Informationsvisualisierung und Navigation im Semantic Web: Eine Analyse bestehender Visualisierungstechniken im Hinblick auf Eignung für das Semantic Web

MICHAEL AUFREITER

BACHELORARBEIT

05/1/0307/043

eingereicht am Fachhochschul-Bachelorstudiengang

SOFTWARE ENGINEERING

in Hagenberg

im Juli 2008

Diese Arbeit entstand im Rahmen des Gegenstands

Data Engineering

im

 $Sommersemester\ 2008$

Betreuer:

Josef Altmann, Mag. Dr.

Inhaltsverzeichnis

Kurzfassung v							
\mathbf{A}	bstra	act	vii				
1	Einleitung						
	1.1	Ausgangssituation	1				
	1.2	Problemstellung	2				
	1.3	Fragestellung	2				
	1.4	Ziel	2				
	1.5	Vorgehensweise	3				
2	Grundlagen 4						
	2.1	Wissensmanagement	4				
	2.2	Implizites und explizites Wissen	4				
	2.3	Das Modell nach Nonaka/Takeuchi	5				
	2.4	Semantic Web	5				
	2.5	Taxonomien	6				
	2.6	Ontologien	6				
	2.7	Repräsentation von semantischen Daten	6				
		2.7.1 RDF	7				
		2.7.2 RDF-Schema (RDFS)	8				
		2.7.3 Web Ontology Language (OWL)	8				
		2.7.4 Alternative Darstellungsformen	9				
	2.8	Abfragen	9				
		2.8.1 Reasoning	9				
		2.8.2 SPARQL	10				
	2.9	Zusammenfassung	11				
3	Info	ormationsvisualisierung	12				
	3.1	Einführung	12				
	3.2	Ontology Visualization	13				
	3.3	Cluster Maps	13				
	3.4	Tree Maps	14				
	3.5	Textual Interfaces	15				

IN	HAL	TSVERZEICHNIS	iv
		3.5.1 Faceted Navigation	16
		3.5.2 Faceted Browser	16
	3.6	Graphs and Trees	17
		3.6.1 Hyperbolic Trees	18
		3.6.2 3D-Graphs	19
		3.6.3 Node-Link Diagrams	20
	3.7	Zusammenfassung	22
4	Kri	terienkatalog	23
	4.1	Typologie nach Shneiderman	23
	4.2	Usability der visuellen Repräsentation	23
		4.2.1 Datentyp (data type)	24
		4.2.2 Aussagekraft (expressiveness)	25
		4.2.3 Effektivität (effectiveness)	25
		4.2.4 Kognitive Komplexität (cognitive complexity)	26
		4.2.5 Räumliche Organisation (spatial organization)	$\frac{27}{27}$
		4.2.6 Informations codierung (information coding)	27
		4.2.7 Beziehungen (relations)	28
		4.2.8 Übergang (state transition)	28
	4.3	Interface Usability	29
	1.0	4.3.1 Orientierung und Hilfe (orientation and help)	29
		4.3.2 Navigation und Abfragen (navigation and querying) .	30
		4.3.3 Reduktion des Datenbestands (data set reduction)	30
		4.3.4 Datenerkundung (data exploration)	31
	4.4	Vorgehensweise bei der Bewertung	32
	4.4	4.4.1 Gewichtung der Kriterien	$\frac{32}{32}$
		4.4.2 Bewertungsskala	$\frac{32}{33}$
	4.5	Zusammenfassung	33
	4.0	Zusammemassung	JJ
5		vertung ausgewählter Werkzeuge	35
	5.1	Aduna AutoFocus	35
		5.1.1 Bewertung	36
		5.1.2 Bewertungstabelle	39
		5.1.3 Fazit	40
	5.2	Longwell RDF Browser	40
		5.2.1 Bewertung	41
		5.2.2 Bewertungstabelle	44
		5.2.3 Fazit	44
	5.3	Zusammenfassung	44
6	Sch	lussbetrachtung	46
	6.1	Ergebnisse	46
	6.2	Ausblick	47

INHALTSVERZEICHNIS	V
Literaturverzeichnis	48
Abbildungsverzeichnis	50

Kurzfassung

Der Anreiz und das Potential von Informationsvisualisierungen wird bereits häufig erkannt und der Wunsch nach deren Anwendung immer stärker. Gerade im Bereich des Wissensmanagements spielt dieses Gebiet eine immer wichtigere Rolle.

Diese Arbeit beschäftigt sich mit Informationsvisualisierung im Semantic Web und vermittelt einen Überblick über aktuelle Entwicklungen zum Thema Knowledge Visualization.

Zunächst werden grundlegende Konzepte der Informationsvisualisierung vorgestellt und deren Bedeutung in Hinblick auf das Wissensmanagement erklärt. Aus den Anforderungen, die das Semantic Web an die Informationsvisualisierungen stellt, lassen sich Kriterien ableiten, die zur Beurteilung von Visualisierungstechniken herangezogen werden können. Die ausgewählten Kriterien werden im Rahmen dieser Arbeit zu einem Kriterienkatalog zusammengefasst.

Schließlich werden ausgewählte Werkzeuge beschrieben, die im Wissensmanagement bereits erfolgreich Anwendung finden. Die einzelnen Untersuchungsobjekte werden nach einer detailierten Beschreibung anhand der ausgewählten Kriterien analysiert und bewertet. Dabei wird besonders auf deren Anwendung im Kontext des Semantic Web eingegangen.

Abstract

Both, the attraction and the potential of information visualization have been widely recognized and its use is requested frequently. Especially in terms of knowledge management, this topic plays a very significant role.

This paper follows the topic of information visualization in the Semantic Web and aims to give an overview of current developments regarding knowledge visualization.

At first, basic concepts of information visualization are described whereas their relevance in context of knowledge management is shown.

The requirements concerning information visualization in the Semantic Web lead to a set of criteria, which can be used for the evaluation of selected tools. The chosen attributes are summarized to a catalogue of criteria to ease the process of evaluation. Finally, selected tools, which are successfully used within the knowledge management sector, are described, analyzed and evaluated using the addressed set of criteria. Especially their applicability in context of the Semantic Web is examined.

Kapitel 1

Einleitung

1.1 Ausgangssituation

Das Semantic Web, welches derzeit vorwiegend Anwendung im Wissensmanagement findet, stellt eine neurartige Form der Datenrepräsentation dar. Im Unterschied zu herkömmlichen Dokumenten, wie Hypertext oder ähnlichen Formaten, die unstrukturiert als Text vorliegen, gibt Semantic Web den publizierten Inhalten eine "Bedeutung". Die vorliegenden Daten werden mit Meta-Informationen angereichert und haben eine strukturierte Form, die auch die Verarbeitung durch Maschinen ermöglicht.

Es liegen bereits Standards und konkrete Formate zur Repräsentation semantischer Daten vor. Ebenso stehen Werkzeuge zur Verfügung, mit deren Hilfe sich Informationen aus bereits vorliegenden Daten extrahieren lassen. Mit Hilfe von Abfragesprachen lassen sich gezielt Information aus großen Datenbeständen gewinnen. Die Möglichkeiten einer herkömmlichen textbasierten Suchmaschine (Vgl. Google PageRank¹ Algorithm) werden dabei bei weitem übertroffen.

Der Grundstein aus technologischer Sicht ist also gelegt. Nun gilt es, dem Endanwender Werkzeuge bereitzustellen, mit deren Hilfe sich semantische Daten darstellen, navigieren und durchsuchen lassen. Nur so kann das Semantic Web einer breiten Nutzerschicht zugänglich gemacht und die Vision des Semantic Web zur Realität werden.

Eine wichtige Rolle kommt hierbei der Informationsvisualisierung zu, deren Aufgabe es ist, vorliegende Daten in ein grafisch aufbereitetes Format zu transformieren. Informationsvisualisierung leistet damit einen wesentlichen Beitrag zur Verbesserung des Wissenstransfers und bildet das zentrale Thema dieser Arbeit.

¹http://www.google.com/technology/

1.2 Problemstellung

Die Anforderungen des Semantic Web stellen das Gebiet der Informationsvisualisierung vor eine neue Herausforderung. Es gilt einerseits bestehende Visualisierungstechniken zu analysieren und deren Potential für die erfolgreiche Anwendung im Semantic Web zu prüfen, andererseits neue Lösungen zu entwickeln, welche den speziellen Anforderungen des Semantic Web gerecht werden.

Zur Beurteilung der Qualität einzelner Visualisierungstechniken müssen Kriterien gefunden werden, anhand derer eine objektive Beurteilung erfolgen kann.

1.3 Fragestellung

Die Arbeit setzt sich unter anderem mit der Frage auseinander, in wie weit bestehende Visualisierungstechniken unter dem Kontext des Semantic Web eingesetzt werden können und auf welche Aspekte dabei besonderes Augenmerk gelegt werden muss.

Semantische Daten haben wenige Einschränkungen im Hinblick auf die Typisierung, um größtmögliche Flexibilität zu gewährleisten. Visualisierungstechniken sollten demnach einen generischen Ansatz verfolgen und auch dann funktionieren, wenn nur Teile der zu Grunde liegenden Datenstruktur bekannt sind. Weiters soll der Benutzer die Möglichkeit haben, auch dieses unbekannte Terrain mit einzubeziehen. Diese Eigenschaft stellt besondere Anfordungen an Usability und Navigierbarkeit von Visualisierungstechniken.

Im Rahmen dieser Arbeit soll zudem die Frage geklärt werden, welche bestehenden Bewertungskriterien, die zur Bewertung von Visualisierungstechniken herangezogen werden, im Kontext des Semantic Web von Bedeutung sind.

1.4 Ziel

Ziel dieser Arbeit ist es, bestehende Vorgehensweisen und Werkzeuge gegenüber zu stellen und hinsichtlich ihrer Qualität zu untersuchen. Diese Techniken werden anschließend exemplarisch präsentiert und anhand zuvor festgelegter Kriterien bewertet. Die Erarbeitung dieser Kriterien stellt dabei ein weiteres implizites Ziel dar. Personen, die mit der Visualisierung von semantischen Inhalten betraut sind, erhalten ein Hilfsmittel, welches sie dabei unterstützt, aus den bestehenden Möglichkeiten zur Informationsvisualisierung die geeignete Methode als Grundlage für ihr Projekt auszuwählen.

1.5 Vorgehensweise

Zunächst wird im Rahmen des Grundlagenkapitels auf Basiskonzepte zum Thema Semantic Web eingegangen. Darauf aufbauend werden Konzepte der Informationsvisualisierung vorgestellt und erklärt, wobei jeweils auf die Besonderheiten, welche bei der Anwendung auf semantische Daten zu berücksichtigen sind, eingegangen wird.

Durch die Analyse der Anforderungen, die das Semantic Web an die Visualisierungstechniken stellt, sowie durch Zuhilfenahme von bestehenden Kriterienkatalogen aus der Literatur des Bereichs der Informationsvisualisierung wird ein Kriterienkatalog erarbeitet, der anschließend als Bewertungsgrundlage dient.

Abschließend werden ausgewählte Visualisierungswerkzeuge vorgestellt und anhand der erarbeiteten Kriterien bewertet. Das Ergebnis ist ein Leitfaden für die erfolgreiche Anwendung des Kriterienkatalogs zur Bewertung von konkreten Visualisierungswerkzeugen.

Kapitel 2

Grundlagen

2.1 Wissensmanagement

Im heutigen Informationszeitalter wird Wissen bereits als vierter Produktionsfaktor (neben Kapital, Arbeit und Boden) gesehen (Vgl. Kleske [9, p.6]). Für Unternehmen ist es besonders wichtig, Wissen zu ihrem Vorteil einzusetzen, um in einer Informations- und Wissensgesellschaft bestehen zu können. Unternehmen werden mit einer immer größer werdenden Informationsflut konfrontiert und stehen vor der Aufgabe Wissen zu strukturieren, zu speichern und möglichst einfach zugänglich zu machen. Das Wissensmanagement beschäftigt sich mit genau diesen Aufgaben und umfasst Praktiken, die den Umgang mit Wissen im Unternehmen vereinfachen. Einen wichtigen Beitrag dazu leisten Informationssysteme, die Mitarbeiter miteinander vernetzen und ihnen die Bereitstellung und den Austausch von Informationen ermöglichen.

2.2 Implizites und explizites Wissen

Als implizites oder stilles Wissen (tacit knowledge) wird jenes Wissen bezeichnet, das nicht in formalisierter Form vorliegt. Damit umfasst es jene Kenntnisse und Fähigkeiten von Menschen, welche in deren Köpfen existieren, jedoch nicht für andere zugänglich sind. Implizites Wissen schließt dabei sowohl Verhalten als auch Kultur mit ein, welche meist nicht bewusst wahrgenommen werden. Der Begriff "tacit knowledge" ist von Michael Polanyi geprägt, der in seinem Buch, The Tacit Dimension (1967) eine berühmte Aussage dazu machte:

"We know more than we can tell."

Im Unterschied dazu bezeichnet explizites Wissen¹ jene Kompetenzen, die bereits eindeutig mittels Zeichen² codiert und somit kommunizierbar sind.

2.3 Das Modell nach Nonaka/Takeuchi

Organisationen (wie Unternehmen) versuchen sowohl implizites, stilles Wissen als auch festgeschriebenes, explizites Wissen nutzbar zu machen, um den Mitarbeitern eine möglichst breite Wissensbasis (knowledge base) bereitzustellen.

Die Japaner Ikujiro Nonaka und Hirotaka Takeuchi stellen mit ihrem SECI-Modell (Socialization, Externalization, Combination, Internalization). ein Modell zur Wissenserzeugung bereit. Dabei entsteht Wissen durch eine kontinuierliche Transformation zwischen explizitem und implizitem Wissen.

Sozialisation ist die Transformation von implizitem zu implizitem Wissen. Wissen wird hierbei schweigend weitergegeben, etwa wenn Menschen Erfahrungen zusammen machen.

Externalisierung bezeichnet die Transformation von implizitem zu explizitem Wissen. Dies passiert etwa, wenn Wissen dokumentiert, niedergeschrieben oder in einer zentralen Wissensdatenbank gespeichert wird.

Kombination wird die Transformation von explizitem zu explizitem Wissen genannt. Durch Kombination von explizitem Wissen kann neues Wissen entstehen.

Internalisierung bezeichnet jene Transformation, in der explizites Wissen wieder zu implizitem Wissen wird. Dadurch wird ein Lernprozess bezeichnet, der die Anwendung von erlerntem Wissen (aus explizitem Wissen) in der Praxis (implizites Wissen) umfasst.

2.4 Semantic Web

Die ursprüngliche Idee des Semantic Web ist eine Erweiterung des bestehenden World Wide Web (WWW) um Inhalte mit Bedeutung, sodass sie im Unterschied zu herkömmlichen Textdokumenten auch von Computern interpretiert und verarbeitet werden können. Ziel dieser mit Semantik versehenen Inhalte ist es, die Repräsentation von Wissen in einer stark strukturierten Form zu gewährleisten, um das Auffinden von Informationen in Zukunft zu vereinfachen. Durch die starke Strukturierung der Inhalte wird es möglich,

¹bezeichnet ausdrückliches bzw. ausführliches Wissen

²wie Sprache oder Schrift

exakt formulierte Abfragen an Stelle von stichwortbasierten Suchanfragen zu verwenden.

Das Semantic Web beruht auf einer Idee von Tim Berners-Lee, die er 2001 in einem Artikel für das Scientific American Magazine veröffentlichte [15]. Er beschreibt darin ein Szenario, in dem Compuer in naher Zukunft in der Lage sind, komplexe Aufgaben für den Menschen zu lösen. Sogenannte Software Agenten könnten beispielsweise völlig eigenständig ein nahegelegenes Restaurant ausfindig machen, das den zuvor festgelegten Wünschen des Benutzers entspricht und abschließend eine Reservierung vornehmen.

2.5 Taxonomien

Taxonomien bezeichnen eine Form der Klassifikation, in der hierarchische Strukturen abgebildet werden. Dabei wird eine betroffene Klasse immer genau einer Oberklasse zugeordnet und die gesamte Hierarchie auf eine Baumstruktur abgebildet. Je tiefer man sich in der Hierarchieebene bewegt, desto spezieller wird die hinterlegte Information.

2.6 Ontologien

In der Informatik und Informationswissenschaft bezeichnen Ontologien ein Datenmodell, das sowohl Begriffe innerhalb eines Wissensbereichs, als auch deren Beziehungen zueinander abbildet. Dabei werden Klassen zur Typisierung von Objekten und Attribute zur Beschreibung von Objekten verwendet. Ontologien werden verwendet um Wissendomänen zu beschreiben und die Bedeutung der unterschiedlichen Objektklassen, sowie deren Beziehungen zueinander festzulegen. Ontologien sind ebenso wie Taxonomien Klassifizierungssysteme, die jedoch nicht die Einschränkungen einer starren hierarchischen Struktur aufweisen. Die einzelnen Objektklassen können zueinander in beliebiger Relation stehen und werden als Graph repräsentiert.

2.7 Repräsentation von semantischen Daten

Das Semantic Web wird durch formale Spezifikationen beschrieben, die vom W3C³ veröffentlicht werden. Die Grundlage bildet dabei das Resource Description Framework (RDF), das ein universelles Datenmodell zur Beschreibung von Metadaten im World Wide Web bereitstellt. Es existieren verschiedene Notationen, die entweder auf RDF basieren, als Datenaustauschformat dienen, oder einen alternativen Ansatz verfolgen.

Im Rahmen des folgenden Abschnitts werden grundlegende Formate zur Repräsentierung von semantischen Daten vorgestellt und deren Unterschiede

 $^{^3 \}mbox{World}$ Wide Web Consortium http://www.w3.org

kurz erläutert, um einen grundsätzlichen Einblick in die Modellierung von semantischem Datenmaterial zu gewähren.

2.7.1 RDF

RDF ist eine auf XML basierende Sprache zur Bereitstellung von Metadaten im World Wide Web. Mit Hilfe von RDF können semantische Netzwerke, die als Graphen vorliegen, auf ein austauschbares Format abgebildet werden.

Dabei werden die Knoten und Kanten des Graphen durch URI's⁴ oder Literale (Zahlen, oder Zeichenfolgen) repräsentiert. Zur Darstellung von Aussagen werden in RDF Subjekte mit Objekten durch Prädikate verbunden. Man spricht von sogenannten "Triples", die den Kern der semantischen Datenrepräsentation bilden.

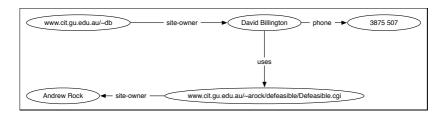


Abbildung 2.1: Semantisches Netz übernommen von Antoniou und Harmelen [7]

Zur Veranschaulichung zeigt Abbildung 2.1, entnommen aus Antoniou und Harmelen 2004 [7], eine Graphenrepräsentation mehrerer Triples. Diese Art der Graphendarstellung wird als semantisches Netz (semantic net) bezeichnet.

Der Graph aus Abbildung 2.1 beinhaltet folgende Triples:

```
[ http://www.cit.gu.edu.au/~db, http://www.mydomain.org/site-owner, "David Billington" ]
[ "David Billington", http://www.mydomain.org/phone, "3875 507" ]
[ "David Billington", http://www.mydomain.org/uses, http://www.cit.gu.edu.au/~arock/defeasible/Defeasible.cgi ]
[ "http://www.cit.gu.edu.au/~arock/defeasible/Defeasible.cgi", http://www.mydomain.org/site-owner, "Andrew Rock"
```

 $^{^4\}mathrm{URI}$ steht für Uniform Resource Identifier und wird für die eindeutige Bezeichnung von Ressourcen verwendet

Eine einfache Modellierung dieses Graphen in RDF-Syntax würde folgendermaßen aussehen:

```
<?xml version="1.0" encoding="UTF-16"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:mydomain="http://www.mydomain.org/my-rdf-ns"
  <rdf:Description rdf:about="http://www.cit.gu.edu.au/~db">
       <mydomain:site-owner>
David Billington
       </mydomain:site-owner>
  </rdf:Description>
  <rdf:Description rdf:about="David Billington">
        <mydomain:uses>
          www.cit.gu.edu.au/~arock/defeasible/Defeasible.cgi
        </mydomain:uses>
        <mydomain:phone>
          3875 507
        </mydomain:phone
  </rdf:Description>
  <rdf:Description rdf:about="www.cit.gu.edu.au/~arock/defeasible/Defeasible.cgi">
         Andrew Rock
  </mydomain:site-owner>
</rdf:Description>
</rdf:RDF>
```

2.7.2 RDF-Schema (RDFS)

RDF-Schema⁵ (RDFS) ist wie RDF eine Empfehlung des W3C. Während RDF konkrete Ressourcen mit deren Attributen definiert, legt RDF-Schema das Vokabular für ein bestimmtes Wissensgebiet fest. RDFS dient also zur Beschreibung von Ontologien und manifestiert den gemeinsamen Sprachgebrauch, also die Syntax für RDF Dokumente.

In einem RDF-Schema Dokument werden Klassen festgelegt, welche abstrakte Objekte beschreiben, die anschließend bei der Erzeugung von Instanzen verwendet werden. RDFS bietet dabei verschiedene Möglichkeiten zur Beschreibung von Klassen, wie die Definition von Attributen, die Festlegung von Beziehungen zwischen Klassen und Eigenschaften und die Einschränkung von Eigenschaften auf einen bestimmten Wertebereich. Mit diesen Mitteln lassen sich einfache Ontologien erstellen, die eine Domäne ausreichend beschreiben und einen gemeinsamen Sprachgebrauch festlegen, der von RDF Dokumenten verwendet werden kann.

2.7.3 Web Ontology Language (OWL)

Die Aussagekraft von RDF und RDF Schema ist in vielen Fällen nicht ausreichend, um komplexe Wissensdomänen angemessen zu beschreiben. RDF/RDFS ist annähernd darauf beschränkt, Hierarchien für Klassen und Eigenschaften zu modellieren, was dazu führte, dass mit OWL eine Beschreibungssprache eingeführt wurde, die RDF/RDFS um weitere Möglichkeiten

⁵http://www.w3.org/1999/02/22-rdf-syntax-ns

zur Modellierung von Ontologien ergänzt. OWL führt dazu neue Sprachkonstrukte ein, die die Formulierung von Ausdrücken ähnlich der Prädikatenlogik ermöglichen. Diese Eigenschaften kommen dann bei dem in Abschnitt 2.8.1 beschriebenen Reasoning zum Einsatz.

2.7.4 Alternative Darstellungsformen

Neben der gängigen XML Notation des Ressource Description Frameworks existieren weitere Formate, die sich für die Darstellung und den Austausch von RDF-Graphen eignen. Das W3C führte beispielsweise das Notation 3⁶ (oder N3) Format ein, das eine tabellarische Darstellung der zugrunde liegenden Triples verwendet.

Die Darstellung lehnt sich an N-Triples⁷ und Turtle⁸ an, ebenfalls zwei alternative RDF Formate, und ist wesentlich kompakter und leichter lesbar als die ursprüngliche XML Notation. N3 ist ausserdem ein beliebtes Datenaustauschformat, ob seiner geringeren Dateigröße.

Weiters existieren verschiedene Techniken, die bei der Speicherung von RDF-Modellen in Datenbanken zum Einsatz kommen. RDF Datenbanken, wie beispielsweise Sesame⁹, zielen darauf ab, eine effiziente Speicherung und Abfrage von RDF-Graphen zu ermöglichen und sind bei der Verwaltung großer Datenmengen unabdingbar.

2.8 Abfragen

Um in einem Datenbestand, der als RDF-Graph vorliegt, Daten lokalisieren zu können, sind Abfragemöglichkeiten und damit geeignete Abfragesprachen notwendig. Gängige Abfragesprachen für XML Dokumente wie XPath sind für diese Aufgabe jedoch ungeeignet, da diese nicht auf die speziellen Eigenschaften des RDF-Datenmodells eingehen. Mit SPARQL (siehe Abschnitt 2.8.2) wird in diesem Kapitel ein mächtiges Instrument für die Abfrage von RDF-Graphen vorgestellt. Damit das Semantic Web auch die oft angesprochene Intelligenz aufweist, werden Abfragen mit Hilfe von Reasoning Techniken um Schlussfolgerungen erweitert.

2.8.1 Reasoning

Reasoning¹⁰ ist im Kontext des Semantic Web ein Mechanismus, der es Computern erlaubt, Informationen durch den Einsatz von Logik zu verarbeiten. Ziel dabei ist es, die Bedeutung von Dingen bezüglich einer Anwendung zu

⁶http://www.w3.org/TeamSubmission/n3/

⁷http://www.dajobe.org/2001/06/ntriples/

⁸ http://www.dajobe.org/2004/01/turtle/

⁹http://www.openrdf.org/

 $^{^{10}\}mathrm{zu}$ Deutsch schließen oder schlussfolgern

nutzen. May 2006 [11, p. 485f] beschreibt Reasoning im Wesentlichen als gesunden Menschenverstand, verstehen und nutzen der Bedeutung von Fakten. Diese menschlichen Fähigkeiten sollen mit Hilfe der Technolgien des Semantic Web auch in Computersystemen Einzug halten und den Benutzern von Wissensmanagementsystemen dabei helfen, Informationen leichter zu lokalisieren. Es wird versucht, den Suchproblemen des herkömmlichen World Wide Web, wie bedeutungsähnliche Worte (Synonyme) oder Worte die verschiedene Bedeutungen haben (Homonyme), zu begegnen. Reasoning arbeitet auf der Ebene von Abfragesprachen (genauer: auf deren Implementierungsebene). Die zugrundeliegenden Formalismen sind dadurch "verdeckt" und für den Benutzer nicht sichtbar. Der Benutzer ist jedoch in der Lage, die vom System getroffenen Schlussfolgerungen nachzuvollziehen.

2.8.2 **SPARQL**

SPARQL¹¹ ist eine Abfragesprache für RDF-Modelle und steht für "SPARQL Protocol and RDF Query Language". SPARQL ist eine Weiterentwicklung von RQL¹² (RDF Query Language), auf die im Rahmen dieser Arbeit nicht näher eingegangen wird. Die Sprache wird von einem W3C Standard festgelegt und liegt seit dem 15.1.2008 offiziell als Empfehlung vor. SPARQL weist eine SQL-ähnliche Syntax auf und kann auf alle Datenquellen, die auf RDF abgebildet werden können, angewendet werden.

Die Syntax basiert auf der Triple Darstellung von RDF. Durch die Verwendung von Variablen und der Angabe von Eigenschaften und deren Belegung können Bedingungen formuliert werden, welche das Resultat einschränken.

Das folgende Beispiel wurde von Danny Ayers, der einige Beispiele für SPARQL Abfragen auf seiner Website¹³ bereitstellt, entliehen und zeigt den grundsätzlichen Aufbau von SPARQL Abfragen. Variablen werden dabei mit einem vorangestellten "?" gekennzeichnet, während sich mit der Direktive "PREFIX" Aliases für Namensräume vergeben lassen, um eine kompakte Schreibweise der Abfrage zu ermöglichen.

 $^{^{11} \}rm http://www.w3.org/TR/rdf\text{-}sparql\text{-}query/$

¹²http://139.91.183.30:9090/RDF/RQL/

¹³http://hyperdata.org/

SPARQL Query

```
PREFIX rdf: <a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://wmlns.com/foaf/0.1/>

SELECT DISTINCT ?name
WHERE {
    ?x rdf:type foaf:Person .
    ?x foaf:name ?name
}
```

Das gezeigte Beispiel sucht innerhalb einer beliebigen SPARQL-Datenquelle nach Elementen vom Typ foaf:Person und liefert als Ergebnis eine Liste von Namen. Die SELECT-Bedingung identifiziert dabei die Variablen, welche in den Ergebnissen erscheinen, während die WHERE-Bedingung die Ergebnisliste anhand von einfachen Mustern einschränkt. Diese Einschränkungen werden mit Hilfe von Triple-Mustern formuliert, wobei hier erneut Variablen eine wichtige Rolle spielen.

Ergebnis

```
name
_______
Luther Blisset
Jane Lambda
Joe Hacker
Nik Jewell
Ben Blakeslee-Drain
Rowland Watkins
Dan Connolly
Stephen Downes
Edd Dumbill
Pete Wilson
```

Das Ergebnis ist eine Liste von Namen, die das Suchkriterium erfüllen.

2.9 Zusammenfassung

Im Rahmen dieses Kapitels wurden grundlegende Konzepte des Semantic Web sowie deren Bedeutung für das Wissensmanagement behandelt. Die vorgestellten Techniken und Standards bilden die Grundlage für die folgenden Kapitel, die sich mit der Informationsvisualisierung im Allgemeinen, dem Finden von Kriterien zur Beurteilung von Visualierungstechniken und der Bewertung konkreter Visualisierungswerkzeuge für das Semantic Web beschäftigen.

Kapitel 3

Informationsvisualisierung

3.1 Einführung

Das Semantic Web bietet eine neuartige Form der Datenrepräsentation, die es ermöglicht, Inhalte und deren Bedeutung generisch abzubilden. Es stellt ein universelles Framework von Ressourcen zur Verfügung, welche von Maschinen interpretiert werden können.

Das folgende Kapitel beschäftigt sich mit Vorgehensweisen und Techniken zur Visualisierung von semantisch annotierten Daten. Die besonderen Eigenschaften dieser Datengrundlage stellen den Bereich der Informationsvisualisierung vor eine große Herausforderung: Die zur Verfügung stehenden Metainformationen, welche auf Ontologien basieren, können sowohl das Erstellen einer graphischen Repräsentation erleichtern, als auch in die endgültige Darstellung einfließen. Der Benutzer einer Visualisierung soll nicht mit der Komplexität der zugrunde liegenden Datenstruktur konfrontiert werden, um sich voll auf seine Aufgaben konzentrieren zu können. Neben der Erstellung einer graphischen Darstellung ist die Bereitstellung einer intuitiv zu bedienenden Benutzerschnittstelle entscheidend, um diese speziellen Anforderungen erfüllen zu können. Die Entwickler von Visualisierungstechniken sind dabei mit mehreren Problemstellungen konfrontiert.

RDF¹ ist ein abstraktes und unstrukturiertes Datenmodell. Durch den hohen Grad an Abstraktion fehlen meist die absoluten Visualisierungsinformationen. Zudem erschweren es fehlende Strukturinformationen, einen visuellen Zusammenhang zu konstruieren. In vielen Fällen hat der Benutzer keine Kenntnis über das zugrunde liegende Datenmodell und muss sich durch eine kompakte Darstellung des gesamten Datenmaterials einen Überblick verschaffen. Die folgenden Methoden stammen aus verschiedenen Bereichen der Informationsvisualisierung und sollen einen grundsätzlichen Einblick in verschiedene Techniken zur Wissensvisualisierung im Semantic Web geben.

¹http://www.w3.org/RDF/

3.2 Ontology Visualization

Die Visualisierungskomponente wird für Semantic Web Werkzeuge immer wichtiger. Fluit et. al [2] beschreiben einige Techniken, die sich mit Ontology Visualization, also mit der Visualisierung von Ongologien beschäftigen. Visualisierungen erweitern bereits Werkzeuge zur Extraktion von Ontologien (OntoLift, Text-to-Onto) oder Ontologie-Editoren (Protègè).

Diese Werkzeuge zielen auf das Visualisieren der Strukturinformationen von Ontologien ab und stellen die Begriffe, sowie deren Beziehungen zueinander dar. Dabei sollen Ontology Visualization Tools Aufgaben wie die Datenanalyse, den Datenvergleich und die Abfrage von Daten anhand der zugrunde liegenden Ontologiestrukturen übernehmen.

Die folgenden Techniken und Beispiele sollen einen groben Einblick geben, wie auf Basis von Ontologien graphische Repräsentierungen erzeugt werden können.

3.3 Cluster Maps

Fluit et. al [2] beschreiben Cluster Maps als eine Visualisierungstechnik. die Endanwendern erlaubt, mit komplexen semantischen Strukturen umzugehen. Cluster Maps wurden vom Schweizer Hersteller Aduna² entwickelt und bieten eine einfache und benutzerfreundliche Darstellung von Informationen. Die Cluster Map Technik konzentriert sich auf die Visualisierung von Instanzen und deren Klassifikationen basierend auf den Begriffen der zugrunde liegenden Ontologie. Ontologien, auch Taxonomien genannt, finden in verschieden Bereichen, beispielsweise in Bibliotheken, Anwendung als Klassifizierungssysteme. Solche Klassifizierungssysteme werden bereits vielfach in Webanwendungen oder Webverzeichnissen angewendet. Semantic Web Standards wie RDF unterstützen diese Klassifizierung in Form von Ontologien. Durch Spezialisierungsbeziehungen lassen sich Sammlungen von Objekten als Hierarchien mit Unter- und Oberklassen abbilden. Sammlungen von Unterklassen sind unvollständig, wenn ihre Vereinigung nicht alle Objekte der Oberklasse enthält. Klassen, die gemeinsame Objekte enthalten, überlappen sich. Der entscheidende Vorteil von Cluster Maps liegt in der Darstellung dieser Eigenschaften.

²http://www.aduna-software.com

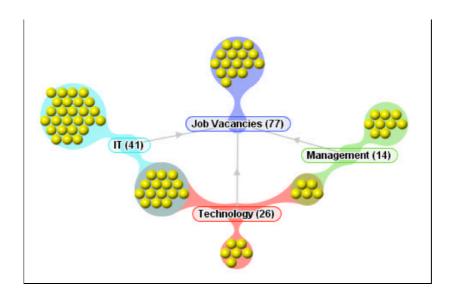


Abbildung 3.1: Beispiel einer Clustermap nach Fluit et. al [2]

Fluit et. al [2] beschreiben anhand eines Beispiels die Cluster Map Darstellung. Das Beispiel in Abbildung 3.1 zeigt eine Sammlung von Jobangeboten basierend auf einer einfachen Ontologie. Instanzen, also Jobangebote, werden durch gelbe Kugeln dargestellt. Die einzelnen Klassen erscheinen als abgerundete Rechtecke und beinhalten den Namen sowie deren Kardinalität. Durch die hellgrauen Pfeile werden Vererbungshierarchien dargestellt. Management ist demnach eine Unterklasse von Jobangebote. Instanzen werden durch ballonförmige Kanten mit ihrer zugeordneten Klasse der untersten Hierarchieebene verbunden. Instanzen, die der selben Klasse angehören werden zu Clustern zusammengefasst.

3.4 Tree Maps

Tree Maps verfolgen das Ziel der Visualisierung von Hierarchien, und stellen eine relevante Methode der Wissensvisualisierung im Semantic Web dar. Shneiderman [13] beschreibt diese Technik Anfang der 90er Jahre erstmals im Zusammenhang mit der Visualisierung der Kapazitätsausnutzung von Dateisystemen. Die Darstellung erfolgt anhand eines Algorithmus, der Gruppen von Rechtecken, welche die zugrunde liegenden Daten repräsentieren, auf einer Anzeigefläche abhängig von deren Relevanz anordnet. Da bei der Darstellung tiefer Hierarchien die Lesbarkeit der Darstellung leidet, wird zunehmend eine Limitierung der dargestellten Hierarchietiefe vorgenommen, um die Übersicht zu gewährleisten.



Abbildung 3.2: Repräsentation einer Tree Map am Beispiel der Newsmap Anwendung

Abbildung 3.2 zeigt Newsmap³, eine Tree Map Darstellung, welche auf den Daten von Google News basiert. Für mehrere Länder werden jeweils Kategorien (Sport, Business, National) mit Farben codiert. Die Artikel selbst werden mit Rechtecken innerhalb der Kategorien dargestellt, wobei die Größe der Rechtecke angibt, wieviele Artikel zu diesem Thema existieren. Bei dieser Darstellung wurde die Anzeige aus Gründen der Übersicht auf zwei Hierarchieebenen (Land, Kategorie) limitiert.

3.5 Textual Interfaces

Ein textuelles Interface stellt die zugrunde liegenden Informationen in strukturierter Form als Text dar und verzichtet dabei zunächst auf eine graphische Repräsentation. Le Grand und Soto [1] geben den Ontologie-Editor Protégé (siehe Abschnitt 3.2) als Beispiel für ein textuelles Interface an.

Im Kontext des Semantic Web bieten Textual Interfaces in der Regel Mechanismen zur Navigation und Suche in strukturierten Datenbeständen. Ihnen kommt eine große Bedeutung im Bereich der Informationsvisualisierung zu, da ein textuelles Interface dem Großteil der Visualisierungswerkzeuge entweder als Basis, oder als unterstützendes Element, wie etwa zur Festlegung von Filterkriterien, dient.

³http://www.marumushi.com/apps/newsmap/

3.5.1 Faceted Navigation

Faceted Navigation beruht auf der Facettenklassifikation, einem Klassifikationssystem, welches anstatt den Wissensbereich in eine starre Baumstruktur einzugliedern, den Inhalt durch mehrere Gruppen (Facetten) klassifiziert. Jede Gruppe enthält dabei konkrete Einfachklassen nach denen die Elemente des Wissensbereichs eingeteilt sind. Faceted Navigation erlaubt es nun dem Benutzer mit Hilfe dieser Facetten einen individuellen Navigationspfad zu erstellen. Der Benutzer wird durch den gesamten Wissensbereich geführt indem er Kenntnis über die zur Verfügung stehenden Kategorien (Facetten) hat, dabei aber selbst vorgibt, welchen Zugang er zuerst wählt. Das derzeit bekannteste Klassifizierungssystem dieser Art ist die Colon-Klassifikation nach Ranganathan [12], die insbesondere auf die Kategorisierung von kleinen Artikeln (Zeitungen, Zeitschriften) ausgerichtet ist.

3.5.2 Faceted Browser

Ein Faceted Browser ist ein Werkzeug, welches Benutzern das Durchsuchen von Informationen ermöglicht. Faceted Browser verwenden das Konzept der Faceted Navigation (siehe. Abschnitt 3.5.1), um den Benutzer durch den zugrunde liegenden Wissensbereich zu führen. Im Semantic Web nehmen Faceted Browser eine entscheidende Rolle ein. Semantisches Datenmaterial enthält in der Regel eine große Anzahl an typisierten Elementen, deren Struktur durch Ontologien (siehe Abschnitt 2.6) festgelegt ist. Aus dieser Struktur ergeben sich viele Facetten, die zur Navigation herangezogen werden können. Die effektivste Navigation dieser Art ist kontextabhängig, dabei werden die zur Verfügung stehenden Facetten, deren Ausprägungen und die Anzahl der Elemente je Ausprägung abhängig von den aktuell selektierten Daten generiert. Hat der Benutzer beispielsweise bereits eine Vorselektion vorgenommen, werden lediglich die Navigationelemente angezeigt, welche die aktuell selektierten Daten betreffen. Durch eine weitere Selektion nach einem bestimmten Kriterium würde sich die Anzahl der Navigationselemente erneut verringern.

Mit Longwell⁴ steht bereits ein leistungsstarker Faceted Browser zur Verfügung, der mit RDF Daten arbeitet.

Moritz Stefaner zeigt mit Elastic Lists⁵ eine weitere Möglichkeit der Umsetzung dieses Konzepts (siehe Abbildung 3.3). Das Beispiel demonstriert die Funktionsweise anhand der Darstellung von Nobelpreisgewinnern, welche aufgrund von unterschiedlichen Facetten gefiltert werden können. Elastic Lists verbessert dabei gängige Faceted Browsing Konzepte durch die Visualisierung von relativen Proportionen der Metadaten (Gewichtung). Dabei

⁴http://simile.mit.edu/longwell/

⁵http://well-formed-data.net/archives/54/elastic-lists

wird die Charakteristik der Gewichtung durch unterschiedliche Helligkeiten dargestellt. Heller dargestellte Ausprägungen haben demnach eine höhere Gewichtung. Im Beispiel sind dies jene Ausprägungen, die in den letzten Jahren eine steigende Anzahl an Nobelpreisträgern aufweisen (abhängig von der aktuellen Auswahl verändert sich auch die Darstellung der Gewichtung). Ausserdem werden Ausprägungen, welche eine größere Anzahl an Elementen beinhalten dementsprechend größer dargestellt. Die Veränderung der Auswahl wird zudem mit animierten Übergängen visualisiert.

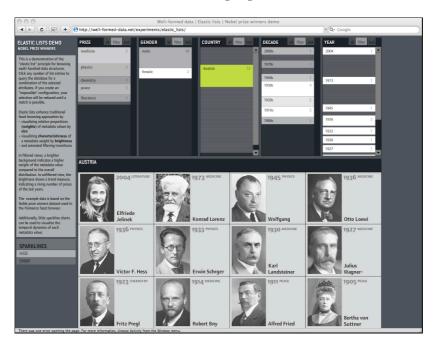


Abbildung 3.3: Elastic Lists Darstellung von Moritz Stefaner

3.6 Graphs and Trees

Eine graphische Repräsentierung in Form von Graphen und Bäumen liegt insofern nahe, da die interne Repräsentierung der Daten des Semantic Web durch eine Graphenstruktur gegeben ist. Während die Transformation in eine Graphen- oder Baumdarstellung bei kleineren Datenmengen durchaus sinnvoll ist, wird es bei zunehmender Größe der Datenstruktur für den Benutzer meist schwierig den Überblick zu behalten und die dargestellten Informationen zu interpretieren.

Le Grand und Soto 2006 [1] sehen den Nutzen der Graphen- und Baumdarstellung hauptsächlich in der Repräsentierung der globalen Struktur von semantischen Graphen. Ausserdem geben sie an, dass sich Bäume besser für das Verständnis von Menschen eignen, da sie eine hierarchische Struktur aufweisen, die sich einfach interpretieren lässt. RDF Graphen und Ontologien sind keine Bäume und bilden in der Regel eine komplexe Struktur, weshalb eine direkte Überführung in eine Baumdarstellung nicht möglich ist.

Nach Le Grand und Soto [1] ist es jedoch sinnvoll kleine Teile aus der semantischen Struktur in Bäume zu tranformieren. Die Herausforderung der Graphen-Visualisierung liegt generell darin, möglichst viele Knoten (Nodes) darzustellen, ohne dass dabei die Lesbarkeit und die Übersicht verloren geht.

3.6.1 Hyperbolic Trees

Eine mögliche Lösung zur Darstellung von Graphen ist die Verwendung von hyperbolischer anstatt der euklidschen Geometrie, wodurch sich eine große Anzahl an Knoten und Beziehungen auf dem Bildschirm darstellen lassen. Ein Hyperbolic Tree (HT) ist eine Visualisierungstechnik, die Informationen abhängig vom Fokus und Kontext in einer einfach zu interpretierenden Baum-Ansicht darstellt. Hyperbolic Trees wurden ursprünglich von Xerox PARC entwickelt und patentiert. Abbildung 3.4 zeigt eine von Musictrails⁶ verwendete Visualisierung von Künstlern und deren Beziehungen zueinander mit Hilfe des JavaScript Hyperbolic Browsers⁷.

⁶Musictrails ist ein Music Recommendations System, das anhand der Musikgewohnheiten von Benutzern Informationen sammelt - http://musictrails.com.ar/

 $^{^7\}mathrm{Der}$ JavaScript Hyperbolic Browser ist eine von Nicolas Garcia Belmonte entwickelte JavaScript Library zur Darstellung von interaktiven Hyperbolic Trees in Webbrowsern - http://hypertree.woot.com.ar/

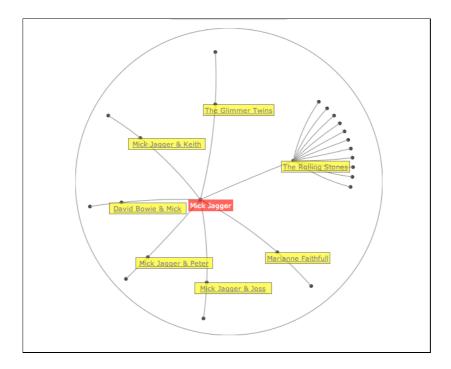


Abbildung 3.4: Darstellung einer Hyperbolic Tree Repräsentation am Beispiel des JavaScript Hyperbolic Browsers

3.6.2 3D-Graphs

Eine weitere Möglichkeit dem Platzmangel am Bildschirm zu begegnen, ist die Repräsentierung des semantischen Graphs in drei Dimensionen. Le Grand und Soto 2006 [1] entwickelten mit UNIVIT⁸ ein Werzeug zur Visualisierung von Topic Maps (siehe Abbildung 3.5). UNIVIT verwendet Virtual Reality Techniken wie 3D, Interaktionen und verschiedene Detailebenen. Durch die Abstände der Knoten zueinander können zusätzliche Informationen abgelesen werden. Eng zueinander in Beziehung stehende Elemente werden durch knappe Abstände der Knoten zueinander signalisiert und Elemente desselben Typs werden zu Clustern zusammengefasst.

⁸UNIVIT steht für Universal Interactive Visualization Tool

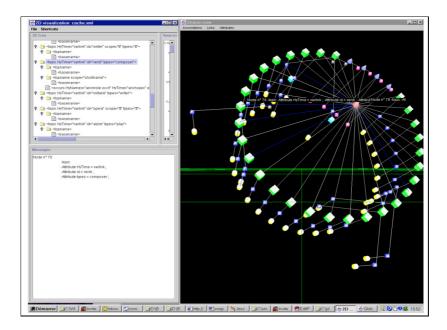


Abbildung 3.5: 3D-Darstellung einer Topic Map Repräsentation mit UNI-VIT

3.6.3 Node-Link Diagrams

Node-Link Diagrams entsprechen der klassischen Darstellung von Graphenstrukturen. Dabei werden Knoten und Beziehungen auf der Anzeigefläche angeordnet. Die Visualisierung komplexer Graphen ist bei dieser Darstellung zwar möglich, doch wird ein grosser Anzeigebereich benötigt und die Navigation erheblich erschwert. Derartige Visualisierungen werden beispielsweise bei Werkzeugen zur Erstellung von Ontologien eingesetzt.

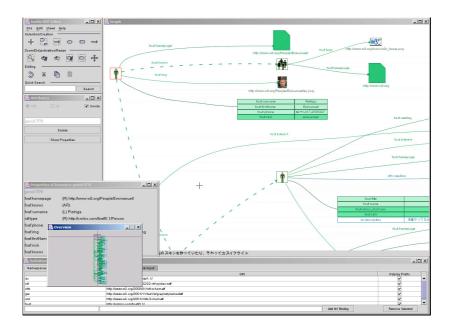


Abbildung 3.6: IsaViz Graph

IsaViz⁹ (siehe Abbildung 3.6) ist eine visuelle Umgebung zur Darstellung und Bearbeitung von RDF Modellen. Sie erlaubt dem Benutzer Ontologien mit Hilfe eines graphischen Werkzeugs zu erstellen und anschließend als RDF-Format zu exportieren.

Die Herausforderung bei der Darstellung von Node-Link Diagrams liegt in der Anordnung der Knoten und deren Verbindungen zueinander. Um ein lesbares Bild zu erhalten, dürfen sich Knoten und Verbindungen nicht überlappen und müssen geeignete Abstände, sowie nach Möglichkeit Symmetrieeigenschaften aufweisen. Um diesen Anforderungen Rechnung zu tragen, sind geeignete Algorithmen notwendig, welche in der Lage sind, eine möglichst optimale Anordnung der Elemente zu berechnen.

Fruchterman und Reingold [6] stellen 1991 den Force-directed Placement Algorithmus¹⁰ vor, der verwendet werden kann um zufällig platzierte Knoten in ein wünschenswertes Layout zu überführen und dabei den Anforderungen an die Ästhetik der visuelle Repräsentation entspricht.

 $^{^9}$ http://www.w3.org/2001/11/IsaViz/

 $^{^{10}}$ auch unter Spring embedding Algorithmus bekannt

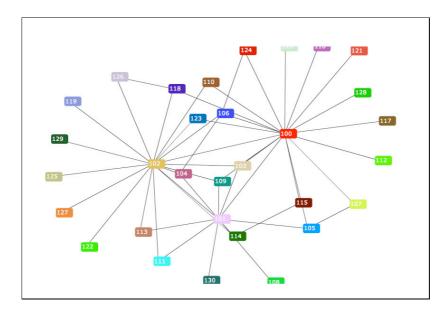


Abbildung 3.7: ActionScript Implementierung des Force-directed placement Algorithmus

Abbildung 3.7 zeigt eine Implementierung des Force-directed placement Algorithmus von Masaki Sawamura 11 unter der Verwendung von ActionScript 12 .

3.7 Zusammenfassung

Mit Cluster Maps, Tree Maps, Graphs und Trees wurden unterschiedliche Techniken zur Erzeugung einer grafischen Repräsentation vorgestellt. Ausserdem wurde mit Faceted Browsern ein alternatives und rein textbasierendes Konzept zur Aufbereitung und Strukturierung von semantischem Datenmaterial gezeigt. Die beschriebenen Techniken kommen in verschiedenen Werkzeugen zum Einsatz und bilden die Grundlage für eine qualitative Informationsvisualisierung. Im folgenden Kapitel werden Kriterien behandelt, anhand derer sich konkrete Visualisierungswerkzeuge gegenüberstellen und bewerten lassen.

¹¹http://sawamuland.com/flash/graph.html

¹²ActionScript ist Bestandteil von Adobe Flash - http://www.adobe.com/flash/

Kapitel 4

Kriterienkatalog

Im Rahmen dieses Kapitels erfolgt eine Auswahl an Kriterien, die zur Bewertung von konkreten Visualisierungstechniken herangezogen werden können. Basierend auf den Erkenntnissen von Edlinger [3], der einen Überblick über Bewertungskriterien der Informationsvisualisierung im Wissensmanagement gibt, werden jene Eigenschaften erhoben und adaptiert, die für die Bewertung von Visualisierungsmethoden im Semantic Web relevant sind.

Die Grundlage des folgenden Kriterienkatalogs bildet das Visual Information Seeking Mantra nach Shneiderman [14, p.2ff]. Ausgehend von diesen Erkenntnissen werden basierend auf den Merkmalen von Dal Sasso Freitas et. al [5, p.42f], die auch Edlinger [3] als Grundlage dienten, Kriterien zur Beurteilung von Visualisierungstechniken vorgestellt.

4.1 Typologie nach Shneiderman

Shneiderman [14, p.1] fasst mit dem Visual Information Seeking Mantra viele Entwürfe zusammen und stellt damit ein Rahmenwerk zur Einteilung und Bewertung von Informationsvisualisierungen bereit.

"Overview first, zoom and filter, then details-on-demand"

Demnach soll eine Visualisierung vorerst einen Überblick über die betreffenden Informationen bieten. Filter- und Zoomfunktionen ermöglichen dann eine speziellere Selektierung und zuletzt können zu jedem dargestellten Element Details eingesehen werden. Durch diese Technik ist es dem Anwender möglich, auch durch große Datenbestände zu navigieren und bei Bedarf Zugang zu allen Details zu haben.

4.2 Usability der visuellen Repräsentation

Dal Sasso Freitas et al. [5] beschäftigen sich mit der Evaluierung von Usability bei Visualisierungstechniken und deklarieren im Zuge dessen Kriterien,

die zur Bewertung von Visualisierungstechniken herangezogen werden können

Dal Sasso Freitas et al. [5] legen drei Kategorien zur Einteilung der behandelten Merkmale fest. Die Usability der visuellen Repräsentation bezieht sich auf die Aussagekraft und Qualität der resultierenden Darstellung. Interface Usability geht auf die Interaktionsmechanismen ein, die das Visualisierungswerkzeug dem Benutzer zur Verfügung stellt. Der Bereich Data Usability widmet sich der Eignung der zugrunde liegenden Daten in Hinblick auf die Unterstützung von Anwenderaufgaben.

Die folgende Auswahl an Kriterien lässt sich auch auf Visualisierungen anwenden, die auf der Grundlage von semantischen Daten arbeiten, wobei auf die ersten beiden Kategorien näher eingegangen wird. Ausserdem soll für jede Eigenschaft geklärt werden, in welchem Ausmaß sie für die Visualisierung von semantischen Daten von Bedeutung ist.

4.2.1 Datentyp (data type)

Shneiderman [14] beschreibt verschiedene Datentypen im Rahmen seiner Task by Data Type Taxonomy. Er geht davon aus, dass der Anwender eine Sammlung von Elementen anzeigen möchte, die wiederum mehrere Attribute beinhaltet. Für alle sieben von Shneiderman beschrieben Datentypen gilt, dass die zugrunde liegenden Elemente Attribute haben und damit eine Selektion nach bestimmten Attributwerten ermöglichen. Anhand des Datentyps ist eine Einordnung des zugrunde liegenden Datenmaterials möglich.

Eine Visualisierungstechnik, die einen vorgegebenen Datentyp nicht unterstützt, muss nicht weiter untersucht werden, da es nicht möglich ist, eine graphische Repräsentation zu erstellen. Angesichts dieser Eigenschaft kommt dem Datentyp bei der Bewertung die größte Bedeutung zu, da es sich um ein Knock-Out-Kriterium handelt.

Shneiderman [14] beschreibt dabei folgende sieben Datentypen: 1-dimensional, 2-dimensional, 3-dimensional, Temporal, Multi-dimensional, Tree und Network.

Messmethoden: Übereinstimmungsprüfung, Situationsanalyse

Messgrößen:

- Unterstützung der geforderten Datentypen: Werden die geforderten Datentypen in vollem Umfang unterstützt?
- Unterstützung alternativer Datentypen: Werden alternativ weitere Datentypen unterstützt und sind diese für die Visualisierungsaufgabe hilfreich?

4.2.2 Aussagekraft (expressiveness)

Die Aussagekraft basiert nach MacKinlay [10] auf einer formalen Analyse von graphischen Sprachen und legt fest, ob eine graphische Repräsentierung exakt die gewünschte Information darstellt.

MacKinlay [10] bezeichnet die Aussagen als Fakten und verwendet für das Datenmaterial die Bezeichnung Set. Um eine hohe Aussagekraft zu erreichen, müssen in jedem Fall alle Fakten des Sets codiert werden. Als zweite Bedingung nennt MacKinlay die Ausschließlichkeit der Fakten des Sets. Sind nicht ausschließlich die Fakten des Sets codiert, kann das nur dann toleriert werden, wenn durch die Darstellung von anderen Informationen die Lesbarkeit weiterhin gegeben ist und die Darstellung durch den Benutzer interpretiert werden kann.

Messmethoden: Zielgruppentests, Stichprobentests

Messgrößen:

- Allgemein: Wird durch die graphische Darstellung die zugrunde liegende Information korrekt wiedergegeben?
- Vollständigkeit der Fakten: Werden alle Fakten des Sets codiert?
- Ausschließlichkeit der Fakten: Werden ausschließlich die Fakten des Sets codiert? Ist die Lesbarkeit auch dann noch gegeben, wenn nicht ausschließlich die Fakten des Sets codiert werden?

4.2.3 Effektivität (effectiveness)

Effektivität bezieht sich auf die Eignung der Maßnahmen zur Zielerreichung. Die Effektivität einer Visualisierung wird herangezogen, um Aussagen über die Wirksamkeit der Darstellung von Informationen treffen zu können. Eine Darstellung ist effektiv, wenn sich der Betrachter rasch einen Überblick über die in den Daten enthaltenen Informationen verschaffen kann. Eine Evaluierung erfolgt auf Basis der unterschiedlichen Designaspekte der Visualisierung, wobei auch die Interpretationsfähigkeit des Benutzers einen Einfluss auf diese Eigenschaft hat. MacKinlay [10] verfolgte mit der Bewertung das Ziel, Visualisierungstechniken hinsichtlich ihrer Effektivität vergleichbar zu machen.

Die Aussagekraft, sowie weitere Kriterien die sich auf die Usability von visuellen Repräsentationen beziehen, sind Voraussetzung für eine effektive Darstellung. Edlinger [3] reduziert dieses Kriterium auf den Kern, der lautet: "Werden die Informationen klar und verständlich dargestellt?" Die Messung der Effektivität kann durch die für die Interpretation benötigte Zeit, sowie den Anteil richtiger Interpretationen in Relation zu den Fehlinterpretationen erfolgen. Um Aussagen über die Effektivität eines Werkzeugs treffen zu

können, sind Tests mit Zielgruppen erforderlich, die Werte für die benötigte Zeit der Interpretation und die Anzahl der Fehlinterpretation erheben.

Messmethoden: Zielgruppentests, Zeitmessung, direkte Erhebung durch Stichproben

Messgrößen:

- Gefahr der Missinterpretation: Werden mindestens 90% der dargestellten Informationen richtig interpretiert? Können Testfragen nach der Betrachtung der Darstellung vom Testsubjekt richtig beantwortet werden?
- Geschwindigkeit der Interpretation: Ist die durchschnittliche Geschwindigkeit der Interpretation zufriedenstellend? Liegt der durchschnittliche Wert unterhalb eines zuvor festgelegten Maximalwerts?

4.2.4 Kognitive Komplexität (cognitive complexity)

Kognitive Komplexität ist nach Bieri et al. [8] durch die Anzahl und Eigenschaften der Dimensionen definiert, die Individuen zur kognitiven Strukturierung von Gegenstandsbereichen verwenden. Fenton und Pfleeger [4, p.245] beschreiben kognitive Komplexität als den Aufwand, der vom Menschen zum Verständnis der Lösung erbracht werden muss. Die kognitive Komplexität einer Darstellung lässt sich mit Hilfe der Datendichte, der Datendimension und der Relevanz der dargestellten Informationen messen. Die Anzahl der visualisierten Datenelemente auf einer bestimmte Anzeigefläche legen die Datendichte fest, während die Datendimension die Anzahl der Dimensionen bestimmt, die gleichzeitig angezeigt werden.

Messmethoden: Überprüfung der Parameter anhand zuvor festgelegter Sollwerte, Zielgruppentests

Messgrößen:

- Datendichte: Wird die vorgegebene minimale Datendichte nicht unterschritten? Wird die maximale Datendichte, die durch die definierte Anzeigefläche festgelegt wird, überschritten?
- Datendimension: Wird die minimale Datendimension, die durch den Datentyp vorgegeben wird, erreicht? Lässt sich die vorgegebene Datendimension mit dem untersuchten Visualisierungswerkzeug sinnvoll darstellen?

4.2.5 Räumliche Organisation (spatial organization)

Die räumliche Organisation hängt mit dem Layout einer visuellen Darstellung zusammen und beschäftigt sich im Wesentlichen mit der Frage, wie einfach es für den Benutzer ist, einzelne Elemente in der Darstellung zu lokalisieren, um sich der ganzheitlichen Darstellung bewusst zu werden.

Die räumliche Organisation umfasst nach Dal Sasso Freitas et. al [5] die logische Reihenfolge der Elemente, den Grad der Verdeckung, sowie die Anzeige von Details und den Referenzkontext. Die Lokalisierung eines einzelnen Elements kann sich schwierig gestalten, sollten einige Objekte von anderen überlagert werden oder sich das Layout nicht an die Struktur der Datenelemente halten. Betrachtet der Benutzer Details zu einem Objekt, so beinflusst die Darstellung des Referenzkontexts die räumliche Orientierung, die es dem Benutzer ermöglicht, sich über die Verteilung der Informationselemente bewusst zu werden.

Messmethoden: Überprüfung der Parameter anhand zuvor festgelegter Sollwerte, Zielgruppentests

Messgrößen:

- Logische Reihenfolge der Elemente: Werden die geforderten logischen Anordnungen (alphabetisch, tabellarisch, hierarchisch, relational oder geographisch) vom Visualisierungswerkzeug unterstützt?
- Grad der Verdeckung: Kann vollständige Verdeckung ausgeschlossen werden? Kann teilweise Verdeckung ausgeschlossen werden? Können Informationsobjekte einfach lokalisiert werden?
- Anzeige von Details: Hat der Benutzer die Möglichkeit, sich bei Bedarf Details zu einzelnen Informationselementen anzeigen zu lassen?
- Referenzkontext: Kann der Referenzkontext vom Benutzer lokalisiert werden, während der Fokus auf der Darstellung von Details zu einem bestimmten Element liegt?

4.2.6 Informationscodierung (information coding)

Dal Sasso Freitas et. al [5] sehen die Informationscodierung als einen weiteren Aspekt, der zur Evaluierung von Visualisierungstechniken herangezogen werden kann. Durch die Verwendung von zusätzlichen Symbolen können alternative Repräsentationen wie Gruppen als Cluster-Darstellung generiert, oder der Benutzer dabei unterstützt werden, eine bessere Vorstellung von den Informationselementen zu bekommen. Die Unterstützung von Informationscodierung kann einen wesentlichen Beitrag zur Usability von visuellen Repräsentationen leisten.

Messmethoden: Überprüfung der Parameter anhand zuvor festgelegter Sollwerte, Zielgruppentests

Messgrößen:

- Anzahl der unterstützten Informationscodierungen: Werden die vom Anwendungsfall geforderten Codierungen (Farbe, Symbole, Textform) unterstützt?
- Qualität der Darstellung: Werden die codierten Informationen übersichtlich und nachvollziehbar dargestellt?

4.2.7 Beziehungen (relations)

Semantische Daten basieren auf einer Sammlung von Objekten die durch Prädikate miteinander in Beziehung gebracht werden. Relationen bilden somit das Herzstück eines semantischen Netzwerks. Der visuellen Repräsentation von Beziehungen zwischen den Informationsobjekten kommt somit große Bedeutung zu, deshalb sollte sie unbedingt in die Evaluierung miteinfließen. Entscheidend ist weiterhin, inwieweit unterschiedliche Beziehungstypen unterstützt werden und wie deren Kennzeichnung erfolgt. Ein Werkzeug, das die Darstellung von Beziehungen nicht unterstützt, ist demnach für die Visualisierung von semantischen Daten in den meisten Fällen ungeeigenet.

Messmethoden: Überprüfung der Parameter anhand zuvor festgelegter Sollwerte, Zielgruppentests

Messgrößen:

- Anzahl der unterstützten Beziehungstypen: Werden alle geforderten Beziehungstypen (hierarchisch, gleichrangig, 1:M, M:N) unterstützt?
- Qualität der Darstellung: Werden Beziehungen klar und übersichtlich dargestellt? Kann man optisch zwischen den unterschiedlichen Beziehungstypen unterscheiden?
- Visualisierung impliziter Zusammenhänge: Werden zusätzliche Möglichkeiten zur Visualisierung von Zusammenhängen geboten, die sich nicht unmittelbar aus dem Datenmaterial ergeben, sondern erst bei der Visualisierung ersichtlich werden?

4.2.8 Übergang (state transition)

Dal Sasso Freitas et. al [5] führen den Übergang zwischen zwei Darstellungen als wesentliches Usability-Kriterium an. Der Übergang ist das Resultat des Neuzeichnens von einzelnen Teilen oder der gesamten visuellen Repräsentation nach einer Benutzeraktion, wie beispielsweise dem Ändern von

Filterkriterien. Die Vorstellung der Informationselemente des Benutzers bei Übergängen hängt von Faktoren wie der Dauer des Übergangs oder den Änderungen in der räumlichen Darstellung ab. Ein nachvollziehbarer Übergang hat in der Regel eine positive Wirkung auf den Benutzer bei der Verwendung.

Messmethoden: Überprüfung der Parameter anhand zuvor festgelegter Sollwerte, Zielgruppentests

Messgrößen:

- Qualität der Darstellung von Übergängen: Behält der Benutzer während des Übergangs die Übersicht und kann die neue Anordnung der Elemente nachvollziehen?
- Arbeitsfluss: Bietet der Übergang dem Benutzer einen Nutzen zur Orientierung, ohne dabei den Arbeitsfluss zu unterbrechen?

4.3 Interface Usability

Neben den im ersten Teil vorgestellten Kriterien ist auch die Qualität der Benutzerschnittstelle von großer Bedeutung. Nachfolgend werden jene Kriterien vorgestellt, die sich nicht mit der graphischen Darstellung an sich, sondern mit jenen Funktionen, die dem Benutzer bei der Erstellung einer Visualisierung unterstützen, auseinandersetzen.

4.3.1 Orientierung und Hilfe (orientation and help)

Die Art der Orientierung und Hilfe, die dem Benutzer zur Verfügung steht, ist ein wichtiges Merkmal, das zur Evaluierung der Usability herangezogen werden kann. Dal Sasso Freitas et. al [5] führen in diesem Zusammenhang Funktionen wie die Kontrolle der Detailebene, Undo / Redo Aktionen und die Repräsentation von zusätzlichen Informationen, wie das Einblenden eines Navigationspfads an. Mit der Unterstützung dieser Funktionen bieten entsprechende Werkzeuge dem Benutzer die Möglichkeit, ihre Aufgaben effizienter lösen zu können, in dem die Handhabung bei der Navigation von großen Datenbeständen erheblich erleichtert wird.

Messmethoden: Überprüfung der Parameter anhand zuvor festgelegter Sollwerte, Zielgruppentests

Messgrößen:

• Kontrolle der Detailebene: Stehen dem Benutzer ausreichende Möglichkeiten zur Kontrolle der Detailebene zur Verfügung?

- Unterstützung von Undo / Redo Funktionen: Unterstützt das Werkzeug das Rückgängigmachen und Wiederherstellen von Aktionen?
- Repräsentation von zusätzlichen Informationen: Ist das Visualisierungswerkzeug in der Lage Zusatzinformationen wie Metainformationen zum betrachteten Element bei Bedarf einzublenden?

4.3.2 Navigation und Abfragen (navigation and querying)

Die Analyse von Funktionen zur Navigation und Abfragen sollte in keiner Bewertung fehlen, zumal dieses Merkmal entscheidend darüber ist, ob überhaupt mit großen Datenbeständen gearbeitet werden kann. Nach Dal Sasso Freitas et. al [5] sollen dafür die Möglichkeit und Einfachheit der Selektion eines Datenelements, die Veränderung des Standpunkts des Benutzers, Manipulation von geometrischen Repräsentationen von Datenelementen, die Suche und Abfrage nach spezifischen Informationen und das Ausdehnen von geclusterten oder versteckten Elementen untersucht werden.

Diese Eigenschaften sind insbesondere bei der Visualisierung von semantischen Daten von Bedeutung, da im Regelfall eine große Menge an typisierten Daten vorliegt, die lediglich durch Abfragen und Gruppierung handhabbar wird.

Messmethoden: Überprüfung der Parameter anhand zuvor festgelegter Sollwerte, Zielgruppentests

Messgrößen:

- Selektion eines Datenelements: Ist es möglich einzelne Datenelemente leicht zu lokalisieren und auszuwählen?
- Veränderung des Standpunkts des Benutzers: Kann der Benutzer auf einfache Weise einen alternativen Standpunkt wählen?
- Manipulation von geometrischen Repräsentationen von Datenelementen: Kann der Benutzer in die Art der Darstellung eingreifen und graphische Elemente zur besseren Lesbarkeit neu anordnen?
- Suche und Abfrage nach spezifischen Informationen: Kann der Benutzer gezielt nach Informationen suchen und eigene Abfragen formulieren?
- Ausdehnen von geclusterten oder versteckten Elementen: Ist es möglich, Details zu geclusterten Bereichen zu erfahren und versteckte Elemente auf Wunsch des Benutzers einzublenden?

4.3.3 Reduktion des Datenbestands (data set reduction)

Die Reduktion des Datenbestands schafft die Möglichkeit, den Fokus auf interessante Bereiche zu legen und die für den Zweck der Visualisierung un-

relevanten Informationen auszublenden. Dal Sasso Freitas et. al [5] geben Filtering, Clustering und Pruning als Varianten der Reduktion des Datenbestands an. Filtern erlaubt die Reduktion der angezeigten Informationen zu einem bestimmten Zeitpunkt, Clustering wird verwendet, um Teilmengen der Datenelementen zu gruppieren und mit speziellen Symbolen zu Clustern zusammenzufassen, während Pruning für das Verständnis der Visualisierung unrelevante Informationen ausschneidet.

Messmethoden: Überprüfung der Parameter anhand zuvor festgelegter Sollwerte, Zielgruppentests

Messgrößen:

- Filtering: Stehen dem Benutzer ausreichende Möglichkeiten zum Filtern des Datenbestands zur Verfügung?
- Clustering: Ist es möglich zusammengehörige Objekte zu Clustern zusammenzufassen?
- Pruning: Lassen sich unrelevante Informationen aus der Betrachtung ausschließen?

4.3.4 Datenerkundung (data exploration)

Christian Fluit et. al [2] beschreiben die Aufgabe der Datenerkundung als einen Prozess der Informationssuche, der für die aktuellen Aktivitäten des Benutzers nicht unmittelbar von Relevanz ist. Datenerkundung unterscheidet sich von der Abfrage von Daten dahingehend, dass keine individuellen Fragen beantwortet werden sollen, sondern der Benutzer sich einen Überblick verschaffen möchte, welche Informationen überhaupt zur Verfügung stehen. Im Unterschied zur allgemeinen Analyse soll der Benutzer bei der Datenerkundung in der Lage sein, den gesamten Datenbestand im Auge zu behalten, er kann sich dabei aber auf die für ihn relevanten Bereiche konzentrieren.

Bei Visualisierungsaufgaben im Semantic Web sind in vielen Anwendungsfällen keine Informationen über den grundsätzlichen Aufbau der Daten verfügbar. Der Benutzer benötigt also einen Einstiegspunkt, um sich zunächst einen Überblick über die vorhandenen Elemente und deren Typinformationen verschaffen zu können, um in weiterer Folge, nach Erlangen der Kenntnis über die für ihn relevanten Datenstrukturen, gezielte Abfragen formulieren zu können.

Im vorherigen Kapitel wurden mit Cluster Maps und textuellen Interfaces Lösungen für diesen Aspekt vorgestellt, der eine entscheidende Rolle im Bereich der Visualisierung von Ontologien darstellt.

Messmethoden: Überprüfung der Parameter anhand zuvor festgelegter Sollwerte, Zielgruppentests

Messgrößen:

- Visualisierung des gesamten Datenbestands: Ist das Werkzeug in der Lage, eine Übersicht über den gesamten Datenbestand zu geben?
- Navigation durch verschiedene Detailebenen: Kann der Benutzer sich auf einfache Art in unterschiedlichen Detailebenen bewegen und dabei die Struktur der Daten erkunden?

4.4 Vorgehensweise bei der Bewertung

Im Folgenden wird beschrieben, wie durch die Gewichtung der Kriterien und deren Bewertung mit Hilfe einer Bewertungsskala eine Visualisierungstechnik hinsichtlich ihrer Eignung für das Semantic Web bewertet werden kann. Die Bewertung der einzelnen Kriterien erfolgt mit geeigneten Messmethoden. Vorschläge für geeignete Messmethoden finden sich unterhalb der Beschreibung der einzelnen Kriterien.

Als Ergebnis der Evaluierung liegt eine Tabelle vor, die Aufschluss darüber gibt, welche Kriterien von einer konkreten Visualisierungstechnik in welchem Umfang unterstützt werden. Aus den einzelnen Werten kann nun unter Berücksichtigung der festgelegten Gewichtung eine Gesamtbewertung abgeleitet werden.

4.4.1 Gewichtung der Kriterien

Da die Eigenschaften einen unterschiedlich starken Einfluss auf die Eignung einer Visualisierungstechnik haben, erfolgt im Vorfeld eine Gewichtung, die eine differenzierte Berechnung der Gesamtsumme ermöglicht.

Edlinger [3] schlägt vor, auf jene Kriterien, die einen Bezug zur untersuchenden Visualisierungstechnik aufweisen, besonders hoch zu gewichten. Dem Datentyp kommt demnach das höchste Gewicht zu, da mit Hilfe der durch das Datenmaterial festgelegten Voraussetzungen bereits früh ungeeignete Techniken und Werkzeuge ausgeschlossen werden können. Nach Edlinger [3] bekommen die Aussagekraft und die Effektivität besonders viel Gewicht, da diese Kriterien die Eignung zur Interpretation durch den Benutzer festlegen. Der Rest der Gewichtung wird auf weitere Kriterien, welche die Usability der visuellen Repräsentation sowie die Interface Usability betreffen, aufgeteilt.

Die folgende Auflistung stellt, angelehnt an Edlinger [3], eine exemplarische Gewichtung der Kriterien dar. In der Praxis ist es wichtig, die Gewichtung entsprechend der individuellen Voraussetzungen vorzunehmen.

 ${\bf Tabelle~4.1:}~{\bf Gewichtung~der~einzelnen~Kriterien}.$

Kriterium	Gewichtung
Usability der visuellen Repräsentation	
Datentyp (data type)	K.O.
Aussagekraft (expressiveness)	25 %
Effektivität (effectiveness)	25 %
Kognitive Komplexität (cognitive complexity)	5,56 %
Räumliche Organisation (spatial organization)	5,56 %
Informationscodierung (information coding)	5,56 %
Beziehungen (relations)	5,56 %
Übergang (state transition)	5,56 %
Interface Usability	
Orientierung und Hilfe (orientation and help)	5,56 %
Navigation und Abfragen (navigation and querying)	5,56 %
Reduktion des Datenbestands (data set reduction)	5,56 %
Datenerkundung (data exploration)	5,56 %
Gesamt	100 %

4.4.2 Bewertungsskala

Die Bewertung erfolgt auf Basis eines Punktesystems. Jedes Kriterium wird auf einer Punkteskala von 0 Punkte (nicht unterstützt) bis 3 Punkte (vollständig unterstützt) bewertet.

Tabelle 4.2: Mögliche Ausprägungen der Werte.

	All gemein	Grad der Unterstütztung	Qualität
3	sehr gut	vollständig	sehr hoch
2	gut	großteils	${\it zufrieden stellend}$
1	ausreichend	ausreichend	ausreichend
0	nicht genügend	unzureichend	unzureichend

4.5 Zusammenfassung

Im Rahmen dieses Kaptitels wurden Kriterien beschrieben, anhand derer eine Bewertung von Visualisierungstechniken für das Semantic Web möglich

ist. Es wurde die Vorgehensweise bei der Bewertung erläutert und dabei auf die individuelle Gewichtung der Eigenschaften sowie auf die Verwendung einer Bewertungsskala eingegangen. Das folgende Kapitel beschäftigt sich nun mit der Anwendung des Kriterienkatalogs anhand ausgewählter Visualisierungswerkzeuge.

Kapitel 5

Bewertung ausgewählter Werkzeuge

Im Rahmen dieses Kapitels werden zwei ausgewählte Visualisierungswerkzeuge vorgestellt und anhand des Kriterienkatalogs bewertet. Dabei fiel die Wahl der Untersuchungsobjekte auf Aduna AutoFocus, das über eine besonders aufwendige graphische Darstellung verfügt und den Longwell RDF Browser, der im Gegensatz dazu gänzlich auf eine graphische Repräsentation verzichtet. Dabei soll gezeigt werden, wie anhand des erstellten Kriterienkatalogs Werkzeuge mit unterschiedlichen Schwerpunkten bewertet werden können.

5.1 Aduna AutoFocus

Mit AutoFocus stellt Aduna¹ ein Werkzeug zur Desktop-Suche bereit, das eine auf Cluster Maps basierenden Ergebnisdarstellung bietet. AutoFocus ermöglicht die Extraktion und Indizierung von Metadaten aus verschiedenen Informationsquellen. Unter anderem werden dabei Verzeichnisse, Websites und E-Mail Accounts unterstützt. AutoFocus baut aus den extrahierten Informationen eine Datenbank auf, die auf Semantic Web Technologien basiert. AutoFocus greift dabei auf Sesame zurück, einem Semantic Web Data Store, der für die effiziente Speicherung und Abfrage des Datenbestands zuständig ist.

 $^{^{1}}$ www.aduna-software.com

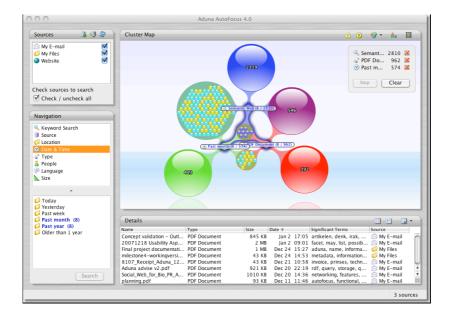


Abbildung 5.1: Aduna AutoFocus

5.1.1 Bewertung

Die Cluster Map Darstellung gilt als eine besonders aussagekräftige, visuelle Informationsdarstellung und schneidet bei entsprechenden Evaluierungen in Bezug auf die Usabilty der visuellen Repräsentation in der Regel ausgesprochen gut ab (Vgl. Fluit et. al 2006 [2]). Die Stärke von Aduna AutoFocus ist im Wesentlichen mit der Cluster Map Darstellung begründet (siehe Abschnitt 3.3). Das einfach gehaltene User Interface ermöglicht dem Benutzer durch den Datenbestand zu navigieren, ohne dabei Kenntnis über die zugrunde liegende Datenstruktur haben zu müssen.

Die Bewertung von Aduna AutoFocus erfolgt nun anhand der in Kapitel 4 vorgestellen Kriterien und der dazugehörigen Bewertungsskala. Die Bewertung basiert auf der Version 4.0 von Aduna AutoFocus.

Datentyp (data type)

AutoFocus unterstützt die Datentypen Tree und Network für Datenmaterial, das bereits in einem semantisch aufbereiteten Format vorliegt. Zusätzlich dazu ist AutoFocus in der Lage mit Hilfe von Informationsextraktion unter bestimmten Voraussetzungen eindimensionale, zweidimensionale und temporale Daten zu behandeln.

Bewertung: Abhängig von Anforderung (K.O. Kriterium)

Aussagekraft (expressiveness)

Aduna AutoFocus codiert alle Fakten des Sets und garantiert ausserdem deren Ausschließlichkeit. Die Aussagekraft ist daher in jedem Fall gegeben.

Bewertung: 3/3 Punkten

Effektivität (effectivness)

Cluster Maps liefern hervorragende Ergebnisse was die Interpretation der Darstellung betrifft. Einzelne Suchbegriffe, sowie deren Beziehungen zueinander werden klar und unmißverständlich dargestellt und können vom Betrachter schnell interpretiert werden (Vgl. Fluit et. al 2006 [2]).

Bewertung: 3/3 Punkten

Kognitive Komplexität (cognitive complexity)

Der Benutzer von AutoFocus hat die volle Kontrolle über die kognitive Komplexität der Darstellung. Werden zu Beginn noch weniger Informationselemente visualisiert, so steigt die Komplexität mit der Anzahl der hinzukommenden Suchbegriffe.

Bewertung: 3/3 Punkten

Räumliche Organisation (spatial organization)

Die Cluster Map Darstellung von AutoFocus ermöglicht durch geeignete Algorithmen eine nahezu optimale Anordnung der Informationselemente. Dabei kann vollständige und teilweise Verdeckung ausgeschlossen werden. Einzelne Informationsobjekte können einfach lokalisiert und bei Bedarf näher untersucht werden. Der Referenzkontext ist im Benutzer Interface zu jedem Zeitpunkt ersichtlich.

Bewertung: 3/3 Punkten

Informations coding (information coding)

AutoFocus verwendet Symbole, Farben und verschiedene Formen zur Informationscodierung. Die Darstellung dieser Codierungen ist dabei stets übersichtlich und nachvollziehbar.

Bewertung: 3/3 Punkten

Beziehungen (relations)

AutoFocus wurde grundsätzlich für die Suche und Visualisierung von Informationen, die aus unterschiedlichen Dokumenten stammen, entwickelt. Das hierbei verwendete Datenmaterial basiert in der Regel nicht auf Ontologien und verfügt deshalb nicht über typisierte Beziehungsinformationen. Auto-

Focus stellt Beziehungen anhand von Suchbegriffen her, die immer dann miteinander in Zusammenhang stehen, wenn die zugrunde liegenden Dokumente diese Begriffe teilen. Die Art der Darstellung dieser Beziehungen ist dabei äußerst aussagekräftig und effektiv. Sie ermöglicht dem Benutzer implizite Zusammenhänge zwischen unterschiedlichen Begriffen zu erkennen, die sich erst durch die Visualisierung des Datenmaterials ergeben.

Bewertung: 2/3 Punkten

Übergang (state transition)

Die Übergänge bei der Veränderung des aktuellen Kontexts werden in AutoFocus durch eine kurze Animation veranschaulicht. Dabei verändert sich die Anordnung der Darstellungselemente in der Art, dass der Benutzer diese Änderungen optisch nachvollziehen kann.

Bewertung: 3/3 Punkten

Orientierung und Hilfe (orientation and help)

Der Benutzer ist in der Lage, zu jedem Zeitpunkt zusätzliche Informationen zu den dargestellten Informationselementen einzuholen. Dazu werden zum einen Tooltips angeboten, zum anderen können entsprechende Meta-Informationen bei Bedarf in einer Tabelle eingesehen werden. AutoFocus unterstützt keine Kontrolle der Detailebene. Das Clustering erfolgt immer anhand von Begriffen bzw. Sucheigenschaften. AutoFocus bietet auch keine direkte Unterstützung für Undo / Redo Aktionen. Der Benutzer kann lediglich durch Hinzufügen oder Entfernen von Suchbegriffen den vorherigen Zustand wiederherstellen.

Bewertung: 1/3 Punkten

Navigation und Abfragen (navigation and querying)

Die Navigation geschieht im Wesentlichen durch Suchanfragen, eine explizite Formulierung von Abfragen ist aber nicht möglich. Zwar können einzelne Datenelemente eines Clusters ausfindig gemacht und untersucht werden, jedoch gibt es keine Möglichkeit, diese in die Darstellung einfließen zu lassen. Weiters fehlt dem Werkzeug eine Funktion, um den Standpunkt des Benutzers zu verändern. AutoFocus bietet dem Benutzer jedoch die Möglichkeit, die dargestellten Informationselemente selbst anzuordnen, um das Ergebnis weiter zu verbessern.

Bewertung: 2/3 Punkten

Reduktion des Datenbestands (data set reduction)

AutoFocus verwendet Clustering um den Datenbestand zu reduzieren und

große Informationsmengen darstellen zu können. Das Werkzeug bietet darüber hinaus jedoch kaum Möglichkeiten, die erhaltenen Suchergebnisse weiter zu filtern und unrelevante Informationen auszuschneiden.

Bewertung: 1/3 Punkten

Datenerkundung (data exploration)

AutoFocus eignet sich hervorragend zur Erkundung von unterschiedlichem Datenmaterial. Bereits ausgehend von einem Suchbegriff bietet das Werkzeug eine Reihe von Möglichkeiten zur sogenannten "Guided Exploration". Nachdem der erste Cluster anhand eines Suchbegriffs platziert wurde, werden in den betroffenen Dokumenten ebenfalls häufig verwendete Begriffe vorgeschlagen. Durch vorgeschlagene Ausprägungen von Meta-Informationen kann das Ergebnis weiter verfeinert und verbessert werden.

Bewertung: 3/3 Punkten

5.1.2 Bewertungstabelle

Tabelle 5.1: Bewertung von Aduna AutoFocus 4.0

Kriterium	Gewichtung	Bewertung
Usability der visuellen Repräsentation		
Datentyp (data type)	K.O.	
Aussagekraft (expressiveness)	25 %	3/3
Effektivität (effectiveness)	25 %	3/3
Kognitive Komplexität (cognitive complexity)	5,56 %	3/3
Räumliche Organisation (spatial organization)	5,56 %	3/3
Informationscodierung (information coding)	5,56 %	3/3
Beziehungen (relations)	5,56 %	2/3
Übergang (state transition)	5,56 %	3/3
Interface Usability		
Orientierung und Hilfe (orientation and help)	5,56 %	1/3
Navigation und Abfragen (navigation and querying)	5,56 %	2/3
Reduktion des Datenbestands (data set reduction)	5,56 %	1/3
Datenerkundung (data exploration)	5,56 %	3/3
Gesamt	100 %	2,67/3

5.1.3 Fazit

Dank der Cluster Map Darstellung bietet Aduna AutoFocus, in Bezug auf die Qualität der graphischen Repräsentation, herausragende Ergebnisse. Beim Visualisierungsvorgang können bisher versteckte Zusammenhänge erkannt und in weiterer Folge beispielsweise für strategische Entscheidungen genutzt werden. Durch die Unterstützung von Informationsextraktion aus gängigen Formaten unter der Zuhilfenahme von Meta-Information eröffnet sich ein äußerst breites Anwendungsgebiet. Von der Analyse von beliebigen Informationsmaterials können sowohl Firmen als auch Bildungseinrichtungen stark profitieren.

5.2 Longwell RDF Browser

Zunächst ist der Longwell RDF Browser kein Werkzeug zur Informationsvisualisierung im herkömmlichen Sinne, da keine graphische Repräsentation erzeugt wird. Das Werkzeug fällt in die Kategorie der Textual Interfaces und ist speziell auf die Navigation und Abfrage von semantischen Daten ausgerichtet. Longwell wurde im Rahmen des SIMILE¹ Projekts am MIT² entwickelt und basiert auf dem Prinzip von Faceted Navigation (siehe Abschnitt 3.5.1)

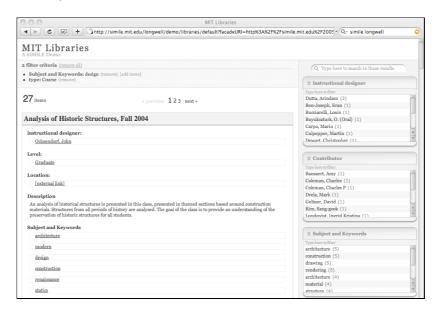


Abbildung 5.2: SIMILE Longwell RDF Browser

¹http://simile.mit.edu

²Massachusetts Institute of Technology - http://www.mit.edu

Die Bewertung eines Werkzeugs dieser Art anhand des Kriterienkatalogs ist insofern interessant, da es einen direkten Vergleich mit graphischen Tools ermöglicht. Dabei soll unter anderem die Frage geklärt werden, inwieweit ein rein textuelles Interface in der Lage ist, mit graphischen Werkzeugen zu konkurrieren.

5.2.1 Bewertung

Datentyp (data type)

Der Longwell RDF Browser arbeitet mit Daten, die im RDF Format vorliegen. Er unterstützt somit die Datentypen Tree und Network. Daten, die nicht als RDF vorliegen, müssen zunächst mit Informationsextraktionsmechanismen in dieses Format übergeführt werden, die jedoch nicht Bestandteil des Werkzeugs sind.

Bewertung: Abhängig von Anforderung (K.O. Kriterium)

Aussagekraft (expressiveness)

Longwell codiert alle Fakten des Sets und garantiert ausserdem deren Ausschließlichkeit. Die Aussagekraft ist daher in jedem Fall gegeben.

Bewertung: 3/3 Punkten

Effektivität (effectivness)

Im Unterschied zu graphischen Repräsentationen werden die Informationselemente nicht mittels Symbolen codiert. Die Interpretation der Informationselemente durch den Benutzer ist daher schnell und einfach möglich.

Bewertung: 3/3 Punkten

Kognitive Komplexität (cognitive complexity)

Die Darstellung enthält zum einen eine Listenansicht der aktuell ausgewählten Ergebnismenge, zum anderen einen Bereich mit unterschiedlichen Facetten zur Verfeinerung des Suchergebnisses. Die Anzahl der gleichzeitig angezeigten Elemente der Ergebnisliste ist begrenzt und nachfolgende Elemente sind mit Hilfe einer Blättern-Funktion zugänglich. Die dargestellte Lösung ist dabei in der Regel verständlich und leicht zu erfassen.

Bewertung: 3/3 Punkten

Räumliche Organisation (spatial organization)

Die Ergebnisse werden bei Longwell tabellarisch dargestellt und geben Aufschluss über Details zu einzelnen Elementen. Bei dieser schlichten Form der Darstellung der Ergebnismenge ist natürlich jede Art von Verdeckung ausge-

schlossen. Der Benutzer kann ausserdem jederzeit Details zu den einzelnen Informationselementen einsehen, während der Referenzkontext immer im Benutzer-Interface ersichtlich ist.

Bewertung: 2/3 Punkten

Informationscodierung (information coding)

Longwell unterstützt keinerlei Formen der Informationscodierung, die Darstellung erfolgt ausschließlich textuell.

Bewertung: 0/3 Punkten

Beziehungen (relations)

Beziehungen sind bei Longwell nur aus dem Kontext auszumachen und werden nicht explizit visualisiert. Da die Ergebnismenge durch den aktuellen Kontext, also die aktuelle Auswahl der Filterkriterien, definiert ist, kann der Benutzer den Bezug der Informationselemente zueinander jedoch zu jedem Zeitpunkt ausmachen. Während dies für die Erkundung des Datenmaterials, also die gezielte Suche nach Informationen, ausreichend ist, ist das Erkennen von impliziten Zusammenhängen jedoch nicht oder nur schwer möglich.

Bewertung: 1/3 Punkten

Übergang (state transition)

Ändert der Benutzer die aktuellen Such- und Filterkriterien, so wird die Ergebnisliste neu berechnet. Longwell bietet bei diesem Vorgang keinerlei graphische Visualisierung des Übergangs, der Benutzer behält jedoch die Übersicht mit Hilfe des aktuellen Referenzkontexts.

Bewertung: 1/3 Punkten

Orientierung und Hilfe (orientation and help)

Der Benutzer sieht zu jedem Zeitpunkt die Detailinformationen zu den einzelnen Informationsobjekten. Dies ist zwar eine informative Darstellung, hat jedoch wenig Flexibilität um beispielsweise größere Datenmengen anzuzeigen und bei Bedarf Details zu expandieren. Longwell bietet ausserdem keine direkte Unterstützung für Undo / Redo Aktionen. Der Benutzer kann lediglich durch Hinzufügen oder Entfernen von Suchbegriffen und Filterkriterien den vorherigen Zustand wiederherstellen.

Bewertung: 1/3 Punkten

Navigation und Abfragen (navigation and querying)

Die Navigation und Abfrage von Daten ist eine der wesentlichen Stärken

von Faceted Browsing, wie es bei Longwell zur Anwendung kommt. Der Benutzer ist in der Lage auf einfache Weise individuelle Abfragen zu erstellen. Die Suchergebnisse werden dabei schrittweise verfeinert, bis das gewünschte Endergebnis vorliegt. Der Benutzer kann ausserdem ausgehend von den dargestellten Elementen gezielt Eigenschaften und Objekte näher untersuchen und dabei seinen Standpunkt verändern.

Bewertung: 3/3 Punkten

Reduktion des Datenbestands (data set reduction)

Das Filtern von Daten auf Basis von Faceted Navigation ist äußerst intuitiv und bietet dem Benutzer besonders vielseitige Möglichkeiten den Datenbestand nach seinen Wünschen zu reduzieren. Die einzelnen Ausprägungen der Facetten (Facets) stellen Cluster dar, was bedeutet, dass auch Clustering unterstützt wird, wenn dies auch nicht direkt in der Ergebnisdarstellung geschieht.

Bewertung: 3/3 Punkten

Datenerkundung (data exploration)

Der Longwell RDF Browser wurde speziell für große Datenmengen entwickelt, die vom Benutzer untersucht werden können. Auch ohne Kenntnisse über das zugrunde liegende Datenmodell zu haben, kann der Benutzer, ausgehend vom gesamten Datenbestand, den Informationsbereich erkunden. Beginnend bei einer initialen Suchanfrage kann er sich mittels Faceted Browsing weiterbewegen und erhält vom Werkzeug ständig Hinweise bezüglich des zugrunde liegenden Datenbestands und dessen Struktur.

Bewertung: 3/3 Punkten

5.2.2 Bewertungstabelle

Tabelle 5.2: Bewertung des SIMILE Longwell RDF Browsers

Kriterium	Gewichtung	Bewertung
Usability der visuellen Repräsentation		
Datentyp (data type)	K.O.	
Aussagekraft (expressiveness)	25 %	3/3
Effektivität (effectiveness)	25 %	3/3
Kognitive Komplexität (cognitive complexity)	5,56 %	3/3
Räumliche Organisation (spatial organization)	5,56 %	2/3
Informationscodierung (information coding)	5,56 %	0/3
Beziehungen (relations)	5,56 %	1/3
Übergang (state transition)	5,56 %	1/3
Interface Usability		
Orientierung und Hilfe (orientation and help)	5,56 %	1/3
Navigation und Abfragen (navigation and querying)	5,56 %	3/3
Reduktion des Datenbestands (data set reduction)	5,56 %	3/3
Datenerkundung (data exploration)	5,56 %	3/3
Gesamt	100 %	2,44/3

5.2.3 Fazit

Anders als graphische Werkzeuge eignet sich Longwell vor allem zur Informationssuche in großen, strukturierten Datenbeständen. Durch die Anwendung von Faceted Navigation ergeben vielseitige Möglichkeiten zur Erkundung von Datenmaterial und dem Auffinden von relevanten Informationen. Im Gegensatz zu den meisten graphischen Werkzeugen ist das Ergebnis keine abstrahierte Veranschaulichung von Datenmaterial, sondern ausschließlich das Finden von Informationen.

5.3 Zusammenfassung

Mit Aduna AutoFocus und dem Longwell RDF Browser wurden zwei bekannte Werkzeuge zur Informationsvisualisierung vorgestellt. Obwohl die genannten Werkzeuge verschiedene Ansätze verfolgen (graphisch vs. textuell) konnten beide ein ausgesprochen gutes Ergebnis bei der Bewertung erreichen. Durch die Wahl zweier derart unterschiedlichen Werkzeuge konnte

ausserdem die erfolgreiche Anwendung des Kriterienkatalogs unter abweichenden Voraussetzungen veranschaulicht werden.

Kapitel 6

Schlussbetrachtung

6.1 Ergebnisse

Die Technologien des Semantic Web stellen für das Wissensmanagement ein ernormes Potential dar. Immer häufiger wollen Unternehmen und Bildungseinrichtungen diese Chance nutzen und mit Hilfe dieser neuen Art der Datenrepräsentation Wissensnetzwerke schaffen.

Damit dieses Vorhaben zum Erfolg führt, muss ein besonders leichter Zugang für eine breite Nutzerschicht ermöglicht werden. Da die Kerntechnologien des Semantic Web komplexe Strukturen aufweisen und der Umgang auf dieser Ebene Spezialisten vorbehalten bleibt, sind Werkzeuge erforderlich, die Benutzern aus verschiedenen Tätigkeitsbereichen die tägliche Arbeit mit diesen neuartigen Wissensnetzwerken ermöglicht.

Personen, die mit der Entwicklung und Betreuung dieser Systeme betraut sind, stehen vor der Aufgabe, die richtigen Werkzeuge für individuelle Problemstellungen auszuwählen. Werkzeuge zur Visualisierung von Informationen nehmen dabei eine wesentliche Rolle ein. Der aus dieser Arbeit resultierende Kriterienkatalog kann dazu verwendet werden, um Werkzeuge hinsichtlich ihrer Eignung für bestimmte Anwendungsgebiete zu bewerten. Es ist gelungen wesentliche Kriterien zu finden, anhand derer eine Bewertung möglich ist. Dabei wurde auf bestehende Kriterienkataloge zur Beurteilung von Techniken zur Informationsvisualisierung zurückgegriffen. Diese wurden ergänzt und unter dem Blickwinkel der Anwendung im Semantic Web betrachtet. Die einzelnen Kriterien beschreiben dabei unterschiedliche Gesichtspunkte nach denen ein Visualisierungswerkzeug bewertet werden kann. Da sich einzelne Eigenschaften verschieden stark auf die Eignung eines Werkzeugs auswirken, muss eine Gewichtung erfolgen. Die Aussagekraft und Effektivität einer Visualisierung werden beispielsweise besonders hoch gewichtet, da dadurch im Wesentlichen die Eignung zur Interpretation durch den Benutzer festgelegt wird. Eine individuelle Gewichtung ist jedoch für unterschiedliche Problemstellungen unbedingt erforderlich, um möglichst aussagekräftige Ergebnisse bei der Bewertung erzielen zu können. Es wurde die Vorgehensweise bei der Bewertung beschrieben und eine Bewertungsskala eingeführt, die bei der Evaluierung herangezogen werden kann.

Um die Anwendung des Kriterienkatalogs zu demonstrieren wurden zwei Werkzeuge exemplarisch untersucht und hinsichtlich ihrer Eignung für das Semantic Web bewertet. Dabei wurde sowohl ein graphisches Visualisierungwerkzeug (Aduna AutoFocus), das auf der Darstellung von Cluster Maps basiert, als auch ein rein textbasiertes Tool (SIMILE Longwell) näher analysiert.

Anhand dieser konkreten Beispiele wurde verdeutlicht, dass für verschiedene Problemstellungen durch den Einsatz differenzierter Technologien und Vorgehensweisen hervorragende Ergebnisse erzielt werden können.

6.2 Ausblick

Das Semantic Web ist derzeit ein äußerst lebendiges Thema. War die Anwendung dieser Technologien vor einigen Jahren noch vorwiegend im akademischen Bereichen angesiedelt, so werden semantische Netzwerke nun immer häufiger in unterschiedlichsten Bereichen für das Wissensmanagement eingesetzt. Die Entwicklung von geeigneten Werkzeugen schreitet voran und das Semantic Web rückt immer näher an den Endbenutzer. Noch hält sich die Verbreitung in Grenzen, jedoch könnte sich dies in den nächsten Jahren schnell ändern. Das Semantic Web könnte unter dem Namen "Web 3.0" eine neue Generation von Anwendungen einläuten. Aktuelle Projekte wie Freebase¹, eine für jeden zugängliche Online-Wissensdatenbank, sowie Twine², einer Plattform für neuartige Formen des Informationsaustausches, könnten dazu beitragen, dass in Zukunft öfter vom "Semantic Web" gesprochen wird.

¹http://www.freebase.com

²http://www.twine.com

Literaturverzeichnis

- [1] BÉNÉDICTE LE GRAND, M. S.: Topic Maps, RDF Graphs, and Ontologies Visualization, 2006.
- [2] Christiaan Fluit, Marta Sabou, F. v. H.: Ontology-Based Information Visualization: Toward Semantic Web Applications, 2006.
- [3] EDLINGER, K.: Informationsvisualisierung im Wissensmanagement Eine Analyse unterschiedlicher Visualisierungstechniken auf ihre Eignung für das Wissensmanagement. URL, http://eprints.rclis.org/archive/00006756/01/AC05381340.pdf, 2006.
- [4] FENTON, N. E. und S. L. PFLEEGER: Software Metrics: A Rigorous and Practical Approach, 1996.
- [5] Freitas, C. M. D. S., P. R. G. Luzzardi, R. A. Cava, M. A. A. Winckler, M. S. Pimenta und L. P. Nedel: *Evaluating Usability of Information Visualization Techniques*.
- [6] FRUCHTERMAN, T. M. J. und E. M. REINGOLD: *Graph Drawing by Force-directed Placement*. Software Practice and Experience, 21(11):1129–1164, 1991.
- [7] GRIGORIS ANTONIOU, F. v. H.: A Semantic Web Primer. The MIT Press, Reading, 2004.
- [8] J. Bieri A. L. Atkins, S. Briar, E. A.: Clinical and social judgment: the discrimination of behavioural information, 2006.
- [9] Kleske, J.: Wissensarbeit mit Social Software, 2006.
- [10] Mackinlay, J.: Automating the design of graphical presentations of relational information. ACM Trans. Graph., 5, 1986.
- [11] MAY, W.: Semantic Web Wege zur vernetzten Wissensgesellschaft, Kap. Reasoning im und für das Semantic Web. Pelligrini u. Blumauer, Springer, 2006.
- [12] RANGANATHAN, S. R.: Colon Classification, 7th edition, 1987.

- [13] Shneiderman, B.: Tree visualization with tree-maps: 2-d space-filling approach. ACM Trans. Graph., 11(1):92–99, 1992.
- [14] Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Techn. Ber. CS-TR-3665, University of Maryland, Department of Computer Science, July 1996.
- [15] TIM BERNERS-LEE, JAMES HENDLER, O. L.: *The Semantic Web*. URL, http://www.sciam.com/article.cfm?articleID= 00048144-10D2-1C70-84A9809EC588EF21, May 2001.

Abbildungsverzeichnis

2.1	RDFGraph	7
3.1	Clustermap	14
3.2	TreeMap	15
3.3	ElasticLists	17
3.4	HyperbolicTree	19
3.5	3DTree	20
	IsaViz	
3.7	FDP	22
5.1	AutoFocus	36
5.2	Longwell	40