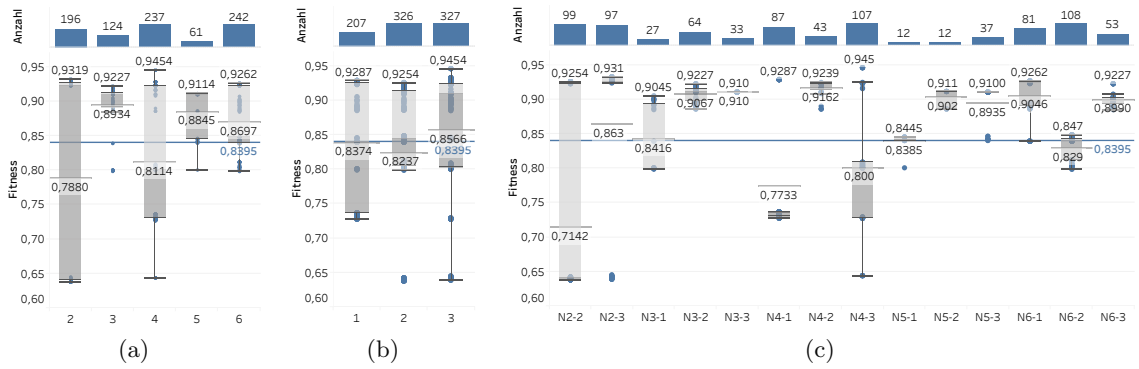


Abbildung 4.32: Anzahl und Verteilung der *Top100* Fitnesswerte aller Anfragen auf a) die Anreicherungsstufen der Netzvarianten, b) die Reduktionsstufen der Netzvarianten und c) die Netzvarianten

und *N1-3*) sind nicht in den *Top100* Konfigurationen vertreten. Zudem ist die unreduzierte Netzvariante mit Anreicherung durch Synonyme *N2-1* nicht in der Auswahl der Besten enthalten. Das unterstützt die Erkenntnisse aus der anfrageübergreifenden Optimierung mit *OZ1*, in denen Netze ohne semantische Anreicherung oder Reduktion nicht in den untersuchten *Top1000* Konfigurationen auftraten. Im Gegensatz zu den Ergebnissen für *OZ1*, sind bei der anfragespezifischen Optimierung mit *OZ2* jedoch einige Netzvarianten ohne Reduktion (Stufe 1) in den *Top100* Konfigurationen vertreten. Das ist ein Zeichen dafür, dass die Bedeutung der semantischen Anreicherung die Bedeutung der Reduktion der Netze überragt. Die Abbildung 4.32b zeigt dennoch, dass die Reduktion sinnvoll ist, da die Reduktionsstufen 2 und 3 häufiger als die Reduktionsstufe 1 sehr gute Ergebnisse erreichen. Bei den Anreicherungsstufen überragen anzahlmäßig die Verwendung erweiterter Keywords ohne und mit Synonymen (Stufen 4 und 6), gefolgt von der Anreicherung durch Synonyme (Stufe 2) und einfachen Keywords.

Zwischenfazit: Semantische Anreicherung hat einen größeren Einfluss auf die Qualität der Ergebnisse als die Reduktion der Netze.

4.4.3.4 Einfluss der Konfigurationsparameter

Die Abbildung 4.33 zeigt, wie sich die *Top100* Konfigurationen aller Anfragen anzahlmäßig auf verschiedene Konfigurationsparameter verteilen. Im Vergleich zu den Ergebnissen der anfrageübergreifenden Optimierung für *OZ1* (Abbildung 4.21, Seite 150) sind hier einige Unterschiede erkennbar. Bei der anfragespezifischen Optimierung bilden die Konfigurationen mit dem Ausbreitungsmodus *FWL* den größten Anteil, während für *OZ1* keine Top-Konfiguration für diesen Modus gefunden werden konnte. Die anfragespezifische Optimierung zeigt zudem eine starke Tendenz für die Ausbreitungsmodi *FW*, *FW-RR* und *FWL-RR*, welche *B* nach Anzahl überragen. Bei den Aktivierungsmodi hat *IDENTITY* erneut den häufigsten

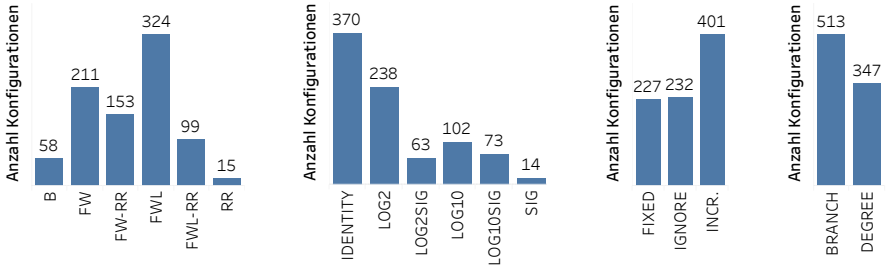


Abbildung 4.33: Anzahl der *Top100* Konfigurationen aller Anfragen aufgeteilt nach a) Ausbreitungsmodi, b) Aktivierungsmodi c) Dämpfungsmodi und d) Verzweigungsmodi

Anteil an den Top-Konfigurationen. Allerdings wird in Abbildung 4.33b deutlich, dass die Aktivierungsmodi mit Beteiligung der Logarithmus- und Sigmoidfunktionen eine wesentlich größere Rolle spielen als in den anfrageübergreifenden Top-Konfigurationen. Die steigende Dämpfung überragt auch hier die konstante Dämpfung. Im Gegenteil zu *OZ1* ist der Verzicht auf Dämpfung nicht mehr die stärkste Option, sondern ist in den *Top100* Konfigurationen wesentlich seltener vertreten als die steigende Dämpfung. Für die Verzweigungsmodi konnte die anfragespezifische Optimierung mehr Top-Konfigurationen mit *BRANCH* als *DEGREE* finden.

Zwischenfazit: Die Ausbreitungsmodi *FWL*, *FW*, *FW-RR* und *FWL-RR* besitzen die größten Anteile an den *Top100* Konfigurationen und überragen den Ausbreitungsmodus *B*, der für die ungehemmte Flutung des Netzes steht. Das zeigt, dass das Konzept der direkten Steuerung des Ausbreitungsverhaltens sinnvoll ist.

Kombination der Konfigurationsparameter: Die Frage, welche Netzvarianten für welche Anfragen und mit welchen Konfigurationen besonders gute Ergebnisse lieferten, beantwortet die Abbildung 4.28 (Seite 161). Wir können erkennen, dass sich die *Top100* Konfigurationen für jede Anfrage jeweils nur auf wenige Netzvarianten verteilen. Beispielsweise wurden alle *Top100* Konfigurationen der Anfrage Q35 (Abbildung 4.28b) mit den Netzen *N2-2* und *N2-3* gefunden, für die Anfrage Q54 (Abbildung 4.28d) beispielsweise mit den Netzen *N4-1* und *N4-3*. Mit sieben Netzvarianten besitzt Q81 (Abbildung 4.28g) die größte Bandbreite. Der Ausschluss ungeeigneter Netzvarianten je Anfrage verringert die Komplexität des großen Parameterraums und erleichtert dadurch die Entscheidung für geeignete Konfigurationen für ein IR-System.

Wir können außerdem erkennen, dass die Netzvarianten sich je Anfrage auf bestimmte Konfigurationsparameter konzentrieren. Beispielsweise sehen wir für die Anfrage Q23 in Abbildung 4.28a, dass *N4-2* für Konfigurationen mit *B* + *LOG2SIG* gute Fitnesswerte erreicht, *N6-1* hingegen durch Konfigurationen mit *FWL* + *IDENTITY*. Das weist darauf hin, dass die Netzvarianten auch unabhängig von den Anfragen mit bestimmten Kombinationen der Konfigurationsparameter besonders gut harmonieren. Die Abbildung 4.29 (Seite 162) aggregiert die Kombinationen der Konfigurationsparameter der *Top100* Konfigurationen aller Anfragen in einem Diagramm, so dass wir besser erkennen können, welche Kombinationen der Konfigurationsparameter gut mit welchen Netzen zusammenpassen. Hier ist deutlich

erkennbar, dass je Netzvariante jeweils nur wenige Kombinationen der möglichen 216 Kombinationen aus den sechs Ausbreitungsmodi, sechs Aktivierungsmodi, zwei Dämpfungsmodi und drei Verzweigungsmodi in den *Top100* vertreten sind. Die maximale Anzahl sehen wir mit 18 Kombinationen für Netz *N2-3*, die kleinste Anzahl mit zwei Kombinationen für *N3-3* und *N5-2*. Das ist ein deutliches Zeichen für eine starke Präferenz bestimmter Kombinationen. Dass die Präferenz der Kombination nicht für jedes Netz gleich ist, unterstreicht unsere Vermutung, dass auch Netzvarianten und Konfigurationen aufeinander abgestimmt werden sollten. Wir erkennen beispielsweise, dass die Netze der Anreicherungsstufe 6 (*N6-1*–*N6-3*) ausschließlich Konfigurationen mit Aktivierungsmodus *IDENTITY* in den *Top100* besitzen, während die Netze der Anreicherungsstufe 2 (*N2-2* und *N2-3*) keine Konfiguration mit *IDENTITY*, sondern ausschließlich die verbleibenden Aktivierungsmodi mit Beteiligung der Logarithmus- und Sigmoidfunktionen enthalten.

Die Abbildung 4.30 (Seite 163) aggregiert die Daten der Abbildung 4.28 (Seite 161) derart, dass wir einen Eindruck über die Stabilität der Kombinationen von Konfigurationsparametern erhalten. Die Stabilität einer Kombination aus Konfigurationsparameter zeigt sich wieder durch die möglichst hohe Anzahl an Konfigurationen, die in den *Top100* enthalten sind. Sehr stabil ist demnach beispielsweise in Abbildung 4.28f für Q64 die Kombination *FWL-RR* + *IDENTITY* + *BRANCH* + *IGNORE*, für die drei Kreise dargestellt sind, zu denen auch die beste Konfiguration für diese Anfrage gehört. In Abbildung 4.30f sehen wir, dass diese Kombination anzahlmäßig (31 Konfigurationen) die stärkste Gruppe für Q64 bildet, was wir als Zeichen für Stabilität der Kombination werten. Die Kombination *FW-RR* + *SIG* + *BRANCH* + *FIXED*, die für Q81 die beste Konfiguration erreicht (Abbildung 4.28g), ist ein Beispiel für eine nicht sehr stabile Kombination, da sie lediglich in sieben Konfigurationen (Abbildung 4.30g) Top-Fitnesswerte erreichen konnte. Wenn auch mit geringfügig kleineren Fitnesswerten erreicht die Kombination *FW* + *LOG2* + *DEGREE* + *INCREASING* mit 63 Konfigurationen wesentlich stabilere gute Werte. Die Kombinationen der Konfigurationsparameter sind meist nicht sehr breit gefächert, sondern stattdessen sehr spezifisch für eine Anfrage. Diese starke Reduktion der möglichen Kombinationen auf eine kleine Auswahl je Anfrage hat den Vorteil, dass wir bei der Auswahl passender Konfigurationen für ein IR-System bereits einen reduzierten Parameterraum vorfinden. Es unterstützt zudem die wichtige Rolle, die die Wahl geeigneter Konfigurationen spielt.

Zwischenfazit: Die Eignung der Netzvarianten und der unterschiedlichen Kombinationen der Konfigurationsparameter variiert für die unterschiedlichen Anfragen. Da die Anzahl geeigneter Kombinationen im Top-Bereich stark reduziert ist, sollten Anfragen, Konfigurationsparameter und Netzvarianten aufeinander abgestimmt sein.

Kantengewichtung: Anders als bei der anfrageübergreifenden Optimierung gibt es in den *Top100* der anfragespezifischen Optimierung Konfigurationen mit der von uns festgelegten Kantengewichtung. Wie wir in Abbildung 4.28 (Seite 161) erkennen können, gilt das für die Anfragen Q46, Q81 und Q85, wobei Konfigurationen mit Kantengewichtung für Q46 und Q81 sogar die besten Konfigurationen stellen, für Q81 sogar die beste Konfiguration der Optimierungsläufe für OZ2. Die Kantengewichtung hat hier also eine größere Bedeutung als in den zufälligen oder anfrageübergreifenden Konfigurationen. In Abbildung 4.34 ist oben zusätzlich zur Verteilung die Anzahl der Konfigurationen dargestellt. Während in Q81 und Q85 nur ein kleiner Anteil (10 und 7 Konfigurationen) mit Kantengewichtung gute Ergebnisse erreicht, ist in Q46 mit 96 Konfigurationen ein sehr großer Anteil mit den gewichteten Kanten erfolgreich.

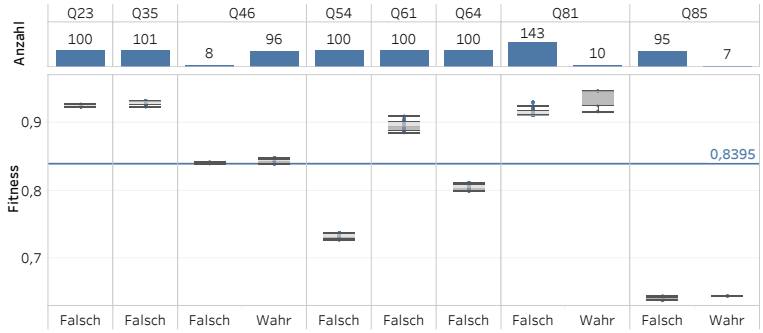


Abbildung 4.34: Anzahl und Verteilung der *Top100* Fitnesswerte auf die Anfragen und den Parameter Kantengewichtung

Zwischenfazit: Die Kantengewichtung stellt für drei Anfragen Konfigurationen unter den *Top100* bereit, darunter die beste Konfiguration aller Optimierungsläufe.

Pulsanzahl: In Abbildung 4.28 (Seite 161) wird die Anzahl der Pulse durch die Größe der Kreise und Sterne für die Konfigurationen dargestellt. In der jeweiligen Legende sind die dargestellten Pulsanzahlen aufgelistet. Beispielsweise sind in den *Top100* für Q23 nur Konfigurationen mit sechs verschiedenen Pulsanzahlen vertreten. Zum einen gibt es Pulsanzahlen zwischen 16 und 19, die sich ausschließlich auf die Konfigurationen mit *FWL + IDENTITY* und ausschließlich auf die Netze *N4-3*, *N6-1* und *N6-3* verteilen. Zum anderen gibt es die Pulsanzahlen 44 und 45, die sich ausschließlich auf die Konfigurationen mit *B + LOG2SIG* und ausschließlich auf das Netz *N4-2* verteilen. Für die anderen Anfragen zeichnet sich ein ähnliches Bild. Ein Zusammenhang zwischen Höhe der Pulsanzahl und erreichter Fitness ist nicht zu erkennen. Wir vermuten, dass die Pulsanzahl eine untergeordnete Rolle in der Parametrisierung spielt und auch spielen sollte.

4.4.3.5 Fazit

In der Analyse anfragespezifischer Konfigurationen führten wir je Netzvariante und Anfrage einen separaten Optimierungslauf mit evolutionären Algorithmen durch und untersuchten insgesamt 317.600 Konfigurationen. Mit der anfragespezifischen Optimierung konnten wir Konfigurationen finden, die höhere Fitnesswerte als die anfrageübergreifenden und zufällig generierten Konfigurationen erreichen. Die Analyse zeigte, dass sich die maximalen Fitnesswerte der einzelnen Anfragen trotz Optimierung stark unterscheiden. Ein Grund dafür können Unterschiede zwischen den zusammengefassten objektiven Relevanzen der Anfragen darstellen. Beispielsweise kann es schwieriger sein, hohe Fitnesswerte zu erreichen, wenn die Anzahl oder der Anteil objektiv relevanter Elemente nur sehr klein ist.

Die Analyse bestätigte und erweiterte Erkenntnisse, die bereits in eingeschobene Zwischenfazits und in die abschließende Zusammenfassung der anfrageübergreifenden Analyse in Abschnitt 4.4.2.4 flossen. Daher fassen wir sie an dieser Stelle kompakt durch die Beantwortung der folgenden Evaluationsfragen aus Tabelle 4.7 (Seite 129) zusammen.

***EQ3a**: Ist der Einsatz der semantischen Anreicherungen und der Reduktionen der Netze sinnvoll? Welchen Einfluss haben sie auf die Qualität der Ergebnisse?*

Die semantischen Anreicherungen und die Reduktionen, die wir in den Abschnitten 3.2.3 und 3.2.4 vorstellten, tragen essenziell zur Qualität der Ergebnisse bei. Daher ist ihr Ein-

satz sinnvoll. Sowohl in den anfrageübergreifenden als auch in den anfragespezifischen Top-Konfigurationen waren keine Netze beteiligt, die auf semantische Anreicherung verzichten. Anfragespezifische Top-Konfigurationen zeigten zwar auch gute Ergebnisse für einige unreduzierte Netze. Die Netze der Reduktionsstufen 2 und 3, die auf ihren semantischen Kern oder ihren erweiterten semantischen Kern reduziert sind, erreichten dennoch häufiger Top-Ergebnisse.

***EQ3b**: Ist der Einsatz der Ausbreitungsmodi sinnvoll? Welchen Einfluss haben sie auf die Qualität der Ergebnisse?*

Der Einsatz der Ausbreitungsmodi, die wir in Abschnitt 3.3.4 als zusätzliches Mittel zur Steuerung der Aktivierungsausbreitung einführten, ist sinnvoll. Wir konnten zeigen, dass die neu definierten Ausbreitungsmodi häufiger Top-Ergebnisse erreichen als der Ausbreitungsmodus *B*, der für die ungebremste Flutung des Netzes steht. Anfrageübergreifend dominiert *FW* deutlich. Anfragespezifisch überragen *FWL*, *FW*, *FW-RR* und *FWL-RR*.

***EQ3c**: Welchen Einfluss haben die Konfigurationsparameter und deren Kombinationen auf die Qualität der Ergebnisse?*

Unsere Analysen konnten zeigen, welche Konfigurationsparameter in den Top-Konfigurationen nach Anzahl und Fitness dominieren. Der Aktivierungsmodus *IDENTITY* überragt alle anderen Aktivierungsmodi. In anfragespezifischen Top-Konfigurationen erreichen allerdings Aktivierungsmodi mit Beteiligung der Logarithmus- und Sigmoidfunktionen wesentlich häufiger gute Ergebnisse als in den anfrageübergreifenden Top-Konfigurationen. Während der Verzweigungsmodus *DEGREE* anfrageübergreifend eine bessere Qualität zeigt als der Verzweigungsmodus *BRANCH*, dominiert *BRANCH* anfragespezifisch. Die Dämpfungsmodi für steigende Dämpfung *INCREASING* und konstante Dämpfung *FIXED* zeigen höhere Fitnesswerte als der Dämpfungsmodus für den Verzicht auf Dämpfung *IGNORE*.

Die Analysen zeigten zudem, dass beispielsweise nicht jede Kombinationen der Konfigurationsparameter Ausbreitungsmodus, Aktivierungsmodus, Dämpfungsmodus und Verzweigungsmodus sinnvoll ist. Stattdessen gibt es ein komplexes Zusammenspiel von Netzvarianten, Kombinationen der Konfigurationsparameter und Anfragen. Wir konnten zeigen, dass unter den Top-Konfigurationen nur noch eine stark reduzierte Anzahl der möglichen Kombinationen vertreten ist. In den anfrageübergreifenden Top-Konfigurationen sind nur noch 29 der 216 möglichen Kombinationen vertreten, in den anfragespezifischen Top-Konfigurationen nur noch maximal 17 für die einzelnen Anfragen und maximal 18 für die einzelnen Netzvarianten. Das bedeutet, eine Vielzahl an Kombinationen erreicht nicht die notwendige Güte. Außerdem zeigt das, dass eine gute Abstimmung der Parameter wichtig ist, um qualitativ gute Ergebnisse zu erzielen.

Einige Konfigurationsparameter spielen eine untergeordnete Rolle. Beispielsweise zeigte die Pulsanzahl keine Auffälligkeiten. Sie besitzt vermutlich wenig Optimierungspotenzial. Diese sollte allerdings plausibel sein und beispielsweise in einem angemessenem Verhältnis zum Durchmesser des Netzes stehen, um alle Bereiche des Netzes potenziell erreichen zu können.

Die Verwendung der festgelegten Kantengewichtung zeigte bei der anfrageübergreifenden Optimierung keinen positiven Effekt auf die Qualität der Ergebnisse. In den anfragespezifischen Top-Konfigurationen ist Kantengewichtung allerdings in drei der acht Anfragen vertreten und stellt sogar die beste Konfiguration aller Optimierungsläufe der Evaluation. Wir vermuten dennoch, dass eine eigene Optimierung der Kantengewichte bezüglich der verschiedenen Kantentypen eine Verbesserung der Qualität bewirken kann.

4.4.4 Relevanz und Nützlichkeit der Ergebnisse des IR-Systems

In diesem Abschnitt verwenden wir die in Abschnitt 4.4 optimierten Konfigurationen, um die Relevanz und Nützlichkeit der Ergebnisse des IR-Systems zur Unterstützung von G&Rs zu bewerten. Wir untersuchen zum einen die optimierte anfrageübergreifende Konfiguration K_Q aus Tabelle 4.14 (Seite 147) und zum anderen die acht optimierten anfragespezifischen Konfigurationen K_{Q23}-K_{Q85} aus den Tabellen 4.15 und 4.16. Dafür betrachten wir die Bewertungsmaße EP und $NDCG_{Min}$, die gleichermaßen in die Fitnesswerte flossen. Um die Ergebnisse trotz ihrer schwachen Ordnung vergleichbar zu machen, haben wir bei der Fitnessberechnung den $NDCG_{Min}$-Wert für die schlechtestmögliche Anordnung auf den Rängen berücksichtigt (siehe Abschnitt 2.1.3.3). Dieser Wert ist sehr diskriminierend. Für die abschließende Bewertung betrachten wir daher ebenfalls die $NDCG_{Max}$-Werte für die bestmögliche Anordnung auf den Rängen.

In Tabelle 4.17 listen wir die Werte für Fitness, EP, $NDCG_{Min}$ und $NDCG_{Max}$ auf, die sich bei Ausführung des IR-Systems mit der anfrageübergreifenden Konfiguration K_Q für die einzelnen Anfragen ergeben. Abbildung 4.35 stellt die entsprechenden EP-Recall-Graphen für jede einzelne Anfrage dar. Die schwarze Kurve entspricht den über die Anfragen gemittelten EP-Recall-Werten. Die vielen Datenpunkte wurden mittels 11-Punkt-Interpolation reduziert und die Kurve durch die Salton-Interpolation zusätzlich begradigt (Abschnitt 2.1.3.3).

Konfiguration	**Anfrage**	**Fitness**	**EP**	**$NDCG_{Min}$**	**$NDCG_{Max}$**
K_Q	Q23	0,8590	0,7931	0,9249	0,9348
K_Q	Q35	0,8164	0,8193	0,8134	0,8140
K_Q	Q46	0,7580	0,6059	0,9101	0,9101
K_Q	Q54	0,6701	0,6421	0,6980	0,7294
K_Q	Q61	0,8157	0,8506	0,7808	0,7808
K_Q	Q64	0,7422	0,5784	0,9060	0,9098
K_Q	Q81	0,8454	0,7297	0,9612	0,9612
K_Q	Q85	0,5434	0,2972	0,7896	0,7949
	∅	**0,7563**	**0,6645**	**0,8480**	**0,8544**

Tabelle 4.17: Qualität der optimierten anfrageübergreifenden Konfiguration K_Q für die einzelnen Anfragen

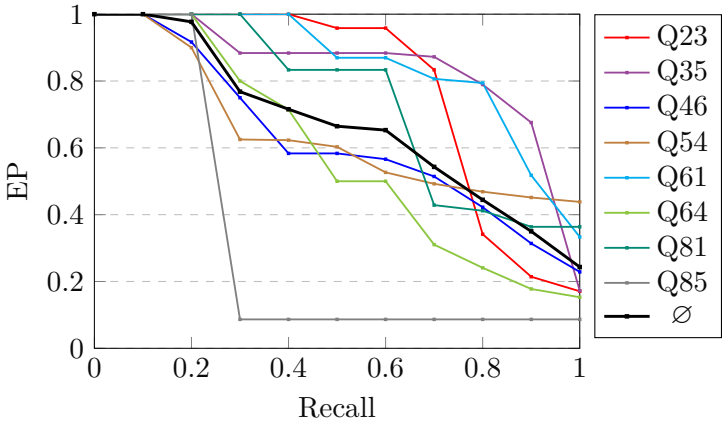


Abbildung 4.35: EP-Recall-Graphen für die anfrageübergreifende Konfiguration K_Q für alle Anfragen

Zur Interpretation stellen wir uns vor, dass das IR-System die Relevanz für eine Menge von Elementen berechnet und diese als sortierte Liste zur Verfügung stellt. Ob die gefundenen

Elemente tatsächlich relevant sind, wird mit der objektiven Relevanz abgeglichen. Tauchen alle tatsächlich relevanten Elemente gleich am Anfang der Liste auf, so hat die Genauigkeit an den Recall-Punkten den Wert 1. Erscheint in der Liste ein nicht relevantes Element, so fällt die Kurve an dem Recall-Punkt leicht ab. Fällt die Kurve stark ab, wie bei Q85, so deutet das darauf hin, dass im Verlauf der Liste auch viele Elemente nicht relevant sind. Für die Nutzer des IR-Systems ist es wichtig, dass insbesondere die Ergebnisse im oberen Bereich der Liste auch tatsächlich relevant sind, das bedeutet hohe EP-Werte vor allem im niedrigen Recall-Bereich. Da das Hauptinteresse der Nutzer ohnehin auf den ersten Treffern liegt, spielen die Vollständigkeit und damit hohe Recall-Werte eine untergeordnete Rolle. Die obersten Elemente der Liste sollten idealerweise die höchsten Relevanzgrade aufweisen, daher sind hohe NDCG-Werte wichtig.

Wir erkennen zunächst, dass alle Kurven im niedrigen Recall-Bereich eine sehr hohe Genauigkeit besitzen, was auf richtige Treffer im oberen Bereich der Ergebnislisten hindeutet. Außer für Q85, zeigen alle Kurven eine gute Trefferquote und nur einen leichten Abfall. Die hohen NDCG-Werte sind ein Zeichen dafür, dass die Elemente mit hoher Relevanz weit oben in der Liste erscheinen. Insgesamt ergibt sich für die EP-Recall-Kurven in Abbildung 4.35 ein positives Bild mit relevanten und nützlichen Ergebnisse für alle Anfragen.

In Tabelle 4.18 listen wir die Werte für Fitness, EP, NDCG_{Min} und NDCG_{Max} auf, die sich bei Ausführung des IR-Systems mit den anfragespezifischen Konfigurationen für die einzelnen Anfragen ergeben. Abbildung 4.36 stellt die entsprechenden EP-Recall-Graphen sowohl für jede einzelne Anfrage als auch gemittelt über die Anfragen dar.

Konfiguration	**Anfrage**	**Fitness**	**EP**	**NDCG_{Min}**	**NDCG_{Max}**
K_{Q23}	Q23	0,9262	0,8944	0,9579	0,9579
K_{Q35}	Q35	0,9319	0,9230	0,9409	0,9467
K_{Q46}	Q46	0,8477	0,7444	0,9510	0,9620
K_{Q54}	Q54	0,7367	0,6340	0,8393	0,8393
K_{Q61}	Q61	0,9084	0,8840	0,9328	0,9328
K_{Q64}	Q64	0,8111	0,6867	0,9355	0,9372
K_{Q81}	Q81	0,9453	0,9139	0,9768	0,9768
K_{Q85}	Q85	0,6446	0,3732	0,9159	0,9159
	∅	**0,8440**	**0,7567**	**0,9313**	**0,9336**

Tabelle 4.18: Qualität der optimierten anfragespezifischen Konfigurationen für die einzelnen Anfragen

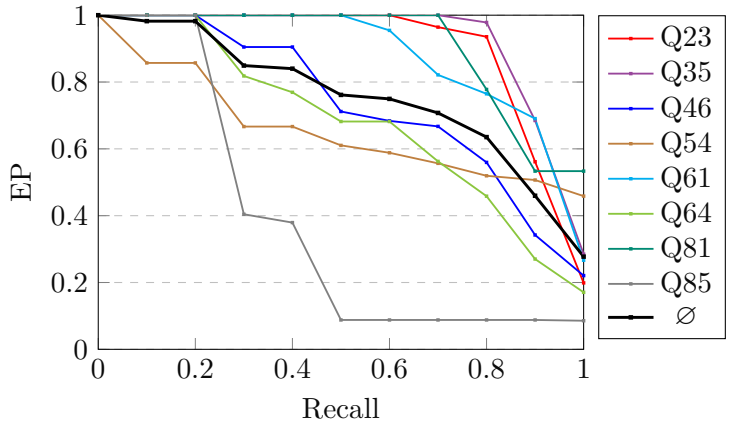


Abbildung 4.36: EP-Recall-Graphen für die anfragespezifischen Konfigurationen K_{Q23}-K_{Q85} für alle Anfragen

Wir konnten bereits im letzten Abschnitt zeigen, dass die Güte der Ergebnisse für die anfragespezifischen Konfigurationen höher ist als für die anfrageübergreifende Konfiguration. Das bestätigt sich zudem in den EP-Recall-Kurven. Sie zeigen ebenfalls hohe Genauigkeiten im oberen Bereich der Ergebnislisten und zudem einen leichteren Abfall der Genauigkeit für alle Anfragen.

***EQ4**: Sind die Ergebnisse des IR-Systems relevant und nützlich?*

Mit den optimierten Konfigurationen erreichen wir hohe Genauigkeiten am Beginn der Ergebnislisten und gute Trefferquoten im Verlauf der Ergebnislisten. Das entspricht den Erwartungen, die Nutzer üblicherweise an die Ergebnisse eines IR-Systems haben. Durch die hohen NDCG-Werte konnten wir zudem zeigen, dass die Ergebnisse zumeist eine sehr hohe Nützlichkeit besitzen.

4.4.5 Auswahl geeigneter Konfigurationen

In diesem Abschnitt setzen wir uns mit der Problematik der Auswahl geeigneter Konfigurationen für ein IR-System auseinander. Wir haben in der Evaluation zwei Optimierungsziele unterschieden, die zwei unterschiedliche Herangehensweisen an die Parametrisierung der Aktivierungsausbreitung in IR-Systemen verkörpern. Zum einen haben wir nach anfrageübergreifenden Konfigurationen gesucht, deren Ziel der universelle Einsatz für alle möglichen Anfragen darstellt. Zum anderen haben wir die Konfigurationen derart optimiert, dass sie anfragespezifisch sind. Wir haben gezeigt, dass die anfragespezifische Konfiguration zu qualitativ besseren Ergebnissen führt als die anfrageübergreifende Konfiguration. In der Praxis ist es beim Aufbau eines IR-Systems allerdings nicht möglich, im Vorfeld bereits optimale Konfigurationen für jede noch nicht bekannte Anfrage zur Befriedigung noch unbekannter Informationsbedürfnisse zu erstellen. Stattdessen ist es erforderlich, eine möglichst geeignete Parametrisierung der Aktivierungsausbreitung und Auswahl des Netzes vorzunehmen.

Vorgehen: Die Vorgehensweise, die in dieser Arbeit zur Evaluation genutzt wurde, kann in der Praxis ein Anhaltspunkt für die Identifikation geeigneter Konfigurationen für ein domänenspezifisches IR-System sein. Der Entwickler des IR-Systems lässt von Domänenexperten mehrere Anfragen bewerten und identifiziert für diese optimierte Konfigurationen mittels evolutionärer Algorithmen. Darstellungen wie in den Abbildungen 4.22 und 4.23 (Seiten 151 und 152) dienen als Hilfsmittel zur Bewertung der Top-Konfigurationen. Sie erleichtern die Auswahl einer geeigneten Konfiguration und Netzvariante. Ist die Konfiguration mit höchstem Fitnesswert beispielsweise ein Ausreißer, so ist sie für die allgemeine Verwendung für andere Anfragen vermutlich nicht geeignet. Daher berücksichtigen wir bei der Bewertung von Kombinationen neben der hohen Fitness und einer passenden Netzvariante auch die Stabilität der Kombination.

Anfragetypspezifische Konfigurationen: Um den positiven Effekt spezifischer Konfigurationen zu nutzen, empfehlen wir, die Anfragen zu gruppieren und zumindest spezifische Konfigurationen für spezifische Eigenschaften der Anfragen zu erstellen. Eine intuitive Gruppierung kann nach Anfragetyp vorgenommen werden. Anfragetypen gruppieren Anfragen mit gleichem Ausgangspunkt und Suchziel. In dieser Arbeit sind beispielsweise die Konfigurationen Q61 und Q64 vom gleichen Anfragetyp, da beide den Ausgangspunkt und das Suchziel teilen und ausgehend von einer Gefährdungssituation nach Effekten suchen (Tabelle 4.2, Seite 123). Für eine Domäne lassen sich zumeist eine begrenzte Menge sinnvoller Anfragetypen

herausfiltern, die den Grundstein für eine solche Analyse legen können. In dem Fall werden Optimierungsläufe für mehrere Anfragen jedes Anfragetyps zusammengefasst. Wir vermuten, dass anfragetypspezifische Konfigurationen durch den gezielten Zuschnitt qualitativ bessere Ergebnisse erzielen können.

Anfragetypspezifische Analysen erfordern aufgrund der notwendigen domänenspezifischen Analyse zur Definition der Anfragetypen vermutlich mehr Aufwand als die anfrageübergreifenden Analysen. Sie haben dafür Potenzial, durch ihren spezifischen Charakter bessere IR-Ergebnisse zu liefern. Ist der zusätzliche Aufwand nicht akzeptabel, so kann alternativ anfrageübergreifend optimiert werden. Bei der Identifikation einer universell einsetzbaren, anfrageübergreifenden Konfiguration entfällt die dafür notwendige Analyse. Dennoch sollten die Anfragen ein gewisses Maß an Diversität aufweisen und die unterschiedlichen Informationsbedürfnisse der Domäne widerspiegeln.

4.5 Fazit und Bewertung

Mit der Evaluation haben wir die Anwendbarkeit und Qualität der in dieser Arbeit vorgestellten Techniken und Konzepte gezeigt. Wir konnten unter anderem zeigen, dass der Einsatz semantischen Anreicherungen und die Reduktionen, die wir in den Abschnitten 3.2.3 und 3.2.4 vorstellten, sinnvoll ist, da sie essenziell zur Qualität der Ergebnisse beitragen. Außerdem konnten wir zeigen, dass der Einsatz der Ausbreitungsmodi, die wir in Abschnitt 3.3.4 als zusätzliches Werkzeug zur Steuerung der Aktivierungsausbreitung eingeführt haben, sinnvoll ist. Mithilfe optimierter Konfigurationen konnten wir zudem die Relevanz und Nützlichkeit der Ergebnisse unseres IR-Systems für G&Rs verdeutlichen.

Die Ergebnisse der Evaluation zeigen auch, dass die Qualität der Empfehlungen stark von den Voraussetzungen abhängt, unter denen der Suchalgorithmus läuft. Dazu gehören, wie die Evaluation zeigt, sowohl die Konfigurationsparameter als auch die Verschiedenheit der Netze. Die Steuerung der Konfigurationsparameter zur Generierung möglichst relevanter Empfehlungen erfordert viel Wissen um die Einflüsse der unterschiedlichen Parameter, vor allem auch in ihren Kombinationen, und den Effekten, die diese auf die Ergebnisse der semantischen Suche durch Aktivierungsausbreitung haben. Mit unserer Vorgehensweise bei der Optimierung der Konfigurationen mittels evolutionärer Algorithmen haben wir bereits eine Möglichkeit aufgezeigt, wie geeignete Konfigurationen für ein domänenspezifisches IR-System identifiziert werden können. Für den Entwickler von IR-Systemen kann dies eine große Herausforderung sein, die zudem mit großem Aufwand verbunden ist. Daher sollten zukünftige Arbeiten eine automatisierte Konfiguration anstreben.

Im folgenden Kapitel unterstützen wir den Entwickler von IR-Systemen mit dem semantischen Netzwerkskelett, deren kompakte Darstellungsform den Umgang mit großen Netzen vereinfacht. Außerdem zeigen wir, wie die Simulation der Aktivierungsausbreitung auf dem Netzwerkskelett zum frühzeitigen Ausschluss ungünstiger Konfigurationen genutzt werden kann.

5 Semantisches Netzwerkskelett

In den vorherigen Kapiteln haben wir gezeigt, wie domänenspezifische IR-Systeme weitestgehend automatisiert erstellt werden können. Zu den nicht automatisierten Aufgaben eines Entwicklers solcher IR-Systeme gehört unter anderem die Auswahl der Netzvarianten und Konfigurationsparameter. Deren geeignete Auswahl trägt allerdings, wie die Evaluation gezeigt hat, zur Qualität der Ergebnisse des IR-Systems bei und ist entsprechend bedeutend. Dieses Kapitel widmet sich daher der Unterstützung der Entwickler von IR-Systemen bei der Bewältigung der damit verbundenen Herausforderungen. Wir präsentieren in diesem Kapitel das semantische Netzwerkskelett, eine neuartige Abstraktion semantischer Netze, welche wesentliche strukturelle Eigenschaften eines semantischen Netzes in kompakter Form abbildet. Dafür werden wir zunächst definieren, wie ein semantisches Netzwerkskelett aufgebaut ist und welche Informationen wir im Skelett konservieren. Anschließend werden wir am Beispiel eines semantischen Netzwerkskelettes für ein G&R-Netz zeigen, wie die übersichtliche Darstellung sonst komplexer Zusammenhänge den Umgang mit großen semantischen Netzen erleichtert. Darüber hinaus eröffnen sich weitere Verwendungsmöglichkeiten der Netzwerkskelette. Wir werden beispielhaft zeigen, wie Aktivierungsausbreitung auf semantischen Netzwerkskeletten simuliert werden kann, und erläutern, welche Analysepotenziale die Simulationsergebnisse besitzen.

5.1 Syntax und Semantik

semantisches Skelett

Ursprungsnetzwerk

Das *semantische Netzwerkskelett* (*engl.:* semantic network skeleton) ist eine abstrakte Repräsentation eines semantischen Netzwerks, dem Ursprungsnetzwerk (*engl.:* source network). Das Skelett fasst die Kanten und Knoten seines Ursprungsnetzwerks derart zusammen, dass eine wesentlich kompaktere Sicht auf das semantische Netz entsteht, grundsätzliche Strukturen allerdings erhalten bleiben, beziehungsweise strukturelle Eigenschaften sogar explizit werden. Dadurch unterstützt die entstehende Skelettstruktur das Verständnis über die Grundstruktur des Ursprungsnetzwerks. Die gruppierten und zusammengefassten Knoten werden im Netzwerkskelett von Knotenrepräsentanten abgebildet, Kanten entsprechend von Kantenrepräsentanten.

Knoten-/ Kantenrepräsentant

Annotationen

Zusätzliche Annotationen an den Knoten- oder Kantenrepräsentanten bereichern das Skelett mit zusätzlichen Informationen. Die Annotationen können grundsätzlich an den Verwendungszweck der Skelette angepasst werden. Wir sehen den Vorteil semantischer Netzwerkskelette in der kompakten Darstellung der strukturellen Eigenschaften großer semantischer Netze und dokumentieren in den Annotationen daher beispielsweise, wie viele Knoten beziehungsweise Kanten von einem Repräsentanten repräsentiert werden. Annotationen erlauben uns die Bewertung des Ausmaßes der Reduktion, der anzahlmäßigen Verteilung der Knoten im semantischen Netz und des Ausmaßes der Konnektivität zwischen bestimmten Arten von Knoten.

In Abbildung 5.1 zeigen wir ein Beispiel für ein Ursprungsnetzwerk mit entsprechendem semantischen Netzwerkskelett. Im oberen rechten Teil der Abbildung führen wir die graphische Notation semantischer Netzwerkskelette ein. Sie ähnelt der Notation semantischer Netze, enthält aber zusätzlich die Annotationen, welche an Knoten, Kanten, Kantenquelle und -ziel dargestellt werden. In Abbildung 5.1 werden beispielsweise die zwei Instanzen vom Typ `hara:Function` zu einem sie repräsentierenden Knoten `hara:Ins_Function` zusammengefasst. Ihr Knotenrepräsentant erhält daher die Knotenannotation 2. Analog dazu werden die drei Instanzen vom Typ `hara:Malfunction` zu einem sie repräsentierenden Knoten

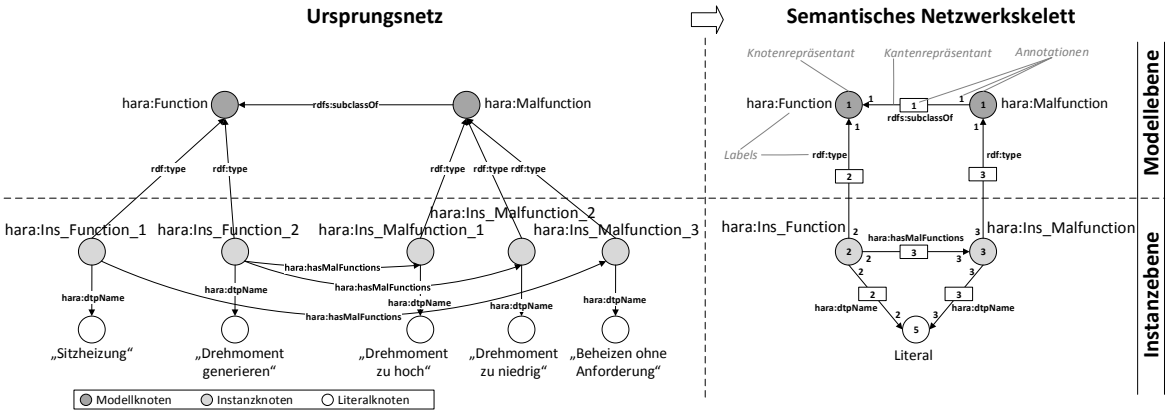


Abbildung 5.1: Beispiel eines semantischen Netzes und seines Netzwerkskelettes

`hara:Ins_Malfunction` zusammengefasst, der die Knotenannotation 3 erhält. Die drei Kanten mit dem Prädikat `hara:hasMalfunctions` verbinden im Ursprungsnetz Instanzknoten vom Typ `hara:Function` mit Instanzknoten vom Typ `hara:Malfunction`, für die im Skelett bereits Knotenrepräsentanten existieren. Wir fassen die drei Kanten zusammen. Ihr Kantenrepräsentant verbindet im Skelett die entsprechenden Knotenrepräsentanten. An der Kantenquelle annotieren wir, dass beide vom Knotenrepräsentant repräsentierten Knoten im Ursprungsnetz mit der entsprechenden Kante verbunden sind, am Kantenziel gilt das für alle drei Knoten.

Das Zusammenspiel der Annotationen konserviert Informationen über die Konnektivität im Ursprungsnetz. Genau genommen repräsentiert jedes Skelett-Tripel einen bipartiten Graphen, in dem die repräsentierten Knoten eines Knotenrepräsentanten jeweils eine Partitionsklasse darstellen. Ohne genauere Details zum Ursprungsnetz zu haben, wird beispielsweise deutlich, dass das Skelett-Tripel `<hara:Ins_Function> <hara:hasMalfunctions> <hara:Ins_Malfunction>` drei Tripel im Ursprungsnetz repräsentiert. In einem vollständig bipartiten Graphen sind maximal sechs Tripel und damit sechs Kanten möglich. Wir können an dem Beispieltripel erkennen, dass jeder repräsentierte Knoten jeweils ein Mal mit einer der repräsentierten Kanten verbunden ist. Mit der Abstraktion geht auch ein Informationsverlust einher. Es ist beispielsweise keine Aussage mehr darüber möglich, welche Knoten miteinander verbunden sind. Im Skelett lässt sich außerdem erkennen, dass im Ursprungsnetz nicht jedes Literal, sondern nur zwei der fünf Literalknoten mit einem Instanzknoten vom Typ `hara:Function` verbunden sind. Im Gegensatz dazu sind die Repräsentationen des Skelett-Tripels `<hara:Malfunction> <rdf:type> <hara:Ins_Malfunction>` vollständig miteinander verknüpft.

Syntax

Definition semantisches Netzwerkskelett: Sei L_S eine Menge von Labels, dann definieren wir das semantische Netzwerkskelett S eines RDF-basierten semantischen Netzes G (siehe Definition in Abschnitt 3.3.1), im Folgenden auch Ursprungsnetz genannt, durch das 5-Tupel $S = (N_S, E_S, s_S, t_S, l_S)$. Hierbei sind N_S eine nichtleere Menge von Knotenrepräsentanten, E_S eine Menge von Kantenrepräsentanten, $s_S : E_S \to N_S$ eine Quellknoten-Abbildung für Kantenrepräsentanten, $t_S : E_S \to N_S$ eine Zielknoten-Abbildung für Kantenrepräsentanten und $l_S : N_S \cup E_S \to L_S$ eine Labelfunktion.

Die Knoten- und Kantenrepräsentanten des Skelettes repräsentieren jeweils Mengen von Knoten beziehungsweise Kanten ihres Ursprungsnetzes. Analog zu Formel 3.1 (Seite 91)

definieren wir, dass die Knoten $n, m \in N_S$ und die sie verbindende Kante $e \in E_S$ ein *Skelett-Tripel* bilden. *Skelett-Tripel*

$$trip_S(n, m, e) \Leftrightarrow (s_S(e) = n \wedge t_S(e) = m) \vee (s_S(e) = m \wedge t_S(e) = n) \tag{5.1}$$

Definition Skelettannotationen: Für ein Skelett S definieren wir die *Skelettannotationen* A_S durch das 4-Tupel $A_S = (A_n, A_e, A_s, A_t)$, mit $A_n : K \times N \to V$ als *Knotenannotation*, $A_e : K \times E \to V$ als *Kantenannotation*, $A_s : K \times E \to V$ als *Kantenannotation an der Kantenquelle* und $A_t : K \times E \to V$ als *Kantenannotation am Kantenziel*. Hier steht K für eine Menge an *Annotationsschlüssel* und V für eine Menge an *Annotationswerten*.

Die verwendeten Annotationsschlüssel sind prinzipiell an die Verwendung des semantischen Netzwerkskelettes anpassbar. In dieser Arbeit nutzen wir die Annotationen, um einen Überblick über potenzielle Pfade im semantischen Netz und das Ausmaß der Konnektivität von Knoten und Kanten darzustellen. Daher verwenden wir in dieser Arbeit die folgenden vier Annotationsschlüssel, die bereits im einleitenden Beispiel in Abbildung 5.1 eingesetzt wurden. Wir annotieren jeden Knotenrepräsentanten mit der Anzahl der Knoten, die er zusammenfasst (*node_count*). Analog dazu annotieren wir jeden Kantenrepräsentanten mit der Anzahl der Kanten, die er zusammenfasst (*edge_count*). Wir annotieren an jeder Quelle eines Kantenrepräsentanten, wie viele Verbindungen des Ursprungsnetzwerks durch die Verbindung der Knoten- und Kantenrepräsentanten repräsentiert werden (*src_rep*). Analog dazu annotieren wir die repräsentierten Verbindungen am Ziel eines Kantenrepräsentanten (*tgt_rep*). Insbesondere die Annotationen an Kantenquelle und Kantenziel ermöglichen die Bewertung der Konnektivität im semantischen Netz.

Typen semantischer Netzwerkskelette

Wir unterscheiden zwei Typen semantischer Netzwerkskelette, die sich bezüglich ihrer Vollständigkeit und ihres Detailgrades unterscheiden: das effektive und das maximale Netzwerkskelett.

Das *effektive Skelett* repräsentiert die spezifische Struktur eines konkreten semantischen Netzes. Es enthält ausschließlich Repräsentanten für Knoten und Kanten, die auch tatsächlich im entsprechenden Ursprungsnetz vorhanden sind. Das bedeutet, dass beispielsweise Properties oder Klassen, die auf Modellebene definiert sind, auf Instanzebene allerdings nicht verwendet oder instanziiert werden, auch keine Repräsentanten auf der Instanzebene des Skelettes erhalten. *effektives Skelett*

Das *maximale Skelett* hingegen enthält auf Instanzebene Knoten- und Kantenrepräsentanten für alle potenziell auftretenden Knoten und Kanten des Ursprungsnetzwerks. Das maximale Skelett ergibt sich aus den Informationen der Modellebene des Ursprungsnetzwerks. Werden also Properties oder Klassen auf Modellebene definiert, so wird auch deren Verwendung oder Instanziierung auf Instanzebene repräsentiert, ohne dass diese tatsächlich verwendet oder instanziiert werden müssen. Das maximale Skelett ist daher vollständig, kann aber ungenutzte Elemente enthalten. Das bedeutet auch, dass das maximale Skelett für alle Instanzen semantischer Netze gilt, die der schematischen Struktur der Modellebene entsprechen. Daher entfallen allerdings die Annotationen, da beispielsweise auf Instanzebene keine Aussage mehr über die Anzahl der repräsentierten Knoten möglich ist. *maximales Skelett*

Im direkten Vergleich lassen sich Vor- und Nachteile beider Typen identifizieren. Das maximale Skelett ist eine allgemeine und vollständige Variante des Skelettes, die alle potenziell möglichen strukturellen Elemente enthält und daher für alle Netzinstanzen für das entsprechende Modell gilt. Daher eignet es sich für die Darstellung struktureller Eigenheiten oder zur Analyse

potenzieller Pfade in Netzen. Der Allgemeingültigkeit stehen die fehlenden spezifischen Informationen, z.B. durch die fehlenden Annotationen, entgegen. Effektive Netze hingegen enthalten diese Informationen für eine konkrete Netzinstanz und erlauben dadurch spezifische Erkenntnisse auch über Größe und Konnektivität im Netz. Effektive Skelette tendieren dazu, kompakter zu sein, da sie ausschließlich die enthaltenen Knoten und Kanten eines einzelnen Ursprungsnetzes repräsentieren.

5.2 Erstellen semantischer Netzwerkskelette

Wir fokussieren beim Zusammenfassen von Knoten und Kanten auf den erweiterten semantischen Kern von Netzen (siehe Abbildung 3.8, Seite 70). Modellknoten und Modellkanten außerhalb des erweiterten semantischen Kerns eignen sich nicht dafür, gruppiert und zusammengefasst zu werden. Zum einen befinden sich auf Modellebene, anders als auf Instanzebene, wenige gleichartige Knoten und Kanten, die derart zusammengefasst werden können, dass die Strukturen im Skelett sicher bestehen bleiben. Zum anderen stellen sie nur einen sehr geringen Teil instanziierter Netze dar, beispielsweise im G&R-Basisnetz lediglich 0,33% aller Knoten und Kanten (siehe Tabelle 3.4, Seite 72). Auf Modellebene fassen wir daher keine Knoten oder Kanten zusammen. Für jeden Knoten und jede Kante wird genau ein Knoten- beziehungsweise ein Kantenrepräsentant im Skelett abgebildet. Entsprechend erhalten alle Knoten- und Kantenannotationen den Wert 1. Auf Instanzebene bedienen wir uns des Vorteils RDF-basierter Netze, wonach die Instanziierung den Eigenschaften genügen muss, die auf Modellebene festgelegt sind. Beispielsweise existieren auf Instanzebene ausschließlich Instanzen der Klassen, die auf Modellebene definiert sind, und es werden nur Eigenschaften zum Verknüpfen von Knoten verwendet, die auf Modellebene beispielsweise als Object Properties oder Datatype Properties definiert sind. Wir fassen daher alle Instanzknoten des gleichen Typs zusammen und erstellen für sie im Skelett jeweils einen Knotenrepräsentanten. Außerdem werden alle Literalknoten durch einen Knotenrepräsentanten abgebildet. Das Vorgehen unterscheidet sich beim Erstellen maximaler und effektiver Skelette, da maximale Skelette sich aus der Modellebene des Ursprungsnetzes ableiten und effektive Skelette vorrangig die Betrachtung der Instanzebene des Ursprungsnetzes erfordern. Daher beschreiben wir das Vorgehen im Folgenden für beide Typen getrennt.

Erstellen effektiver Skelette: Folgendes Vorgehen beschreibt die Ableitung eines effektiven Skelettes.

1. Jeder Knoten der Modellebene erhält einen Knotenrepräsentanten im Skelett.
2. Für jeden Typ werden dessen Instanzen, sofern vorhanden, von einem Knotenrepräsentanten zusammengefasst.
3. Alle Literale werden von einem Knotenrepräsentanten zusammengefasst.
4. Für jedes Tripel des Ursprungsnetzes wird, sofern noch nicht vorhanden, ein Kantenrepräsentant für das Prädikat zwischen den Knotenrepräsentanten von Subjekt und Objekt des Tripels erstellt.

Während des Erstellungsprozesses werden zudem die Werte für die Knoten- und Kantenannotationen erfasst und dem Skelett hinzugefügt.

Erstellen maximaler Skelette: Folgendes Vorgehen beschreibt die Ableitung eines maximalen Skelettes.

1. Jeder Knoten der Modellebene erhält einen Knotenrepräsentanten im Skelett. Jede Kante der Modellebene erhält einen Kantenrepräsentanten im Skelett.
2. Jede definierte Klasse erhält einen Knotenrepräsentanten für ihre potenziellen Instanzen im Skelett.
3. Existiert mindestens ein Datatype Property, so erhält das Skelett einen Knotenrepräsentanten für Literale.
4. Jedes Object Property und jedes Datatype Property erhält Kantenrepräsentanten zwischen den Knotenrepräsentanten für die Instanzen des Definitionsbereiches und den Knotenrepräsentanten für die Instanzen des Wertebereiches im Skelett. Dabei wird der Gültigkeitsbereich auf die Subklassen propagiert, was dazu führen kann, dass für ein Property mehrere Kantenrepräsentanten im Skelett entstehen.

Die von uns festgelegten Annotationen entfallen im maximalen Skelett, da es eine Abstraktion aller potenzieller Netzinstanzen darstellt und das Ausmaß der Instanziierung nicht limitiert ist. Daher sind hier keine Aussagen über die anzahlmäßige Verteilung auf Instanzebene möglich.

5.3 Verwendung als Hilfsmittel

In Abbildung 5.2 ist das effektive semantische Netzwerkskelett der G&R-Netzvariante N4-2 dargestellt. Das Ursprungsnetz N4-2 ist auf den erweiterten semantischen Kern reduziert, da es sich um ein Netz der Reduktionsstufe 2 handelt. Im oberen Bereich der Abbildung ist die Modellebene mit den Klassenknoten dargestellt, im unteren Bereich die Instanzebene. Aufgrund der Abstraktion erhalten wir eine kompakte Darstellung, die einen Überblick über das vom Skelett repräsentierte Ursprungsnetz bietet. Die 240.733 Tripel des Ursprungsnetzes werden von nur 87 Skelett-Tripeln abgebildet, aus 47.679 Knoten werden 37 Knotenrepräsentanten (siehe Tabelle 4.2, Seite 121). Auf einen Blick sind grundlegende strukturelle Eigenschaften erkennbar, z.B. welche Klassen im Ursprungsnetz enthalten sind, welche Eigenschaften ihre Instanzen besitzen und welche Verbindungen dadurch zu anderen Instanzen oder zu Literalen entstehen. Darüber hinaus wird die Anzahl der Instanzen jeder Klasse im Ursprungsnetz sowie die Anzahl von Verbindungen explizit.

Neben den Vorteilen, die die Übersichtlichkeit dieser Art der Darstellung bietet, sind grundsätzlich auch weitere Anwendungen denkbar. Beispielsweise erlauben semantische Netzwerkskelette eine Analyse der Algorithmen, die auf den von ihnen repräsentierten Netzen angewendet werden. Daher kann das Netzwerkskelett auch als Hilfsmittel im Umgang mit Herausforderungen der Konfiguration der Aktivierungsausbreitung dienen. Eine Idee ist, das Netzwerkskelett zu nutzen, um die Aktivierungsausbreitung zu simulieren und dadurch bereits vorab Konfigurationen zu identifizieren, die zu unerwünschten Effekten führen. Im folgenden Abschnitt präsentieren wir einen Ansatz zur Simulation der Aktivierungsausbreitung auf semantischen Netzwerkskeletten.

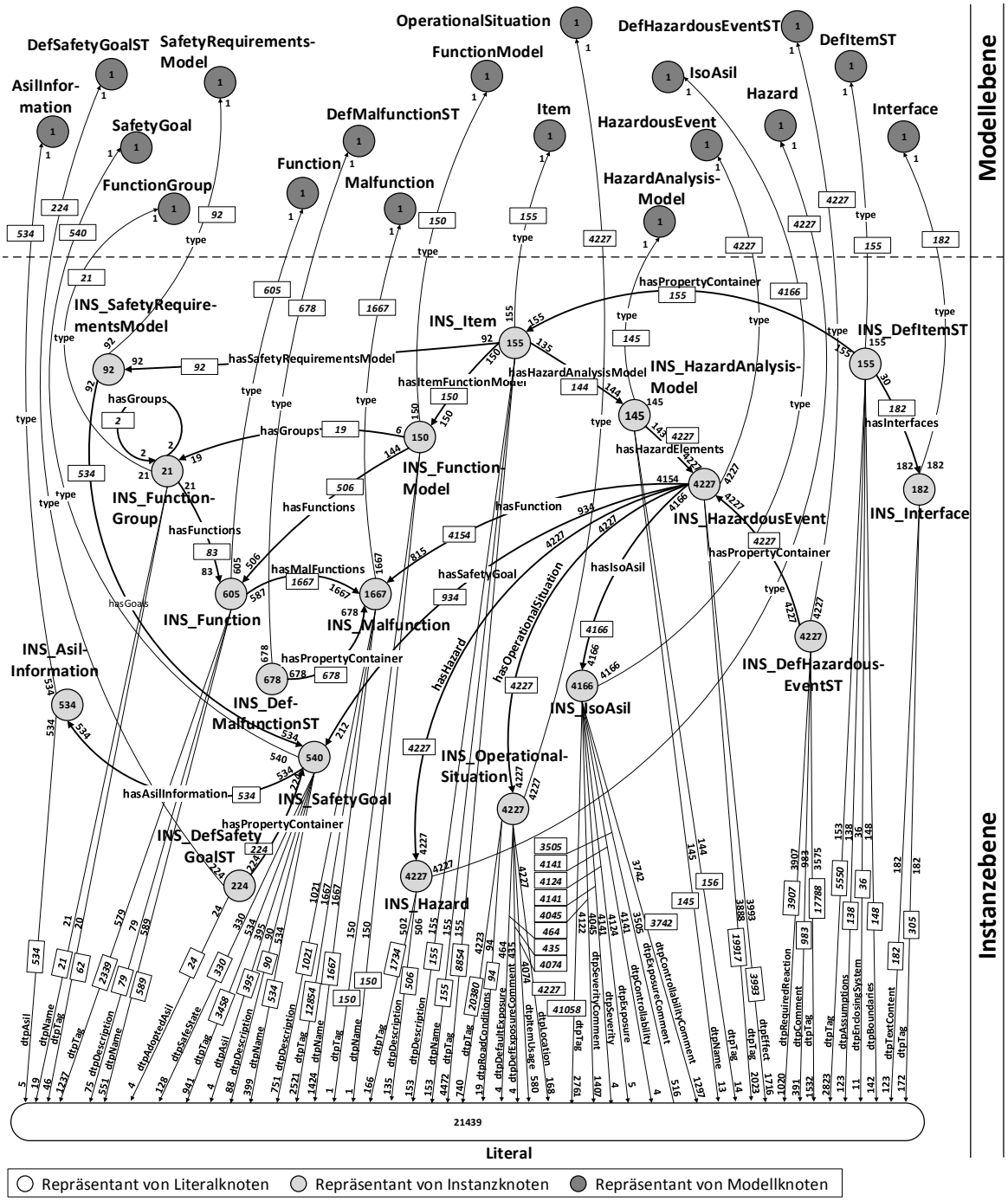


Abbildung 5.2: Effektives semantisches Netzwerkskelett der Netzvariante N4-2

5.4 Simulation der Aktivierungsausbreitung auf dem Netzwerkskelett

Die Struktur semantischer Netze und die Konfiguration der Aktivierungsausbreitung haben Einfluss auf die Ergebnisse der Aktivierungsausbreitung. Einflussfaktoren können spezielle Effekte bewirken, die sowohl erwünscht als auch unerwünscht sein können. Ein Beispiel eines unerwünschten Effektes ist der Abbruch der Aktivierungsausbreitung, weil Knoten als Senken agieren und aufgrund ungünstiger Konfigurationseinstellungen die Weiterverteilung der Aktivierung einstellen. Das Wissen um Einflussfaktoren und deren Effekte im semantischen Netz ermöglicht eine Vorab-Analyse der Konfigurationen mit dem Ziel, unerwünschte Effekte zu erkennen und zu umgehen. Da semantische Netzwerke oftmals sehr groß sind, kann eine ausführliche Analyse vieler Konfigurationen eine sehr zeitaufwendige Aufgabe sein. In diesem Abschnitt präsentieren wir, wie Aktivierungsausbreitung auf den wesentlich kleineren semantischen Netzwerkskeletten simuliert werden kann. Im Folgenden stellen wir die Grundidee vor und beschreiben das Vorgehen bei der Simulation sowie die Analysepotenziale, welche die Simulationsergebnisse besitzen.

5.4.1 Vorgehen

Die Grundidee besteht darin, die Aktivierungsausbreitung auf der komprimierten Abstraktion semantischer Netzwerke, den semantischen Netzwerkskeletten, auszuführen. Da die Knoten und Kanten des semantischen Netzwerkskelettes Repräsentanten für potenziell mehrere Knoten und Kanten des repräsentierten Netzes sind, passen wir die Aktivierungsverteilungsfunktionen derart an die neue Graphstruktur an, dass wir eine Näherung der Ergebnisse für jeden Simulationsschritt erhalten. Abbildung 5.3 stellt die Grundidee dar, die im Folgenden detaillierter ausgeführt wird.

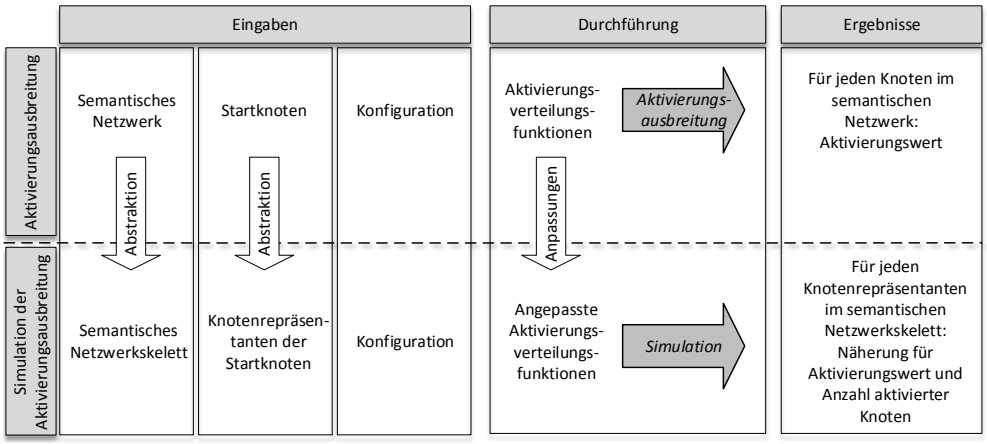


Abbildung 5.3: Grundidee der Simulation der Aktivierungsausbreitung auf dem semantischen Netzwerkskelett

5.4.1.1 Eingaben für die Simulation

Als erste Eingabe benötigen wir ein semantisches Netzwerk, für welches die Aktivierungsausbreitung simuliert werden soll. Die Simulation erfolgt auf dem semantischen Netzwerkskelett dieses semantischen Netzwerks, welches durch Abstraktion erstellt wird. Die Startknoten für die Simulation entsprechen der Menge der Knotenrepräsentanten der Startknoten für die

Aktivierungsausbreitung auf dem semantischen Netz. Die dritte Eingabe ist die Konfiguration der Aktivierungsausbreitung, die keine weitere Anpassung benötigt.

5.4.1.2 Ergebnisse der Simulation

Während der Simulation verfolgen wir zwei Eigenschaften, die von besonderem Interesse sind und für die wir daher Näherungswerte berechnen. Der erste Aspekt ist die erwartete Anzahl aktivierter Knoten, die von einem Knotenrepräsentanten repräsentiert werden. Bei der Aktivierungsausbreitung im semantischen Netzwerk wird diese Eigenschaft nicht explizit berechnet, sondern ist implizit durch die Höhe der Aktivierungswerte vorhanden. Wir nennen diese Eigenschaft die *Ausbreitungsstärke* (*engl.:* spreading strength). Sie beschreibt, wie schnell sich Aktivierung im semantischen Netzwerk verteilt. Dafür wird pulsweise die erwartete Anzahl aktivierter repräsentierter Knoten jedes Knotenrepräsentanten berechnet. Dies erlaubt wiederum eine Abschätzung des Aktivierungsstatus der Knoten im gesamten Netz.

Ausbreitungsstärke

Der zweite Aspekt betrifft die erwarteten Aktivierungswerte der repräsentierten Knoten. Diese Eigenschaft nennen wir *Aktivierungsstärke* (*engl.:* activation strength). Sie beschreibt, wie schnell und in welchem Ausmaß sich die Aktivierungswerte im Ursprungsnetz vermutlich ändern.

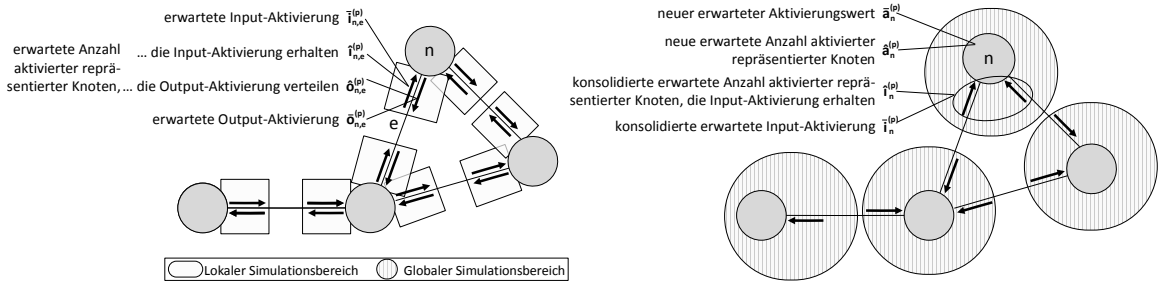
Aktivierungsstärke

Für jeden Knotenrepräsentanten berechnen wir in jedem simulierten Puls sowohl Näherungswerte für die erwartete Anzahl aktivierter Knoten als auch für die erwarteten Aktivierungswerte.

5.4.1.3 Durchführung der Simulation

Simulation eines Pulses - Lokale und globale Simulationsschritte

Jeder simulierte Puls $p \in \mathbb{N}_{>0}$ besteht aus einer Abfolge von zuerst *lokalen* und anschließend *globalen* Simulationsschritten. Diese entsprechen der Abfolge der Aktivierungsverteilungsfunktionen. Abbildung 5.4 stellt lokale und globale Simulationsschritte schematisch dar und bildet zudem die verschiedenen Werte für Ausbreitungs- und Aktivierungsstärke ab, die während der Simulation berechnet werden.



(a) Lokal (b) Global

Abbildung 5.4: Simulationsschritte

lokaler Schritt

Zunächst simulieren wir lokal (Abbildung 5.4a) die Output- und die Input-Aktivierung für jedes Knoten-Kanten-Paar aller Skelett-Tripel. Für die Aktivierungsstärke berechnen wir, analog zu den Aktivierungsverteilungsfunktionen, die *erwartete Output-Aktivierung* $\bar{o}^{(p)}_{n,e}$ und die *erwartete Input-Aktivierung* $\bar{i}^{(p)}_{n,e}$ eines Knotens $n \in N_S$ über Kante $e \in E_S$. Für die

Ausbreitungsstärke berechnen wir die *erwartete Anzahl aktivierter repräsentierter Knoten* $\hat{o}^{(p)}_{n,e}$ und $\hat{i}^{(p)}_{n,e}$, die Output-Aktivierung verteilen beziehungsweise Input-Aktivierung erhalten. Da ein Knotenrepräsentant Aktivierung über mehrere Kanten erhalten kann, folgt nach den lokalen Simulationsschritten eines Pulses für jeden Knotenrepräsentanten des Skelettes ein globaler Simulationsschritt. Wir simulieren global (Abbildung 5.4b) die Konsolidierung der Input-Aktivierung für jeden Knoten und die Berechnung der neuen Aktivierungswerte. Für die Aktivierungsstärke berechnen wir die *konsolidierte erwartete Input-Aktivierung* $\bar{i}^{(p)}_n$ eines Knotens n über alle Kanten und den *neuen erwarteten Aktivierungswert* $\bar{a}^{(p)}_n$. Für die Ausbreitungsstärke berechnen wir für jeden Knoten n die *konsolidierte erwartete Anzahl aktivierter repräsentierter Knoten* $\hat{i}^{(p)}_n$, die Input-Aktivierung erhalten sowie die *neue erwartete Anzahl aktivierter repräsentierter Knoten* $\hat{a}^{(p)}_n$.

globaler Schritt

Anpassung der Aktivierungsverteilungsfunktionen

Während bei der Aktivierungsausbreitung exakte Aktivierungswerte für Knoten berechnet werden, bestimmen wir bei der Simulation der Aktivierungsausbreitung Näherungswerte für Aktivierungswerte von Knotenrepräsentanten. Die Information, wie viele Knoten aktiviert sind, ist in den Ergebnissen der Aktivierungsausbreitung explizit enthalten. In der Simulation erfordert dies allerdings eine gesonderte Betrachtung, da die erwarteten Aktivierungswerte diese Information nicht enthalten. Das erfordert Anpassungen aller Aktivierungsverteilungsfunktionen zur Bestimmung der Aktivierungs- und Ausbreitungsstärke. Eine detaillierte formale Anpassung von Aktivierungsverteilungsfunktionen[52] veröffentlichten wir in einem Journalbeitrag [HK17]. An dieser Stelle fokussieren wir auf die Besonderheiten, die bei der Anpassung der Aktivierungsverteilungsfunktionen für die Simulation von Bedeutung sind. Diese stellen wir im Folgenden anhand eines Beispiels für lokale Simulationsschritte vor.

Aktivierungsstärke: Die Abbildung 5.5 zeigt ein Skelett-Tripel mit den folgenden Annotationen: $A_n(\text{node\_count}, n1) = 10$, $A_n(\text{node\_count}, n2) = 8$, $A_e(\text{edge\_count}, e) = 12$, $A_s(\text{src\_rep}, e) = 8$ und $A_t(\text{tgt\_rep}, e) = 4$. Außerdem sind die Simulationsergebnisse von Puls 3 an den Knoten vermerkt. Am Beispiel von Knoten n1 entsprechen die Werte der Schätzung, dass bereits vier der zehn von n1 repräsentierten Knoten aktiviert wurden, wobei die Knoten einen erwarteten Aktivierungswert von durchschnittlich 1,2 besitzen. In der Realität ist es unwahrscheinlich, dass alle von n1 repräsentierten Knoten den gleichen Aktivierungswert besitzen. Dieser Wert ist eine Näherung an die tatsächlichen Werte und ermöglicht einen Rückschluss auf die Dimension der Aktivierungswerte. Im unteren Teil der Abbildung sind die repräsentierten Knoten und Kanten dargestellt. Jeder Simulationsschritt im Skelett-Tripel simuliert die Aktivierungsausbreitung für den zugrundeliegenden bipartiten Graphen. Mithilfe der expliziten Annotationen können wir Rückschlüsse darauf ziehen, wie viele der repräsentierten Knoten nicht mit einer repräsentierten Kante verbunden sind und daher unerreichbar bleiben und nicht an der Aktivierungsausbreitung teilnehmen. Das trifft im Fall von n1 für zwei Knoten zu. Daraus berechnen wir eine *Konnektivitätsrate* con (*engl.:* connection rate), die den Anteil der repräsentierten Knoten beschreibt, die überhaupt mit mindestens

Konnektivitätsrate

[52] In [HK17] haben wir Aktivierungsverteilungsfunktionen verwendet, die sich geringfügig von denen in Abschnitt 3.3.1 unterscheiden. Ein Unterschied ist beispielsweise, dass die Aktivierungsmodi (siehe Tabelle 3.8, Seite 93) noch nicht in den Aktivierungsverteilungsfunktionen integriert sind. Die Aktivierungsmodi erlauben die Reduktion von Aktivierungswerten, was in den Anpassungen an die Aktivierungsverteilungsfunktionen berücksichtigt werden muss.

einer Kante verbunden sind. Wir können allerdings aus dem annotierten Skelett-Tripel keine Aussagen über die genaue Zusammensetzung der repräsentierten Tripel machen, also darüber, welche Knoten miteinander verbunden sind. Daher berechnen wir, mit wie vielen Kanten ein erreichbarer Knoten durchschnittlich verbunden ist, den *Ausbreitungsfaktor* fac (*engl.:* spread factor). Die Konnektivitätsrate beträgt 0,8 für Knoten n1, da acht der zehn repräsentierten Knoten mit mindestens einer Kante verbunden sind. Da die acht erreichbaren Knoten mit 12 Kanten verbunden sind, beträgt der Ausbreitungsfaktor 1,5.

Ausbreitungsfaktor

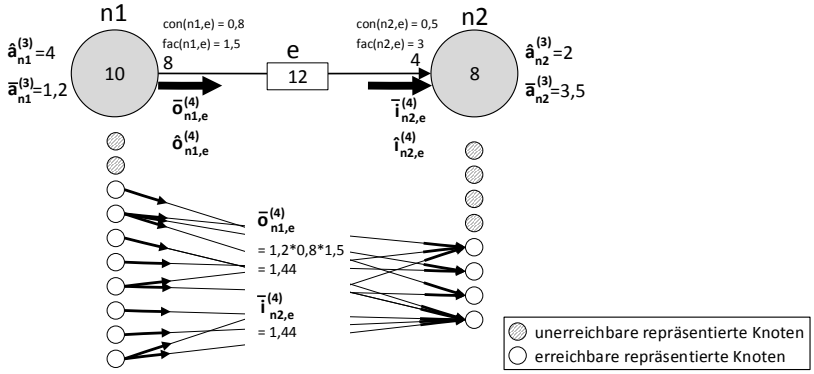


Abbildung 5.5: Beispiel eines lokalen Simulationsschrittes zur Bestimmung der Aktivierungsstärke

In den Anpassungen der Aktivierungsverteilungsfunktionen verwenden wir die Konnektivitätsrate und den Ausbreitungsfaktor, um beispielsweise einen Näherungswert für die erwartete Output-Aktivierung zu ermitteln. Im einfachsten Fall der unbeschränkten Aktivierungsausbreitung geschieht dies durch die Multiplikation des Aktivierungswertes von n1 mit Konnektivitätsrate und Ausbreitungsfaktor. Die erwartete Output-Aktivierung kann interpretiert werden als die Output-Aktivierung, die jeder Knoten durchschnittlich verteilt, im Beispiel für Knoten n1 der Wert 1,44. Ohne Kantengewichtung entspricht dieser Wert, analog zu den Aktivierungsverteilungsfunktionen der Aktivierungsausbreitung, der erwarteten Input-Aktivierung von n2. Weitere Anpassungen haben wir, sofern notwendig, für die Beschränkungen durch Verzweigung, Dämpfung oder Schwellwerte vorgenommen. Beispielsweise müssen bei der Verwendung der Verzweigungsbeschränkung für $\bar{o}^{(p)}_{n1,e}$ die Ausbreitungsfaktoren aller verbundenen Kantenrepräsentanten berücksichtigt werden.

Ausbreitungsstärke: Die Besonderheit bei der Abschätzung der Anzahl aktivierter repräsentierter Knoten ist die mögliche Überlappung im Aktivierungsstatus der Knoten. Wir müssen beispielsweise berücksichtigen, wie viele der Knoten, die Aktivierung erhalten, vermutlich bereits in der Menge der aktivierten Knoten enthalten sind. Außerdem sind kombinatorische Betrachtungen bei der Anpassung der Aktivierungsverteilungsfunktionen erforderlich. Dies zeigen wir am Beispiel der Abbildung 5.6.

Unter jedem Knotenrepräsentanten des Skelett-Tripels ist die Menge der repräsentierten Knoten als Fläche dargestellt. Den oberen Bereich der Fläche von n1 bilden die zwei repräsentierten, aber nicht erreichbaren Knoten. Darunter sind die erreichbaren Knoten dargestellt. Die erwartete Anzahl aktivierter repräsentierter Knoten des vorherigen Pulses $\hat{a}^{(3)}_{n1}$ beträgt den Wert 4. Da auch nicht verbundene Knoten aktivierte Knoten sein können, muss die Konnektivitätsrate bei der Berechnung der erwarteten Anzahl aktivierter repräsentierter Knoten, die Output-Aktivierung verteilen, berücksichtigt werden. Im Beispiel erwarten wir daher, dass 3,2

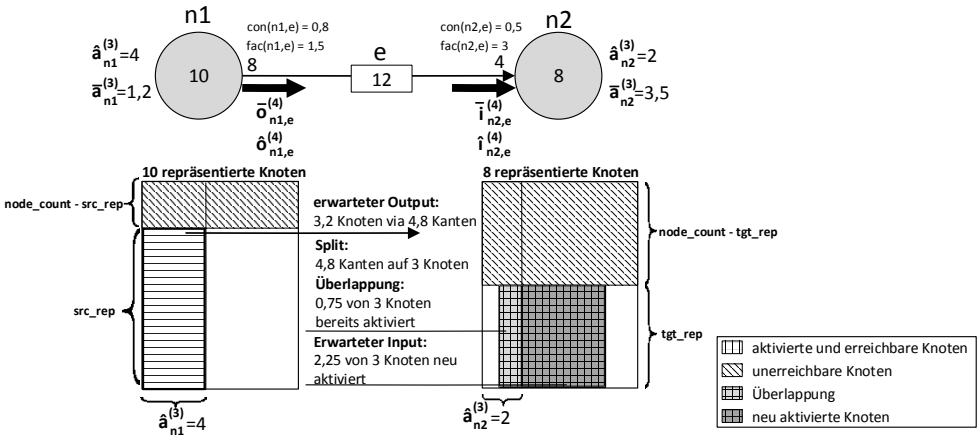


Abbildung 5.6: Beispiel eines lokalen Simulationsschrittes zur Bestimmung der Ausbreitungsstärke

der von n1 repräsentierten Knoten Aktivierung verteilen. Durch den Ausbreitungsfaktor (1,5) können wir abschätzen, dass sie die Aktivierung über 4,8 Kanten verteilen. Da wir im Skelett keine Kenntnis über die tatsächlichen Verbindungen zwischen repräsentierten Knoten und Kanten besitzen, existiert eine Vielzahl potenzieller Kombinationen, wie die 4,8 Aktivierung transportierenden Kanten mit den vier erreichbaren Knoten verbunden sein können. Daher verwenden wir die Split-Funktion in Formel 5.2 zur Abschätzung, wie viele repräsentierte Knoten von einer Anzahl Aktivierung transportierender Kanten erreicht werden. Sie basiert auf dem kombinatorischen Ansatz zur Bestimmung der durchschnittlichen Anzahl an Eimern (y), die einen Ball erhalten, wenn wir x Bälle verteilen.

$$\text{split}(x, y) := \frac{1 - (\frac{y-1}{y})^x}{1 - (\frac{y-1}{y})} \tag{5.2}$$

Durch Anwendung der Split-Funktion ergibt sich am Beispiel, dass die 4,8 eingehenden, Aktivierung transportierenden Kanten (x), die sich auf die vier erreichbaren Knoten (y) verteilen können, wahrscheinlich drei Knoten erreichen, da split(4,8, 4) = 3. Zusätzlich muss die mögliche Überlappung dieser drei Knoten mit bereits aktivierten Knoten berücksichtigt werden. Daraus ergibt sich eine erwartete Anzahl neuer repräsentierter Knoten, die Input-Aktivierung erhalten, von 2,25. In den globalen Simulationsschritten verfolgen wir ähnliche Betrachtungen zu möglichen Überlappungen. Diese können im Journalbeitrag [HK17] nachgelesen werden.

5.4.2 Experimente

In Experimenten konnten wir zum einen zeigen, dass die Ergebnisse der Simulation eine angemessene Genauigkeit im Vergleich zur Durchführung der Aktivierungsausbreitung auf den Ursprungsnetzen erreichen und zum anderen, dass mit der Simulation deutliche Zeitvorteile erreicht werden können. Wir stellen im Folgenden beide Experimente genauer vor.

Angemessenheit der Ergebnisse

Die Qualität der Simulationsergebnisse hängt eng mit der Genauigkeit ihrer Vorhersage zusammen. Zur Bewertung der Angemessenheit vergleichen wir jeweils die Ergebnisse der

Simulation eines Skelett-Tripels mit den Ergebnissen der Aktivierungsausbreitung auf dem entsprechend repräsentierten Netzwerk. Dafür haben wir 5.000 Paare aus jeweils einem Skelett-Tripel $T_S = (n1, e, n2)$ und einem dazu passenden Ursprungsnetz generiert. Wir haben große Skelett-Tripel mit $A_n(node\_count, n1) = 100$, $A_n(node\_count, n2) = 130$ und zufällig generierten Kantenannotationen $A_e(edge\_count, e)$ verwendet, um eine hohe Anzahl unterschiedlicher Fälle abzudecken. Außerdem erhielten die repräsentierten Knoten initiale Aktivierungswerte, aus denen sich die initialen Werte sowohl für die erwartete Anzahl aktivierter repräsentierter Knoten $\hat{a}_{n1}^{(p-1)}$ und $\hat{a}_{n2}^{(p-1)}$ als auch für die erwarteten Aktivierungswerte für die Knotenrepräsentanten $\bar{a}_{n1}^{(p-1)}$ und $\bar{a}_{n2}^{(p-1)}$ berechnen.

Während der Simulation eines Pulses p wurden die lokalen und globalen Simulationsschritte durchgeführt und diverse Zwischenergebnisse der angepassten Aktivierungsverteilungsfunktionen für die erwartete Output- und Input-Aktivierung beobachtet. Die relevanten Endergebnisse sind die neu berechnete erwartete Anzahl aktivierter repräsentierter Knoten $\hat{a}_{n2}^{(p)}$ und der erwartete Aktivierungswert $\bar{a}_{n2}^{(p)}$ für den Zielknoten n2.

In fünf Experimentreihen E1 bis E5 untersuchten wir verschiedene Beschränkungen, die wir in E1 bis E4 zunächst getrennt voneinander betrachteten. Dadurch war eine genauere Analyse der Qualität einzelner Anpassungen der Aktivierungsverteilungsfunktionen möglich. In E1 verwendeten wir keine Beschränkungen und untersuchten damit die pure Aktivierungsausbreitung. Die Experimentreihe E2 untersuchte Konfigurationen mit Verzweigungsbeschränkung, während sich E3 der Dämpfung und E4 den Schwellwerten widmeten. In der Experimentreihe E5 wurden alle Beschränkungen kombiniert. In Tabelle 5.1 sind für jede Experimentreihe die gemittelten Werte $\bar{a}_{n2}^{(p)}$ und $\hat{a}_{n2}^{(p)}$ aller Durchläufe aufgelistet. Wir stellen zudem die gemittelte Differenz zwischen Aktivierungsausbreitung und Simulation dar, die dem mittleren Fehler (*engl.:* mean error) entspricht, und die gemittelte prozentuale Differenz der Simulation im Verhältnis zum tatsächlichen Wert der Aktivierungsausbreitung.

Experiment	**Wert**	**Aktivierungs-ausbreitung ∅**	**Simulation ∅**	**Differenz**	**Differenz in %**
E1	$\bar{a}_{n2}^{(p)}$	529,567	529,949	0,382	0,072%
(ohne Beschränkungen)	$\hat{a}_{n2}^{(p)}$	73,801	70,313	-3,488	-4,727%
E2	$\bar{a}_{n2}^{(p)}$	303,203	303,273	0,070	0,023%
(mit Verzweigung)	$\hat{a}_{n2}^{(p)}$	73,912	70,356	-3,557	-4,812%
E3	$\bar{a}_{n2}^{(p)}$	330,468	330,159	-0,309	-0,094%
(mit Dämpfung)	$\hat{a}_{n2}^{(p)}$	74,037	70,365	-3,672	-4,959%
E4	$\bar{a}_{n2}^{(p)}$	495,542	494,037	-1,504	-0,304%
(mit Schwellwerten)	$\hat{a}_{n2}^{(p)}$	62,666	60,187	-2,479	-3,956%
E5	$\bar{a}_{n2}^{(p)}$	209,716	209,390	-0,327	-0,156%
(alle Beschränkungen)	$\hat{a}_{n2}^{(p)}$	62,527	60,152	-2,375	-3,798%

Tabelle 5.1: Experimentreihen E1 bis E5: Differenz zwischen Aktivierungsausbreitung und Simulation

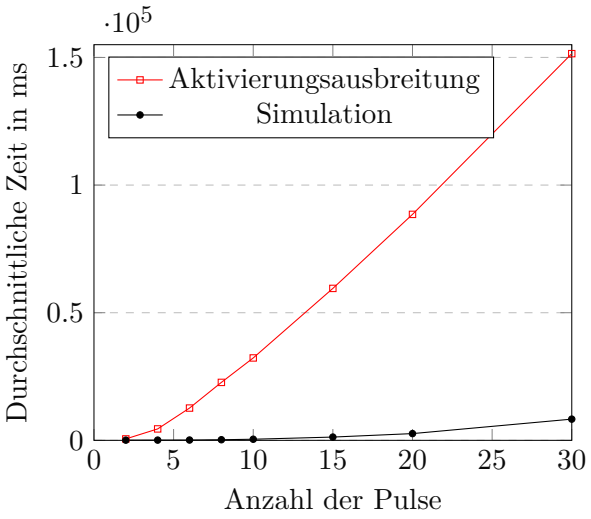


Abbildung 5.7: Durchschnittliche Laufzeit der Aktivierungsausbreitung auf einem G&R-Netz und der Simulation auf dem entsprechenden Netzwerkskelett

Die Experimente zeigen, dass die Aktivierungswerte $\bar{a}_{n2}^{(p)}$ durch die Simulation in hoher Präzision vorhergesagt werden können. Der mittlere Fehler beträgt für alle Experimentreihen deutlich unter 1%. Die Simulationsergebnisse für die Anzahl aktivierter repräsentierter Knoten $\hat{a}_{n2}^{(p)}$ erreichen nicht die Präzision der Aktivierungswerte, liegen aber dennoch in einem akzeptablen Bereich unter 5%. Im Vergleich zur Basisreihe E1 verschlechtert sich der mittlere Fehler für $\hat{a}_{n2}^{(p)}$ bei Verwendung von Verzweigung in E2 und Dämpfung in E3. Beide Parameter bewirken, dass die Knoten in der Regel weniger Aktivierung verteilen. Die Schwellwerte in E4 bewirken hingegen in der Regel, dass weniger Knoten Aktivierung verteilen. Die Unterschiede der Ungenauigkeiten für $\hat{a}_{n2}^{(p)}$ in den Experimentreihen E2 bis E4 werden indirekt bewirkt. Die Beschränkungen haben in den angepassten Aktivierungsverteilungsfunktionen keinen Einfluss auf die Berechnung der Anzahl der erwarteten aktivierten Knoten, sondern ausschließlich auf die Berechnung der erwarteten Aktivierungswerte. Die erwarteten Aktivierungswerte fließen allerdings in die Berechnung der Anzahl der erwarteten aktivierten Knoten ein, so dass sich Ungenauigkeiten hier dennoch auswirken können.

Die Ergebnisse zeigen eine sehr gute Qualität der Vorhersage von Aktivierungswerten und eine gute Qualität der erwarteten Anzahl aktivierter repräsentierter Knoten. Insbesondere die Verwendung aller untersuchter Beschränkungen in E5 führte zu akzeptablen Ergebnissen.

Zeitvorteile

Neben der Angemessenheit der Ergebnisse untersuchten wir die Zeitvorteile der Simulation. Dafür verglichen wir die Laufzeit der Aktivierungsausbreitung mit der Laufzeit der Simulation für insgesamt 72 Konfigurationen, bestehend aus verschiedenen Kombinationen der verfügbaren Beschränkungen. Ursprungsnetz ist ein G&R-Netz aus mehr als 257.000 Kanten und mehr als 61.800 Knoten. Das entsprechende semantische Netzwerkskelett enthält hingegen nur 136 Skelett-Tripel und besteht aus 45 Knotenrepräsentanten und 136 Kantenrepräsentanten. In Abbildung 5.7 ist die durchschnittliche Laufzeit der Durchläufe sowohl für die Aktivierungsausbreitung als auch für die Simulation dargestellt.

Die bessere Performanz der Simulationen ist offensichtlich. Im Durchschnitt ist die Laufzeit

nach 30 Pulsen etwa 18 Mal so schnell wie die Laufzeit der Aktivierungsausbreitung. Der Hauptgrund hierfür ist die Reduktion der notwendigen Operationen durch die deutlich reduzierte Anzahl an Knoten und Kanten im Netzwerkskelett.

Dieses Ergebnis unterstützt unsere Vermutung, dass die Kompaktheit semantischer Netzwerkskelette bei Herausforderungen im Umgang mit großen Netzen vorteilhaft sein kann. Sofern die Näherungswerte der Simulation als Ergebnisse für eine spezielle Analyse oder Anwendung ausreichen, z.B. bei Effektanalysen im Präkonfigurationskontext, kann die Verwendung des Netzwerkskelettes und der Simulation darauf zu Zeitvorteilen führen.

5.4.3 Analysepotenziale der Simulationsergebnisse

Wie bereits einleitend beschrieben, ermöglicht die Kenntnis über Einflussfaktoren und deren Effekte bei der Aktivierungsausbreitung diverse Vorab-Analysen. Ziel dieser Vorab-Analysen ist es, Konfigurationen auszuschließen, die zu unerwünschten Effekten führen, und Konfigurationen zu identifizieren, die zu erwünschten Effekten führen. Daher kann eine Untersuchung des Zusammenspiels von Einflussfaktoren und Effekten die Parametrisierung der Aktivierungsausbreitung vereinfachen. Dies gilt insbesondere, da das aufwendige Einbeziehen potenzieller Nutzer oder Experten, wie in der Evaluation des IR-Systems in dieser Arbeit, nicht immer möglich ist. Die Vorab-Analysen funktionieren unabhängig von Beurteilungen durch Nutzer, da nicht die Genauigkeit und Nützlichkeit der Ergebnisse untersucht werden, sondern das Vorhandensein definierter Effekte.

Die Verwendung der Simulationsergebnisse für derartige Analysen zielt auf die Nutzung der enormen Zeitvorteile ab. Da Näherungswerte oftmals ausreichen, erlaubt eine Vorab-Analyse auf den Simulationsergebnissen eine schnelle Auswertung. Weiterhin ist neben der Untersuchung einzelner Konfigurationen auch eine Optimierung durch evolutionäre Algorithmen möglich, die aufgrund der dafür notwendigen hohen Anzahl an Ausführungen von der schnelleren Laufzeit profitiert. Sind Näherungswerte nicht ausreichend, können Einfluss- und Effektanalysen grundsätzlich auch auf dem Ursprungsnetz durchgeführt werden.

Die Definition von Effekten und die Untersuchung von Einflussfaktoren sind potenzielle zukünftige Erweiterungen dieser Arbeit. Die Beschreibung der Effekte kann allgemeiner oder anwendungsspezifischer Natur sein. Ein Beispiel für einen Effekt, der in der Regel als unerwünscht gilt, ist die Existenz von Senken.

Senken

Ein Knoten verhält sich in einem Puls als *Senke*, wenn er zwar aktiviert ist und daher einen Aktivierungswert >0 besitzt, allerdings über keine Kante Aktivierung verteilt. Analog dazu verhält sich ein Knoten-Kanten-Paar als Senke, wenn der Knoten aktiviert ist, aber keine Aktivierung über die Kante verteilt. Einen unerwünschten Effekt stellen Senken dar, wenn sie sich über viele Pulse nicht mehr an der Aktivierungsausbreitung beteiligen. Die Grenze sollte anwendungsspezifisch definiert werden und den Durchmesser des Netzes beziehungsweise des Netzwerkskelettes berücksichtigen. Der Durchmesser ist ein Indikator für die Anzahl an Pulsen, die mindestens notwendig ist, um alle Knoten eines Netzes mindestens ein Mal zu erreichen. Der Senkeneffekt korreliert mit der Konnektivität der Knoten und bestimmten Konfigurationsparametern. Als Beispiel kann ein restriktiver Verzweigungsmodus in Kombination mit hohem Aktivierungsschwellwert bei stark verzweigten Knoten dazu führen, dass der Schwellwert schwer überschritten werden kann.

Erste Experimente zeigten die Machbarkeit der Identifikation von Senken mithilfe der Simulationsergebnisse. Die Erkenntnisse sind eine Basis für zukünftige Arbeiten an Vorab-Analysen, die eine optimale Parametrisierung ohne menschliche Beurteilungen ermöglichen sollen.

5.5 Zusammenfassung und Beitrag

In diesem Abschnitt haben wir das semantische Netzwerkskelett präsentiert. Das Netzwerkskelett ist eine Abstraktion semantischer Netze, die gleichartige Teile des Netzes versteckt, die strukturellen Eigenschaften dennoch explizit abbildet. Das ermöglicht eine kompakte und übersichtliche Darstellung semantischer Netze, was insbesondere den Umgang mit großen Netzen erleichtert. Mithilfe der Abstraktion ließ sich beispielsweise ein G&R-Netz aus ursprünglich 240.733 Tripel auf nur noch 87 Skelett-Tripel reduzieren und somit in dieser Arbeit abbilden (siehe Abbildung 5.2, Seite 180).

Als weitere Verwendungsmöglichkeit haben wir gezeigt, wie Netzwerkskelette als Hilfsmittel bei der Konfiguration der Aktivierungsausbreitung dienen können. Dafür haben wir ein Näherungsverfahren zur Simulation der Aktivierungsausbreitung auf dem Netzwerkskelett entwickelt, mit dem sich die erwartete Ausbreitungsstärke und die erwartete Aktivierungsstärke in jedem Simulationsschritt bestimmen lassen. Die Simulation profitiert von der wesentlich geringeren Größe der Netzwerkskelette im Vergleich zu den größeren Ursprungsnetzen und weist in Experimenten neben einer angemessenen Genauigkeit der Vorhersage auch wesentliche Zeitvorteile auf. Wir haben beschrieben, welche Analysepotenziale die Simulationsergebnisse besitzen, beispielsweise um bereits vorab ungeeignete Konfigurationen zu identifizieren, die zu unerwünschten Effekten führen. Diese Grundidee haben wir am Beispiel von Senken verdeutlicht und damit eine Basis für die Definition und Identifikation weiterer Effekte und für weitere Vorab-Analysen gelegt.

6 | Zusammenfassung und Ausblick

In diesem Kapitel fassen wir die Beiträge und Ergebnisse dieser Arbeit zusammen und geben einen Ausblick auf zukünftige Arbeiten.

6.1 Zusammenfassung der Ergebnisse

Erfahrungs- und expertenbasierte Aufgaben wie die Erstellung von Gefährdungs- und Risikoanalysen (G&Rs) sind meist mit großem Aufwand verbunden. Die Automatisierung, beispielsweise durch Wiederverwendung von Erfahrungswissen, stellt eine große Herausforderung dar. Sie bietet aber auch Chancen, um Aufwände einzusparen. In dieser Arbeit haben wir ein System entwickelt, das dazu in der Lage ist, Wissen früherer G&Rs zu archivieren, es zielgerichtet zu durchsuchen und bei der Erstellung neuer G&Rs zur Verfügung zu stellen. Dadurch haben wir die Wiederverwendung von Wissen ermöglicht, was die Konsistenz von G&Rs steigern kann und eine effizientere Durchführung unterstützt. Wir haben eine automatisierte und in ein Entwicklungswerkzeug integrierbare Lösung vorgestellt, so dass die gewonnene Unterstützung nicht mit zusätzlichem Aufwand verbunden ist und Domänenexperten entlastet werden. Das ist ein zentraler Punkt für die Nutzerakzeptanz. Wie bereits im Einführungskapitel dargestellt, leistet diese Arbeit drei große Beiträge.

- Beitrag A: Entwicklung eines Rahmenwerks mit generischen und konfigurierbaren Lösungen für den automatisierten Aufbau domänenspezifischer IR-Systeme
- Beitrag B: Methode zum Aufbau domänenspezifischer IR-Systeme unter Verwendung des Rahmenwerks sowie zusätzliche Hilfsmittel wie das semantische Netzwerkskelett für die Entwickler
- Beitrag C: Aufbau eines IR-Systems für die Domäne der G&Rs auf Basis industrieller Daten und Nachweis seiner Güte

Im Folgenden fassen wir die wesentlichen Teilergebnisse für jeden dieser Beiträge zusammen.

A | Entwicklung eines Rahmenwerks mit generischen und konfigurierbaren Lösungen für den Aufbau domänenspezifischer IR-Systeme

Der Schwerpunkt dieser Arbeit liegt im praktischen Anwendungsfeld der funktionalen Sicherheit im Automobilbereich, in dem wir die erfahrungsbasierte Durchführung von G&Rs als zentralen Anwendungsfall betrachten. Da ähnliche Anwendungsfälle auch in anderen Domänen und für andere Artefakte existieren, haben wir in der Auseinandersetzung mit dem konkreten Anwendungsfall Lösungen entwickelt, die universell sind und prinzipiell auf andere Domänen übertragen werden können.

Generischer Rahmen: Der erste Beitrag dieser Arbeit besteht in der Entwicklung einer Lösungsstrategie, die den generischen Rahmen für die Lösung der untersuchten Herausforderungen und Zielvorstellungen absteckt. Wir haben den zu lösenden Anwendungsfall als IR-Problem betrachtet und ein IR-System entwickelt, dessen assoziativ agierende Bestandteile harmonisch aufeinander abgestimmt sind. Für dieses IR-System haben wir einen grundsätzlichen Aufbau und eine Arbeitsweise entwickelt, die in einem Rahmenwerk generische und konfigurierbare Lösungen vereinen. Diese sind so konzipiert, dass sie für den Aufbau domänenspezifischer IR-Systeme genutzt werden können. Die drei Bestandteile sind, wie

in IR-Systemen üblich, die Wissensrepräsentation, die Retrievalstrategie und die Ergebnisaufbereitung, die wir zusätzlich mit optionalen Erklärungen der Ergebnisse angereichert haben. Das Besondere unserer Lösungen ist, dass wir den Aufbau neuer domänenspezifischer IR-Systeme weitestgehend automatisieren konnten, so dass deren Entwicklung mit geringem Aufwand möglich ist. Dafür ist es erforderlich, dass ein IR-Problem vorliegt, welches von der Identifikation relevanter Elemente im Datensatz und deren Wiederverwendung profitieren kann. Eine weitere Voraussetzung ist, dass die Daten entsprechend strukturiert sind, um vom datenmodellbasierten Ansatz zu profitieren. Die Bestandteile werden im Folgenden einzeln erläutert.

Wissensrepräsentation: Wir haben eine datenmodellbasierte Erzeugung semantischer Basisnetze vorgestellt. Damit haben wir uns zunutze gemacht, dass die Daten strukturiert sind, und konnten direkte Transformationen verwenden, um die strukturierten Daten in semantische Netze zu überführen. Dadurch konnten wir die bisher sehr aufwendige Wissensrepräsentation weitestgehend automatisieren. Zur Aufwertung der Basisnetze haben wir zusätzliche Techniken zu ihrer semantischen Anreicherung vorgestellt, die eine Extraktion von Keywords aus Literalen mit komplexen Textinhalten erlauben und den Umgang mit bekannten Synonymen und Akronymen ermöglichen. Außerdem haben wir Techniken zur Reduktion der Netze vorgestellt, mit denen ein Netz auf seinen (erweiterten) semantischen Kern beschränkt wird. Die verschiedenen Anreicherungs- und Reduktionsstufen ermöglichen die domänenspezifische Anpassung der Netze. In der Evaluation konnten wir zeigen, dass die semantische Anreicherung und die Reduktion der Netze maßgeblich zur Qualität der Ergebnisse beitragen (siehe Evaluationsfrage EQ3a, Seite 168). Zusätzlich haben wir ein Konzept zur graphbasierten Repräsentation strukturierter Anfragen vorgestellt, welches auf die ebenfalls graphbasierte Repräsentation des Wissens abgestimmt ist. Dieses erlaubt uns bei Ausgabe der Ergebnisse eine suchzielspezifische Eingrenzung.

Retrievalstrategie: Mithilfe der Aktivierungsausbreitung als graphbasiertes Retrievalverfahren konnten wir die gegenseitige Relevanz von Knoten im semantischen Netz beurteilen und dadurch relevante Zusammenhänge erkennen. Wir haben die Aktivierungsverteilung durch bekannte Aktivierungsverteilungsfunktionen reproduzierbar formalisiert und Beschränkungen konzeptualisiert. Entstanden ist eine erweiterbare Liste verschiedener Modi zur Konfiguration der Beschränkungen bestehend aus Aktivierungs-, Dämpfungs- und Verzweigungsmodi. Zudem haben wir die Aktivierungsausbreitung um ein neuartiges Konzept erweitert, das nicht die Berechnung von Aktivierungswerten, sondern das Ausbreitungsverhalten im Graphen steuert und damit strukturelle Eigenheiten der semantischen Netzwerke berücksichtigt. Das Ergebnis ist eine erweiterbare Liste verschiedener Sende- und Ausbreitungsmodi. Durch die Konzeptualisierung und Formalisierung bestehender Techniken haben wir unseren Ansatz reproduzierbar gemacht. Die Modi machen den Parameterraum übersichtlich und vereinfachen die Konfiguration. Die Verwendung der Ausbreitungsmodi hat sich in der Evaluation als sinnvoll erwiesen (siehe Evaluationsfrage EQ3b, Seite 169).

Ergebnisse und Erklärungen: Bei der Aufbereitung der Ergebnisse profitieren wir von unserem datenmodellbasierten Ansatz, der uns eine suchzielspezifische Filterung im semantischen Netz erlaubt. Das Ergebnis des IR-Systems ist eine dem Typ des Suchziels entsprechende, nach Relevanz sortierte und gekürzte Ergebnisliste. Wir haben zudem ein Verfahren zur automatisierten Erklärung der Ergebnisse des IR-Systems vorgestellt. Das Verfahren basiert auf einem ebenfalls graphbasierten Ansatz, mit dem auf Basis der Aktivierungshistorie re-

konstruiert wird, welche Aussagen im Netz maßgeblich an der Aktivierung eines Ergebnisses beteiligt waren. Die Verfügbarkeit von Erklärungen sorgt für Transparenz, was wiederum Vertrauen in das System und dessen Ergebnisse schafft und daher die Nutzerakzeptanz erhöht.

Der Wert unseres ersten Beitrags liegt im harmonischen Zusammenspiel der einzelnen Bestandteile und darin, dass diese aufgrund des datenmodellbasierten Ansatzes weitestgehend automatisiert erstellt werden können. Insbesondere haben wir damit eine große Hürde in der Wissensrepräsentation überwunden, da ein manueller Aufbau der Netze und eine kontinuierliche Wartung entfällt.

B | Methode zum Aufbau domänenspezifischer IR-Systeme unter Verwendung des Rahmenwerks

Als zweiten Beitrag haben wir eine aus vier Schritten bestehende Methode präsentiert, welche das Vorgehen beim Aufbau eines domänenspezifischen IR-Systems aus dem generischen Rahmenwerk beschreibt. Mithilfe der Methode trifft der Entwickler eines domänenspezifischen IR-Systems nach initialer Analyse von Daten und Domäne sukzessive Entscheidungen für die Auswahl und Konfiguration der Komponenten sowie für deren Optimierung. Das vorhandene generische Rahmenwerk und die darin enthaltenen Lösungen vereinfachen den Aufbau neuer IR-Systeme erheblich.

Die Evaluation zeigt, dass die Auswahl von Netzvarianten und von Konfigurationsparametern der Aktivierungsausbreitung im untersuchten Anwendungsfall großen Einfluss auf die Qualität der Ergebnisse haben (siehe Evaluationsfrage EQ2, Seite 144). Für den Entwickler domänenspezifischer IR-Systeme besteht die Schwierigkeit vor allem darin, geeignete Konfigurationen im Vorfeld festzulegen. Wir haben in unserer Evaluation bereits Erkenntnisse über geeignete Kombinationen von Konfigurationen für das G&R-IR-System identifiziert (siehe Evaluationsfrage EQ3c, Seite 169). Die Auswahl guter Konfigurationen für neue domänenspezifische IR-Systeme verbleibt aber eine anspruchsvolle und mitunter aufwendige Aufgabe für den Entwickler des IR-Systems. Daher haben wir mit dem semantischen Netzwerkskelett ein Hilfsmittel präsentiert, welches durch Abstraktion eine kompakte und übersichtliche Darstellung großer, oft unübersichtlicher Netze erlaubt. Durch die Visualisierung vereinfachen wir den Umgang mit großen Netzen nicht nur für das Themengebiet IR. Außerdem haben wir ein Näherungsverfahren vorgestellt, mit dem die Aktivierungsausbreitung auf dem Netzwerkskelett simuliert werden kann. Für die Simulation konnten wir in ersten Experimenten eine angemessene Genauigkeit und einen Zeitvorteil im Vergleich zur Aktivierungsausbreitung zeigen. Daher eignet sich die Simulation beispielsweise für die Detektion von Konfigurationen, die zu unerwünschten Effekten führen. Diese können dann frühzeitig ausgeschlossen werden. Die Analysepotenziale haben wir am Beispiel der Detektion von Senken erläutert. Damit konnten wir eine Basis für die Definition weiterer unerwünschter oder erwünschter Effekte sowie für deren Identifikation schaffen.

C | Aufbau eines IR-Systems für die Domäne der Gefährdungs- und Risikoanalysen

Der dritte Beitrag der Arbeit ist der Aufbau eines IR-Systems für G&Rs, das auf den Lösungen des generischen Rahmenwerks basiert und den Schritten der Methode zum Aufbau domänenspezifischer IR-Systeme folgt. Damit führen wir die Machbarkeit der Methode und des Lösungsansatzes vor. Genau genommen hat die Auseinandersetzung mit dem konkreten Anwendungsfall der Unterstützung von G&Rs maßgeblich zur Entwicklung der unterschiedli-

chen alternativen Lösungen beigetragen, die das generische Rahmenwerk ausmachen. Für den Aufbau der semantischen Netze haben wir industrielle Daten aus 155 abgeschlossenen G&Rs verwendet und damit Netze in der Größenordnung zwischen circa 80.000 und 277.000 Tripeln erhalten. Das Ergebnis ist ein domänenspezifisches IR-System, das für Informationsbedürfnisse, die während der Durchführung von G&Rs auftreten, relevante Zusammenhänge im vorhandenen Datensatz erkennt und dem Nutzer als Empfehlung zur Verfügung stellt.

Prototypische Umsetzung: Wir haben alle vorgestellten generischen Konzepte und Techniken prototypisch in einem eigenständigen IR-System implementiert. Dafür haben wir das Entwicklungswerkzeug nachgebildet und die Generierung von Empfehlungen so in das Werkzeug integriert, dass die Anforderung von Empfehlungen und Erklärungen für die Ergebnisse direkt über das Kontextmenü angesteuert werden können. Dadurch erfordert die Nutzung keinen manuellen Zusatzaufwand und keine relevanten Vorkenntnisse. Diese Umsetzung ist im praktischen Anwendungsfeld auf gute Nutzerakzeptanz getroffen, da die zusätzliche Integration des IR-Systems das etablierte Vorgehen nicht stört, sondern minimalinvasiv zusätzliche Unterstützung anbietet.

Qualitative Bewertung der Ergebnisse des IR-Systems: Die Evaluation hat gute bis sehr gute Werte für die Genauigkeit und Nützlichkeit der Ergebnisse des IR-Systems gezeigt (siehe Evaluationsfrage EQ4, Seite 172). Für die qualitative Bewertung haben wir optimierte Konfigurationen verwendet, die wir durch die Anwendung von evolutionären Algorithmen erzeugt haben. Zum einen haben wir eine optimierte anfrageübergreifende Konfiguration erhalten, mit der wir für die acht untersuchten Anfragen eine durchschnittliche Fitness von circa 0,76 erreichen konnten. Das entspricht einer erwarteten Genauigkeit von 0,66 und einer Nützlichkeit von circa 0,85. Zum anderen haben wir für jede der acht Anfragen eine eigene optimierte Konfiguration erzeugt. Mit diesen anfragespezifischen Konfigurationen konnten wir eine durchschnittliche Fitness von circa 0,84 erreichen mit einer erwarteten Genauigkeit von 0,76 und einer Nützlichkeit von circa 0,93. Das Gesamtbild zeigt hohe Genauigkeiten am Beginn der Ergebnislisten und gute Trefferquoten in ihrem Verlauf. Das entspricht den Erwartungen, die Nutzer an die Ergebnisse eines IR-Systems haben. Es wurde deutlich, dass die Güte der Ergebnisse für die anfragespezifischen Konfigurationen höher ist (11,6%) als für die anfrageübergreifende Konfiguration. Da in der praktischen Anwendung Netzvarianten und Konfigurationsparameter festgelegt werden bevor die Anfragen bekannt sind, ist die Verwendung anfragespezifischer Konfigurationen vorerst nicht praktikabel. Die Ergebnisse legen allerdings die Vermutung nahe, dass ihre Qualität von spezifischen Konfigurationen für die verschiedenen Anfragetypen profitieren könnten.

6.2 Ausblick

Im Folgenden stellen wir Anknüpfungspunkte und Ideen für weitere Forschung im Themengebiet dieser Arbeit vor.

Weiterentwicklung der Bestandteile des generischen IR-Systems

Das vorgestellte IR-System stellt eine gute Grundlage dar und bietet Möglichkeiten für Erweiterungen, die wir im Rahmen dieser Arbeit nicht umgesetzt haben, da sie außerhalb unseres Fokus lagen. Wir stellen daher einige Verfeinerungen vor, von denen die einzelnen Bestandteile profitieren könnten.

Aufbau der Netze: Für eine weitere Aufwertung der Netze können weitere NLP-Techniken wie Lemmatisierung zum Einsatz kommen. Es ist ebenfalls möglich, den gesamten Textkorpus aller Literale nicht nur linguistisch sondern auch statistisch auszuwerten. Dadurch könnte die Qualität und Aussagekraft der Keywords durch eine höhere Relevanz und Domänenspezifizität verbessert werden. Es könnten statistische Filter zum Einsatz kommen ähnlich dem Ansatz, den wir für die automatisierte Extraktion von Glossartermen in großen Anforderungsspezifikationen [GCHV18] präsentierten. Dort haben wir zur Bewertung der Relevanz von Termen etablierte Maße wie das tf-idf-Maß für Vorkommenshäufigkeit und inverse Dokumenthäufigkeit (*engl.:* term frequency and inverse document frequency) verwendet. Für die Bewertung der Domänenspezifizität lassen sich die Terme beispielsweise mit großen, allgemeinen Textkorpora vergleichen.

Der Umgang mit Wörtern oder Wortgruppen gleicher semantischer Bedeutung stellt eine große Herausforderung im Aufbau einer domänenspezifischen Wissensrepräsentation dar. In dieser Arbeit haben wir ausschließlich explizite und durch Domänenexperten verfügbar gemachte Synonymlisten in die Netze integriert. Zusätzlich zu den domänenspezifischen Synonymlisten lassen sich auch bekannte Synonyme aus anderen Domänen heranziehen. Als Beispiel sei hier GermaNet [HF97, HH10] erwähnt, eine in verschiedenen Forschungsprojekten entstandene deutsche Variante des WordNet. Unter anderem enthält GermaNet über sogenannte SynSets Informationen über Synonyme. Die Sichtung der Synonymdaten sowie Experimente mit der Integration in die G&R-Netze haben allerdings ergeben, dass die darin enthaltenen Synonyme aufgrund fehlender Domänenspezifizität keine Bereicherung für das domänenspezifische Netz darstellten. Durch statistische Filterung nach Domänenspezifizität, ähnlich dem oben erläuterten Ansatz für Keywords, könnten relevante SynSets identifiziert und nutzbar gemacht werden. Eine automatisierte Extraktion von gleicher Bedeutung aus einem Textkorpus ist Gegenstand aktueller Forschung. Eine Möglichkeit ist der Einsatz von Word Embedding. Die Grundidee ist, Wörter eines Textkorpus vektorbasiert darzustellen und über eine Berechnung der Nähe von Vektoren im Vektorraum auf die Nähe der repräsentierten Wörter und auch auf Synonyme zu schließen. Dafür werden oftmals verschiedene Ansätze des maschinellen Lernens wie neuronale Netze eingesetzt [LVDv16].

Anfragen: Anfragen werden im vorgestellten Konzept durch den Ausgangspunkt der semantischen Suche und das Suchziel repräsentiert. Der Ausgangspunkt entspricht im Suchalgorithmus einem oder mehreren Startknoten. Üblicherweise sind das jene Instanzknoten, für welche das Informationsbedürfnis formuliert wird, und zusätzliche Knoten, die im Kontext des Informationsbedürfnisses eine gewisse Bedeutung besitzen. Das kann beispielsweise die Navigationshistorie im Entwicklungswerkzeug, relevante Keywords oder ein expliziter Bezug zum Item sein. Bisher erlauben die vorgestellten Konzepte zwar die Integration zusätzlichen Kontextes in die Anfrage, dieser muss allerdings manuell identifiziert und hinzugefügt werden. Die Entscheidung für eine automatisierte, kontextbezogene Ausweitung der Ausgangspunkte wird vom Entwickler des IR-Systems nach Analyse der Domäne und Daten getroffen. Hier ist ein systematisches Verfahren zur automatischen Identifikation und Integration von angemessenem Kontext in die Anfrage denkbar. Das kann sowohl anfragespezifisch als auch für die verschiedenen Anfragetypen eine sinnvolle Ergänzung sein. Dafür ist vorerst eine Untersuchung notwendig, welche Art von Kontext besonders hilfreich ist, um den Suchalgorithmus sinnvoll zu unterstützen und die Qualität der Ergebnisse zu steigern.

Aktivierungsausbreitung: Diverse Verfeinerungen könnten die semantische Suche optimieren. Beispielsweise haben wir keinen Schwerpunkt auf die Auswahl geeigneter Kanten-

gewichte gelegt. Interessant wäre, ob und inwiefern eine optimierte Kantengewichtung die Ergebnisse verbessern kann. Dafür könnten Techniken entwickelt werden, die die Relevanz der Kanten automatisiert bestimmen.

Das vorgestellte IR-System verzichtet auf die Verwertung von Nutzerfeedback. Ein Grund ist der kleine Nutzerkreis bei der Erstellung von G&Rs. Außerdem birgt insbesondere aktives Nutzerfeedback aufgrund des Mehraufwands einer zusätzlichen Interaktion mit der Feedback-Komponente die Gefahr geringer Nutzerakzeptanz [Fuh13]. Bei Anwendung in anderen Domänen mit größerem Nutzerkreis könnte Relevanzfeedback die Qualität der Ergebnisse verbessern. Insbesondere passives Feedback, beispielsweise durch das Tracken des Nutzerverhaltens und das Nachverfolgen von akzeptierten und modifizierten Ergebnissen, könnte Aufschluss über die Nützlichkeit der Ergebnisse geben. Relevanzfeedback eröffnet Möglichkeiten, die Ergebnisse durch adaptives IR stetig zu verbessern und besser an die Nutzer anzupassen [MWH00].

Integration in Entwicklungswerkzeuge

In dieser Arbeit haben wir die Konzepte in einem eigenständigen Prototyp umgesetzt. Eine praktische Anschlussarbeit könnte die Integration in das Entwicklungswerkzeug ANSYS medini analyze darstellen. Die direkte Anbindung an die Datenbank erlaubt die Verwendung direkter Mappings der relationalen Daten zu RDF-Graphen, beispielsweise durch Werkzeuge wie die D2RQ-Plattform [D2RQ], teilweise ohne diese replizieren zu müssen.

Neben der Unterstützung von G&Rs erlaubt der generische Ansatz die Übertragung auf andere Domänen. Voraussetzung ist, dass diese ein ähnliches IR-Problem besitzen, welches von der Identifikation relevanter Elemente im Datensatz und deren Wiederverwendung profitieren kann. Um vom datenmodellbasierten Ansatz zu profitieren, ist es zudem notwendig, dass die Daten entsprechend strukturiert sind. Ein mögliches Anwendungsgebiet in der Funktionssicherheit ist beispielsweise die Sicherheitsanalyse FMEA. Der Einsatz ist allerdings nicht auf Funktionssicherheit beschränkt und auch außerhalb des Themenbereichs denkbar.

Konfiguration der Aktivierungsausbreitung

Beim Aufbau eines domänenspezifischer IR-Systems muss dessen Entwickler unter anderem Entscheidungen bezüglich der Auswahl der Netzvarianten und der Konfigurationsparameter für die Aktivierungsausbreitung treffen. Die Konfiguration der Aktivierungsausbreitung besitzt, wie wir zeigen konnten, starken Einfluss auf die Qualität der Ergebnisse. Wir sehen zwei Familien von Einflussfaktoren, die in ihrem Zusammenspiel Effekte auf die Ergebnisse und deren Qualität besitzen. Diese sind zum einen die Netzeigenschaften und zum anderen die Konfigurationsparameter, die einen komplexen Parameterraum aufspannen. In unserer Evaluation konnten wir mithilfe menschlicher Experteneinschätzungen und evolutionärer Algorithmen zeigen, welche Kombinationen der Konfigurationsparameter sich für bestimmte Netzvarianten besonders eignen. In der Realität ist dieser Ansatz vor Inbetriebnahme eines IR-Systems sehr aufwendig, zumal die Netze nicht statisch sind, sondern sich durch neue Daten ständig verändern können. Daher ist unsere Vision eine automatisierte Konfiguration der Aktivierungsausbreitung. Idealerweise sollten sich Konfigurationen an die Netzeigenschaften und beispielsweise auch an die Eigenschaften der Anfrage anpassen. Ein interessanter Ansatzpunkt hierfür könnten Erkenntnisse aus der Untersuchung von Netzwerkeigenschaften, beispielsweise aus der sozialen Netzwerkanalyse, liefern. Zukünftige Arbeiten, die auf eine automatisierte Konfiguration abzielen, könnten zunächst durch eine systematische Analyse

von Einflussfaktoren und deren Effekte auf die Ergebnisse Richtlinien zur Konfiguration ableiten. Stehen ausreichend aussagekräftige Richtlinien zur Verfügung, so ist eine automatische, selbstkonfigurierende Aktivierungsausbreitung gemäß dieser Richtlinien möglich. Diese sollte derart erfolgen, dass die Aktivierungsausbreitung nicht im Vorfeld konfiguriert werden muss, sondern sich an wechselnde und wachsende Netze sowie veränderte Anfragen anpasst.

Literaturverzeichnis

[Á12] José María Álvarez Rodríguez. *Sistemas y servicios informáticos para internet.* Dissertation, Universidad de Oviedo, 2012. Zitiert auf Seite 49.

[And83] John R. Anderson. A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22(3), Seiten 261–295, 1983. Zitiert auf Seite 32.

[ÁPJ+11] José María Álvarez, Luis Polo, Weena Jimenez, Pablo Abella, Jose Emilio Labra. Application of the spreading activation technique for recommending concepts of well-known ontologies in medical systems. In: *Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine (BCB)*, Seiten 626–635, New York, NY, USA, 2011. ACM Press. Zitiert auf Seite 49.

[AS11] Khalid M. Albarrak, Edgar H. Sibley. A survey of methods that transform data models into ontology models. In: *IEEE International Conference on Information Reuse & Integration (IRI)*, Seiten 58–65. IEEE SMCS, Piscataway, NJ, USA, 2011. Zitiert auf Seite 30.

[BA05] Hannes Bohring, Sören Auer. Mapping XML to OWL ontologies. In: *Proceedings of 13. Leipziger Informatik-Tage (LIT)*, Seiten 147–156. Lecture Notes in Informatics (LNI), 2005. Zitiert auf den Seiten 30, 31 und 212.

[BB04] Albert-László Barabási, Eric Bonabeau. Skalenfreie Netze. *Spektrum der Wissenschaft*, 2004(Juli), Seiten 62–69, 2004. Zitiert auf Seite 71.

[BBK+09] Michael R. Berthold, Ulrik Brandes, Tobias Kötter, Martin Mader, Uwe Nagel, Kilian Thiel. Pure spreading activation is pointless. In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM)*, Seiten 1915–1918, New York, NY, USA, 2009. ACM. Zitiert auf Seite 90.

[BDM03] Helmut Berger, Michael Dittenbach, Dieter Merkl. Activation on the move: Querying tourism information via spreading activation. In: *Database and Expert Systems Applications*, Volume 2736 of *Lecture Notes in Computer Science*, Seiten 474–483. Springer Berlin Heidelberg, 2003. Zitiert auf Seite 49.

[BF16] Stephan Baumgart, Joakim Fröberg. Functional safety in product lines – A systematic mapping study. In: *Proceedings of the 42th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, Seiten 313–322. IEEE Computer Society, 2016. Zitiert auf Seite 48.

[BGH] Bundesgerichtshof. *Urteil vom 16.06.2009, Aktenzeichen VI ZR 107/08.* Zitiert auf Seite 36.

[BH08] Frada Burstein, Clyde W. Holsapple (Hrsg.). *Handbook on decision support systems 1: Basic themes.* Springer Berlin Heidelberg, 2008. Zitiert auf Seite 14.

[BHCH17] Kristian Beckers, Dominik Holling, Isabelle Côté, Denis Hatebur. A structured hazard analysis and risk assessment method for automotive systems – A descriptive study. *Reliability Engineering & System Safety*, 158, Seiten 185–195, 2017. Zitiert auf Seite 47.

[BHFH13] Kristian Beckers, Maritta Heisel, Thomas Frese, Denis Hatebur. A struc-

tured and model-based hazard analysis and risk assessment method for automotive systems. In: *IEEE 24th International Symposium on Software Reliability Engineering (ISSRE)*, Seiten 238–247. IEEE Computer Society, 2013. Zitiert auf Seite 47.

[BLB08] Jürgen Bortz, Gustav A. Lienert, Klaus Boehnke. *Verteilungsfreie Methoden in der Biostatistik*. Springer-Lehrbuch. Springer Heidelberg, 3. Auflage, 2008. Zitiert auf den Seiten 23 und 24.

[BLR14] Benjamin Behringer, Martina Lehser, Steffen Rothkugel. Towards feature-oriented fault tree analysis. In: *IEEE 38th International Computer Software and Applications Conference Workshops (COMPSACW)*, Seiten 522–527, Piscataway, NJ, USA, 2014. IEEE Computer Society. Zitiert auf Seite 48.

[Bör15] Josef Börcsök. *Funktionale Sicherheit: Grundzüge sicherheitstechnischer Systeme*. VDE-Verlag, Berlin, 4. Auflage, 2015. Zitiert auf den Seiten 36 und 37.

[BP81] R. L. Brennan, D. J. Prediger. Coefficient kappa: some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3), Seiten 687–699, 1981. Zitiert auf den Seiten 22 und 23.

[Bro06] Manfred Broy. Challenges in automotive software engineering. In: *Proceedings of the 28th International Conference on Software Engineering (ICSE)*, New York, NY, USA, 2006. ACM. Zitiert auf Seite 1.

[BYR99] Ricardo Baeza-Yates, Berthier de Araújo Neto Ribeiro (Hrsg.). *Modern information retrieval*. ACM Press. Addison-Wesley [u.a.], 1. Auflage, 1999. Zitiert auf den Seiten 3, 11 und 12.

[CC04] Gerardo Canfora, Luigi Cerulo. A taxonomy of information retrieval models and tools. *Journal of Computing and Information Technology*, 12(3), Seiten 175–194, 2004. Zitiert auf den Seiten 10 und 11.

[CEE+10] Kai-Uwe Carstensen, Christian Ebert, Cornelia Ebert, Susanne J. Jekat, Ralf Klabunde, Hagen Langer (Hrsg.). *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Spektrum Akademischer Verlag, Heidelberg, 3. Auflage, 2010. Zitiert auf den Seiten 12 und 80.

[CF90] Domenic V. Cicchetti, Alvan R. Feinstein. High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, 43(6), Seiten 551–558, 1990. Zitiert auf Seite 22.

[CL75] Allan M. Collins, Elizabeth F. Loftus. A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), Seiten 407–428, 1975. Zitiert auf den Seiten 32 und 49.

[CL00] Fabio Crestani, Puay Leng Lee. Searching the web by constrained spreading activation. *Information Processing & Management*, 36(4), Seiten 585–605, 2000. Zitiert auf Seite 49.

[Cle67] Cyril Cleverdon. The Cranfield tests on index language devices. *Aslib Proceedings*, 19(6), Seiten 173–194, 1967. Zitiert auf Seite 15.

[CLEF] CLEF Initiative. *Conference and Labs of the Evaluation Forum*, `http://www.clef-initiative.eu/`, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 16.

[Coh60] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational*

and Psychological Measurement, 20(1), Seiten 37–46, 1960. Zitiert auf den Seiten 22 und 23.

[ConQAT] CQSE. *ConQAT – Continuous Quality Assessment Toolkit*, https://www.cqse.eu/en/blog/conqat-end-of-life/, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 56.

[Cre97] Fabio Crestani. Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review*, 11(6), Seiten 453–482, 1997. Zitiert auf den Seiten 4, 9, 11, 13, 26, 32, 33, 34, 35, 49, 50 und 212.

[D2RQ] D2RQ. *The D2RQ Platform – Accessing Relational Databases as Virtual RDF Graphs*, http://d2rq.org/, Letzter Zugriff: 14.01.2019. Zitiert auf den Seiten 30 und 196.

[dBM+15] André L. de Oliveira, Rosana T. V. Braga, Paulo C. Masiero, Yiannis Papadopoulos, Ibrahim Habli, Tim Kelly. Supporting the automated generation of modular product line safety cases. In: *Theory and Engineering of Complex Systems and Dependability*, Advances in Intelligent Systems and Computing, Seiten 319–330. Springer International Publishing, Cham, 2015. Zitiert auf Seite 48.

[dBM+18] André L. de Oliveira, Rosana T. V. Braga, Paulo C. Masiero, Yiannis Papadopoulos, Ibrahim Habli, Timothy Patrick Kelly. Variability management in safety-critical software product line engineering. In: *Proceedings of the 17th International Conference on Software Reuse (ICSR)*, Lecture Notes in Computer Science, Seiten 3–22. Springer, 2018. Zitiert auf Seite 48.

[Den12] Andreas Dengel (Hrsg.). *Semantische Technologien: Grundlagen – Konzepte – Anwendungen*. Spektrum Akademischer Verlag, 1. Auflage, 2012. Zitiert auf den Seiten 26 und 28.

[DINEN61508-1] Deutsches Institut für Normung. *Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme – Teil 1: Allgemeine Anforderungen (DIN EN 61508-1 (VDE 0803-1):2011-2*, 2011. Zitiert auf Seite 36.

[DINEN61508-Bbl] Deutsches Institut für Normung. *Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme – Teil 0: Funktionale Sicherheit und die IEC 61508 (DIN EN 61508 Beiblatt 1 (VDE 0803 Beiblatt 1):2005-10*, 2011. Zitiert auf den Seiten 36 und 37.

[DKL+10] Alan Dix, Akrivi Katifori, Giorgos Lepouras, Costas Vassilakis, Nadeem Shabir. Spreading activation over ontology-based resources: From personal context to web scale reasoning. *International Journal of Semantic Computing*, 4(1), Seiten 59–102, 2010. Zitiert auf Seite 49.

[DL09] Josh Dehlinger, Robyn R. Lutz. Evaluating the reusability of product-line software fault tree analysis assets for a safety-critical system. In: *Formal Foundations of Reuse and Domain Engineering*, Volume 5791 of *Lecture Notes in Computer Science*, Seiten 160–169. Springer Berlin, 2009. Zitiert auf Seite 48.

[DS05] M. Duerst, M. Suignard. *RFC 3987: Internationalized Resource Identifiers (IRIs)*, 2005, http://www.ietf.org/rfc/rfc3987.txt, Letzter Zugriff:

14.01.2019. Zitiert auf Seite 27.

[FKL+13] Christoph Fehling, Tobias Kollmann, Richard Lackes, Frank Leymann, Markus Siepermann. *Kompakt-Lexikon Wirtschaftsinformatik*. Springer Fachmedien Wiesbaden, 2013. Zitiert auf Seite 13.

[Fle71] Joseph L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), Seiten 378–382, 1971. Zitiert auf den Seiten 22 und 23.

[Fuh06] Norbert Fuhr. *Information Retrieval: Skriptum zur Vorlesung im SS 06*. 2006. Zitiert auf den Seiten 10, 15, 17, 18, 19 und 20.

[Fuh13] Norbert Fuhr. Modelle im Information Retrieval. In: *Grundlagen der praktischen Information und Dokumentation*, Seiten 321–332. 2013. Zitiert auf den Seiten 11, 13 und 196.

[GCHV18] Tim Gemkow, Miro Conzelmann, Kerstin Hartig, Andreas Vogelsang. Automatic glossary term extraction from large-scale requirements specifications. In: *IEEE 26th International Requirements Engineering Conference (RE)*, Seiten 412–417, Piscataway, NJ, USA, 2018. IEEE Computer Society. Zitiert auf Seite 195.

[Geb13] Vera Gebhardt. *Funktionale Sicherheit nach ISO 26262: Ein Praxisleitfaden zur Umsetzung*. dpunkt.verlag, Heidelberg, 1. Auflage, 2013. Zitiert auf Seite 39.

[GG17] László Grad-Gyenge. On the optimal setting of spreading activation parameters to calculate recommendations on the knowledge graph. In: *Information Technologies – Applications and Theory (ITAT), CEUR Workshop Proceedings*, Seiten 210–217. 2017. Zitiert auf Seite 50.

[GGF16] Laszlo Grad-Gyenge, Peter Filzmoser. Recommendation techniques on a knowledge graph for email remarketing. In: *The 8th International Conference on Information, Process, and Knowledge (eKNOW)*, Seiten 51–56, Venice, Italy, 2016. IARIA. Zitiert auf Seite 49.

[GGFW15] László Grad-Gyenge, Peter Filzmoser, Hannes Werthner. Recommendations on a knowledge graph. In: *1st International Workshop on Machine Learning Methods for Recommender Systems (MLRec)*, Seiten 13–20. Vancouver, Kanada, 2015. Zitiert auf Seite 49.

[GGWF16] László Grad-Gyenge, Hannes Werthner, Peter Filzmoser. Knowledge graph based recommendation techniques for email remarketing. *International Journal on Advances in Intelligent Systems*, 9(3&4), 2016. Zitiert auf Seite 50.

[GHS10] Maurice Grinberg, Vladimir Haltakov, Hristo Stefanov. Approximate spreading activation for efficient knowledge retrieval from large datasets. In: *Proceedings of the 20th Italian Workshop on Neural Nets (WIRN)*. IOS Press, 2010. Zitiert auf Seite 49.

[GLS10] Carolina Gómez, Peter Liggesmeyer, Ariane Sutor. Variability management of safety and reliability models: An intermediate model towards systematic reuse of component fault trees. In: *Computer Safety, Reliability, and Security: 29th International Conference (SAFECOMP) Proceedings*, Seiten 28–40. Springer Berlin Heidelberg, 2010. Zitiert auf Seite 48.

[GSV+07] Kevin Gary, Bradley Szabo, Lavanya Vijayan, Braden Chapman, Jayavarshini Radhakrishnan, Aishwarya Sivaraman. JMaPSS: Spreading activation search for the semantic web. In: *IEEE International Conference on Information Reuse and Integration*, Seiten 104–109, Piscataway, NJ, USA, 2007. Zitiert auf Seite 49.

[GT06] Holger Giese, Matthias Tichy. Component-based hazard analysis: Optimal designs, product lines, and online-reconfiguration. In: *Computer Safety, Reliability, and Security: 25th International Conference (SAFECOMP) Proceedings*, Volume 4166 of *Lecture Notes in Computer Science*, Seiten 156–169. Springer Berlin, 2006. Zitiert auf Seite 48.

[GTS04] Holger Giese, Matthias Tichy, Daniela Schilling. Compositional hazard analysis of UML component and deployment models. In: *Computer Safety, Reliability, and Security: 23rd International Conference (SAFECOMP) Proceedings*, Volume 3219 of *Lecture Notes in Computer Science*, Seiten 166–179. Springer Berlin Heidelberg, 2004. Zitiert auf Seite 47.

[Gwe08] Kilem Li Gwet. Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, Seiten 29–48, 2008. Zitiert auf den Seiten 22 und 23.

[Gwe14] Kilem Li Gwet. *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters.* Advanced Analytics LLC, Gaithersburg, MD, 4. Auflage, 2014. Zitiert auf den Seiten 22 und 23.

[Har13] Donna Harman. TREC-style evaluations. In: *Information Retrieval Meets Information Visualization: PROMISE Winter School*, Seiten 97–115. Springer Berlin Heidelberg, 2013. Zitiert auf Seite 16.

[Hel13] Ekkehard Helmig. Produkthaftung braucht Reform. *QZ, Hanser Verlag, München*, 58(1), Seiten 22–24, 2013. Zitiert auf Seite 36.

[Hen08] Andreas Henrich. *Information Retrieval 1: Grundlagen, Modelle und Anwendungen.* Lehrbuch, Otto-Friedrich-Universität Bamberg, 2008. Zitiert auf den Seiten 3, 9, 10, 11, 13, 14 und 212.

[HF97] Birgit Hamp, Helmut Feldweg. GermaNet - a lexical-semantic net for German. In: *Proceedings of ACL Workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, Seiten 9–15, 1997. Zitiert auf den Seiten 80, 84 und 195.

[HH10] Verena Henrich, Erhard Hinrichs. GernEdiT – The GermaNet editing tool. In: *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC)*, Seiten 2228–2235. European Language Resources Association, 2010. Zitiert auf den Seiten 80, 84 und 195.

[HK17] Kerstin Hartig, Thomas Karbe. Spreading activation simulation with semantic network skeletons. *International Journal on Advances in Intelligent Systems*, 10(1&2), Seiten 1–13, 2017. Zitiert auf den Seiten 183 und 185.

[HKRS08] Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, York Sure. *Semantic Web: Grundlagen.* eXamen.press. Springer Berlin Heidelberg, 2008. Zitiert auf den Seiten 26, 28 und 29.

[ISO26262] International Organization for Standardization. *Road Vehicles – Functional Safety, Part 1-10 (ISO 26262:2011)*, 2011. Zitiert auf den Seiten 1 und 36.

[ISO26262-1] International Organization for Standardization. *Road Vehicles – Functional Safety, Part 1: Vocabulary (ISO 26262-1:2011)*, 2011. Zitiert auf den Seiten 36, 37 und 38.

[ISO26262-2] International Organization for Standardization. *Road Vehicles – Functional Safety, Part 2: Management of functional safety (ISO 26262-2:2011)*, 2011. Zitiert auf Seite 38.

[ISO26262-3] International Organization for Standardization. *Road Vehicles – Functional Safety, Part 3: Concept phase (ISO 26262-3:2011)*, 2011. Zitiert auf den Seiten 39, 40, 41, 42, 43, 45, 47, 59 und 212.

[ISO26262-4] International Organization for Standardization. *Road Vehicles – Functional Safety, Part 4: Product development at the system level (ISO 26262-4:2011)*, 2011. Zitiert auf Seite 39.

[ISO26262-7] International Organization for Standardization. *Road Vehicles – Functional Safety, Part 7: Production and operation (ISO 26262-7:2011)*, 2011. Zitiert auf Seite 40.

[ISO26262-8] International Organization for Standardization. *Road Vehicles – Functional Safety, Part 8: Supporting processes (ISO 26262-8:2011)*, 2011. Zitiert auf Seite 47.

[Jena] Apache. *Jena – A free and open source Java framework for building Semantic Web and Linked Data applications*, `https://jena.apache.org/`, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 117.

[JGraphT] JGraphT. *JGraphT: A Java library of graph theory data structures and algorithms*, `https://jgrapht.org/`, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 117.

[JK02] Kalervo Järvelin, Jaana Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, 20(4), Seiten 422–446, 2002. Zitiert auf Seite 20.

[Kem15] Sören Kemmann. *SAHARA: A structured approach for hazard analysis and risk assessments*. Dissertation, Technische Universität Kaiserslautern, 2015. Zitiert auf Seite 45.

[KG90] Maurice George Kendall, Jean Dickinson Gibbons. *Rank correlation methods*. Arnold, London, 5. Auflage, 1990. Zitiert auf Seite 24.

[KLM03] Bernhard Kaiser, Peter Liggesmeyer, Oliver Mäckel. A new component concept for fault trees. In: *Proceedings of the 8th Australian Workshop on Safety Critical Systems and Software (SCS)*, Seiten 37–46. Australian Computer Society, Sydney, 2003. Zitiert auf Seite 47.

[KMMK$^+$] Serhiy Kosinov, Stephane Marchand-Maillet, Igor Kozintsev, Carole Dulong, Thierry Pun. Dual diffusion model of spreading activation for content-based image retrieval. In: *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval (MIR)*, Seiten 43–50, New York, NY, USA. ACM. Zitiert auf Seite 49.

[KS39] M. G. Kendall, B. Babington Smith. The problem of m rankings. *The Annals of Mathematical Statistics*, 10(3), Seiten 275–287, 1939. Zitiert

auf Seite 24.

[Kur04] Dominik Kuropka. *Modelle zur Repräsentation natürlichsprachlicher Dokumente: Ontologie-basiertes Information-Filtering und -Retrieval mit relationalen Datenbanken*, Volume 10 of *Advances in information systems and management science*. Logos-Verlag, Berlin, 2004. Zitiert auf den Seiten 11 und 13.

[Lan79] F. Wilfrid Lancaster. *Information retrieval systems characteristics, testing and evalution*. John Wiley and Sons, New York, NY, USA, 2. Auflage, 1979. Zitiert auf Seite 10.

[LK77] J. Richard Landis, Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1), Seiten 159–174, 1977. Zitiert auf den Seiten 23 und 215.

[LPP10] Peter Löw, Roland Pabst, Erwin Petry. *Funktionale Sicherheit in der Praxis: Anwendung von DIN EN 61508 und ISO/DIS 26262 bei der Entwicklung von Serienprodukten*. dpunkt.verlag, 1. Auflage, 2010. Zitiert auf den Seiten 36, 37 und 40.

[LSL13] Gi Hyun Lim, Il Hong Suh, Luís Seabra Lopes. The representation of weighted action-coupled semantic network and spreading activation model for improvisational action. In: *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Seiten 4054–4059. 2013. Zitiert auf Seite 49.

[LVDv16] Artuur Leeuwenberg, Mihaela Vela, Jon Dehdari, Josef van Genabith. A minimally supervised approach for synonym extraction with word embeddings. *The Prague Bulletin of Mathematical Linguistics*, Seiten 111–142, 2016. Zitiert auf Seite 195.

[MBA17] Diana Maynard, Kalina Bontcheva, Isabelle Augenstein. *Natural language processing for the semantic web*. Synthesis lectures on the semantic web. Morgan & Claypool, 2017. Zitiert auf den Seiten 11 und 12.

[Medini] ANSYS. *Medini Analyze – Automotive Functional Safety Analysis*, `https://www.ansys.com/de-de/products/systems/ansys-medini-analyze`, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 55.

[MG11] Nicolas Marie, Fabien Gandon. Social objects description and recommendation in multidimensional social networks: OCSO ontology and semantic spreading activation. In: *1st International Workshop on Social Object Networks (SocialObjects)*, Seiten 1415–1420. Boston, USA, 2011. Zitiert auf Seite 49.

[MH16] Vanessa Nicole Michalke, Kerstin Hartig. Explanation retrieval in semantic networks – Understanding spreading activation based recommendations. In: *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KDIR)*, Seiten 291–298. SciTePress – Science and Technology Publications, 2016. Zitiert auf Seite 108.

[Mic16] Vanessa Nicole Michalke. *Automatische Erklärung der Resultate eines auf „Spreading Activation“ basierenden Vorschlagssystems für semantische Netze*. Masterarbeit, Technische Universität Berlin, 2016. Zitiert auf den Seiten 108, 113, 118 und 212.

[Moo50] Calvin N. Mooers. Coding, information retrieval, and the rapid selector. *American Documentation*, 1(4), Seiten 225–229, 1950. Zitiert auf Seite 9.

[MRS08] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze. *An introduction to information retrieval*. Cambridge University Press, Cambridge, England, 2008. Zitiert auf den Seiten 9, 10, 16, 17 und 19.

[MSB13] Christoph Maier, Alexander Schloske, Steffen Bothe. *Studie zur Funktionalen Sicherheit in der Automobilbranche (ISO 26262)*. Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA, Stuttgart, 2013. Zitiert auf den Seiten 1, 2 und 43.

[MSB+14] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, David McClosky. The Stanford CoreNLP natural language processing toolkit. In: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Seiten 55–60, Stroudsburg, PA, USA, 2014. Association for Computational Linguistics. Zitiert auf den Seiten 12 und 117.

[MWH00] Thomas Mandl, Christa Womser-Hacker. Ein adaptives Information-Retrieval-Modell für Digitale Bibliotheken. In: *Informationskompetenz – Basiskompetenz in der Informationsgesellschaft: Proceedings des 7. Internationalen Symposiums für Informationswissenschaft (ISI)*, Schriften zur Informationswissenschaft, Seiten 1–16. Univ.-Verl. Konstanz, 2000. Zitiert auf Seite 196.

[NW15] Mattias Nyberg, Jonas Westman. Failure propagation modeling based on contracts theory. In: *11th European Dependable Computing Conference (EDCC) Proceedings*, Seiten 108–119, Piscataway, NJ, USA, 2015. Zitiert auf Seite 47.

[Ontmalizer] Software Research & Development and Consultancy Corporation. *Ontmalizer*, `https://github.com/srdc/ontmalizer/`, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 30.

[OntoSpread] Álvarez Rodríguez, José María. *OntoSpread*, `https://code.google.com/archive/p/ontospread/`, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 49.

[OpenGALEN] OpenGALEN Foundation. *OpenGALEN*, `http://www.opengalen.org/`, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 49.

[OpenNLP] Apache. *OpenNLP*, `https://opennlp.apache.org/`, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 117.

[OSP14] Markus Oertel, Michael Schulze, Thomas Peikenkamp. Reusing a functional safety concept in variable system architectures. In: *Proceedings of the 7th International Workshop on Model-based Architecting and Construction of Embedded Systems (MoDELS)*, Volume 1250 of *CEUR Workshop Proceedings*, Seiten 16–25. 2014. Zitiert auf Seite 55.

[Pas03] Gabriella Pasi. Intelligent information retrieval: some research trends. In: *Advances in Soft Computing*. Springer, London, 2003. Zitiert auf Seite 14.

[Pau07] Wilfried Paulus. Herleitung der Sicherheitsklasse (ASIL) nach dem Entwurf der »Automobil-Sicherheitsnorm« ISO WD 26262: Ein Erfahrungsbericht. In: *Moderne Elektronik im Kraftfahrzeug II*, Haus der

Technik – Fachbuchreihe, Seiten 104–117. expert Verlag, Renningen, 2007. Zitiert auf Seite 43.

[PM95] David Lorge Parnas, Jan Madey. Functional documents for computer systems. *Science of Computer Programming*, 25(1), Seiten 41–61, 1995. Zitiert auf Seite 46.

[PMSH01] Y. Papadopoulos, J. McDermid, R. Sasse, G. Heiner. Analysis and synthesis of the behaviour of complex programmable electronic systems in conditions of failure. *Reliability Engineering & System Safety*, 71(3), Seiten 229–247, 2001. Zitiert auf Seite 47.

[PRK] Maciej Piasecki, Radosław Ramocki, Michał Kaliński. Information spreading in expanding wordnet hypernymy structure. In: *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP)*, Seiten 553–561. Zitiert auf Seite 49.

[ProdHaftG] *Gesetz über die Haftung für fehlerhafte Produkte (Produkthaftungsgesetz – ProdHaftG)*, 2017. Zitiert auf Seite 36.

[PTFG] Yiannis Papadopoulos, Audrey Tran, Jean-Marc Faure, Christian Grante. Component failure behaviour: Patterns and reuse in automated system safety analysis. *SAE International Journal of Passenger Cars – Mechanical Systems*, Seiten 1296–1311. Zitiert auf Seite 47.

[Qui68] M. Ross Quillian. Semantic memory. In: *Semantic Information Processing*, Seiten 216–270. MIT Press, 1968. Zitiert auf den Seiten 32 und 49.

[RBJ89] Vijay Raghavan, Peter Bollmann, Gwang S. Jung. A critical investigation of recall and precision as measures of retrieval system performance. *ACM Transactions on Information Systems*, 7(3), Seiten 205–229, 1989. Zitiert auf den Seiten 18, 19 und 20.

[Rei10] Klaus Reichenberger. *Kompendium Semantische Netze: Konzepte, Technologie, Modellierung*. Springer Berlin Heidelberg, 1. Auflage, 2010. Zitiert auf Seite 26.

[Ros14] Hans-Leo Ross. *Funktionale Sicherheit im Automobil: ISO 26262, Systemengineering auf Basis eines Sicherheitslebenszyklus und bewährten Managementsystemen*. Hanser, München, Wien, 2014. Zitiert auf den Seiten 36 und 37.

[Sal68] Gerard Salton. *Automatic information organization and retrieval*. McGraw-Hill Computer Science Series. McGraw Hill, New York [u.a.], 7. Auflage, 1968. Zitiert auf Seite 13.

[Salus] Salus. *Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies*, `http://www.srdc.com.tr/projects/salus/`, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 30.

[SC12] Mark Sanderson, W. Bruce Croft. The history of information retrieval research. *Proceedings of the IEEE*, Seiten 1444–1451, 2012. Zitiert auf den Seiten 9 und 10.

[Sch16] Kinga Schumacher. *Hybride semantische Suche – eine Kombination aus Fakten- und Dokumentretrieval*. Dissertation, Universität Potsdam, 2016. Zitiert auf Seite 50.

[Sei18] Christopher Seidel. *Klonerkennung in Gefährdungs- & Risikoanalysen*. Bachelorarbeit, Technische Universität Berlin, 2018. Zitiert auf den

Seiten 56, 214 und 218.

[SGHP15] Shengtao Sun, Jibing Gong, Jijun He, Siwei Peng. A spreading activation algorithm of spatial big data retrieval based on the spatial ontology model. *Cluster Computing*, 18(2), Seiten 563–575, 2015. Zitiert auf Seite 49.

[SKSB09] Leo Sauermann, Malte Kiesel, Kinga Schumacher, Ansgar Bernardi. Semantic desktop. In: *Social Semantic Web: Web 2.0 – Was nun?*, X.media.press, Seiten 337–362. Springer Berlin Heidelberg, 2009. Zitiert auf Seite 50.

[Šlj15] Irfan Šljivo. *Facilitating reuse of safety case artefacts using safety contracts*. Dissertation, Mälardalen University, 2015. Zitiert auf Seite 47.

[SM83] Gerard Salton, Michael J. MacGill. *Introduction to modern information retrieval*. McGraw-Hill, New York, NY, USA, 1983. Zitiert auf den Seiten 9, 10 und 19.

[SMB13] Michael Schulze, Jan Mauersberger, Danilo Beuche. Functional safety and variability: Can it be brought together? In: *Proceedings of the 17th International Software Product Line Conference (SPLC)*, Seiten 236–243, New York, NY, USA, 2013. ACM Press. Zitiert auf den Seiten 48 und 55.

[SNOMED] SNOMED International. *SNOMED CT Browser*, `https://browser.ihtsdotools.org/`, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 49.

[SSA13] Christoph Seidl, Ina Schaefer, Uwe Aßmann. Variability-aware safety analysis using delta component fault diagrams. In: *Proceedings of the 17th International Software Product Line Conference (SPLC) co-located Workshops*, Seiten 2–9. ACM Press, 2013. Zitiert auf Seite 48.

[SSS08] Kinga Schumacher, Michael Sintek, Leo Sauermann. Combining fact and document retrieval with spreading activation for semantic desktop search. In: *Proceedings of the 5th European Semantic Web Conference (ESWC)*, Volume 5021 of *Lecture Notes in Computer Science*, Seiten 569–583. Springer Berlin Heidelberg, 2008. Zitiert auf Seite 50.

[SSZ+17] Shengtao Sun, Weijing Song, Albert Y. Zomaya, Yang Xiang, Kim-Kwang Raymond Choo, Tejal Shah, Lizhe Wang. Associative retrieval in spatial big data based on spreading activation with semantic ontology. *Future Generation Computer Systems*, 76, Seiten 499–509, 2017. Zitiert auf Seite 49.

[Stä10] Tobias Ständer. *Eine modellbasierte Methode zur Objektivierung der Risikoanalyse nach ISO 26262*. Dissertation, Technische Universität Braunschweig, 2010. Zitiert auf Seite 45.

[STST99] Anne Schiller, Simone Teufel, Christine Stöckert, Christine Thielen. *Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset)*. Technischer Report, Universität Stuttgart, Universität Tübingen, 1999. Zitiert auf Seite 12.

[SZ16] Jörg Schäuffele, Thomas Zurawka (Hrsg.). *Automotive Software Engineering: Grundlagen, Prozesse, Methoden und Werkzeuge effizient einsetzen*. Springer Vieweg, Wiesbaden, 6. Auflage, 2016. Zitiert auf Seite 2.

[Thi12] Kilian Thiel. *Knotenähnlichkeiten aus Aktivierungsausbreitungen*. Dissertation, Universität Konstanz, 2012. Zitiert auf Seite 35.

[TM11] Nana Tintarev, Judith Masthoff. Designing and evaluating explanations

for recommender systems. In: *Recommender systems handbook*. Springer, 2011. Zitiert auf Seite 113.

[TopBraid] TopQuadrant. *TopBraid Composer*, https://www.topquadrant.com/tools/IDE-topbraid-composer-maestro-edition/, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 31.

[TREC] TREC. *Text REtrieval Conference (TREC)*, https://trec.nist.gov/, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 16.

[Uri08] Filemon A. Uriarte. *Introduction to knowledge management: A brief introduction to the basic elements of knowledge management for non-practitioners interested in understanding the subject*. ASEAN Foundation, Jakarta, Indonesia, 2008. Zitiert auf den Seiten 13 und 14.

[VDA] Verband der Automobilindustrie. *VDA 702 – Situationskatalog E-Parameter nach ISO 26262-3*, VDA-Empfehlung, 2015. Zitiert auf den Seiten 41 und 46.

[W3C04] World Wide Web Consortium W3C. *OWL Web Ontology Language – Overview*, W3C Recommendation, 2004, https://www.w3.org/TR/owl-features/, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 28.

[W3C12a] World Wide Web Consortium W3C. *A Direct Mapping of Relational Data to RDF*, W3C Recommendation, 2012, https://www.w3.org/TR/rdb-direct-mapping/, Letzter Zugriff: 14.01.2019. Zitiert auf den Seiten 28, 30 und 212.

[W3C12b] World Wide Web Consortium W3C. *OWL 2 Web Ontology Language – Direct Semantics (Second Edition)*, 2012, https://www.w3.org/TR/owl2-direct-semantics/, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 27.

[W3C12c] World Wide Web Consortium W3C. *OWL 2 Web Ontology Language – Mapping to RDF Graphs (Second Edition)*, W3C Recommendation, 2012, https://www.w3.org/TR/owl2-mapping-to-rdf/, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 27.

[W3C12d] World Wide Web Consortium W3C. *OWL 2 Web Ontology Language – Structural Specification and Functional-Style Syntax (Second Edition)*, 2012, https://www.w3.org/TR/owl2-syntax/, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 27.

[W3C14a] World Wide Web Consortium W3C. *RDF 1.1 Concepts and Abstract Syntax*, W3C Recommendation, 2014, https://www.w3.org/TR/rdf11-concepts/, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 27.

[W3C14b] World Wide Web Consortium W3C. *RDF 1.1 N-Triples – A line-based syntax for an RDF graph*, W3C Recommendation, 2014, https://www.w3.org/TR/n-triples/, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 29.

[W3C14c] World Wide Web Consortium W3C. *RDF 1.1 Semantics*, W3C Recommendation, 2014, https://www.w3.org/TR/rdf11-mt/, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 27.

[W3C14d] World Wide Web Consortium W3C. *RDF Schema 1.1*, W3C Recommendation, 2014, https://www.w3.org/TR/rdf-schema/, Letzter Zugriff: 14.01.2019. Zitiert auf Seite 27.

[Win13] Hermann Winner. Challenges of automotive systems engineering for

industry and academia. In: *Automotive systems engineering*, Seiten 3–15. Springer, 2013. Zitiert auf den Seiten 1 und 2.

[WN18] Jonas Westman, Mattias Nyberg. Conditions of contracts for separating responsibilities in heterogeneous systems. *Formal Methods in System Design*, 52(2), Seiten 147–192, 2018. Zitiert auf Seite 47.

[WNT13] Jonas Westman, Mattias Nyberg, Martin Törngren. Structuring safety requirements in ISO 26262 using contract theory. In: *Computer Safety, Reliability, and Security: 32nd International Conference (SAFECOMP) Proceedings*, Volume 8153 of *Lecture Notes in Computer Science / Programming and Software Engineering*, Seiten 166–177. Springer Berlin Heidelberg, 2013. Zitiert auf Seite 47.

[WP50] Carl S. Wise, James W. Perry. Multiple coding and the rapid selector. *American Documentation*, 1(2), Seiten 76–83, 1950. Zitiert auf Seite 9.

[Yük13] Mustafa Yüksel. *A semantic interoperability framework for reinforcing post market safety studies*. Dissertation, Middle East Technical University, 2013. Zitiert auf den Seiten 30 und 217.

[Zob98] Justin Zobel. How reliable are the results of large-scale information retrieval experiments? In: *Proceedings of the 21st Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR)*, Seiten 307–314, New York, NY, USA, 1998. ACM. Zitiert auf Seite 17.

[ZWL12] Songhao Zhu, Baoyun Wang, Yuncai Liu. Spreading activation theory based image annotation. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seiten 1041–1044. Ieee, 2012. Zitiert auf Seite 49.

Abkürzungsverzeichnis

AIS	Abbreviated Injury Scale ... 41
AP	Average Precision ... 19
ASIL	Automotive Safety Integrity Level ... 2
CLEF	Conference and Labs of the Evaluation Forum ... 16
ConQAT	Continuous Quality Assessment Toolkit ... 56
DCG	Discounted Cumulated Gain ... 20
E/E-Systeme	elektrische und elektronische Systeme ... 36
EP	Expected Precision ... 20
FMEA	Fehlermöglichkeits- und Einflussanalyse (*engl.:* failure mode & effects analysis) ... 40
FTA	Fehlerbaumanalyse (*engl.:* fault tree analysis) ... 47
FuSi	Funktionale Sicherheit ... 36
G&R	Gefährdungs- und Risikoanalyse ... iv
HARA	hazard analysis and risk assessment ... v
HARvESTer	Hazard Analysis and Risk assessment dEcision Support Tool ... 115
IR	Information-Retrieval ... iv
IRI	Internationalized Resource Identifier ... 27
ISS	Information Severity Score ... 41
MAIS	Maximum Abbreviated Injury Scale ... 41
NDCG	Normalized Discounted Cumulated Gain ... 20
NLP	Natural Language Processing ... 12
OCL	Object Constraint Language ... 47
OWL	Web Ontology Language ... 26
PoS	Part-of-Speech ... 12
PRR	Probability of Relevance ... 19
RCP	Rich Client Platform ... 115
RDF	Resource Description Framework ... 26
SIL	Safety Integrity Level ... 37
SPARQL	SPARQL Protocol and RDF Query Language ... 30
STTS	Stuttgart-Tübingen-Tagset ... 12
TREC	Text Retrieval Conference ... 16
UML	Unified Modeling Language ... 47
URI	Uniform Resource Identifier ... 27
VDA	Verband der Automobilindustrie ... 41
W3C	World Wide Web Consortium ... 27
WWW	World Wide Web ... 26
XML	Extensible Markup Language ... 30
XSD	XML Schema Definition ... 30
XSLT	Extensible Stylesheet Language Transformations ... 30

Abbildungsverzeichnis

1.1 Beiträge dieser Arbeit im Überblick 5

2.1 Allgemeine Arbeitsweise eines IR-Systems nach [Hen08] 11
2.2 Grundsätzlicher Evaluationsaufbau im IR 16
2.3 Dimensionen der komparativen Evaluation im IR 16
2.4 Beispiel für das Zusammenführen von Systemrelevanz und objektiver Relevanz 18
2.5 RDF-Graph - Definition des abstrakten Datenmodells [W3C12a] 28
2.6 RDF-Tripel 28
2.7 Beispiel eines RDF-Graphen 29
2.8 Ablauf der Transformation von XML/XSD zu OWL [BA05] 31
2.9 Ablauf des Retrievals durch Aktivierungsausbreitung nach [Cre97] 32
2.10 Schematischer Ablauf der Aktivierungsausbreitung in Pulsen 33
2.11 Phasen des Sicherheitslebenszyklus mit FuSi-Management 38
2.12 Schadensausmaßklassen [ISO26262-3] 41
2.13 Kontrollierbarkeitsklassen [ISO26262-3] 41
2.14 Auftretenswahrscheinlichkeitsklassen [ISO26262-3] 41
2.15 Risikograph zur ASIL-Bestimmung [ISO26262-3] 42
2.16 Beispielhafter Ausschnitt einer G&R 43

3.1 Beispiel für Wiederverwendungspotenziale 54
3.2 Zuordnung der Typen/Prädikate des Datenmodells zu G&R-Teilartefakten . 56
3.3 Anwendungsfall - Zusammenhänge erkennen 57
3.4 Generisches IR-System zur Anfragebearbeitung 60
3.5 Methode zum Aufbau domänenspezifischer IR-Systeme 61
3.6 Integration der IR-Komponente mit dualer Datenhaltung 64
3.7 RDF-Graph nach *XSD2OWL*- und *XML2OWL*-Transformationen 68
3.8 Schematische Struktur des Basisnetzes 70
3.9 Durchschnittliche Tokenanzahl je Prädikat gruppiert nach Teilartefakten . . . 75
3.10 Beispiel zweier Instanzknoten mit Literalen und gemeinsamen Keywords . . . 76
3.11 NLP-Pipeline zur Extraktion von Keywords 77
3.12 Beispiel zweier Instanzknoten mit Literalen und verwendeten Synonymen . . 79
3.13 Singuläre Synonyminstanzen mit additiver und integrativer Integration 81
3.14 Gruppierte Synonyminstanzen mit additiver und integrativer Integration . . . 82
3.15 Lokaler und globaler Gültigkeitsbereich bei Synonymverknüpfung 83
3.16 Beispieltripel mit semantisch äquivalenten Aussagen 91
3.17 Sende-, Empfangs- und Transportperspektive 97
3.18 Ausbreitungsverhalten im Sendemodus *B* 100
3.19 Ausbreitungsverhalten im Sendemodus *RR* 101
3.20 Ausbreitungsverhalten im Sendemodus *FW* 102
3.21 Ausbreitungsverhalten im Sendemodus *FWL* 103
3.22 Ausbreitungsverhalten im Ausbreitungsmodus *FW-RR* 104
3.23 Ausbreitungsverhalten im Ausbreitungsmodus *FWL-RR* 104
3.24 Methode zur automatischen Erklärung der Empfehlungen für G&Rs [Mic16] . 108
3.25 Ausgangsgraph zu Ausbreitungsgraph am Beispiel 109
3.26 Backtracking im Ausbreitungsgraphen und Minimierung 110
3.27 Ziel- und Startgruppen des minimierten Ausbreitungsgraphen 112
3.28 Identifizierbare Muster und deren Priorisierung 112

3.29 Architektur des entwickelten Prototyps 116

4.1 Dimensionen der komparativen Evaluation in dieser Arbeit 120
4.2 Erstellen objektiver Relevanzbewertungen 123
4.3 Bewertungsszenario 125
4.4 Selbsteinschätzung der Rater bezüglich ihres Experten- und Konfidenzlevels . 126
4.5 Anteilige Bewertungen für die Anfragen je Rater 128
4.6 95%-Konfidenzintervalle des κ_{BP}-Koeffizienten 131
4.7 95%-Konfidenzintervalle des Gwet-AC_1-Koeffizienten 131
4.8 Verteilung der berechneten Werte für Fitness, EP und $\mathrm{NDCG}_{\mathrm{Min}}$ 138
4.9 Verteilung der Fitness auf die Netzvarianten 139
4.10 Verteilung der *Top100* auf die Netzvarianten 140
4.11 Verteilung der Fitness auf die Modi 141
4.12 Verteilung der *Top100* auf die Modi 142
4.13 Verteilung der Fitness auf die Kantengewichtung 142
4.14 Verteilung der Fitness auf Dämpfungsfaktor und Aktivierungsschwellwert . . 142
4.15 Verteilung der Fitness auf die Anzahl der Pulse 143
4.16 Verteilung der Fitness auf die Anfragen 143
4.17 Entwicklung der maximalen Fitness bei evolutionärer Optimierung (OZ1) . . 146
4.18 Anzahl der Konfigurationen OZ1 147
4.19 Verteilung der berechneten Werte für Fitness, EP und $\mathrm{NDCG}_{\mathrm{Min}}$ 148
4.20 Anzahl und Verteilung der *Top1000* auf die Netzvarianten 149
4.21 Anzahl und Verteilung der *Top1000* auf die Modi 150
4.22 Verteilung der *Top1000* auf Netzvarianten und Modi 151
4.23 Anzahl der Kombinationen von Modi in den *Top1000* 152
4.24 Anzahl und Verteilung der *Top1000* auf die Pulsanzahl 153
4.25 Verteilung der *Top1000* je Anfrage auf die Modi 154
4.26 Anzahl der Konfigurationen OZ2 157
4.27 Verteilung der berechneten Werte Fitness, EP und $\mathrm{NDCG}_{\mathrm{Min}}$ je Anfrage . . . 158
4.28 Verteilung der *Top100* auf Netzvarianten und Konfigurationsparameter (1/2) 160
4.28 Verteilung der *Top100* auf Netzvarianten und Konfigurationsparameter (2/2) 161
4.29 Anzahl der Kombinationen für die Modi in den *Top100* je Netzvariante . . . 162
4.30 Anzahl der Kombinationen für die Modi in den *Top100* je Anfrage 163
4.31 Anzahl der Kombinationen für die Modi in den *Top100* 164
4.32 Anzahl und Verteilung der *Top100* auf die Netzvarianten 165
4.33 Anzahl der *Top100* Konfigurationen aller Anfragen aufgeteilt nach Modi . . . 166
4.34 Anzahl und Verteilung der *Top100* auf Anfragen und Kantengewichtung . . . 168
4.35 EP-Recall-Graphen für die anfrageübergreifende Konfiguration 170
4.36 EP-Recall-Graphen für die anfragespezifischen Konfigurationen 171

5.1 Beispiel eines semantischen Netzes und seines Netzwerkskelettes 176
5.2 Effektives semantisches Netzwerkskelett der Netzvariante N4-2 180
5.3 Simulation der Aktivierungsausbreitung auf dem Netzwerkskelett 181
5.4 Simulationsschritte 182
5.5 Lokaler Simulationsschritt zur Bestimmung der Aktivierungsstärke 184
5.6 Lokaler Simulationsschritt zur Bestimmung der Ausbreitungsstärke 185
5.7 Laufzeit von Aktivierungsausbreitung und Simulation 187

A.1 XSD2OWL-Mapping der Java-Bibliothek Ontmalizer 217

B.1 Analyse einfacher Klonklassen [Sei18] 218
B.2 Analyse kombinierter Klonklassen [Sei18] 218
B.3 Durchschnittliche Tokenanzahl je Prädikat, Typ, Teilartefakt 219
B.4 Durchschnittliche Tokenanzahl je Prädikat 220

C.1 Streudiagramm der $NDCG_{Min}$- und $NDCG_{Max}$-Werte jedes Durchlaufes . . . 229
C.2 Streudiagramm der EP- und NDCG-Werte jedes Durchlaufes 229
C.3 Streudiagramm der Fitness, $NDCG_{Min}$- und EP-Werte jedes Durchlaufes . . 229

Tabellenverzeichnis

2.1 Interpretation der Kappa-Werte nach Landis und Koch [LK77] 23
2.2 Überblick über die Bewertungsmaße für die Evaluation in dieser Arbeit . . . 25
2.3 Abkürzung der IRI durch Präfixe 27

3.1 Kategorisierung der Systeme im Datensatz 55
3.2 RDF-Graph nach *XSD2OWL*-Transformation am Beispiel 67
3.3 Namenskonvention zur Benennung domänenspezifischer Elemente 68
3.4 Konkrete Eigenschaften des G&R-Basisnetzes 72
3.5 Verteilung der Tokenanzahl aller verwendeten Literale 74
3.6 Keyword-spezifische Knoten und Tripel im erweiterten G&R-Basisnetz 79
3.7 Synonymspezifische Tripel im erweiterten G&R-Basisnetz 84
3.8 Aktivierungsmodi mit Transferfunktionen und Funktionsgraphen 93
3.9 Dämpfungsmodi mit Dämpfungsfunktionen und Funktionsgraphen 95
3.10 Verzweigungsmodi mit Verzweigungsfunktionen 95
3.11 Erweiterung der Verzweigungsmodi 100
3.12 Resultierende Ausbreitungsmodi aus beteiligten Sendemodi 103

4.1 Netzvarianten für die Evaluation 121
4.2 Eigenschaften der evaluierten Netzvarianten 121
4.3 Evaluierte Informationsbedürfnisse 122
4.4 Basisset verwendeter Konfigurationseinstellungen 124
4.5 Anzahl der Bewertungselemente für die einzelnen Stufen des Poolings 125
4.6 Anzahl der als relevant markierten Bewertungselemente 128
4.7 Evaluationsfragen 129
4.8 Ergebnisse der Brennan-Prediger-Koeffizient-Analyse 130
4.9 Ergebnisse der Gwet-AC_1 130
4.10 Ergebnisse der Konkordanzanalyse nach Kendall 133
4.11 Konkordanzkoeffizient nach Kendall für jedes Raterpaar 134
4.12 Ergebnisse der Brennan-Prediger-Koeffizient-Analyse (dichotomisiert) 135
4.13 Wertebereiche der Konfigurationsparameter für den Zufallsgenerator 137
4.14 Parametrisierung der besten anfrageübergreifenden Konfiguration 147
4.15 Parametrisierung der besten anfragespezifischen Konfigurationen (1/2) 156
4.16 Parametrisierung der besten anfragespezifischen Konfigurationen (2/2) 156
4.17 Qualität der optimierten anfrageübergreifenden Konfiguration 170
4.18 Qualität der optimierten anfragespezifischen Konfigurationen 171

5.1 Differenz zwischen Aktivierungsausbreitung und Simulation 186

C.1 Ergebnisse der linear gewichteten κ_{BP}-Analyse 221
C.2 Ergebnisse der quadratisch gewichteten κ_{BP}-Analyse 221
C.3 Ergebnisse der linear gewichteten Gwet-AC_2-Analyse 222
C.4 Ergebnisse der quadratisch gewichteten Gwet-AC_2-Analyse 222
C.5 Ergebnisse der κ_{BP}-Analyse (reduzierter Pool) 223
C.6 Ergebnisse der linear gewichteten κ_{BP}-Analyse (reduzierter Pool) 223
C.7 Ergebnisse der quadratisch gewichteten κ_{BP}-Analyse (reduzierter Pool) . . . 223
C.8 Ergebnisse der Gwet-AC_1-Analyse (reduzierter Pool) 224
C.9 Ergebnisse der linear gewichteten Gwet-AC_2-Analyse (reduzierter Pool) . . . 224
C.10 Ergebnisse der quadratisch gewichteten Gwet-AC_2-Analyse (reduzierter Pool) 224

C.11 Ergebnisse der Konkordanzanalyse nach Kendall (reduzierter Pool) 225
C.12 Konkordanzkoeffizient nach Kendall für jedes Rater-OR-Paar 226
C.13 Brennan-Prediger-Koeffizient-Analyse für jedes Rater-OR-Paar 227

A | Transformationen

In Abbildung A.1 dieses Anhangs ergänzen wir die Beschreibung der Java-Bibliothek *Ontmalizer* aus Abschnitt 2.2.3 um die von Yüksel [Yük13, S. 46] veröffentlichten *XSD2OWL*-Transformationen, die im *Ontmalizer* zum Einsatz kommen. Wir verwenden die *Ontmalizer*-Bibliothek in Abschnitt 3.2.2 bei der Erstellung unserer Basisnetze.

#	**XSD Constructs**	**OWL Constructs**
1	xs:simpleType	rdfs:Datatype with the suffix "Datatype" on datatype name.
2	xs:simpleType with xs:enumeration	rdfs:Datatype with the suffix "Datatype" on datatype name. In addition, for the enumerations, an owl:Class as a subclass of EnumeratedValue is created. Instances are created for every enumerated value. An instance of Enumeration, referring to all the instances, is created as well as the owl:oneOf union over the instances. These are mostly informative, as they are not used directly during the XML data to RDF/OWL transformation.
3	xs:complexType over xs:complexContent	owl:Class
4	xs:complexType over xs:simpleContent	owl:Class
5	xs:element (global) with complex type	owl:Class and subclass of the class generated from the referenced complex type
6	xs:element (global) with simple type	owl:Datatype
7	xs:element (local to a type)	owl:DatatypeProperty or owl:ObjectProperty depending on the element type. OWL Restrictions are built for the occurrence.
8	xs:group	owl:Class
9	xs:attributeGroup	owl:Class
10	Anonymous Complex Type	As for Complex Type except a URI is constructed as Anon_#. Also, the class is defined as a subclass of Anon.
11	Anonymous Simple Type	As for Simple Type except a URI is constructed as Anon_#.
12	Substitution Groups	Subclass statements are generated for the members.
13	xsi:type on an XML element	Overrides the schema abstract type with the specified type.

Abbildung A.1: XSD2OWL-Mapping der Java-Bibliothek Ontmalizer [Yük13, S. 46]

B Zusätzliche Ergebnisse der Datenanalyse

In diesem Anhang stellen wir zusätzliche Ergebnisse der Analysen des G&R-Datensatzes aus den Abschnitten 3.1.1.3 und 3.2.3.1 vor.

B.1 Auswertung der Redundanz im G&R-Datensatz

Die folgenden Abbildungen enthalten detaillierte Ergebnisse aus der explorativen Studie zu Ausmaß und Arten von Redundanz im G&R-Datensatz, die in Abschnitt 3.1.1.3 vorgestellt wurde. In Abbildung B.1 ist das Ausmaß der Redundanz für einfache Klonklassen verteilt auf die einzelnen Attribute dargestellt, in Abbildung B.2 hingegen für kombinierte Klonklassen über die Attributgrenzen hinweg.

Attribut (A)	Lücke=0				Lücke>0			
	Gesamtanzahl (A^0_{ges})	davon Anzahl einfach	erw. einfach	Anteil A^0_{ges} / A_{ges}	Gesamtanzahl (A^1_{ges})	davon Anzahl einfach	erw. einfach	Anteil A^1_{ges} / A_{ges}
HE;Effect	4847	4645 (96%)	202 (4%)	84,16%	912	815 (89%)	97 (11%)	15,84%
IsoAsil;ConCom	4445	4066 (91%)	379 (9%)	80,05%	1108	921 (83%)	187 (17%)	19,95%
IsoAsil;SevCom	3483	3263 (94%)	220 (6%)	85,68%	582	544 (93%)	38 (7%)	14,32%
DHES;RR	3040	2901 (95%)	139 (5%)	83,36%	607	523 (86%)	84 (14%)	16,64%
MF;Desc	2164	1387 (64%)	777 (36%)	68,79%	982	834 (85%)	148 (15%)	31,21%
OS;IU	2659	1803 (68%)	856 (32%)	89,56%	310	248 (80%)	62 (20%)	10,44%
IsoAsil;ExpCom	1333	1200 (90%)	133 (10%)	84,63%	242	216 (89%)	26 (11%)	15,37%
MF;Name	832	602 (72%)	230 (28%)	63,51%	478	400 (84%)	78 (16%)	36,49%
DHES;Com	1089	877 (81%)	212 (19%)	87,33%	158	95 (60%)	63 (40%)	12,67%
OS;Loc	1106	1097 (99%)	9 (1%)	100,00%	0	0	0	0,00%
Hazard;Desc	520	498 (96%)	22 (4%)	81,38%	119	112 (94%)	7 (6%)	18,62%
SG;Name	464	230 (50%)	234 (50%)	76,19%	145	101 (70%)	44 (30%)	23,81%
DHES;SGImport	428	190 (44%)	238 (56%)	86,99%	64	32 (50%)	32 (50%)	13,01%
PI;Desc	175	167 (95%)	8 (5%)	45,10%	213	205 (96%)	8 (4%)	54,90%
DPIS;Bound	158	147 (93%)	11 (7%)	60,08%	105	103 (98%)	2 (2%)	39,92%
SG;SafeState	133	127 (95%)	6 (5%)	89,26%	16	16 (100%)	0 (0%)	10,74%
DPIS;Assump	76	57 (75%)	19 (25%)	54,29%	64	51 (80%)	13 (20%)	45,71%
SG;Desc	72	47 (65%)	25 (35%)	62,07%	44	28 (64%)	16 (36%)	37,93%
IF;Name	63	51 (81%)	12 (19%)	76,83%	19	19 (100%)	0 (0%)	23,17%
IF;Desc	31	28 (90%)	3 (10%)	44,29%	39	31 (79%)	8 (21%)	55,71%
OS;RC	13	2 (15%)	11 (85%)	100,00%	0	0	0	0,00%
Interface;TextContent	2	2 (100%)	0 (0%)	100,00%	0	0	0	0,00%
	∑=27133	∑=23387	∑=3746	Ø=77,43%	∑=6207	∑=5294	∑=913	Ø=22,57%

Abbildung B.1: Analyse einfacher Klonklassen [Sei18, S. 36]

Attributkombination (AK)	Anzahl	Anteil	Lücke=0	Lücke>0
OS;IU#IsoAsil;ExpCom	3031	21,13%	90,94%	9,06%
MF;Name#MF;Desc	2296	16,01%	76,76%	23,24%
IsoAsil;SevCom#HE;Effect	1380	9,62%	73,26%	26,74%
SG;Name#DHES;SGImport	1235	8,61%	89,27%	10,73%
IsoAsil;ConCom#DHES;RR	980	6,83%	81,48%	18,52%
SG;Desc#MF;Desc	879	6,13%	89,87%	10,13%
OS;DECom#IsoAsil;ExpCom	810	5,65%	100,00%	0,00%
IsoAsil;ConCom#DHES;Com	614	4,28%	94,46%	5,54%
MF;Desc#DHES;Com	331	2,31%	95,32%	4,68%
SG;Name#DHES;Com	308	2,15%	71,10%	28,90%
SG;Name#SG;Desc	307	2,14%	88,11%	11,89%
SG;Name#MF;Desc	234	1,63%	89,53%	10,47%
MF;Desc#IsoAsil;SevCom	186	1,30%	45,70%	54,30%
IsoAsil;SevCom#IsoAsil;ConCom	121	0,84%	93,80%	6,20%
MF;Desc#IF;Desc	103	0,72%	52,43%	47,57%
HE;Effect#DHES;RR	101	0,70%	72,28%	27,72%
Top 16 AKs	∑=12916	∑=90,05%	Ø=84,79%	Ø=15,21%
Alle AKs	∑=14343	∑=100%	Ø=83,56%	Ø=16,44%

Abbildung B.2: Analyse kombinierter Klonklassen [Sei18, S. 38]

B.2 Auswertung der Tokenanzahl in Tripeln mit Literalen

Die folgenden Diagramme stellen eine Ergänzung zu der in Abschnitt 3.2.3.1 präsentierten Auswertung der Tokenanzahl in den Textinhalten der Literalknoten des Basisnetzes dar. Abbildung B.3 zeigt die durchschnittliche Tokenanzahl der verwendeten Prädikate mit zugehörigen Typen, welche nach den ihnen zugeordneten Teilartefakten gruppiert sind.

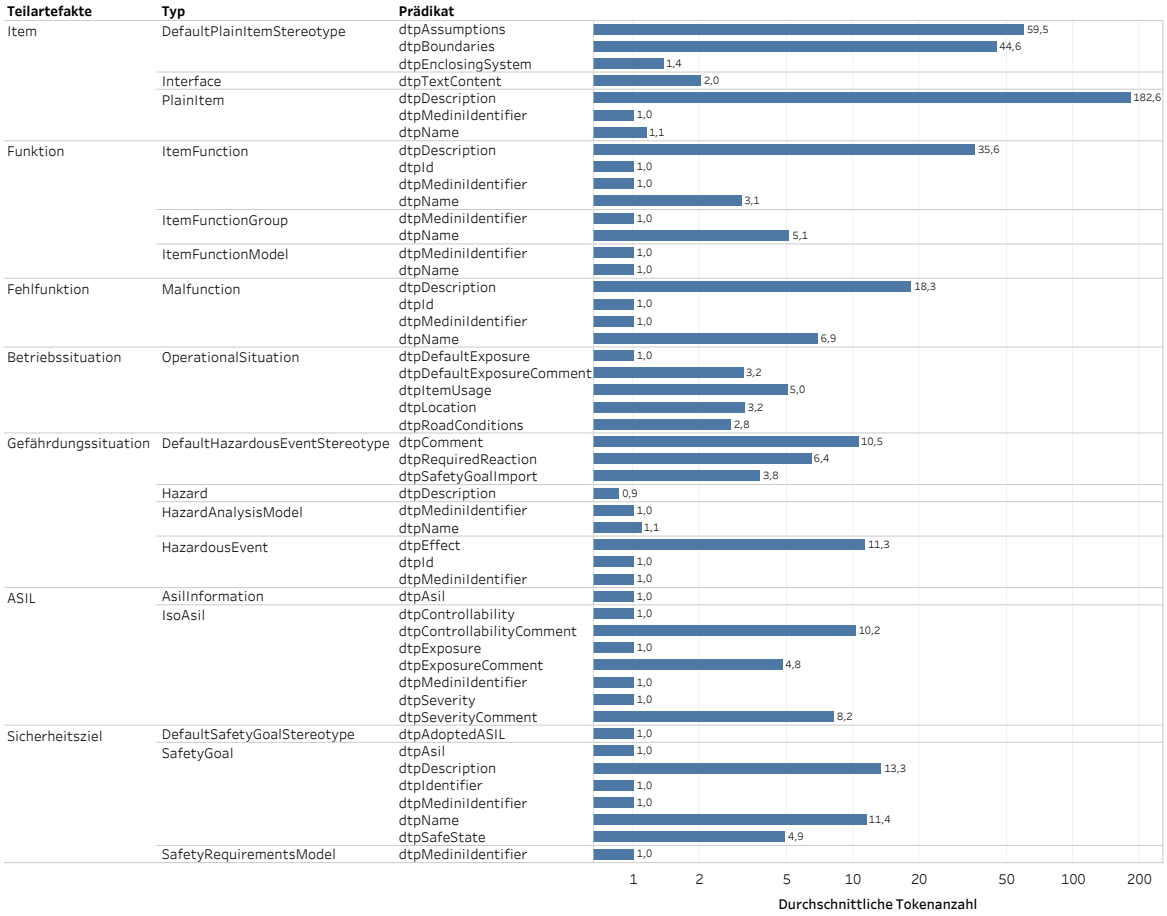


Abbildung B.3: Durchschnittliche Tokenanzahl je Prädikat gruppiert nach Typen und zugeordneten Teilartefakten

Im Vergleich zu Abbildung 3.9 (Seite 75), in der das Vorkommen der Prädikate der verschiedenen Typen aggregiert dargestellt ist, detailliert diese Abbildung das typspezifische Vorkommen. Die Unterschiede sind allerdings gering. Neben einigen Identifizierern, die die Tokenanzahl 1 besitzen und die daher keine Veränderung bei der Aufhebung der Aggregation verursachen, ist der Hauptunterschied die detailliertere Betrachtung des Prädikats `dtpName` im Teilartefakt Funktion. Während es in Funktionsgruppen mit einer durchschnittlichen Tokenanzahl von 5,1 auftaucht, haben Funktionen durchschnittlich lediglich 3,1 und Funktionsmodelle nur 1 Token. Abbildung B.4 aggregiert die Tokenanzahl für jedes Prädikat.

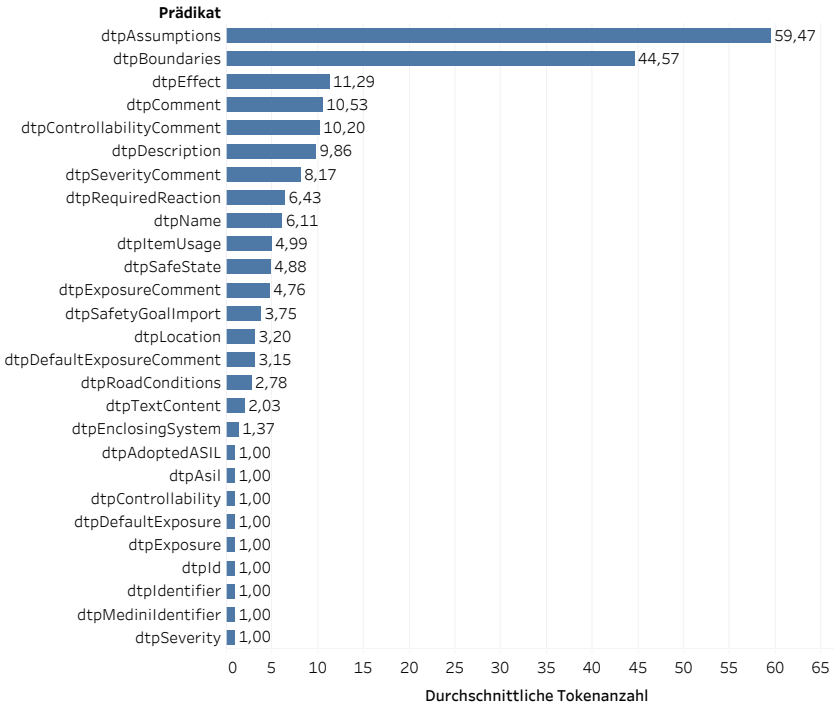


Abbildung B.4: Durchschnittliche Tokenanzahl je Prädikat

C | Zusätzliche Evaluationsergebnisse

In diesem Anhang bilden wir zusätzliche Ergebnisse aus der Evaluation ab.

C.1 Urteilerübereinstimmung

In Abschnitt C.1.1 stellen wir als Ergänzung zu der Übereinstimmungsanalyse in Abschnitt 4.2.1 die Resultate der Analysen unter Berücksichtigung der Ratings als ordinalskalierte Daten mit linearer und quadratischer Gewichtungsmatrix dar. Zudem bilden wir in Abschnitt C.1.2 die Ergebnisse der Übereinstimmungsanalyse für reduzierte Pools ab. Dafür interpretieren wir die Ratings ebenfalls zum einen als nominalskaliert und verwenden zum anderen eine lineare und quadratische Gewichtungsmatrix für ordinalskalierte Daten. In Abschnitt C.1.3 präsentieren wir die Ergebnisse der Konkordanzanalyse für reduzierte Pools.

C.1.1 BP- und Gwet-Koeffizienten für nicht reduzierte Poolgröße

Anfrage	**pa**	κ_{BP}	**se**	$\mathbf{CI}^{-}$	$\mathbf{CI}^{+}$	**t**	$\mathbf{t}_{1-\alpha,\mathrm{N}-1}$
Q23	0.9539	0.8894	0.0193	0.8515	0.9274	46.1591	1.6513
Q35	0.9543	0.8904	0.0152	0.8605	0.9202	58.629	1.6491
Q46	0.9281	0.8275	0.0265	0.7754	0.8797	31.2663	1.6514
Q54	0.8443	0.6264	0.0359	0.5556	0.6971	17.4683	1.6531
Q61	0.9433	0.8639	0.0252	0.8143	0.9136	34.3009	1.6527
Q64	0.9567	0.896	0.0224	0.8518	0.9402	40.0104	1.6533
Q81	0.9972	0.9934	0.0028	0.9879	0.9989	352.182	1.6476
Q85	0.9776	0.9496	0.0079	0.9341	0.9651	120.5327	1.6473

Tabelle C.1: Ergebnisse der linear gewichteten κ_{BP}-Analyse

Anfrage	**pa**	κ_{BP}	**se**	$\mathbf{CI}^{-}$	$\mathbf{CI}^{+}$	**t**	$\mathbf{t}_{1-\alpha,\mathrm{N}-1}$
Q23	0.974	0.9064	0.0199	0.8672	0.9456	45.5878	1.6513
Q35	0.979	0.9244	0.0127	0.8994	0.9495	72.5481	1.6491
Q46	0.9568	0.8445	0.0271	0.7912	0.8979	31.1854	1.6514
Q54	0.9089	0.672	0.0398	0.5935	0.7505	16.8813	1.6531
Q61	0.9621	0.8636	0.0292	0.8061	0.9211	29.6176	1.6527
Q64	0.9744	0.9077	0.0231	0.8622	0.9532	39.3738	1.6533
Q81	0.9988	0.9958	0.0022	0.9915	1	461.5558	1.6476
Q85	0.9882	0.9659	0.0075	0.9512	0.9805	129.0862	1.6473

Tabelle C.2: Ergebnisse der quadratisch gewichteten κ_{BP}-Analyse

Anfrage	**pa**	**Gwet AC_2**	**se**	**CI^-**	**CI^+**	**t**	**$t_{1-\alpha,N-1}$**
Q23	0.9539	0.9464	0.0106	0.9256	0.9672	101.3145	1.6513
Q35	0.9543	0.9471	0.0082	0.931	0.9632	129.3035	1.6491
Q46	0.9281	0.9118	0.0154	0.8814	0.9421	67.3598	1.6514
Q54	0.8443	0.7769	0.0274	0.7228	0.8309	36.2771	1.6531
Q61	0.9433	0.934	0.0136	0.9071	0.9608	76.4011	1.6527
Q64	0.9567	0.951	0.0116	0.9282	0.9738	90.1581	1.6533
Q81	0.9972	0.9972	0.0012	0.9949	0.9996	839.1937	1.6476
Q85	0.9776	0.9762	0.0039	0.9685	0.9838	262.3971	1.6473

Tabelle C.3: Ergebnisse der linear gewichteten Gwet-AC_2-Analyse

Anfrage	**pa**	**Gwet AC_2**	**se**	**CI^-**	**CI^+**	**t**	**$t_{1-\alpha,N-1}$**
Q23	0.974	0.9685	0.0075	0.9538	0.9832	144.8605	1.6513
Q35	0.979	0.9748	0.0047	0.9655	0.984	228.9748	1.6491
Q46	0.9568	0.944	0.0111	0.922	0.9659	96.6949	1.6514
Q54	0.9089	0.8544	0.0223	0.8105	0.8983	48.3458	1.6531
Q61	0.9621	0.9541	0.0109	0.9325	0.9756	97.1554	1.6527
Q64	0.9744	0.9701	0.0082	0.954	0.9862	129.8399	1.6533
Q81	0.9988	0.9988	6e-04	0.9976	1	1644.0829	1.6476
Q85	0.9882	0.9873	0.0029	0.9817	0.9929	355.2811	1.6473

Tabelle C.4: Ergebnisse der quadratisch gewichteten Gwet-AC_2-Analyse

C.1.2 BP- und Gwet-Koeffizienten für reduzierte Poolgröße

Anfrage	**pa**	κ_{BP}	**se**	$\mathbf{CI}^{-}$	$\mathbf{CI}^{+}$	**t**	$\mathbf{t}_{1-\alpha,\mathrm{N}-1}$
Q23	0.5125	0.35	0.0704	0.2075	0.4925	4.9683	1.6849
Q35	0.5185	0.358	0.0714	0.2141	0.5019	5.0138	1.6802
Q46	0.7822	0.7096	0.0556	0.5988	0.8205	12.7536	1.6657
Q54	0.7778	0.7037	0.0509	0.6024	0.805	13.8373	1.6657
Q61	0.7815	0.7087	0.0493	0.6105	0.807	14.376	1.666
Q64	0.6833	0.5778	0.0664	0.4444	0.7111	8.7062	1.6766
Q81	0.9407	0.921	0.0381	0.8441	0.9978	24.1503	1.6802
Q85	0.7933	0.69	0.0683	0.5526	0.8274	10.0954	1.6766

Tabelle C.5: Ergebnisse der κ_{BP}-Analyse (reduzierter Pool)

Anfrage	**pa**	κ_{BP}	**se**	$\mathbf{CI}^{-}$	$\mathbf{CI}^{+}$	**t**	$\mathbf{t}_{1-\alpha,\mathrm{N}-1}$
Q23	0.7806	0.4733	0.0693	0.3331	0.6136	6.8263	1.6849
Q35	0.8123	0.5496	0.0586	0.4315	0.6677	9.3779	1.6802
Q46	0.8815	0.7156	0.0585	0.5991	0.832	12.2416	1.6657
Q54	0.8844	0.7227	0.0539	0.6152	0.8301	13.4035	1.6657
Q61	0.8881	0.7315	0.0498	0.6323	0.8307	14.6953	1.666
Q64	0.8422	0.6213	0.068	0.4847	0.7579	9.141	1.6766
Q81	0.9753	0.9407	0.0304	0.8794	1	30.918	1.6802
Q85	0.9022	0.78	0.0574	0.6646	0.8954	13.5823	1.6766

Tabelle C.6: Ergebnisse der linear gewichteten κ_{BP}-Analyse (reduzierter Pool)

Anfrage	**pa**	κ_{BP}	**se**	$\mathbf{CI}^{-}$	$\mathbf{CI}^{+}$	**t**	$\mathbf{t}_{1-\alpha,\mathrm{N}-1}$
Q23	0.8769	0.5567	0.0863	0.3821	0.7312	6.4516	1.6849
Q35	0.916	0.6978	0.062	0.5727	0.8228	11.2456	1.6802
Q46	0.922	0.7191	0.0692	0.5812	0.857	10.3911	1.6657
Q54	0.9279	0.7404	0.0634	0.6141	0.8668	11.6746	1.6657
Q61	0.9302	0.7486	0.0582	0.6326	0.8647	12.857	1.666
Q64	0.9067	0.664	0.074	0.5153	0.8127	8.9741	1.6766
Q81	0.9885	0.9585	0.0254	0.9074	1	37.7923	1.6802
Q85	0.9407	0.8286	0.0621	0.7037	0.9534	13.3348	1.6766

Tabelle C.7: Ergebnisse der quadratisch gewichteten κ_{BP}-Analyse (reduzierter Pool)

Anfrage	**pa**	**Gwet AC_1**	**se**	**CI^-**	**CI^+**	**t**	**$t_{1-\alpha,N-1}$**
Q23	0.5125	0.3832	0.0758	0.2298	0.5366	5.7317	1.6849
Q35	0.5185	0.4045	0.0713	0.2608	0.5482	6.1076	1.6802
Q46	0.7822	0.7544	0.0509	0.653	0.8558	16.029	1.6657
Q54	0.7778	0.7531	0.0465	0.6605	0.8458	17.7744	1.6657
Q61	0.7815	0.75	0.0463	0.6577	0.8422	17.7228	1.666
Q64	0.6833	0.6258	0.0661	0.493	0.7585	10.6384	1.6766
Q81	0.9407	0.9378	0.0305	0.8763	0.9993	31.2579	1.6802
Q85	0.7933	0.7581	0.0589	0.6397	0.8764	14.2106	1.6766

Tabelle C.8: Ergebnisse der Gwet-AC_1-Analyse (reduzierter Pool)

Anfrage	**pa**	**Gwet AC_2**	**se**	**CI^-**	**CI^+**	**t**	**$t_{1-\alpha,N-1}$**
Q23	0.7806	0.5704	0.079	0.4105	0.7302	10.0841	1.6849
Q35	0.8123	0.6609	0.0561	0.5478	0.774	14.9749	1.6802
Q46	0.8815	0.8389	0.0398	0.7596	0.9182	25.3363	1.6657
Q54	0.8844	0.8494	0.0356	0.7785	0.9202	29.0006	1.6657
Q61	0.8881	0.8414	0.0366	0.7684	0.9144	28.6159	1.666
Q64	0.8422	0.7539	0.0581	0.6371	0.8707	17.0639	1.6766
Q81	0.9753	0.9723	0.0148	0.9425	1	68.3325	1.6802
Q85	0.9022	0.8708	0.0396	0.7913	0.9504	25.8266	1.6766

Tabelle C.9: Ergebnisse der linear gewichteten Gwet-AC_2-Analyse (reduzierter Pool)

Anfrage	**pa**	**Gwet AC_2**	**se**	**CI^-**	**CI^+**	**t**	**$t_{1-\alpha,N-1}$**
Q23	0.8769	0.6877	0.083	0.5198	0.8556	11.3138	1.6849
Q35	0.916	0.8122	0.0478	0.716	0.9084	21.0639	1.6802
Q46	0.922	0.884	0.0341	0.816	0.952	30.9318	1.6657
Q54	0.9279	0.8987	0.0296	0.8397	0.9576	36.2987	1.6657
Q61	0.9302	0.8901	0.0315	0.8273	0.9528	34.9585	1.666
Q64	0.9067	0.832	0.0492	0.7332	0.9308	22.491	1.6766
Q81	0.9885	0.9867	0.0084	0.9697	1	121.1502	1.6802
Q85	0.9407	0.917	0.0341	0.8484	0.9856	30.4713	1.6766

Tabelle C.10: Ergebnisse der quadratisch gewichteten Gwet-AC_2-Analyse (reduzierter Pool)

C.1.3 Konkordanzkoeffizient nach Kendall für reduzierte Poolgröße

Anfrage	**# Rater m**	**# Bewertungs-elemente N**	**W**	χ^2_r	$\chi^2_{1-\alpha,N-1}$	**p-Wert**
Q23	4	40	0.6141	95.7945	54.5722	1.11e-06
Q35	3	45	0.6352	83.8452	60.4809	0.0003
Q46	3	75	0.7209	160.042	95.0815	2.78e-08
Q54	3	75	0.6495	144.1941	95.0815	1.96e-06
Q61	4	74	0.6948	202.8748	93.9453	3.62e-14
Q64	4	50	0.688	134.8464	66.3386	5.98e-10
Q81	3	45	0.7419	97.9297	60.4809	5.58e-06
Q85	3	50	0.6258	91.994	66.3386	0.0002

Tabelle C.11: Ergebnisse der Konkordanzanalyse nach Kendall (reduzierter Pool)

C.2 Paarweise Übereinstimmung zwischen zusammengefasster objektiver Relevanz und Raterbewertung

In diesem Abschnitt werden die Ergebnisse für die in Abschnitt 4.2.3 vorgestellte Übereinstimmungsanalyse dokumentiert. Tabelle C.12 enthält die Ergebnisse für die Rangreihenübereinstimmung zwischen der zusammengefassten gerankten objektiven Relevanz (OR) und den Beurteilungen der einzelnen Rater. Tabelle C.13 enthält die Ergebnisse für die Übereinstimmung zwischen der zusammengefassten dichotomisierten objektiven Relevanz (ORD1 und ORD2) und den ebenfalls dichotomisierten Beurteilungen der einzelnen Rater.

Anfrage	**Raterpaar**	**W**	χ^2_r	$\chi^2_{1-\alpha,Fg}$	**p-Wert**
Q23	R1 OR	0,9642	460,8692	276,0624	3,07e-16
	R2 OR	0,8465	404,6233	276,0624	1,17e-10
	R3 OR	0,7926	378,8619	276,0624	2,08e-08
	R4 OR	0,9713	464,3049	276,0624	1,32e-16
Q35	R1 OR	0,9683	695,2694	404,1821	1,01e-23
	R3 OR	0,5745	412,4724	404,1821	0,0269
	R4 OR	0,9777	702,0203	404,1821	1,92e-24
Q46	R1 OR	0,9134	427,4843	270,6839	1,76e-13
	R3 OR	0,7432	347,8387	270,6839	1,91e-06
	R4 OR	0,9983	467,1873	270,6839	1,14e-17
Q54	R1 OR	0,7418	275,9359	218,8205	2,03e-05
	R2 OR	0,8732	324,8412	218,8205	1,26e-09
	R4 OR	0,9345	347,6416	218,8205	6,71e-12
Q61	R1 OR	0,9234	358,2741	227,4964	6,90e-12
	R2 OR	0,8531	331,0116	227,4964	3,16e-09
	R3 OR	0,8024	311,3218	227,4964	1,80e-07
	R4 OR	0,9033	350,4957	227,4964	4,20e-11
Q64	R1 OR	0,7507	271,7649	213,3906	1,45e-05
	R2 OR	0,8392	303,7808	213,3906	2,92e-08
	R3 OR	0,9089	329,0159	213,3906	1,11e-10
	R4 OR	0,9454	342,2466	213,3906	4,84e-12
Q81	R1 OR	0,9279	1046,6852	620,357	2,25e-31
	R3 OR	0,7017	791,5309	620,357	7,19e-10
	R4 OR	0,8343	941,117	620,357	2,30e-21
Q85	R1 OR	0,6592	809,5075	672,7548	1,74e-07
	R3 OR	0,6872	843,8883	672,7548	1,79e-09
	R4 OR	0,9755	1197,963	672,7548	4,84e-40

Tabelle C.12: Konkordanzkoeffizient nach Kendall für jedes Rater-OR-Paar

Anfrage	**Raterpaar**	**pa**	$\kappa_{\mathbf{BP}}$	**se**	$\mathbf{CI}^{-}$	$\mathbf{CI}^{+}$	**t**	$\mathbf{t}_{1-\alpha,N-1}$
Q23	ORD1 R1	0,9792	0,9583	0,0185	0,9219	0,9947	51,8654	1,6513
	ORD1 R2	0,9292	0,8583	0,0332	0,793	0,9237	25,8618	
	ORD1 R3	0,9125	0,825	0,0366	0,753	0,897	22,5685	
	ORD1 R4	0,9833	0,9667	0,0166	0,934	0,9993	58,3675	
	ORD2 R1	0,9833	0,9667	0,0166	0,934	0,9993	58,3675	
	ORD2 R2	0,9583	0,9167	0,0259	0,8657	0,9676	35,4591	
	ORD2 R3	0,9333	0,8667	0,0323	0,8031	0,9302	26,8564	
	ORD2 R4	0,9875	0,975	0,0144	0,9467	1	67,8344	
Q35	ORD1 R1	0,9806	0,9611	0,0146	0,9324	0,9898	65,9412	1,6491
	ORD1 R3	0,8583	0,7167	0,0368	0,6443	0,7891	19,4703	
	ORD1 R4	0,9861	0,9722	0,0124	0,9479	0,9965	78,7021	
	ORD2 R1	0,9833	0,9667	0,0135	0,9401	0,9932	71,5351	
	ORD2 R3	0,8889	0,7778	0,0332	0,7125	0,843	23,4461	
	ORD2 R4	0,9833	0,9667	0,0135	0,9401	0,9932	71,5351	
Q46	ORD1 R1	0,9362	0,8723	0,032	0,8094	0,9353	27,2945	1,6514
	ORD1 R3	0,8553	0,7106	0,046	0,62	0,8013	15,451	
	ORD1 R4	1	1	0	1	1	-	
	ORD2 R1	0,9915	0,983	0,012	0,9593	1	81,8459	
	ORD2 R3	0,9106	0,8213	0,0373	0,7478	0,8948	22,0201	
	ORD2 R4	0,9447	0,8894	0,0299	0,8305	0,9482	29,7561	
Q54	ORD1 R1	0,6631	0,3262	0,0693	0,1895	0,4629	4,7063	1,6531
	ORD1 R2	0,7701	0,5401	0,0617	0,4184	0,6618	8,7525	
	ORD1 R4	0,9251	0,8503	0,0386	0,7741	0,9264	22,0311	
	ORD2 R1	0,9251	0,8503	0,0386	0,7741	0,9264	22,0311	
	ORD2 R2	0,9358	0,8717	0,0359	0,8008	0,9426	24,2552	
	ORD2 R4	0,7807	0,5615	0,0607	0,4418	0,6812	9,2544	
Q61	ORD1 R1	0,959	0,9179	0,0285	0,8618	0,9741	32,2298	1,6527
	ORD1 R2	0,9231	0,8462	0,0383	0,7707	0,9216	22,1143	
	ORD1 R3	0,9026	0,8051	0,0426	0,7211	0,8891	18,9076	
	ORD1 R4	0,9487	0,8974	0,0317	0,835	0,9599	28,335	
	ORD2 R1	0,959	0,9179	0,0285	0,8618	0,9741	32,2298	
	ORD2 R2	0,9333	0,8667	0,0358	0,796	0,9373	24,1964	
	ORD2 R3	0,9231	0,8462	0,0383	0,7707	0,9216	22,1143	
	ORD2 R4	0,959	0,9179	0,0285	0,8618	0,9741	32,2298	
Q64	ORD1 R1	0,9066	0,8132	0,0433	0,7278	0,8985	18,7977	1,6533
	ORD1 R2	0,9286	0,8571	0,0383	0,7816	0,9327	22,3882	
	ORD1 R3	0,956	0,9121	0,0305	0,852	0,9722	29,9294	
	ORD1 R4	0,9725	0,9451	0,0243	0,8971	0,993	38,8925	
	ORD2 R1	0,9396	0,8791	0,0354	0,8092	0,949	24,8162	
	ORD2 R2	0,9725	0,9451	0,0243	0,8971	0,993	38,8925	
	ORD2 R3	0,978	0,956	0,0218	0,913	0,999	43,865	
	ORD2 R4	0,9615	0,9231	0,0286	0,8667	0,9795	32,2887	
Q81	ORD1 R1	0,9965	0,9929	0,005	0,9831	1	198,5195	1,6476
	ORD1 R3	0,9894	0,9788	0,0086	0,9618	0,9957	113,3844	
	ORD1 R4	0,9929	0,9858	0,0071	0,972	0,9997	139,6218	
	ORD2 R1	0,9929	0,9858	0,0071	0,972	0,9997	139,6218	
	ORD2 R3	1	1	0	1	1	-	
	ORD2 R4	0,9965	0,9929	0,005	0,9831	1	198,5195	
Q85	ORD1 R1	0,9317	0,8634	0,0204	0,8234	0,9034	42,4079	1,6473
	ORD1 R3	0,935	0,8699	0,0199	0,8308	0,909	43,7063	
	ORD1 R4	0,9935	0,987	0,0065	0,9743	0,9997	152,1221	
	ORD2 R1	1	1	0	1	1	-	
	ORD2 R3	0,9935	0,987	0,0065	0,9743	0,9997	152,1221	
	ORD2 R4	0,935	0,8699	0,0199	0,8308	0,909	43,7063	

Tabelle C.13: Brennan-Prediger-Koeffizient-Analyse für jedes Rater-OR-Paar

C.3 Auswertung der Evaluationsergebnisse für zufällig generierte Konfigurationen

C.3.1 Wahl des kombinierten Bewertungsmaßes

In den Abschnitten 4.3 und 4.4 verwenden wir ein kombiniertes Bewertungsmaß, welches gleichberechtigt EP und $NDCG_{Min}$ berücksichtigt. Bei der Wahl des kombinierten Maßes (siehe Formeln 4.1 und 4.2, Seite 137) bilden folgende Überlegungen die Grundlage.

In Abschnitt 2.1.3.3 haben wir bereits erläutert, dass aufgrund identischer Werte für die Systemrelevanz von verschiedenen Bewertungselementen eine schwache Ordnung der Distributionen erwartet werden muss. Das bedeutet, dass ein Rang mehrere Bewertungselemente enthalten kann, deren Reihenfolge für die Berechnung des NDCG-Wertes für die Nützlichkeit nicht hierarchisiert werden kann. Daher haben wir die Berechnung der pessimistischen und optimistischen NDCG-Werte $NDCG_{Min}$ und $NDCG_{Max}$ eingeführt, welchen jeweils die ungünstigste und die bestmögliche Reihenfolge innerhalb der Ränge zugrunde liegt. Wir verdeutlichen die Zusammenhänge zwischen den Bewertungsmaßen im Folgenden am Beispiel der Ergebnisse, die wir mit zufällig generierten Konfigurationen erhalten haben (siehe Abschnitt 4.3). Abbildung C.1 stellt das Streudiagramm der $NDCG_{Min}$- und $NDCG_{Max}$-Werte aller Konfigurationen dar. Es ist deutlich erkennbar, dass viele Durchläufe identische $NDCG_{Min}$- und $NDCG_{Max}$-Werte und damit keine Differenz aufweisen. Die Werte oberhalb der Linie zeigen, dass große NDCG-Differenzen existieren, im extremsten Fall 0,87, wobei eine Verdichtung im oberen $NDCG_{Min}$- und $NDCG_{Max}$-Bereich erkennbar ist.

Eine Gegenüberstellung der berechneten EP-Werte und der Werte $NDCG_{Min}$, $NDCG_{Max}$ und $NDCG_{Diff}$ zeigen in Abbildung C.2 eine höhere Anfälligkeit für Ausreißer bei den $NDCG_{Max}$-Werten als bei den $NDCG_{Min}$-Werten. Die dunkelgrauen Linien stellen Trendlinien der linearen Regression dar. Durchläufe, die einen sehr geringen $NDCG_{Diff}$-Wert besitzen, verfügen gleichzeitig über EP-Werte des gesamten Wertebereichs, während sehr hohe $NDCG_{Diff}$-Werte mit sehr geringen EP-Werten einhergehen. Im mittleren $NDCG_{Diff}$- und EP-Bereich hingegen werden maximal mittelhohe Werte erreicht. Das Streudiagramm deutet darauf hin, dass der Mittelwert aus $NDCG_{Min}$ und $NDCG_{Max}$ ungeeignet ist für eine aussagekräftige Bewertung eines Durchlaufes. Ebenso ist erkennbar, dass sehr hohe $NDCG_{Max}$-Werte auch mit sehr geringen EP-Werten assoziiert werden. Die Streuung der $NDCG_{Min}$-Werte in Bezug auf assoziierte EP-Werte zeigt eine stärkere Korrelation. Daher verwenden wir die berechneten $NDCG_{Min}$-Werte in Kombination mit den EP-Werten für eine gesamtheitliche Bewertung einzelner Durchläufe. Damit diskriminieren wir insbesondere Distributionen, die zum einen viele Bewertungselemente auf identischen Rängen vereinen und die zum anderen sehr unterschiedliche Relevanzgrade besitzen. Die Streudiagramme in Abbildung C.3 zeigen die resultierende Korrelation zwischen der berechneten Fitness und den $NDCG_{Min}$- und EP-Werten.

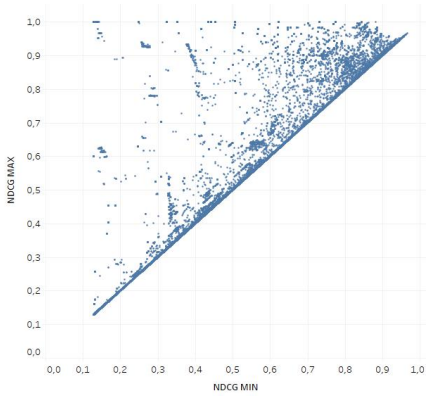


Abbildung C.1: Streudiagramm der NDCG_{Min}- und NDCG_{Max}-Werte jedes Durchlaufes

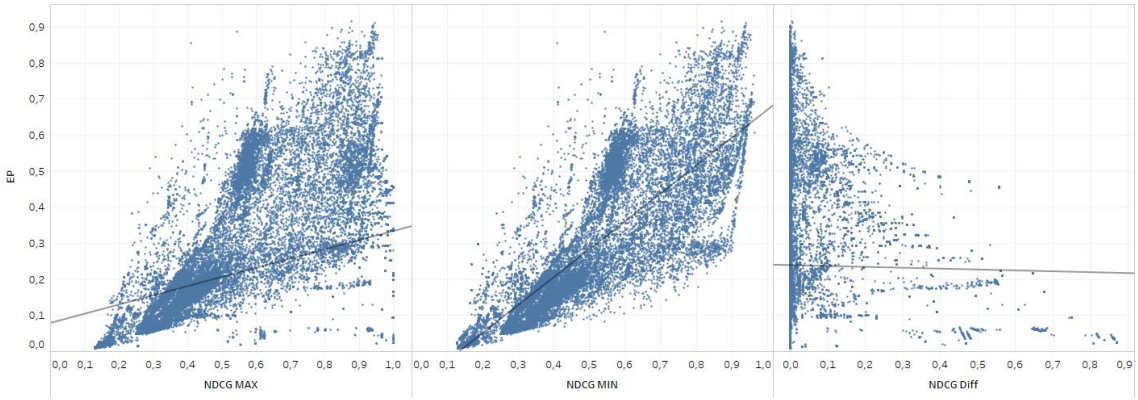


Abbildung C.2: Streudiagramm der EP- und verschiedenen NDCG-Werte jedes Durchlaufes

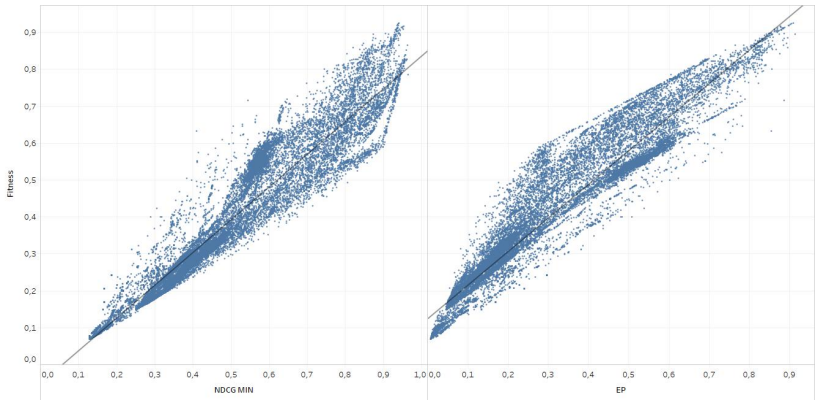


Abbildung C.3: Streudiagramm der Fitness und NDCG_{Min}- und EP-Werte jedes Durchlaufes