

# Knowledge Discovery durch Text Mining

Einsatz intelligenter Systeme  
zur Akquisition, Darstellung und Verteilung  
von textbasiertem Wissen

Diplomarbeit  
Studiengang Informationswirtschaft

Fakultät für Informations- und Kommunikationswissenschaft  
Fachhochschule Köln  
University of Applied Science Cologne

vorgelegt am 2004-07-05 von Stefan Koch  
Matrikelnummer: 11023811  
email: [diplom@stefkoch.de](mailto:diplom@stefkoch.de)  
www: <http://stefkoch.de/diplom/>

Prof. Dr.phil. K. Lepsky  
Erstprüfer

Prof. Dipl.-Math. W. Gödert  
Zweitprüfer

"Like distant islands sundered by the sea,  
We had no sense of one community.  
We lived and worked apart and rarely knew  
That others searched with us for knowledge, too. [...]  
But, could these new resources not be shared?  
Let links be built; machines and men be paired!  
Let distance be no barrier! They set  
That goal: design and built the ARPANET! [...]  
The second node, the NIC, was soon installed.  
The Network Info Center, it was called.  
Hosts and users, services were touted:  
To the NIC was network knowledge routed."<sup>1</sup>

"The first phase of the Web is  
human communication through shared knowledge.  
We have a lot of work to do before we have  
an intuitive space in which we can  
put down our thoughts and build our understanding of  
what we want to do and how and why we will do it.  
The second side to the Web, yet to emerge,  
is that of machine-understandable information.  
As this happens, the day-to-day mechanisms of  
trade and bureaucracy will be handled by  
agents, leaving humans to provide  
the inspiration and the intuition."<sup>2</sup>

---

1 [Cerf1969]

2 [Berners-Lee1997]

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>4</b>
1.1	Motivation der Arbeit	4
1.2	Zielsetzung	5
1.3	Eingrenzung	6
1.4	Vorgehensweise	7
<b>2</b>	<b>Begriffsklärung</b>	<b>8</b>
<b>3</b>	<b>Extraktion</b>	<b>15</b>
3.1	Wissensumfeld	16
3.2	Information Retrieval	20
3.3	Information Extraction	21
3.4	Knowledge Extraction	23
<b>4</b>	<b>Analyse</b>	<b>25</b>
4.1	Normierung	26
4.2	Indexierung	27
4.3	Ähnlichkeit	30
4.4	Relevanz	33
<b>5</b>	<b>Abbildung</b>	<b>34</b>
5.1	Klassifikation	36
5.2	Clustering	39
5.3	Topic Maps	42
5.4	Thesaurus	43
5.5	Ontologien	44
<b>6</b>	<b>XML-Standards</b>	<b>49</b>
6.1	Struktur	50
6.2	Extraktion	51
6.3	Abbildung	55
<b>7</b>	<b>Transfer</b>	<b>60</b>
<b>8</b>	<b>Integration</b>	<b>65</b>
<b>9</b>	<b>Zusammenfassung und Ausblick</b>	<b>69</b>
<b>10</b>	<b>Anhang</b>	<b>72</b>
10.1	Text Mining Tools	72
10.2	DAML-OWL-Beispiel	74
	Abkürzungen	76
	Abbildungen	77
	Literatur	77
	Eidesstattliche Erklärung	87

## 10 Anhang

### 10.1 Text Mining Tools

Diese Liste ist eine Zusammenstellung der in der Arbeit erwähnten Anwendungen und Hilfsmittel. Sie sind entsprechend ihres Auftretens in der Arbeit angeordnet.

#### Content Management Systeme

- Wiki: <http://wiki.org>
- Twiki <http://twiki.org>

#### Knowledgebases

- Cyc Inc., OpenCyc: <http://www.opencyc.org>, <http://www.cyc.com>
- MIT, OpenMind,: <http://commonsense.media.mit.edu>
- Stanford, TAP: <http://tap.semanticweb.org>, <http://tap.stanford.edu/tapkb/>

#### Information Extraction

- ANNIE: <http://gate.ac.uk/annie/index.jsp>

#### Knowledge Extraction

- Natural Language Toolkit: <http://nltk.sourceforge.net>
- Megaputer, TextAnalyst, PolyAnalyst: <http://www.megaputer.com/products/>

#### Klassifikationen

- NACE: <http://www.fifoost.org/database/nace/>
- eCl@ss: <http://www.eclass.de>
- DDC: <http://www.oclc.org/dewey/>, <http://www.ddc-deutsch.de>
- UDC: <http://www.udcc.org>

#### Thesauri

- Openthesaurus: <http://www.openthesaurus.de>, <http://openthesaurus.sf.net>
- Standard Thesaurus Wirtschaft: <http://www.gbi.de/thesaurus/>

#### Topic Maps

- Mindjet, MindManager: <http://www.mindjet.com>

## Ontologien

- SUMO: <http://ontology.teknowledge.com>
- Dun & Bradstreet: <http://www.ksl.stanford.edu/projects/DAML/UNSPSC.daml>
- DAML Verzeichnis: <http://www.daml.org/ontologies/>
- WordNet als OWL: <http://taurus.unine.ch/GroupHome/kowler/wordnet.html>

## Ontologietools

- KAON: <http://sourceforge.net/projects/kaon/>
- TextToOnto: <http://sourceforge.net/projects/texttoonto/>
- Protégé, <http://protege.stanford.edu>
- Chimaera: <http://www.ksl.stanford.edu/software/chimaera/>
- OntoSQL: <http://www.aifb.uni-karlsruhe.de/WBS/aeb/ontosql/>
- OntoAgent: <http://www.i-u.de/schools/eberhart/ontoagent/>
- OntoView <http://ontoview.org>
- Ontoprise (diverse): <http://www.ontoprise.de>

## Transfer

- SCORM Conformance Test Suite:  
<http://www.adlnet.org/index.cfm?fuseaction=SCORDown&listing=Software>
- Mindswap, OntoLink:  
<http://www.mindswap.org/2004/OntoLink/>

## Integration

- Teknowledge: SemanticWord, Briefing Associate  
<http://mr.teknowledge.com/DAML>
- WEKA:  
<http://sourceforge.net/projects/weka/>, <http://www.cs.waikato.ac.nz/~ml/>
- GATE:  
<http://gate.ac.uk>
- Intelligent Miner for Text, IBM:  
<http://www.ibm.com/software/data/iminer/fortext/>
- SAS®, Text Miner:  
<http://www.sas.com/technologies/analytics/datamining/textminer/>
- SPSS, Text Mining for Clementine®:  
[http://www.spss.com/lexiquest/text\\_mining\\_for\\_clementine.htm](http://www.spss.com/lexiquest/text_mining_for_clementine.htm)
- g.a.d.t GmbH:  
<http://gadt.de>
- moresophy GmbH, L4:  
<http://www.moresophy.de>

## 10.2 DAML-OWL-Beispiel

Um ein Beispiel für in Ontologien abgebildetes Wissen zu geben, werden im Folgenden zwei mögliche Schreibweisen desselben Zusammenhangs gegeben. Beide drücken aus, dass jedes Betriebssystem Software darstellt und das Betriebssystem Linux kostenlos ist.

DAML

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
  xmlns:ns0="http://stefkoch.de/diplom/linux.daml"
  xmlns:xsd="http://www.w3.org/2000/10/XMLSchema#"
  xml:base="http://stefkoch.de/diplom/">

  <daml:Class rdf:about="linux.daml#Software">
    <rdfs:label>Software</rdfs:label>
  </daml:Class>
  <daml:Class rdf:about="linux.daml#Betriebssystem">
    <rdfs:label>Betriebssystem</rdfs:label>
    <rdfs:subClassOf>
      <daml:Class rdf:about="linux.daml#Software"/>
    </rdfs:subClassOf>
  </daml:Class>

  <daml:ObjectProperty rdf:about="linux.daml#kostenlos">
    <rdfs:label>kostenlos</rdfs:label>
  </daml:ObjectProperty>

  <rdf:Description rdf:about="linux.daml#Linux">
    <rdf:type>
      <daml:Class rdf:about="linux.daml#Betriebssystem"/>
    </rdf:type>
    <ns0:kostenlos rdf:resource="linux.daml#Linux"/>
  </rdf:Description>

</rdf:RDF>
```

## OWL

```
<?xml version="1.0" encoding="UTF-8"?>
<owls:Ontology xmlns:owls="http://www.w3.org/2003/OWL/XMLSchema"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xml:base="http://stefkoch.de/diplom/">

  <owls:Class owls:name="linux.owl#Software"/>
  <owls:Class owls:name="linux.owl#Betriebssystem"/>

  <owls:SubClassOf>
    <owls:super>
      <owls:Class owls:name="linux.owl#Software"/>
    </owls:super>
    <owls:sub>
      <owls:Class owls:name="linux.owl#Betriebssystem"/>
    </owls:sub>
  </owls:SubClassOf>

  <owls:ObjectProperty owls:name="linux.owl#kostenlos"/>

  <owls:Individual owls:name="linux.owl#Linux">
    <owls:type owls:name="linux.owl#Betriebssystem"/>
    <owls:ObjectPropertyValue owls:property="linux.owl#kostenlos"/>
  </owls:Individual>

</owls:Ontology>
```

Der wesentliche Unterschied zwischen beiden Schreibweisen ist, dass in OWL die Hierarchiebeziehungen von Klassen separat in einem "owls:SubClassOf"-Tag und Instanzen in einem "owls:Individual"-Tag definiert werden können, während DAML hierfür auf die RDF-Definitionen zurückgreift. Die Zuweisung von Properties muss damit bei DAML über den XML-Namespace erfolgen, während bei OWL "owls:ObjectPropertyValue" verwendet werden kann.

## Abkürzungen

ACM – Association for Computing Machinery  
ADL – Advanced Distributed Learning Initiative  
AI – Artificial Intelligence, auch KI  
AICC – Aviation Industry CBT Committee  
AIEE – American Institute of Electrical Engineers  
ARPA – Advanced Research Projects Agency  
ASCII – American Standard Code for Information Interchange  
CBT – Computer Based Training  
CMS – Content Management System  
DAML – Darpa Agent Markup Language  
DARPA – Defense Advanced Research Projects Agency  
DBMS – Datenbank Management System  
DDC – Dewey Decimal Classification  
DTD – Document Type Definition  
EML – Educational Modeling Language  
ERP – Enterprise Resource Planning  
FAQ – Frequently Asked Questions  
HTML – HyperText Markup Language  
ISIC – International Standard Industrial Classification of all Economic Activities  
KD – Knowledge Discovery  
KDD – Knowledge Discovery in Databases  
KI – Künstliche Intelligenz, auch AI  
KIF – Knowledge Interchange Format  
KML – Knowledge Modelling Language  
KQML – Knowledge Query and Manipulation Language  
KRL – Knowledge Representation Language  
KRRS – Knowledge Representation and Reasoning System  
KRS – Knowledge Representation System  
MIS – Management Information System  
MIT – Massachusetts Institute of Technology  
NACE – Wirtschaftszweigklassifikation der Europäischen Union  
OCR – Optical Character Recognition  
OIL – Ontology Inference Language  
OKBC – Open Knowledge Base Connectivity  
OLAP – OnLine Analytical Processing  
OML – Ontology Modelling Language  
OSD – Office of the Secretary of Defense  
RDF – Resource Description Framework  
SCORM – Sharable Content Object Reference Model  
SHOE – Simple HTML Ontology Extensions  
SQL – Structured Query Language  
SWRL – Semantic Web Rule Language  
TM – Text Mining  
TREC – Text Retrieval Conferences  
UDC – Universal Decimal Classification  
W3C – World Wide Web Consortium  
WBT – Web Based Training  
WWW – World Wide Web  
OWL – Web Ontology Language  
OML – Ontology Modelling Language  
XML – eXtensible Markup Language  
XOL – XML Ontology Exchange Language  
XQL – XML Query Language  
XIRQL – XML Information Retrieval Language  
XML – eXtensible Markup Language  
XSD – XML Schema Definition  
XXL – FleXible XML Search Language



## Abbildungen

Abb. 1: KAON-Beispiel "BibTeX Ontologie"	45
Abb. 2: Protégé-Beispiel "Suche"	46
Abb. 3: Chimaera-Beispiel "Superklasse hinzufügen"	47
Abb. 4: Schichten des Semantic Web nach T. Berners-Lee	63

## Literatur

- [Aamodt1995]: Agnar Aamodt, Mads Nygård, "Different roles and mutual dependencies of data, information, and knowledge - an AI perspective on their integration", Data Knowledge Engineering, North-Holland Elsevier, vol.16, 1995, S.191-222, <http://citeseer.ist.psu.edu/aamodt95different.html>.
- [Abiteboul1997]: Serge Abiteboul, Dallan Quass, Jason McHugh, Jennifer Widom, Janet L. Wiener, "The Lorel query language for semistructured data", International Journal on Digital Libraries, Vol. 1, 1997, <http://citeseer.ist.psu.edu/abiteboul97lorel.html>.
- [Agrawal1995]: Rakesh Agrawal, Giuseppe Psaila, "Active Data Mining", 1st International Conference on Knowledge Discovery and Data Mining (KDD-95), <http://citeseer.ist.psu.edu/agrawal95active.html>, S.1
- [Ahonen-Myka2002]: Helena Ahonen-Myka, "Discovery of Frequent Word Sequences in Text", The ESF Exploratory Workshop on Pattern Detection and Discovery in Data Mining, Imperial College, London, 2002, <http://citeseer.ist.psu.edu/534626.html>.
- [Ahonen1997]: Helena Ahonen, Oskari Heinonen, Mika Klemettinen, A. Inkeri Verkamo, "Applying Data Mining Techniques in Text Analysis", 1997, <http://citeseer.ist.psu.edu/ahonen97applying.html>, S. 9ff..
- [Ahonen1998]: Helena Ahonen, Oskari Heinonen, Mika Klemettinen A. Inkeri Verkamo, "Applying Data Mining Techniques for Descriptive Phrase Extraction in Digital Document Collections", Advances in Digital Libraries, 1998, <http://citeseer.ist.psu.edu/ahonen98applying.html>, S.2-11.
- [Aitken2000], "Evaluation of an Ontology-Based Information Retrieval Tool", Workshop on the Applications of Ontologies and Problem-Solving Methods, European Conference on Artificial Intelligence, Berlin, 2000 <http://www.aiai.ed.ac.uk/~stuart/Papers/ontologyeval.pdf>.
- [Al-Khalifa2003]: Shurug Al-Khalifa, Cong Yu, H. V. Jagadish, "Querying Structured Text in an XML Database", SIGMOD, 2003, <http://citeseer.ist.psu.edu/alkhalifa03querying.html>.
- [Ankerst2000]: Mihael Ankerst, Martin Ester, Hans-Peter Kriegel, "Towards an Effective Cooperation of the Computer and the User for Classification", Proceedings of the 6th Int. Conference on Knowledge Discovery and Data Mining (KDD) 2000, Boston, MA, <http://citeseer.ist.psu.edu/ankerst00towards.html>.
- [Baeza-Yates1999]: Ricardo Baeza-Yates, Berthier Ribeiro-Neto, "Modern Information Retrieval", ACM Press, Addison Wesley, New York, 1999, S.124ff..
- [Beier2003]: H. Beier, "Intelligente Informationsstrukturierung und TextMining mit Semantischen Netzen. Intelligent information structuring and text mining with semantic networks", Proceedings of Competence in Content: 25. Online-Tagung der DGI, Hrsg.: R. Schmidt, Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis (DGI), Frankfurt am Main, DE, 2003, S.78-87, Quelle: INFODATA, FIZ Technik.

- [Berners-Lee1997]: Tim Berners-Lee, "Realising the Full Potential of the Web", Based on a talk presented at the W3C meeting, London, <http://www.w3.org/1998/02/Potential.html>
- [Bernatzki1996] A. Bernatzki, W. Eppler, H. Gemmeke, "Interpretation of Neural Networks for Classification Tasks", Proceedings of EUFIT 1996, Aachen, Germany, <http://citeseer.ist.psu.edu/903.html>.
- [Besançon1998]: Martin Rajman, Romaric Besançon, "Text Mining - Knowledge extraction from unstructured textual data", 6th Conference of International Federation of Classification Societies (IFCS-98), Rome, 1998, <http://citeseer.ist.psu.edu/besanon98text.html>.
- [Bobrov1977]: D. G. Bobrov, R. Uinograd "An Overview of KRL, a knowledge Representation Language.", Cognitive Science, 1977, S. 46ff. <ftp://reports.stanford.edu/pub/cstr/reports/cs/tr/76/581/CS-TR-76-581.pdf>.
- [Bollacker1998]: Kurt D. Bollacker, Steve Lawrence, C. Lee Giles, "CiteSeer: An Autonomous Web Agent for Automatic Retrieval and Identification of Interesting Publications", Proceedings of the Second International Conference on Autonomous Agents, 1998 <http://citeseer.ist.psu.edu/bollacker98citeseer.html>.
- [Borkar2001]: Vinayak Borkar, Kaustubh Deshmukh, Sunita Sarawagi, "Automatic segmentation of text into structured records", Indian Institute of Technology - (DATAMOLD-System), Bombay, ACM SIGMOD 2001, <http://citeseer.ist.psu.edu/borkar01automatic.html>, S.175.
- [Broekstra2003]: Jeen Broekstra, Marc Ehrig, Peter Haase, Frank van Harmelen, Arjohm Kampman, Marta Sabou, Ronny Siebes, Steen Staab, Heiner Stuckenschmidt, Christoph Tempich, "A Metadata Model for Semantics-Based Peer-To-Peer Systems ", 2003 <http://citeseer.ist.psu.edu/584933.html>.
- [Callan2003]: Robert Callan, "Neuronale Netze im Klartext", München, Pearson, 2003, S.27.
- [Cerf1969]: Vint Cerf, "Requiem for the Arpanet", <http://www.etext.org/Politics/Essays/arpanet>
- [Chamberlin2000]: Don Chamberlin, Jonathan Robie, Daniela Florescu, "Quilt: An XML Query Language for Heterogeneous Data Sources", Lecture Notes in Computer Science, IBM Almaden Research Center, 2000, <http://citeseer.ist.psu.edu/chamberlin00quilt.html>.
- [Chaudhri1998a]: V. K. Chaudhri, A. Farquhar, R. Fikes, P. D. Karp, J. P. Rice, "Open Knowledge Base Connectivity 2.0.3 Proposed ", Artificial Intelligence Center SRI International, Knowledge Systems Laboratory Stanford University, 1998, <http://www-ksl-svc.stanford.edu:5915/doc/release/okbc/okbc-spec/okbc-2-0-3.pdf>, S.1.
- [Chaudhri1998b]: V. K. Chaudhri, A. Farquhar, R. Fikes, P. D. Karp, J. Rice, "OKBC: A programmatic foundation for knowledge base interoperability", Proceedings of the 15th National Conference on Artificial Intelligence (AAAI'98), S.600-607, 1998, <http://citeseer.ist.psu.edu/chaudhri98okbc.html>.
- [Codd1993]: E.F. Codd, S.B. Codd, C.T. Sally, "Providing OLAP (On-Line Analytical Processing) to User-Analysts - an IT mandat." White paper E.F.,Codd & Associates, 1993.
- [Compton1996]: P. Compton, P. Preston, G. Edwards, B. Kang, "Knowledge Based Systems That Have Some Idea of Their Limits", 1996, <http://citeseer.ist.psu.edu/compton96knowledge.html>.
- [Corcho2000a]: Oscar Corcho, Asunción Gómez-Pérez, "A Roadmap to Ontology Specification Languages", Proceedings of the 12th International Conference on

- Knowledge Engineering and Knowledge Management (EKAW'00), Juan-les-Pins France, October 2000, <http://delicias.dia.fi.upm.es/articulos/ocorcho/ekaw2000-corcho.pdf>.
- [Corcho2000b]: Oscar Corcho, Asunción Gómez-Pérez, "Evaluating Knowledge Representation and Reasoning Capabilities of Ontology Specification Languages", Proceedings of the ECAI'00 Workshop on Applications of Ontologies and Problem Solving Methods, Berlin Germany, 2000, <http://citeseer.ist.psu.edu/corcho00evaluating.html>., S.5.
  - [Cui2003]: Z. Cui, J. W. Shepherdson, Y. Li, "An ontology-based approach to eCatalogue management", BT Technology Journal, Vol 21 No 4, October 2003, <http://www.kluweronline.com/article.asp?PIPS=5254790>.
  - [Dasigi1996]: Venu Dasigi, Reinhold Mann, "Neural Net Learning Issues in Classification of Free Text Documents ", AAAI Spring Symposium on Machine Learning in Information Access Technical Papers, 1996, <http://citeseer.ist.psu.edu/dasigi96neural.html>.
  - [Davenport1998]: Thomas Davenport, Laurenc Prusak, "Wenn ihr Unternehmen wüßte, was es alles weiß. Das Praxisbuch zum Wissensmanagement", Landsberg/Lech., 1998. S.186.
  - [Davis1993]: Randall Davis, Howard Shrobe, Peter Szolovits, "What is a Knowledge Representation?", AI Magazine, 14, S.17-33, MIT AI Lab, 1993, <http://medg.lcs.mit.edu/ftp/psz/aimag-final.ps>.
  - [DeBra1994]: Paul De Bra, Geert-Jan Houben, Joep De Vocht, Yoram Kornatzky, "Retrieval of Hypertext Structures", Proceedings of Stinfon-94 Conference, Tilburg, 1994, <http://citeseer.ist.psu.edu/108018.html>.
  - [Deutsch1998]: Alin Deutsch, Mary Fernandez, Daniela Florescu, Alon Levy, Dan Suciu, "XML-QL: A Query Language for XML", Proceedings of WWW The Query Language Workshop QL, Cambridge, MA, 1998, <http://citeseer.ist.psu.edu/390950.html>.
  - [Desmontils2001]: Emmanuel Desmontils, Christine Jacquin, "Indexing a Web Site with a Terminology Oriented Ontology", Proceedings of SWWS'01, The first Semantic Web Working Symposium, Stanford University, 2001, <http://citeseer.ist.psu.edu/et02indexing.html>, S.7.
  - [Ding2002]: Anne Denton, Qiang Ding, Qin Ding, William Perrizo, "Efficient Hierarchical Clustering of Large Data Sets Using P-trees", Proceedings of 15th International Conference on Computer Applications in Industry and Engineering (CAINE'02), San Diego, CA, Nov. 2002, S. 138-141, [http://cs.hbg.psu.edu/~ding/publications/CAINE\\_109.pdf](http://cs.hbg.psu.edu/~ding/publications/CAINE_109.pdf).
  - [Dixon1997]: M. Dixon, "An Overview of Document Mining Technology", 1997, <http://citeseer.ist.psu.edu/dixon97overview.html>., S.1.
  - [Doan2002]: AnHai Doan, Jayant Madhavan, Pedro Domingos, Alon Halevy, "Learning to Map between Ontologies on the Semantic Web." , Proceedings of the 11th International World Wide Web Conference (S.662-673), 2002. Honolulu, ACM Press, <http://www.cs.washington.edu/homes/pedrod/papers/www02.pdf>, S.4.
  - [Eberhart2002]: Andreas Eberhart, "OntoAgent: A Platform for the Declarative Specification of Agents", Proceedings of the ISWC 2002 Rule Markup Languages for Business Rules on the Semantic Web, <http://citeseer.ist.psu.edu/eberhart02ontoagent.html>.
  - [Ester1998]: Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, "Clustering for Mining in Large Spatial Databases", Special Issue on Data Mining, KI-Journal, ScienTec Publishing, Vol. 1, 1998,

- [http://www.cs.helsinki.fi/u/gionis/seminar\\_papers/ester98clustering.pdf](http://www.cs.helsinki.fi/u/gionis/seminar_papers/ester98clustering.pdf), S.2
- [Ester2000]: Mihael Ankerst, Martin Ester, Hans-Peter Kriegel, "Cooperative Classification: A Visualization-Based Approach of Combining the Strengths of the User and the Computer", Data Mining and Knowledge Discovery Journal, Kluwer Academic Publishers. <http://citeseer.ist.psu.edu/455891.html>.
  - [Fayyad1996a]: Usama Fayyad , Gregory Piatetsky-Shapiro, Padhraic Smyth, "The KDD Process for Extracting Useful Knowledge from Volumes of Data", Communication of the ACM, Vol.29, 1996, <http://citeseer.ist.psu.edu/fayyad96kdd.html>.
  - [Fayyad1996b]: Usama Fayyad , Gregory Piatetsky-Shapiro, Padhraic Smyth, "From Data Mining to Knowledge Discovery in Databases", AI Magazine, American Association for Artificial Intelligence, 0738-4602-1996, 1996, S.39, <http://citeseer.ist.psu.edu/fayyad96from.html>.
  - [Feldman1998]: Ronen Feldman, Moshe Fresko, Yakkov Kinar, Yehuda Lindell, Orly Liphstat, Martin Rajman , Yonatan Schler, Oren Zamir, "Text Mining at the Term Level", Principles of Data Mining and Knowledge Discovery, 1998, <http://citeseer.ist.psu.edu/feldman98text.html>.
  - [Fensel2000]: Dieter Fensel, Ian Horrocks, Frank van Harmelen, Stefan Decker, Michael Erdmann, and Michel C. A. Klein, "OIL in a Nutshell", Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling, and Management (EKAW'00)", Springer, 2000, <http://www.cs.vu.nl/~ontoknow/oil/down/oilnutshell.pdf>
  - [Fensel2001]: Dieter Fensel, "Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce.", Springer-Verlag, 2001, <http://citeseer.ist.psu.edu/413498.html>.
  - [Franke2003]: Ingrid Renz, Jürgen Franke, "Text Mining", Text Mining - Theoretical Aspects and Applications, Physica, 2003, S.1f.
  - [Freitag1998]: Dayne Freitag, "Information Extraction from HTML: Application of a General Machine Learning Approach", AAAI, 1998, <http://citeseer.ist.psu.edu/freitag98information.html>.
  - [Fuhr2001]: Norbert Fuhr, Kai Großjohann, "XIRQL: A Query Language for Information Retrieval in XML Documents", 2001 <http://citeseer.ist.psu.edu/fuhr01xirql.html>.
  - [Fuhr2002]: Mohammad Abolhassani, Norbert Fuhr, Norbert Gövert, Kai Großjohann, "HyREX: Hypermedia Retrieval Engine for XML", Research Report an der University of Dortmund, Department of Computer Science, 2002, [http://www.is.informatik.uni-duisburg.de/bib/xml/Fuhr\\_etal\\_02b.html](http://www.is.informatik.uni-duisburg.de/bib/xml/Fuhr_etal_02b.html).
  - [Fuhr2003]: Norbert Fuhr, Kai Großjohann, S. Kriewel, "A Query Language and User Interface for XML Information Retrieval", Intelligent XML Retrieval Vol. 2818, Springer, 2003, [http://www.is.informatik.uni-duisburg.de/bib/xml/Fuhr\\_etal\\_03a.html](http://www.is.informatik.uni-duisburg.de/bib/xml/Fuhr_etal_03a.html)
  - [Fuhr2004]: Norbert Fuhr, Kai Großjohann, "XIRQL: An XML Query Language Based on Information Retrieval Concepts", ACM Transactions on Information Systems, Volume 22, 2004, S.313—356, [http://www.is.informatik.uni-duisburg.de/bib/xml/Fuhr\\_Grossjohann\\_04.html](http://www.is.informatik.uni-duisburg.de/bib/xml/Fuhr_Grossjohann_04.html), S.36.
  - [Gemert2000]: Jan van Gemert, "Text Mining Tools on the Internet - An overview", Intelligent Sensory Information Systems (ISIS), University of Amsterdam, 2000, [http://carol.science.uva.nl/~jvgemert/mia\\_page/textminingtools.pdf](http://carol.science.uva.nl/~jvgemert/mia_page/textminingtools.pdf)
  - [Gil2000]: Yolanda Gil, Varun Ratnakar, "A Comparison of (Semantic) Markup Languages", American Association for Artificial Intelligence, 2000, <http://www.isi.edu/expect/web/semanticweb/paper.pdf>.

- [Giles1998]: C. Lee Giles, Kurt D. Bollacker, Steve Lawrence, "CiteSeer: An Automatic Citation Indexing System", Digital Libraries 98 - The Third ACM Conference on Digital Libraries, 1998, <http://citeseer.ist.psu.edu/108208.html>.
- [Gödert1998]: Elisabeth Sachse, Martina Liebig, Winfried Gödert, "Automatische Indexierung unter Einbeziehung semantischer Relationen: Ergebnisse des Retrievaltests zum MILOS II-Projekt.", Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft, Band 14, 1998, <http://www.fbi.fh-koeln.de/institut/papers/kabi/volltexte/band014.pdf>, S.37.
- [Gómez-Pérez2004]: Asunción Gómez-Pérez, Mariano Fernández-López, Oscar Corcho, "Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web ", Springer Informatik, 2004.
- [Gonzalo1998]: Julio Gonzalo, Felisa Verdejo, Irina Chugur, Juan Cigarran, "Indexing with WordNet synsets can improve Text Retrieval", Proceedings of the COLING/ACL '98 Workshop on Usage of WordNet for NLP, <http://citeseer.ist.psu.edu/gonzalo98indexing.html>
- [Guha2003]: R. Guha, Rob McCool, "TAP: A Semantic Web Platform", Computer Networks: The International Journal of Computer and Telecommunications Networking , Volume 42 , Special issue: The Semantic Web: an evolution for a revolution, 2003, <http://tap.stanford.edu/tap.pdf>.
- [Hahn1997]: Udo Hahn, Klemens Schnattinger, "A Qualitative Growth Model for Real-World Text Knowledge Bases", RIAO`97 – Proceedings of the 5th Conference on Computer-Assisted Information Searching on the Internet, Montreal, Quebec, Canada, 1997, S.578-597 <http://citeseer.ist.psu.edu/hahn97qualitative.html>, S.3.
- [Heikkinen2000]: B. Heikkinen, "Generalization of Document Structures and Document Assembly", 2000, <http://citeseer.ist.psu.edu/heikkinen00generalization.html>, S.1.
- [Hoeren1998]: T. Hoeren, "Internet und Recht – Neue Paradigmen des Informationsrechts.", Neue Juristische Wochenschrift 51, 1998, S.2854, aus: [Kuhlen1999], S.367
- [Hönig1998]: Thomas Hönig, "Data Warehousing, Data Mining OLAP", Hrsg.: Wolfgang Martin, International Thomson Publishing, Bonn, 1998.
- [Honkela1997]: Timo Honkela "WEBSOM Self-Organizing Maps of Document Collections" , Proceedings of Workshop on Self-Organizing Maps WSOM'97, Espoo, Finland, <http://citeseer.ist.psu.edu/honkela97websom.html>.
- [Horrocks2000]: Ian Horrocks, "A Denotational Semantics for OIL-Lite and Standard Oil", Department of Computer Science University of Manchester, UK, 2000, <http://citeseer.ist.psu.edu/337591.html>.
- [Hotho2002]: Andreas Hotho, Alexander Maedche, Steffen Staab, "Ontology-based Text Document Clustering", [http://www.aifb.uni-karlsruhe.de/WBS/aho/pub/Ontology\\_based\\_Text\\_Document\\_Clustering\\_2002.pdf](http://www.aifb.uni-karlsruhe.de/WBS/aho/pub/Ontology_based_Text_Document_Clustering_2002.pdf).
- [Hotho2003a]: A. Hotho, S. Staab, G. Stumme, "WordNet improves text document clustering", Proceedings of the SIGIR 2003 Semantic Web Workshop, [http://www.aifb.uni-karlsruhe.de/WBS/aho/pub/hothoetal\\_sigir\\_ws\\_sem\\_web.pdf](http://www.aifb.uni-karlsruhe.de/WBS/aho/pub/hothoetal_sigir_ws_sem_web.pdf)
- [Hotho2003b]: A. Hotho, S. Staab, G. Stumme, "Ontologies Improve Text Document Clustering", Proceedings of the ICDM 03, The 2003 IEEE International Conference on Data Mining, [http://www.aifb.uni-karlsruhe.de/WBS/aho/pub/hothoa\\_icdm\\_poster03.pdf](http://www.aifb.uni-karlsruhe.de/WBS/aho/pub/hothoa_icdm_poster03.pdf).
- [Kalfoglou2002]: Yannis Kalfoglou, Harith Alani, Kieron O'Hara, Nigel Shadbolt, "Initiating

- Organizational Memories using Ontology Network Analysis.", Knowledge Management and Organizational Memories workshop, 15th European Conference on Artificial Intelligence, Lyon, France, 2002, <http://citeseer.ist.psu.edu/kalfoglou02initiating.html>.
- [Karvounarakis2002]: Greg Karvounarakis, Sofia Alexaki, Vassilis Christophides, Dimitris Plexousakis, Michel Scholl, "RQL: A Declarative Query Language for RDF", The 11th Intl. World Wide Web Conference (WWW2002), <http://www.ai.mit.edu/people/jimmylin/papers/Karvounarakis02.pdf>
  - [Kifer1990]: Michael Kifer, Georg Lausen, "F-Logic: A Higher-Order Language for Reasoning about Objects, Inheritance, and Scheme", 1990, <http://citeseer.ist.psu.edu/kifer90flogic.html>.
  - [Kirchner1998]: Joachim Kirchner, "Data Warehousing, Data Mining- OLAP", Hrsg.: Wolfgang Martin, International Thomson Publishing, Bonn, 1998, S. 151.
  - [Kirschner2003]: Paul A. Kirschner, 2003, "Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making.", <http://www.visualizingargumentation.info>
  - [Klein2002]: Michel Klein, Dieter Fensel, Atanas Kiryakov, Damyan Ognyanov, "Ontology versioning and change detection on the Web", 2002, <http://citeseer.ist.psu.edu/klein02ontology.html>.
  - [Kohonen2001]: T. Kohonen, "Self-Organizing Maps", Springer Series in Information Sciences, Vol. 30, Springer, Berlin, Heidelberg, New York, 2001.
  - [König1998]: Andreas König, "A Survey of Methods for Multivariate Data Projection, Visualisation and Interactive Analysis", Proceedings of the 5th International Conference on Soft Computing and Information/Intelligent Systems IIZUKA'98, S. 55-59, Iizuka, Fukuoka, Japan, October 1998. <http://citeseer.ist.psu.edu/87942.html>.
  - [Kosala2002]: Raymond Kosala, Jan Van den Bussche, Maurice Bruynooghe, Hendrik Blockeel, "Information extraction in structured documents using tree automata induction.", Proceedings of the the 6<sup>th</sup> European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), 2002, <http://citeseer.ist.psu.edu/article/kosala02information.html>.
  - [Kuhlen1999]: Rainer Kuhlen, "Die Konsequenzen von Informationsassistenzen – Was bedeutet informationelle Autonomie oder wie kann Vertrauen in elektronische Dienste in offenen Informationsmärkten gesichert werden?", Suhrkamp, Frankfurt am Main, 1999
  - [Kurz1998]: Andreas Kurz, "Data Warehousing, Data Mining OLAP", Hrsg.: Wolfgang Martin, International Thomson Publishing, Bonn, 1998, S. 252.
  - [Kushmerick2001]: Nicholas Kushmerick, Edward Johnston, Stephen McGuinness, "Information Extraction By Text Classification", 2001, <http://citeseer.ist.psu.edu/kushmerick01information.html>.
  - [Labrou1999]: Yannis Labrou, Tim Finin, "Yahoo! as an Ontology Using Yahoo! Categories to Describe Documents", CIKM, 1999, <http://citeseer.ist.psu.edu/labrou99yahoo.html>.
  - [Langley1992]: Pat Langley, Wayne Iba, Kevin Thompson, "An Analysis of Bayesian Classifiers" National Conference on Artificial Intelligence, NASA Ames Research Center, 1992, <http://citeseer.ist.psu.edu/langley92analysis.html>, S. 223-228.
  - [Lanquillon2001]: C. Lanquillon, "Enhancing Text Classification to Improve Information Filtering", Dissertation an der Uni Magdeburg, 2001, <http://citeseer.ist.psu.edu/lanquillon01enhancing.html>.
  - [Lawrence1999]: Steve Lawrence, C. Lee Giles, Kurt Bollacker, "Digital Libraries and Autonomous Citation Indexing", IEEE Computer, Volume 32, Number 6, 1999,

- <http://citeseer.ist.psu.edu/lawrence99digital.html>, S.67-71.
- [Lem1964]: Stanislaw Lem, "Summa technologiae", Suhrkamp, S.140.
  - [Lepsky1997]: Klaus Lepsky, "Auf dem Weg zur automatischen Inhaltserschließung? Das DFG-Projekt MILOS und seine Ergebnisse.", Mitteilungen der Gesellschaft für Bibliothekswesen und Dokumentation des Landbaues, Heft 53, 1997, S.46-52.
  - [Lepsky1998]: Winfried Gödert, Klaus Lepsky, "Semantische Umfeldsuche im Information Retrieval.", Zeitschrift für Bibliothekswesen und Bibliographie 45, Heft 4, 1998, S.401-423.
  - [Lin2001]: Dekang Lin, Patrick Pantel, "DIRT Discovery of Inference Rules from Text", Knowledge Discovery and Data Mining, 2001, <http://citeseer.ist.psu.edu/lin01dirt.html>.
  - [Lin2002]: Dekang Lin, Patrick Pantel, "Concept Discovery from Text", Department of Computing Science University of Alberta, 2002, <http://citeseer.ist.psu.edu/lin02concept.html>.
  - [Loh2003]: Stanley Loh, José Palazzo M. de Oliveira, Mauricio A. Gameiro, "Knowledge Discovery in Texts for Constructing Decision Support Systems", Applied Intelligence 18, S. 357-366, Kluwer Academic Publishers, 2003, <http://www.kluweronline.com/article.asp?PIPS=5119087>
  - [Lugo2002]: Gustavo A. Giménez-Lugo, Analia Amandi, Jaime Sichman, Daniela Godoy, "Enriching Information Agents' Knowledge by Ontology Comparison: A Case Study", 2002, <http://citeseer.ist.psu.edu/572772.html>.
  - [Lusti1990]: Markus Lusti, "Wissensbasierte Systeme", Hrsg.: Karl Heinz Böhling, Mannheim/Wien/Zürich, BI Wissenschaftsverlag, 1990.
  - [Martin1998]: Wolfgang Martin, "Data Warehousing, Data Mining - OLAP", Hrsg.: Wolfgang Martin, International Thomson Publishing, Bonn, 1998, S. 418.
  - [May2001]: Wolfgang May, "Integration of XML Data in XPathLog", DIWeb, 2001, <http://citeseer.ist.psu.edu/may01integration.html>.
  - [Maynard2003]: D. Maynard , "Information Extraction - why Google doesn't even come close ", Natural Language Processing Group, University of Sheffield, UK, BCS meeting, 25 September 2003, <http://gate.ac.uk/sale/talks/bcs-03-cheltenham.ppt>.
  - [McEntire2000]: Robin McEntire, Peter Karp, Neil Abernethy David Benton, Gregg Helt Matt DeJongh, Robert Kent Anthony Kosky, Suzanna Lewis, Dan Hodnett, Eric Neumann Frank Olken, Dhiraj Pathak Peter, "An Evaluation of Ontology Exchange Languages for Bioinformatics", 2000, <http://citeseer.ist.psu.edu/mcentire00evaluation.html>.
  - [Michie1994]: D. Michie, D.J. Spiegelhalter, C.C. Taylor , "Machine Learning, Neural and Statistical Classification", Ellis Horwood, 1994, <http://citeseer.ist.psu.edu/michie94machine.html>, S.216
  - [Mladenic1998]: , Dunja, 1998, "Turning Yahoo into an Automatic Web-Page Classifier", 13<sup>th</sup> European Conference on Artificial Intelligence Young Researcher Paper, John Wiley & Sons, Ltd. <http://citeseer.ist.psu.edu/mladenic98turning.html>
  - [Mooney2002]: U. Y. Nahm, R. J. Mooney, "Text Mining with Information Extraction", AAAI-2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases, 2002, Department of Computer Sciences, University of Texas, <http://www.cs.utexas.edu/users/ml/papers/discotex-aaaisymp-02.pdf>, S.2.
  - [Mooney2003]: R.J. Mooney, "Intelligent Information Retrieval and Web Search", Kursmaterial University of Texas, 2003, <http://www.cs.utexas.edu/users/mooney/ir-course/slides/TextCategorization.ppt>, S.16.
  - [Müller1998]: Müller, Hausdorf, Schneeberge, "Data Mining", Hrsg.: Gholamreza

- Nakhaeizadeh, Physica-Verlag, 1998.
- [Myka1992]: Andreas Myka, F. Sarre, Ulrich Güntzer, "Rulebased machine learning of hypertext links", 1992, <http://citeseer.ist.psu.edu/myka92rulebased.html>.
  - [Myka1995]: Andreas Myka, Ulrich Güntzer, "Automatic Hypertext Conversion of Paper Document Collections", 1995, <http://citeseer.ist.psu.edu/myka95automatic.html>.
  - [Myka1996a]: Andreas Myka, Ulrich Güntzer, H. Argenton, "Towards Automatic Hypertextual Representation of Linear Texts", PODP, 1996, <http://citeseer.ist.psu.edu/43202.html>.
  - [Myka1996b]: Andreas Myka, Ulrich Güntzer, "Fuzzy Full-Text Searches in OCR Databases", Advances in Digital Libraries, 1996 <http://citeseer.ist.psu.edu/myka96fuzzy.html>.
  - [Myka1996c]: Andreas Myka, Ulrich Güntzer, "Processing Hypertext Link Descriptions", 1996, <http://citeseer.ist.psu.edu/43066.html>, S.4.
  - [Myka1997]: Andreas Myka, Ulrich Güntzer, "On Automatic Similarity Linking in Digital Libraries", Proceedings of DEXA'97 Workshop", S.278-283, 1997, <http://citeseer.ist.psu.edu/myka97automatic.html>
  - [Nakhaeizadeh1998]: Gholamreza Nakhaeizadeh, "Data Mining" Hrsg.: Gholamreza Nakhaeizadeh, Physica-Verlag 1998.
  - [Nejdl2003]: Wolfgang Nejdl, Martin Wolpers, Wolf Siberski, Christoph Schmitz, Mario Schlosser, Ingo Brunkhorst, Alexander Löser, "Super-Peer-Based Routing and Clustering Strategies for RDF-Based Peer-to-Peer Networks", "Proceedings of the 12th International World Wide Web Conference", Budapest, Ungarn, 2003 <http://citeseer.ist.psu.edu/nejdl03superpeerbased.html>
  - [Nilsson1980]: Nils J. Nilsson, "Principles of Artificial Intelligence", Tioga, Palo Alto, 1980.
  - [Noy2000]: N. Fridman Noy, R. W. Ferguson, M. A. Musen, "The knowledge model of Protégé-2000: combining interoperability and flexibility", [http://www-smi.stanford.edu/pubs/SMI\\_Reports/SMI-2000-0830.pdf](http://www-smi.stanford.edu/pubs/SMI_Reports/SMI-2000-0830.pdf).
  - [Pantel2002]: Patrick Pantel, Dekang Lin, "Discovering Word Senses from Text", University of Alberta Department of Computing Science Edmonton, Canada, SIGKDD'02, July 23-26, 2002 ACM 1-58113-567-X/02/0007 <http://citeseer.ist.psu.edu/570332.html>.
  - [Pinto2004]: So a Pinto, Steffen Staab, York Sure, and Christoph Tempich, "OntoEdit empowering SWAP: A case study in supporting Distributed, Loosely-controlled and evolving Engineering of oNTologies (DILIGENT)", [http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2004\\_esws\\_diligent.pdf](http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2004_esws_diligent.pdf).
  - [Quinlan1986]: J.R. Quinlan, "Induction of Decision Trees. Machine Learning", Machine Learning, Kluwer Journals, 1986, <http://www.kluweronline.com/article.asp?PIPS=422606>
  - [Rajman1997]: Martin Rajman, Romaric Besancon, "Text Mining: Natural Language Techniques and Text Mining Applications", Proceedings of the 7th IFIP 2.6 Working Conference on Database Semantics (DS-7), Chapam & Hall IFIP, 1997, S.4 <http://citeseer.ist.psu.edu/rajman97text.html>.
  - [Ray2001]: Erik T. Ray, "Learning XML", 1st Edition, O'Reilly & Associates, 2001.
  - [Reategui1997]: Eliseo B. Reategui, John A. Campbell, Beatriz F. Leao, "A Case-Based Model that Integrates Specific and General Knowledge in Reasoning" Applied Intelligence Vol. 7, 1997, Kluwer Academic Publishers, Netherlands, <http://www.kluweronline.com/article.asp?PIPS=121979>, S.79-90
  - [Rich1983]: Elaine Rich, "Artificial Intelligence", McGraw-Hill Book, New York, 1983, S.14
  - [Riloff1994]: Ellen Riloff, Wendy Lehnert, "Information Extraction as a Basis for High-



- Precision Text Classification", 1994, ACM Transactions on Information Systems, <http://citeseer.ist.psu.edu/riloff94information.html>, S. 296.
- [Rötzer1999]: Florian Rötzer, "Megamaschine Wissen - Vision: Überleben im Netz", Campus Verlag, Frankfurt, New York, 1999, S.176.
  - [Salton1983]: Gerard Salton, Michael J. McGill, "Information Retrieval. Grundlegendes für Informationswissenschaftler." McGraw-Hill, Originaltitel: "Introduction to Modern Information Retrieval.", 1983, S.68-69.
  - [Savoy1991]: Jacques Savoy, Daniel Desbois, "Information retrieval in hypertext systems: an approach using Bayesian networks", Electronic Publishing/Origination, Dissemination, and Design, 1991, <http://citeseer.ist.psu.edu/savoy91information.html>.
  - [Shannon1980]: Claude F. Shannon, Warren Weaver, "The Mathematical Theory of Communication", University of Illinois Press, 1980.
  - [Sherif2002]: Yacoub Sherif, "Bootstrapping Semantic Web Languages using a UML Meta- Modeling Approach," Information Infrastructure Laboratory HP Laboratories Palo Alto, 2002, <http://www.hpl.hp.com/techreports/2002/HPL-2002-200.html>.
  - [Stenmark2001]: Dick Stenmark, "The Relationship between Information and Knowledge", <http://citeseer.ist.psu.edu/stenmark01relationship.html>.
  - [Stock2000a]: Wolfgang G. Stock, "Informationswirtschaft: Management externen Wissens", München/Wien, Oldenbourg, 2000.
  - [Stock2000b]: Wolfgang G. Stock, "Textwortmethode", PASSWORD 07 und 08/2000, [http://www.phil-fak.uni-duesseldorf.de/infowiss/admin/public\\_dateien/files/1/1078740450password\\_7.pdf](http://www.phil-fak.uni-duesseldorf.de/infowiss/admin/public_dateien/files/1/1078740450password_7.pdf).
  - [Su2002] Vgl. Xiaomeng Su, Lars Ilebrikke, "A Comparative Study of Ontology Languages and Tools", Hrsg.: A. Banks Pidduck, Springer-Verlag Berlin Heidelberg, 2002, <http://citeseer.ist.psu.edu/556209.html>.
  - [Sure1999]: "On-To-Knowledge Methodology - Employed and Evaluated Version", York Sure, Rudi Studer, University of Karlsruhe On-To-Knowledge EU IST-1999-10132 Project Deliverable D16 (WP5), <http://citeseer.ist.psu.edu/sure99toknowledge.html>, S.48-49.
  - [Sure2003]: York Sure, J. Angele, S. Staab, "OntoEdit: Multifaceted inferencing for ontology engineering", Journal on Data Semantics, LNCS 2800, 2003, S.128-152, <http://www.aifb.uni-karlsruhe.de/WBS/sst/Research/Publications/ontoedit-data-semantics.pdf>,
  - [Sure2004]: Marc Ehrig, York Sure, "Ontology Mapping - An Integrated Approach", Institute AIFB, University of Karlsruhe, 2004, [http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2004\\_esws\\_mapping.pdf](http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2004_esws_mapping.pdf).
  - [Takeda1995]: Hideaki Takeda, Kenji Iino, Toyooki Nishida, "Agent organization and communication with multiple ontologies.", International Journal of Cooperative Information Systems, 4, 1995, S.321-337, <http://ai-www.aist-nara.ac.jp/papers/takeda/ps/ijicis.ps.gz>.
  - [Theobald2002a]: Anja Theobald, Gerhard Weikum, "The XXL Search Engine: Ranked Retrieval of XML Data Using Indexes and Ontologies", ACM SIGMOD 2002, <http://ranger.uta.edu/~alp/ix/readings/theobaldXXL-sigmod02.pdf>.
  - [Theobald2002b]: Anja Theobald, Gerhard Weikum, "The index-based XXL search engine for querying XML data with relevance ranking", EDBT, 2002, <http://ranger.uta.edu/~alp/ix/readings/theobaldXXL-EDBT02.pdf>.
  - [Theobald2003]: Anja Theobald, "An Ontology for Domain-oriented Semantic Similarity Search on XML Data", 10th Conference on Database Systems for Business, Technology

- and Web (BTW), Leipzig, Germany, 2003,  
[http://www.mpi-sb.mpg.de/units/ag5/publications/theobald\\_btw2003.pdf](http://www.mpi-sb.mpg.de/units/ag5/publications/theobald_btw2003.pdf).
- [Ultsch2003]: Alfred Ultsch, "U\*-Matrix: a Tool to visualize Clusters in high dimensional Data", 2003, [http://www.mathematik.uni-marburg.de/forschung/publikationen/paper\\_info/bfi36.pdf](http://www.mathematik.uni-marburg.de/forschung/publikationen/paper_info/bfi36.pdf).
  - [Vega1998]: Julio César Arpírez Vega, Asunción Gómez-Pérez, Adolfo Lozano Tello, Helena Sofia Andrade N. P. Pinto, "(ONTO)2Agent: An ontology-based WWW broker to select ontologies", 13th European Conference on Artificial Intelligence ECAI'98, Brighton, England, 1998, <http://citeseer.ist.psu.edu/34215.html>, S.1, S.7.
  - [Vizine-Goetz2001]: Diane Vizine-Goetz, "Exploiting LCSH, LCC, and DDC to Retrieve Networked Resources", Library of Congress, 2001, [http://www.loc.gov/catdir/bibcontrol/vizinegoetz\\_paper.html](http://www.loc.gov/catdir/bibcontrol/vizinegoetz_paper.html)
  - [Wedekind1998]: Hartmut Wedekind, "Datenorganisation", 3. Aufl., de Gruyter, 1989.
  - [Wilde2001]: Klaus Wilde, "Data Warehouse, OLAP und Data Mining im Marketing", Handbuch Data Mining im Marketing, Wiesbaden: Vieweg, Gabler 2001, S.14
  - [Winston1987]: P. Winston, "Künstliche Intelligenz", Addison-Wesley, 1987, S.21.
  - [Wittgenstein1949]: Wittgenstein, Ludwig, "Philosophische Untersuchungen", zitiert aus: Hans Joachim Störig, "Kleine Weltgeschichte der Philosophie", Fischer Taschenbuch Verlag, Frankfurt am Main, 1993, S. 658.
  - [Wong1986]: S.K.M. Wong, W. Ziarko, "A machine learning approach to information retrieval" Department of Computer Science, University of Regina, Regina. Saskatchewan. Canada, 1986 ACM Conference on Research and Development in Information Retrieval, <http://portal.acm.org/citation.cfm?id=253217>.
  - [Yang1997]: Yiming Yang, "An Evaluation of Statistical Approaches to Text Categorization", Information Retrieval, Kluwer Academic Publishers, 1997, <http://citeseer.ist.psu.edu/yang97evaluation.html>.
  - [Zavrel2000]: Jakub Zavrel, Peter Berck, Willem Lavrijsen, "Information Extraction by Text Classification: Corpus Mining for Features", 2000, <http://citeseer.ist.psu.edu/zavrel00information.html>.