

Fatih Karakol

Konzeption einer ontologiebasierten Schnittstelle zur Integration von verteilt vorliegenden Informationsquellen

A stylized illustration of a cityscape with various buildings and a large crowd of people in the foreground, rendered in shades of orange and red. The buildings are simplified geometric shapes with some windows. The crowd consists of many small, simplified human figures. The overall style is modern and graphic.

disserta

Verlag

Karakol, Fatih: Konzeption einer ontologiebasierten Schnittstelle zur Integration von verteilt vorliegenden Informationsquellen. Hamburg, disserta Verlag, 2016

Buch-ISBN: 978-3-95935-244-4

PDF-eBook-ISBN: 978-3-95935-245-1

Druck/Herstellung: disserta Verlag, Hamburg, 2016

Covermotiv: © laurine45 – Fotolia.com

Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Bearbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Die Informationen in diesem Werk wurden mit Sorgfalt erarbeitet. Dennoch können Fehler nicht vollständig ausgeschlossen werden und die Diplomica Verlag GmbH, die Autoren oder Übersetzer übernehmen keine juristische Verantwortung oder irgendeine Haftung für evtl. verbliebene fehlerhafte Angaben und deren Folgen.

Alle Rechte vorbehalten

© disserta Verlag, Imprint der Diplomica Verlag GmbH
Hermannstal 119k, 22119 Hamburg
<http://www.disserta-verlag.de>, Hamburg 2016
Printed in Germany

Inhaltsverzeichnis

| | |
|---|-----------|
| Abkürzungsverzeichnis..... | VII |
| Abbildungsverzeichnis | VIII |
| Tabellenverzeichnis..... | X |
| 1 Einleitung..... | 1 |
| 1.1 Ausgangslage..... | 1 |
| 1.2 Motivation..... | 2 |
| 1.3 Zielsetzung..... | 4 |
| 1.4 Vorgehen..... | 4 |
| 2 Problematik des Informationsaufkommens..... | 6 |
| 2.1 Suche nach Informationen | 8 |
| 2.2 Intelligente Suche nach Informationen | 10 |
| 2.3 Der Weg vom Web zum Semantic Web | 10 |
| 2.3.1 Semantic Web..... | 15 |
| 2.3.2 Semantik im Semantic Web | 16 |
| 2.3.3 Prinzipien und Standards des Semantic Web | 17 |
| 2.3.4 Semantic Web und Informationsintegration..... | 19 |
| 2.4 Metadaten | 20 |
| 2.5 Terminologie..... | 21 |
| 2.5.1 Integrationsbegriff..... | 21 |
| 2.5.2 Was ist Information | 24 |
| 3 Ontologien..... | 31 |
| 3.1 Was ist eine Ontologie? | 31 |
| 3.1.1 Ontologie in der Philosophie | 32 |
| 3.1.2 Ontologie in der Informatik..... | 32 |
| 3.1.3 Definitionen in der Informatik..... | 34 |
| 3.2 Bestandteile einer Ontologie..... | 37 |
| 3.3 Klassifizierung von Ontologien..... | 40 |

| | | |
|----------|--|-----------|
| 3.3.1 | Unterscheidung nach der Allgemeingültigkeit..... | 41 |
| 3.3.2 | Unterscheidung nach dem Formalisierungsgrad..... | 42 |
| 3.3.3 | Unterscheidung nach der Komplexität..... | 43 |
| 3.4 | Warum werden Ontologien benötigt?..... | 45 |
| 3.5 | Ontologiesprachen..... | 48 |
| 3.5.1 | Anforderungen an Sprachen..... | 48 |
| 3.5.2 | Unterscheidung der Sprachen..... | 49 |
| 3.5.3 | Beschreibung der Sprachen..... | 52 |
| 3.6 | Auswahl einer Sprache..... | 56 |
| 3.7 | Ontologieentwicklung..... | 59 |
| 3.7.1 | Drei Stufen der Ontologieentwicklung..... | 61 |
| 3.7.2 | Methoden der Ontologieentwicklung..... | 62 |
| 3.8 | Werkzeuge..... | 66 |
| 4 | Entwurf des Konzepts der Schnittstelle..... | 70 |
| 4.1 | Anforderungsanalyse..... | 71 |
| 4.1.1 | Zielpublikum der Schnittstelle..... | 72 |
| 4.1.2 | Aufgaben und Strategie der Schnittstelle..... | 72 |
| 4.1.3 | Account eines Anwenders..... | 72 |
| 4.1.4 | Ontologien laden und graphisch darstellen..... | 73 |
| 4.2 | Allgemeine Anforderungen an die Schnittstelle..... | 76 |
| 4.3 | Ontologiebasierte Informationsintegration..... | 77 |
| 4.3.1 | Analyse der Informationsintegration..... | 79 |
| 4.3.2 | Operationen der Datenintegration..... | 80 |
| 4.4 | Klassifikation von Integrationskonflikten..... | 81 |
| 4.5 | Integrationsarchitekturen..... | 83 |
| 4.6 | Ansätze zur Integration von Ontologien..... | 85 |
| 4.7 | Integrationsprozess..... | 87 |
| 5 | Grundlegende Überlegungen..... | 90 |
| 5.1 | Informationsquellen..... | 91 |
| 5.2 | Arbeitsschritte bei der Informationsintegration..... | 91 |

| | | |
|----------|---|------------|
| 5.2.1 | Zugriff auf Informationsquelle | 92 |
| 5.2.2 | Hochladen eines Metadatenformats | 93 |
| 5.2.3 | Hochladen einer Ontologie..... | 94 |
| 5.2.4 | Zusammenführen zu einer globalen Ontologie | 95 |
| 5.2.5 | Nächster Schritt: Matching, Mapping und Merge | 100 |
| 5.2.6 | Integrationsaufgabe der Schnittstelle | 108 |
| 5.2.7 | Lösung der semantischen Heterogenität..... | 109 |
| 5.2.8 | Nutzungsszenarien der Schnittstelle..... | 112 |
| 5.2.9 | Zusammenfassung | 113 |
| 6 | Evaluation der Schnittstelle | 115 |
| 6.1 | Definition der Evaluation..... | 115 |
| 6.2 | Evaluationformen..... | 116 |
| 6.3 | Anwendbarkeit des gewählten Evaluierungstyps | 118 |
| 6.4 | Evaluationskriterien | 119 |
| 6.5 | Evaluationsmittel | 119 |
| 6.6 | Betrachtung der Implementierung gegen das Konzept..... | 120 |
| 6.6.1 | Durchführung der Evaluation auf den implementierten Prototypen | 120 |
| 6.6.2 | Testscenario..... | 122 |
| 7 | Zusammenfassung und Ausblick | 126 |
| 8 | Literaturverzeichnis..... | 128 |
| 9 | Anhang | 151 |

Abkürzungsverzeichnis

| | |
|--------|--|
| COMA | Combining Match Algorithm |
| DBMS | Database Management System |
| DAML | DARPA Agent Markup Language |
| DL | Description Logic |
| FOAM | Framework for Ontology Alignment and Mapping |
| HPKB | High Performance Knowledge Base |
| HTML | Hyper Text Markup Language |
| IKT | Informations- und Kommunikationstechnik |
| KI | künstliche Intelligenz |
| NLP | Natural Language Processing |
| OIL | Ontology Inference Layer |
| OKBC | Open Knowledge Base Connectivity |
| OWL | Ontology Web Language |
| RDF | Resource Description Framework |
| RDFS | RDF Schema |
| SAT | Propositional SATisfiability Techniques |
| SHOE | Simple HTML Ontology Extensions |
| SKOS | Simple Knowledge Organisation System |
| SPARQL | Simple Protocol And RDF Query Language |
| u.a. | unter anderen; unter anderem |
| URI | Uniform Resource Identifier |
| W3C | World Wide Web Consortium |
| WWW | World Wide Web |
| XML | Extensible Markup Language |
| XOL | Ontology Exchange Language |

Abbildungsverzeichnis

| | |
|---|----|
| Abbildung 2.1: Prozess der Internetentwicklung | 11 |
| Abbildung 2.2: Entwicklung der Webtechnologien | 14 |
| Abbildung 2.3: Themen in Beziehung zum Semantic Web | 16 |
| Abbildung 2.4: Architektur des Semantic Web..... | 18 |
| Abbildung 2.5: Dimensionen der Integrationsformen..... | 22 |
| Abbildung 2.6: Integrationsepochen | 23 |
| Abbildung 2.7: Handlungstheoretische Abgrenzung des Wissensbegriffes..... | 25 |
| Abbildung 2.8: Begriffshierarchie..... | 26 |
| Abbildung 2.9: Differenzierungsmerkmale..... | 27 |
| Abbildung 2.10: Bedeutungsdreieck | 28 |
| Abbildung 3.1: Formalisierung einer Ontologie | 34 |
| Abbildung 3.2: Gebräuchlichste Ontologiedefinition mit Erläuterungen | 35 |
| Abbildung 3.3: Struktur einer Ontologie..... | 38 |
| Abbildung 3.4: Beispiel einer Ontologie..... | 40 |
| Abbildung 3.5: Gruppen der relevanten Verfahren | 40 |
| Abbildung 3.6: Klassifikation von Ontologien nach Guarino | 42 |
| Abbildung 3.7: Ontologiespektrum..... | 44 |
| Abbildung 3.8: der semantischen Reihenfolge für Ontologien | 45 |
| Abbildung 3.9: Semantische Konflikte | 47 |
| Abbildung 3.10: Phasen der Ontologieentwicklung..... | 61 |
| Abbildung 3.11: Dreiteilung der Entwicklung einer Ontologie | 61 |
| Abbildung 3.12: Entwicklung einer Ontologie | 65 |
| Abbildung 3.13: Ontobroker-Architektur..... | 68 |
| Abbildung 4.1: Phasen im Web Engineering | 71 |
| Abbildung 4.2: Darstellung der Ontologien (einzeln und zusammengeführt) | 75 |
| Abbildung 4.3: Grundarchitektur der Informationsintegration | 78 |

| | |
|--|-----|
| Abbildung 4.4: Ontologiebasierte Datenintegrationsarchitektur | 79 |
| Abbildung 4.5: Grundidee..... | 79 |
| Abbildung 4.6: Architekturansätze mit Ontologien | 84 |
| Abbildung 4.7: Integrationsprozess von Ontologien..... | 88 |
| Abbildung 5.1: Komponenten der Architektur | 90 |
| Abbildung 5.2: Registrieren..... | 92 |
| Abbildung 5.3: Informationsmodell als Metadaten hochladen | 94 |
| Abbildung 5.4: Ontologie hochladen | 95 |
| Abbildung 5.5: Merge und Mapping der Ontologien..... | 95 |
| Abbildung 5.6: Klassifikation von Ontology Heterogenitäten..... | 97 |
| Abbildung 5.7: Schamtischer Matching Prozess | 100 |
| Abbildung 5.8: Klassifikation von Matching-Techniken..... | 102 |
| Abbildung 5.9: Zusammenhang zwischen Matching, Merging und Mapping..... | 105 |
| Abbildung 5.10: Matching-Prozess von Coma++ | 107 |
| Abbildung 5.11: Architektur der Schnittstelle | 114 |
| Abbildung 6.1: Formative und summative Evaluation in der Softwareentwicklung | 117 |
| Abbildung 9.1: Erfassungsbereich der Suchwerkzeuge im Internet | 151 |
| Abbildung 9.2: Vergleich von Ontologiesprachen..... | 153 |

Tabellenverzeichnis

| | |
|---|----|
| Tabelle 3.1: Bestandteile einer Ontologie | 39 |
| Tabelle 3.2: Kenngrößen für die Auswahl von Ontologiesprachen | 49 |
| Tabelle 3.3: Unterscheidung Beschreibungslogik | 52 |
| Tabelle 4.1: Taxonomie von semantischen Konflikten | 83 |
| Tabelle 5.1: Arten von Heterogenitätsproblemen im Überblick | 97 |

1 Einleitung

Wissen resultiert aus der Verarbeitung von Informationen durch das Bewusstsein [Albr93, S.228] und wird von Individuen zur Handlungssteuerung bzw. zur Lösung von Problemen verwendet. Entscheidend für den Wert des Wissens ist die semantische Anreicherung und Abstimmung zwischen dem speichernden und wieder auslesenden System. Um semantische Konflikte vermeiden und die Wiederverwendbarkeit sicherstellen zu können, bedarf es eines Konsens in der Verwendung gleicher Symbole, der Zuordnung der Symbole denselben Konzepten und Beimessung derselben Bedeutung von Konzepten.[Bode05, S.123;Hald04, S.15]

In diesem Einleitungskapitel werden zunächst die Motivation und der Hintergrund, die zur Umsetzung dieser Diplomarbeit führen, erläutert. Im Anschluss wird die genaue Vorgehensweise beschrieben.

1.1 Ausgangslage

Die Menschen leben heute in einem von Informationen geprägten Zeitalter. Daten und Informationen, in jeglicher medialer Ausprägung, sind überall und allgegenwärtig. Die Notwendigkeit einer effektiven Verwaltung charakterisiert den Umgang mit Wissen und Informationen. Historisch betrachtet ist diese Tatsache nicht überraschend. Der Versuch Wissen zu archivieren und dieses anderen zugänglich zu machen liegt in der Natur des Menschen. Seit der Vorzeit gehen die Menschen dem Bedürfnis nach Informationen zu sammeln und ihr Wissen aufzuzeichnen. Mit der Entwicklung der Sprache, gelang es den Menschen, Dinge mit Worten zu beschreiben und sich mit seinen Mitmenschen zu unterhalten [GaHP08, S.15]. Bis zur Erfindung der Schrift, galten die direkte Kommunikation mit weiteren Personen und das Gedächtnis eines Menschen als die einzige Möglichkeit Wissen zu konservieren und es weiterzugeben. Erst die Erfindung der Schrift ermöglichte es bis heute Informationen ohne die direkte Begegnung von Personen über Distanzen und längere Zeiträume hinweg weiterzugeben und zu archivieren. Aufbauend auf die Schrift, ermöglichten die Erfindung der Schriftrolle und später der moderne Buchdruck um 1450, die Vervielfältigung von Schriftstücken im großen Umfang und Ausweitung der Verbreitung von Informationen.[Wart99; Kunz05, S.16ff]

Mit dem verstärkten Aufkommen des Kopierens, Speicherns und der zunehmenden Menge von Schriftstücken, ging schon zu dieser Zeit das Bedürfnis einher, aus der Menge an Schriftstücken diejenigen auszuwählen, die wichtige Informationen enthielten. Für Archive und Bibliotheken, die dieses Wissen öffentlich bis heute verwalten, wurde es daher immer wichtiger, abgelegte Inhalte systematisch zu ordnen, damit sie wieder aufgefunden werden konnten. Für das Finden

und Arbeiten mit den Schriftstücken wurden Erschließungssysteme unerlässlich. Das erste Erschließungssystem, stellte ein einfaches Klassifikationssystem dar und benutzte Etikettentäfelchen, die an Tafelbehältnisse angelehnt wurden und den Inhalt der Werke thematisch umschrieben.[Wart99; Kunz05, S.16ff]

Eine Vielzahl unterschiedlicher Erfindungen und technischen Entwicklungen führten zu den heutigen Medien (Zeitung, Radio, Fernsehen) und letztendlich zum globalen Netzwerk „Internet“, wodurch den Menschen viel schneller, einfacher und mehr an Informationen übermittelt und Wissen zugänglich wurde. Heutzutage sind die Menschen in der Lage, über das Internet auf sehr große Datenbestände und weltweit verteilte Ressourcen von elektronisch gespeicherten Informationen zuzugreifen. Neben dem Zugriff auf sehr große digitale Datenbestände, lassen sich heute zugleich immer schneller Daten produzieren. Die Menge an Informationen, die Weltweit täglich produziert wird, ist enorm und steigt stetig. So wurden allein 2004 *„mehr Informationen als in der bisherigen Menschengeschichte produziert“* [Trku05, S.23]. Doch es macht nur dann einen Sinn große Mengen an neuen Daten zu erzeugen, wenn auf diese zur erwünschten Zeit zugegriffen werden kann. Es lässt sich erkennen, dass neben dem Sammeln von Informationen auch die strukturierte Speicherung, der Austausch und besonders das wirksame Wiederauffinden von Daten oder Informationen ein besonderes Problem und eine besondere Relevanz in jeder Epoche und auch in unserer Gesellschaft hat.

1.2 Motivation

Seit ihrer Einführung und Vorstellung gegenüber der Öffentlichkeit ist es Rechenmaschinen möglich digitale Daten zu verarbeiten. Seither können Daten in verschiedensten technischen oder inhaltlichen Formen gesammelt und in einem verteilten Umfeld abgelegt werden. Gegenwärtig können sie verschiedenen Organisationen gehören und in verschiedenen Formen vorliegen.

Im Zuge von Entwicklungen der Informationstechnik führen neue Trends und Technologien zu einem enormen Anstieg von digitalen Daten, die im Allgemeinen unstrukturiert und heterogen vorliegen. In Folge des fortwährenden Zuwachses an Daten ergeben sich einige große Probleme. Daher bedarf es neuer Zugangswege und neuer Techniken, um die darin enthaltenen Informationen und das mit ihnen verbundene Wissen effizient und zielgerecht nutzen zu können.

Da Daten der Grundstoff und Ausgangspunkt für die Erzeugung von Informationen und Wissen sind, ist es nicht verwunderlich, dass derzeit ein effektiver und unmittelbarer Zugriff auf Information und Wissen, die Forschungen der Wissenschaft und Unternehmen kennzeichnet und einen unumstrittenen Konsens bildet. Dementsprechend sind sehr viele Überlegungen geführt

worden, um Standards für schnellere und intelligendere Informationslagerungs- und Wiederauf-
findungssysteme zur Verfügung stellen zu können.

Stellvertretend für die Entwicklung des immer reichhaltiger und unüberschaubarer werdenden
Informationsangebots steht das Internet. Mit ihrer Unmenge an verteilt vorliegenden und ver-
fügbaren Daten unterschiedlichen Typs und Struktur, sowie Diensten bildet sie heute eine nicht
nur eine große Informationsplattform, sondern ebenso eine Kommunikationsplattform. Derzeit
gilt sie als die meist genutzte Technologie und effektivste Möglichkeit der Informationsbeschaf-
fung. Angesichts der physikalisch verteilten Lage von Informationen und der fehlenden Seman-
tik, stellt sich hier ebenfalls das Problem der präzisen und effizienten Wiederauffindung von
benötigten Informationen. Zwar gelingt es mit unterschiedlichen Suchhilfen dem Anwender
Informationen bereitzustellen. Doch bieten sie trotz eigener Werkzeuge, Algorithmen und Stra-
tegien meist nicht die Ergebnisse, die vom Nutzer letztendlich erwünscht werden und enden
meist in einer erneuten oder erfolglosen Suche.

Mit dem Ansatz der Weiterentwicklung zum Semantic Web und den damit resultierenden se-
mantischen Technologien sollen im Internet befindende Informationen derart präsentiert wer-
den, dass mit Hilfe ihrer Techniken diese durch Maschinen verstanden und verarbeitet werden
können. Dieses wird erzielt, indem Informationen eine Semantik bzw. semantische Beschrei-
bungen verliehen werden. Dadurch geht eine Steigerung des Informationsgehaltes einher, wo-
durch es u.a. Suchmaschinen gelingt, anhand der dahinterliegenden Bedeutung des Netzinhaltes,
Informationen automatisch auffinden und bereitstellen zu können.

Der Kern bzw. die Schlüsseltechnologie dieser Technologien bilden Ontologien, die ein formal
spezifiziertes Vokabular zur Verfügung stellen und der maschinellen Interpretierbarkeit bzw.
Mensch-Maschine oder Maschine-Maschine-Kommunikationsmöglichkeit unterstützend beitra-
gen. Mit Hilfe der Ontologien gelingt es den Inhalt einer Datenquelle mit Semantik zu beschrei-
ben und semantische Äquivalenzen zwischen unterschiedlichen Objekten bzw. Konzepten ent-
decken zu können. So können Ontologien dazu verwendet werden, Funktionalitäten von Systeme-
n durch eine Schnittstelle semantisch zu beschreiben, in ihnen enthaltene Informationen zur
Verfügung zu stellen und zur Nutzung bereitzustellen. Als übergeordnete konzeptuelle Modelle
stellen sie ein gemeinsames semantisches Vokabular über mehrere Informationsquellen hinweg
bereit, dass ein integrieren von Informationen aus heterogenen Quellen ermöglicht¹. Überset-
zungs- und Transformationsvorschriften können mit Ontologiesprachen bequem formuliert
werden, was die übergeordnete Konsistenzüberprüfung und Anfragebearbeitung erlaubt².

¹ Vgl. <http://www.theseus.joint-research.org/assets/Weitere-Informationen/Theseus-Flyer-Druck.pdf> zuletzt aufgeru-
fen 22.03.2011

² Vgl. Ebd.

1.3 Zielsetzung

Informationen, die durchaus speziell erhoben sein können, werden meist in großen Datenbeständen auf verschiedenen, physikalisch verteilten und potentiell relevanten Datenquellen gehalten. Oftmals führen unzureichende Strukturierungen dazu, dass gewünschte und gezielt benötigte Informationen nicht ermittelt oder effizient bereitgestellt werden können. Angesichts der physikalisch verteilten Lage von Informationen und der fehlenden Semantik, ist es das Ziel dieser Diplomarbeit auf die Möglichkeiten von semantischen Technologien im Bereich der ontologiebasierten Integration einzugehen und eine ontologiebasierte Schnittstelle zu konzipieren. Diese soll das Zusammensetzen von Informationsbeständen aus verschiedenen Informationsquellen an einer zentralen Stelle zulassen. Damit einem Nutzer ein erleichtertes und präzises Auffinden von erwünschten Daten oder Informationen ermöglicht werden kann. Zu der Aufgabenstellung gehört daher das Zusammentragen, die Bildung und das Ergänzen von Informationen, die für die Realisierung und der Konstruktion eines Prototyps dienlich sind.

Der Fokus liegt auf der wissenschaftlichen Analyse des „Semantic Web“- Bereichs und fachlich benachbarten Technologien. Parallel dazu werden bestehende Datenbestände, unterstützende Werkzeuge sowie mögliche einzubindende Dienste mit Blick auf die Verwendung innerhalb der Implementierung analysiert. Im Zuge der Konzeptentwicklung müssen Aktivitäten des Vergleichens und Zusammenführens für (zu entwickelnde) Ontologien berücksichtigt werden. Da das Zusammenfassen von mehreren existierenden Ontologien zu einer der Hauptaufgaben der Schnittstelle gehört.

1.4 Vorgehen

Angeregt durch die Entwicklung der semantischen Technologien und der Notwendigkeit nach effektiver Verwaltung von Informationen, lässt sich die Diplomarbeit an diesem Punkt ansetzen. Im Fokus steht die Konzeption und mögliche prototypische Implementierung einer Schnittstelle, mit deren Hilfe eine präzise und zuverlässige Informationsbereitstellung bzw. Informationsnutzung gesichert werden kann.

Nachdem in diesem Kapitel der Hintergrund, die Motivation und Ziele dieser Diplomarbeit erläutert wurden, wird an dieser Stelle die weitere Struktur der Arbeit beschrieben. Den Ausgangspunkt bildet das Kapitel 2. Zu der Vorgehensweise gehört es die Zielsetzungen der Arbeit zu erreichen. Dazu wird die Arbeit in zwei Bereiche unterteilt: dem Untersuchungs- und Konzeptionsbereich.

Zu Beginn jeder Aufgabe wird grundlegendes Wissen benötigt. Daher bildet Kapitel 2 den Ausgangspunkt des Untersuchungsbereichs. Es soll einen Überblick über die wichtigsten Grundla-

gen dieser Arbeit geben. Dafür werden zu diesem Kontext entsprechende und häufig genutzte Begrifflichkeiten verdeutlicht. Sowie die Grundlagen und Trends zu diesem Bereich geschildert. Eine wichtige Voraussetzung für das Konzept sind die Entwicklungen und die weiterführende Ideen des Internet. Daher werden auf die Entwicklungen eingegangen und eine Gegenüberstellung der Netze im Internet durchgeführt. Hierbei werden Vorläufer und mögliche Nachfolger des Semantic Web beschrieben. Letztendlich führt die Betrachtung des WWW zu einem besseren Verständnis der Bemühungen um das Semantic Web und ihrer Kerntechnologie, der Ontologien.

Die in dieser Arbeit zugrunde liegende Vorgehensweise folgt dem Ansatz der Verwendung von Ontologien. Daher bildet den Anfang der Konzeption, die Analyse existierender Methodologien sowie in der Praxis übliche Verfahren zur Entwicklung und Zusammenführung von Ontologien und ihre heutigen Anwendungsgebiete. Auf Grundlage von Untersuchungen werden in Kapitel 3 vorhandene Methoden für die genannten Zielsetzungen herausgearbeitet. Ferner werden in Kapitel 4 die Anforderungen an den Gestaltungsprozess der Schnittstelle erhoben, die Konkretisierungen für die Zielsetzungen beinhalten.

Nachdem ein grundlegendes Verständnis über das Themengebiet und der verwendeten Technologien geschaffen wird, gilt das Kapitel 5 zum Konzeptbereich. Dementsprechend wird in Kapitel 5 ein Konzept für den Lösungsweg beschrieben und auf den Entwurf für die Implementierung des Prototyps eingegangen. Es wird ein Konzept einer Schnittstelle vorgeschlagen, die das Zusammenfassen von aktuellen auf physikalisch verteilt vorliegenden Informationen ermöglicht. Dazu ist es erforderlich auf geforderte Merkmale, die Architektur des Konzepts und die vorhandenen Möglichkeiten von semantischen Technologien im Bereich der Integrationsmöglichkeiten einzugehen.

Darauf aufbauend wird in Kapitel 6 die mögliche Evaluation der Lösung beschrieben. Um das Ziel der Zusammenführung von verteilt vorliegenden Daten beispielhaft darzustellen, wird ein Beispielszenario eingeführt.

Als Abschluss folgen in Kapitel 7 eine Zusammenfassung der Arbeit und ein Ausblick auf die Möglichkeiten, die sich mit dem Prototypen ergeben.

2 Problematik des Informationsaufkommens

Wie bereits einleitend beschrieben, gehen Menschen dem Bedürfnis nach, Informationen und Wissen zu sammeln und über kommunikative Wege miteinander zu teilen. Eine Vielzahl verschiedener Entwicklungen ermöglichte es bis heute, Informationen aufzunehmen und festhalten zu lassen, sowie anderen zur Verfügung zu stellen. Seit der Erfindung der Schrift wuchs die Zahl der abrufbaren Informationen stark an. Treibende Kraft für die Sammlung und Erzeugung von Informationen in den letzten Jahrzehnten, war die Erfindung der Rechenmaschinen [Demm10, S.50]. Im Gegensatz zu früher, als Informationen ausschließlich papierbasiert zur Verfügung standen, stieß die Erfindung der Rechenmaschinen einen digitalen Wandel an. Indem sie eine neue Infrastruktur anboten und neue Wege eröffneten, mit denen sich auf einfachste Weise Informationen und Daten in digitaler Form erfassen, speichern, verarbeiten und abrufen ließen.

Vor zwanzig Jahren ohne Programmierkenntnisse nicht benutzbar, vor zehn Jahren von den Arbeitsplätzen nicht mehr wegzudenken [Wurm01, S. 182], sind Computer heute in unterschiedlichster Form allgegenwärtig und unterstützen den Menschen bei seinen Tätigkeiten oder befreien ihn von teils komplexen Aufgaben. Ihr Einsatz in allen Bereichen gilt als wichtige Grundlage der heutigen Gesellschaft [Kell98, S.22f]. Mit ihrem Siegeszug seit den 80er-Jahren, als Selbstverständlichkeit in einem modernen Haushalt und Werkzeug im Alltag [HaNi02, S.465] haben sie die Arbeits- und Lebenswelt radikaler verändert als sonst eine Erfindung in den letzten Jahrzehnten wie bspw. das Telefon oder Fernsehen [Bueh00, S.14]. Seither können Informationen in Form von digitalen Daten und verschiedensten technischen oder inhaltlichen Formen gesammelt und in einem verteilten Umfeld abgelegt werden. Dieser technologische Fortschritt ermöglichte eine umfassende Informatisierung der Welt [Matt03, S.2]. Im Zuge dieser Informatisierung sind digitale Daten allgegenwärtig geworden.

Heute befinden sich die Menschen in einem Informationszeitalter, zu dessen Merkmalen es gehört von Informationen umgeben zu sein [vgl. Mits09, S.1]. Der technologische und rasche Fortschritt in der Informations- und Kommunikationstechnik (IKT), neue Trends, mehr Speicherkapazitäten und -technologien, sowie schnellere Geräte lassen es zu digitale Informationen bzw. digitale Daten mit leichter Handhabung zu produzieren, zu verarbeiten und bereitzustellen [Lang09, S.8]. Ständig werden neue Daten und Informationsquellen hinzugefügt. Infolgedessen stehen heute den Menschen Zugänge zu gewaltigen Informationsmengen der unterschiedlichsten Art zur Verfügung, auf die fast zu jeder Zeit und von jedem Ort zugegriffen werden kann. Dies ergibt nicht nur eine Massenproduktion an Information, sondern auch eine Informationsproduktion durch die Massen.[LyVa00]

Aus Entwicklungen und Errungenschaften in der Informationstechnologie, des damit verbundenen leichteren Speicherns mit riesigem Speicherplatz und Verbreitens von Information fasst [Krcm10] folgende Merkmale zusammen:

- zunehmende Verbreitung von weltweit vernetzten Informationssystemen³
- Vereinfachtes Erstellen und Kopieren von Informationen
- vereinfachte Informationsverbreitung
- Fülle an Informationskanälen
- Vernachlässigbare Kosten für Verbreitung, Manipulation und Erstellung
- Verstärkte Kommunikation von Informationen [Krcm10: S. 55ff; Demm10, S.51]

Der Mensch benötigt Wissen um qualitative Entscheidungen im alltäglichen Leben, sei es sozial oder beruflich, treffen zu können. Dazu sind vielfältige, vorhandene oder neue Informationen erforderlich, die Verfügbar gehalten werden müssen um in individuelles gebrauchsfertiges Wissen umgesetzt werden zu können [FrGa02, S.99]. Die hierfür benötigten Informationen können auf verschiedenen Systemen oder Komponenten liegen. Es ist daher erforderlich, die auf den Systemen liegenden verschiedenartigen Daten zusammen zu tragen, zu analysieren und für die Problembewältigung zur Verfügung zu stellen.

Die Verfügbarkeit von Informationen, der effektive Umgang, die qualitative Aufbereitung und der schnelle Zugriff auf Informationen stellen zu überwindende Probleme dieser Gesellschaft dar und haben unterdessen einen hohen Stellenwert eingenommen. Indem sie einen immer größer werdenden Anteil des gespeicherten Wissens der heutigen Menschheit darstellen [Braid08]. Für verschiedene Bereiche wie Unternehmen, Behörden, wissenschaftliche Einrichtungen etc. als auch Privatpersonen wird die Lösung dieser Probleme heute immer wichtiger und gehört ebenso zu den zentralen Herausforderungen der Zukunft [vgl. Demm10 S.3]. Desweiteren sieht sich die heutige Gesellschaft durch den technologischen Fortschritt einem fortschreitenden und ständigem Informationsbedarf gegenüber. Da, vor allem die in den Industriestaaten lebenden Menschen auf Informationen „on-demand“ oder „just-in –Time“ zugreifen wollen und müssen.

Waren früher Informationen aufgrund ihrer Erstellung langlebig, ist der Zeitwert bzw. die Lebensdauer von Informationen heute sehr kurz, wodurch sie schneller veralten und schneller irrelevant werden [Nadi99, S.13]. Daraus resultiert eine immer fortwährende Suche nach aktuellen und digital verfügbaren Informationen. Daher haben digitale Informationen mittlerweile dermaßen eine wichtige Stellung in der Gesellschaft, dass sie neben Boden, Arbeit und Kapital zu den

³ Informationssysteme sind Systeme, deren primärer Zweck die Informationsbereitstellung ist. Sie leisten eine Vielzahl von Einzelaufgaben, wie z.B. Erfassung, Bearbeitung, Verarbeitung, Speicherung, Transport und Ausgabe von Informationen.