To appear in: Proceedings of Lernen-Wissen-Adaption 2007 (http://lwa07.informatik.uni-halle.de/)

Information Retrieval on the Semantic Web - Does it exist?

Peter Scheir[1,2], Viktoria Pammer[1,2], Stefanie N. Lindstaedt[2]
[1]Graz University of Technology, Austria
peter.scheir@tugraz.at
[2]Know-Center Graz, Austria
vpammer,slind@know-center.at

Abstract

Plenty of contemporary attempts to search exist that are associated with the area of Semantic Web. But which of them qualify as information retrieval for the Semantic Web? Do such approaches exist?

To answer these questions we take a look at the nature of the Semantic Web and Semantic Desktop and at definitions for information and data retrieval. We survey current approaches referred to by their authors as information retrieval for the Semantic Web or that use Semantic Web technology for search.

1 Introduction

Although Semantic Web research is still a young discipline a fair amount of research exists and progress into the direction of the vision of [Berners-Lee *et al.*, 2001] is made. A major category in this area of research is search and retrieval of information in this new type of web. Within this paper we are going to take a closer look at current approaches to search in the Semantic Web and on the Semantic Desktop[1].

The paper is structured as follows: in section 2 we aim at identifying characteristics of an information retrieval system for the Semantic Web. In section 3 we survey present approaches to search in the Semantic Web, we focus on systems and models for searching documents and ontological concepts (section 3.1) and systems for searching ontologies (section 3.2). In section 4 we try to classify the surveyed systems according to common properties. We conclude with section 5.

2 Information Retrieval for the Semantic Web

Various approaches to search associated with the area of Semantic Web exist. Diverse techniques are employed addressing a variety of problems. However, the notion of information retrieval in the context of Semantic Web seems to be rather diffuse.

We propose the following characteristics of an information retrieval system for the Semantic Web:

- **Criterion 1:** The system operates on the Semantic Web
- **Criterion 2:** The system is based on technology for the Semantic Web
- **Criterion 3:** The system performs information retrieval and not data retrieval

[1]Methods used by the systems presented here also appear in other field of research, such as databases, XML retrieval, geographic retrieval, etc. Within our work we focus on finding a working definition for information retrieval on the Semantic Web and only take a closer look at systems that meet this definition.

2.1 Ad criterion 1: Semantic Web vs. Semantic Desktop

The Semantic Web [Berners-Lee *et al.*, 2001] is not indented to be a new web but an extension to the current one. In the current form of the web information gathering is a task performed mainly by humans using a web browser. The Semantic Web shall provide an infrastructure so that this task can be performed by computer programs. To allow for processing the information on the web by machines, information is annotated with machine-interpretable data. The Semantic Desktop [Sauerman *et al.*, 2005] paradigm aims at applying technologies developed for the Semantic Web to desktop computing to finally provide for a closer integration between (semantic) web and (semantic) desktop.

In order for a system to qualify as operating on the Semantic Web we require it to search resources that are publicly available on the web and to potentially search the whole web.

At present most search approaches based on semantic technologies are for the Semantic Desktop. The incubation of search approaches in a desktop environment is a known phenomenon in information retrieval and current web search engines are based on techniques originally development for non-web environments.

2.2 Ad criterion 2: Ontology-driven information retrieval vs. information retrieval for the Semantic Web

To our understanding, a central element of information retrieval approaches for the Semantic Web is that they use technologies developed for the Semantic Web. Examples of such technologies are the standards RDF , RDF Schema and OWL.

Ontology-driven information retrieval approaches utilize ontologies for retrieval purposes, e.g. to increase retrieval performance by using the information modeled in ontologies. Many information retrieval systems for the Semantic Web use ontologies for retrieval purposes as well. However ontology-driven information retrieval does not necessarily target the Semantic Web.

2.3 Ad criterion 3: Data Retrieval vs. Information Retrieval

For pointing out the difference between Data Retrieval and Information Retrieval we refer to [Baeza-Yates and Ribeiro-Neto, 1999]:

A data retrieval language aims at retrieving all objects which satisfy clearly defined conditions ... a single erroneous object among a thousand retrieved objects means total failure. For an information retrieval system, however, the retrieved objects might be inaccurate and small errors are likely to go unnoticed.

Some of the current approaches to search in the Semantic Web are data retrieval approaches in the sense of the definition above (c.f. [Castells *et al.*, 2007] for a discussion). The most prominent ones are the SQL like query languages for the Semantic Web as SPARQL Query Language for RDF (SPARQL)[2].

3 Survey

This section gives an overview of existing approaches[3] to search in the context of the Semantic Web or the Semantic Desktop. We here differentiate between two search approaches that retrieve information in the Semantic Web or on the Semantic Desktop:

1. Systems or models that *search for information in the form of documents or ontological elements* (see section 3.1)
2. Approaches that *search for ontologies* (being a special type of information on the Semantic Web) (see section 3.2)

For listing approaches in this section one of the following criteria had to be fulfilled:

1. Technologies for the Semantic Web are used for retrieval purposes (e.g. RDF or OWL)
2. The approach is referred to as *for the Semantic Web* by its authors

3.1 Search for documents or ontological elements

In this section we survey systems and models for searching resources on the Semantic Web and on the Semantic Desktop.

SHOE: *SHOE Knowledge Annotator* allows for embedding semantic markup into HTML pages. This markup can be searched using *SHOE Search* [Heflin and Hendler, 2000] which provides a graphical user interface for building complex queries based on an ontology. *Expos* a web-crawler searches for web-pages with SHOE markup, extracts the markup and stores it in a local knowledge base.

SEAL: SEAL [Stojanovic *et al.*, 2001] is a framework developing semantic portals. Part of this framework is ranking of search results. While all results to a search in a semantic portal are equally relevant, SEAL ranks those results with least inference steps needed from the original knowledge base highest.

OWLIR: OWLIR [Shah *et al.*, 2002] indexes RDF triples together with document content. OWLIR treats distinct RDF triples as indexing terms. RDF triples are generated by natural language processing techniques based on textual content. Search can be performed based on words and RDF triples with wildcards. Shah et al. [2002] report an increase of Average Precision using an approach taking semantic information into account opposed to a text-only approach.

[2] `http://www.w3.org/TR/rdf-sparql-query/`

[3] We, the authors of this paper are thankful for every information on existing approaches to search on the Semantic Web, that we have overlooked and thus are not listed in this section.

SCORE: The Semantic Content Organization and Retrieval Engine (SCORE) [Sheth *et al.*, 2002] uses classification and information-extraction techniques to extract metadata from textual sources. This metadata is later used for semantic search. A user can issue a query by specifying the category of document and one or more attribute values of the metadata to the document.

QuizRDF: QuizRDF [Davies *et al.*, 2002] combines keyword-based search with search and navigation through RDF(S) based annotations. Indexing in QuizRDF is based on *content descriptors*, which can be terms from the documents and literals from RDF statements. A query is formulated in using terms, in addition the type of the resource that should be returned can be specified. Search results can be filtered by property values of resources.

PISTA: Aleman-Meza et al. [2003] present an approach to *Semantic Association Ranking* in the Semantic Web. Starting from two entities in a RDF graph they aim at finding semantic associations between them and rank these semantic associations based on their importance. Semantic associations are based on relations between the two entities in the RDF graph and on the types of the relations.

TAP: TAP [Guha *et al.*, 2003] aims at enhancing search results from the WWW with data from the Semantic Web. It performs a graph based search on a RDF graph from the web. It starts at one or more *anchor nodes* in the RDF graph, which have to be mapped to query terms. A breath first search is performed in the RDF graph, collecting a predefined amount of triples. Optionally only links of a certain type are followed in traversing the RDF graph.

[Stojanovic *et al.*, 2003]: Stojanovic et al. [2003] - as in [2001] - aim at ranking search results of a semantic portal. In [Stojanovic *et al.*, 2003] they use the *specifity* of the instance of a relation, which is higher the less often the instances of the concepts in the relation are present in other instances of relations. In addition the inference process of the statements is taken into account for ranking results. This time the deduction of rules is use for ranking search results.

BioPatentMiner: Bamba and Mukherjea [2004] present an approach to *ranking Semantic Web query results*. The set of triples returned by a RDQL query is ranked based on various factors. A central aspect for weighting nodes in the results graphs is an adapted version of Kleinberg's HITS algorithm. Instead of hub and authority scores *subjectivity* and *objectivity* scores are calculated and the type of links between resources is taken into account as well. In addition the position in the class hierarchy or the super class of a resource is taken into account for weighing nodes. Edges are weighted by inverse property frequency of a property in the results graph. For every result graph all node and edge weights are combined to produce a relevance value.

KIM: KIM [Kiryakov *et al.*, 2004] relies on information extraction and relates words in documents with concepts from an ontology. Before indexing, documents are enriched with identifiers for the ontological concepts the words in the document represent. These identifers are directly inserted into the indexed text. For homonyms the same identifier is used. Queries are formed using concepts and relations from the ontology.

InWiss: [Priebe *et al.*, 2004] present a *Search Engine for RDF Metadata* implemented in the InWiss knowledge portal. Before search is performed the RDF graph is extended by following transitive properties and directly relating the resources found this way with the resources the traversal originated from. The same action is performed for the query. In an additional step a set theoretic approach to ranking search results is performed, where the number of matching properties of a query with a resource is divided by the total number of properties of the query.

[Rocha *et al.*, 2004]: Rocha et al. [2004] present a *Hybrid Approach for Searching in the Semantic Web* that combines full-text search with spreading activation search in an ontology. Search starts with a keyword based query. Results to the fulltext search are instances from the ontology. Those instances are used to initiate a spreading activation search in the ontology to find additional instances.

[Bangyong *et al.*, 2005]: Bangyong et al. [2005] present a preliminary approach to association search in the Semantic Web. They propose to generate a Bayesian network from an ontology and use it to find instances not found by the initial query. Initial search is done by traditional web search technology which returns a set of instances. Based on the results of this search, associated instances are searched using the Bayesian network. The transformation of the ontology into a Bayesian network is not explained.

[Song *et al.*, 2005]: Song et al. [2005] present a *Ontology-Based Information Retrieval Model for the Semantic Web*. They suggest to use *semantic index terms* for indexing documents, which means to use the same index terms for synonyms and different indexing terms for homonyms. Also, the query should be represented using semantic index terms. How the indexing process should take place and how a query should be formulated is not addressed.

[Zhang *et al.*, 2005]: Zhang et al. [2005] present an *enhanced model for searching in semantic portals*. They integrate text based search with a fuzzy version of the description logic $\mathcal{ALC}$. A result set to a query is represented as a class in the knowledge base. The retrieval status values (RSV) of the documents retrieved by text-based search for a query are used as the fuzziness degrees for the instances of this class. Zhang et al. [2005] provide a Semantic Web service for search which is queried programmatically.

CORESE: CORESE [Corby *et al.*, 2006] is an ontology-based search engine which operates on conceptual graphs internally. COROSE is queried using one or a combination of triples. The query language is similar to SPARQL, SeRQL or RDQL but allows for approximate search. Approximate search is based on semantic distance of two classes in a common hierarchy and the `rdfs:seeAlso` property. The relevance of a result is measured by the similarity to the query.

OntoSearch: Jiang and Tan [2006] call OntoSearch a *Full-Text Search Engine for the Semantic Web*. Search starts with a term-based query which yields to a set of documents, from these documents semantic metadata is extracted and used for a spreading activation search in an ontology. The extended set of concepts is used to rank the search results of the term based search. Ranking is done using the cosine measure with concepts from the ontology being introduced as additional dimensions in the vector space.

Beagle++: Beagle++ [Chirita *et al.*, 2006] extends the opensource search engine Beagle[4] by indexing functionality for RDF triples. For every document the predicate and object of the RDF triple the document is the subject in are indexed. Additionally so called *predicate paths* are indexed. These are paths of those predicates in the RDF graph that are traversed starting from the document node. Beagle++ is queried using terms.

[Choudhury and Phon-Amnuaisuk, 2006]: Choudhury and Phon-Amnuaisuk [2006] present a search system similar to the one presented by Rocha et al. [2004]. They implement a subset of the weighting functionality of the system presented in [Rocha *et al.*, 2004] and test their system on a smaller data set.

MESH: Castells et al. [2007] combine SPARQL based search with full text search. For ranking results of a SPARQL query they weight annotations of documents with concepts from an ontology using an tf*idf-like measure. Then they combine the results of a full-text-search with the ranked list obtained via the SPARQL query using the CombSUM strategy. For performing the full-text search they extract certain parts of the SPARQL query and use them as query terms.

3.2 Search for ontologies

Ontology search engines are systems that retrieve information in the form of ontologies. The search results are either entire ontologies, ontology modules or ontology elements like statements or single concepts.

Various applications of searching for (parts of) ontologies are possible, such as agent support, ontology reuse or pure information retrieval.

OntoKhoj: OntoKhoj [Patel *et al.*, 2003] returns a ranked list of ontologies for a given WordNet-sense. OntoKhoj searches according to WordNet synonyms and hypernyms of the input. Appropriate ontologies are ranked according to their interconnectivity, i.e. ontologies that are referred to more often are ranked higher. Relationships considered by OntoKhoj are e.g. an *rdfs:subClass* relationship to an element of another ontology. OntoKhoj differently weights different relations and also considers chained relationships.

Swoogle: Swoogle[5] [Ding *et al.*, 2004] is again a search engine for ontologies (Semantic Web documents in Swoogle terminology). Ding et al. implemented various crawlers, e.g. one using Google, another monitoring given websites and a third analyzing retrieved ontologies and spreading out using outgoing hyperlinks. Retrieved ontologies are indexed by keywords and metadata like ontology language and ranked. An ontology's rank is defined by where the ontology lies on a continuous line between "database" and "schema".

[Biddulph, 2004]: In [Biddulph, 2004], the author implemented a Semantic Web crawler which was designed to serve as an aggregation service for external software (agents). The main difference to the engines presented above is that Biddulph merges all retrieved statements into one model stored on and queried from a Joseki[6] server. Biddulph assumes all collected models as partial aspects of the same world and makes inferences on the merged model.

OntoSelect: OntoSelect[7] [Buitelaar *et al.*, 2004] is an ontology library that monitors the web for changes in ontologies and crawls for new ontologies. Available ontologies can be browsed by ontology, keywords (labels), classes and properties. The library can also be searched for ontologies containing certain keywords. The search is either performed directly on the given keywords or with whole sites, i.e. given a Wikipedia topic or any other website, OntoSelect will select one ore more ontologies that match the given page. For this, OntoSelect first extracts relevant keywords from the page and then searches for ontologies matching these keywords.

[4] `http://www.beagle-project.org/`

[5] `http://swoogle.umbc.edu/`

[6] `http://www.joseki.org/`

[7] `http://olp.dfki.de/ontoselect`

OntoSearch2: OntoSearch2[8, 9] [Pan *et al.*, 2006] provides the functionality to query ontologies in a repository. OntoSearch2 relies on an external agent (human or machine) to add ontologies to the repository. It then preprocesses these ontologies for querying. OntoSearch2 supports SPARQL queries and queries all ontologies in the repository. Return results are exact matches of the SPARQL queries and can be either whole ontologies or parts of them. OntoSearch2 supports DL-Lite, which is a sublanguage of OWL DL. Incoming ontologies are preprocessed by translation into DL-Lite.

Watson: Watson[10] [dAquin *et al.*, 2007] is the latest of the here presented ontology search engines. Watson crawls the web for semantic documents (OWL, RDF(S), DAML+OIL). The ontologies are analysed in a first step to detect new locations of ontologies. Then a number of metadata are calculated, e.g. expressivity and level of axiomatization. Watson also checks for "semantic" duplicates, e.g. the same ontology replicated at different locations, the same ontology in another language etc. Watson retrieves single ontology elements like classes or instances and corresponding ontologies.

4 Classification of existing approaches

For our work, we adopt the two top level categories introduced by [Esmaili and Abolhassani, 2006], *Semantic Search Engines* and *Ontology Search Engines*, and present two distinct classifications, one for *approaches to search for documents or ontological elements* (section 4.1) and one for *approaches to search for ontologies* (section 4.2).

4.1 Classification of approaches to search for documents or ontological elements

[Tomassen, 2006] classifies current approaches to ontology-based information retrieval into *Knowledge Base* and *vector space model* driven approaches. More generally speaking one can classify current approaches to Semantic Web information retrieval into (1) approaches that operate on top of knowledge bases and (2) approaches that operate on top of information retrieval systems. Approaches on top of knowledge bases model documents as elements in the knowledge base, for example as instances of a special class in the knowledge base representing documents. Queries are formulated using special query languages and reasoning mechanisms are employed in the knowledge based to retrieve relevant documents. Approaches on top of information retrieval systems index document metadata originating from the ontology together with document content.

The following abbreviations are used for column heads of Table 1:

W ... (Semantic) Web Search is performed in a web environment. Related to criterion 1 from section 2.

D ... (Semantic) Desktop Search is performed in a desktop. Related to criterion 1 from section 2. environment

CBQ ... concept based query The query is formed of concepts or entities stemming from a knowledge representation (e.g. as in SPARQL).

TBQ ... term based query The query is formed of words (terms) as in common search engines.

KBS ... knowledge-based system A knowledge based system is used, a knowledge representation is searched. Based on [Tomassen, 2006].

IRS ... information retrieval system An information retrieval system is used, a document representation is searched. Based on [Tomassen, 2006].

DR ... data retrieval All results are relevant and equally relevant. Related to criterion 3 from section 2.

IR ... information retrieval A ranked list of results is returned. Related to criterion 3 from section 2.

KR ... knowledge representation used Type (format) of the knowledge representation used.

4.2 Classification of approaches to search for ontologies

We think that ontology search engines can effectively be classified by the kind of input they accept. Further distinction is possible by which kind of metadata about the ontologies is generated and available for search, and whether search results are entire ontologies, modules or ontology elements like statements or concepts.

We summarise *Ontology Search Engines* in Table 2. The following characteristics are used:

Input Type of input: free text (F), keyword (K), formal element e.g. WordNet sense, ontology concept/property (FE), formal structure e.g. SPARQL (FS).

Crawler included: Yes/No

Storage of ontologies: Yes/No

Index Yes / No

KR Supported knowledge representations: OWL, DAML+OIL, RDF, RDFS

API Does an API (or any access-point for software like e.g. a webservice) to the search engine exist?

Online available: Yes / No

5 Conclusions

Under critical examination none of the surveyed approaches to *search for documents and ontological concepts* was able to fulfill all of the three characterizations introduced in section 2. Most of the presented systems operate in a desktop environment and some of the presented approaches are data retrieval systems that do not return a relevance value for a search result. Not all of the surveyed systems actually use technology for the Semantic Web. As their authors referred to these systems as being for the Semantic Web, they probably had a different conception about information retrieval in the Semantic Web. Another explanation could be that they have used ontology-based information retrieval and information retrieval for the Semantic Web synonymously.

We remark that the diversity of *ontology search engines* is not overly impressive, but on the other hand ontology search engines mostly adhere to all three criteria for information retrieval on the Semantic Web as given in section 2.

Finally, all of the presented system operate on the lower levels of the Semantic Web, making use of knowledge representation standards as RDF and OWL. Work in progress topics in Semantic Web research as *proof* and *trust* are not addressed in the presented systems.

Acknowledgments

We thank the anonymous reviewers of our submission for their constructive feedback. This work has been partially funded under grant 027023 in the IST work programme of

[8] `http://www.ontosearch.org/`

[9] OntoSearch2 and its predecessor OntoSearch mentioned in this section are not related to OntoSearch in section 3.1.

[10] `http://watson.kmi.open.ac.uk/Overview.html`

the European Community. The Know-Center is funded by the Austrian Competence Center program Kplus under the auspices of the Austrian Ministry of Transport, Innovation and Technology (www.ffg.at/index.php?cid=95) and by the State of Styria.

References

[Aleman-Meza *et al.*, 2003] B. Aleman-Meza, Ch. Halaschek-Wiener, I. B. Arpinar, and A. P. Sheth. Context-aware semantic association ranking. In *Proc. of SWDB'03, 1st Int. Workshop on Semantic Web and Databases*, 2003.

[Baeza-Yates and Ribeiro-Neto, 1999] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., 1999.

[Bamba and Mukherjea, 2004] B. Bamba and S. Mukherjea. Utilizing resource importance for ranking semantic web query results. In *Semantic Web and Databases, Second Int. Workshop, SWDB 2004*, 2004.

[Bangyong *et al.*, 2005] L. Bangyong, T. Jie, and . Juanzi. Association search in semantic web: search + inference. In *WWW '05: Special interest tracks and posters of the 14th int. conf. on World Wide Web*, 2005.

[Berners-Lee *et al.*, 2001] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 2001.

[Biddulph, 2004] M. Biddulph. Semantic web crawling. In *XML Europe 2004*, 2004.

[Buitelaar *et al.*, 2004] P. Buitelaar, Th. Eigner, and Th. Declerck. Ontoselect: A dynamic ontology library with support for ontology selection. In *In: Proc. of the Demo Session at the Int. Semantic Web Conf.*, 2004.

[Castells *et al.*, 2007] P. Castells, M. Fernndez, and D. Vallet. An adaptation of the vector-space model for ontology-based information retrieval. *IEEE Trans. Knowl. Data Eng.*, 19:261–272, 2007.

[Chirita *et al.*, 2006] P. A. Chirita, S. Costache, W. Nejdl, and R. Paiu. Beagle++: Semantically enhanced searching and ranking on the desktop. In *The Semantic Web: Research and Applications*, 2006.

[Choudhury and Phon-Amnuaisuk, 2006] S. Choudhury and S. Phon-Amnuaisuk. Semantic retrieval with spreading activation. In *Proc. of the MMU Int. Symposium on Information and Communications Technologies M2USIC 2006*, 2006.

[Corby *et al.*, 2006] O. Corby, R. Dieng-Kuntz, C. Faron-Zucker, and F. Gandon. Searching the semantic web: Approximate query processing based on ontologies. *IEEE Intelligent Systems*, 21(1):20–27, 2006.

[dAquin *et al.*, 2007] M. dAquin, M. Sabou, M. Dzbor, C. Baldassarre, L. Gridinoc, S. Angeletou, and E. Motta. Watson: A gateway for the semantic web. In *European Semantic Web Conf., ESWC 2007 (poster)*, 2007.

[Davies *et al.*, 2002] J. Davies, U. Krohn, and R. Weeks. Quizrdf: search technology for the semantic web. In *WWW2002 workshop on RDF & Semantic Web Applications, 11th Int. WWW Conf.*, 2002.

[Ding *et al.*, 2004] L. Ding, T. Finin, A. Joshi, R. Pan, R. Scott Cost, Y. Peng, P. Reddivari, V. C. Doshi, and J. Sachs. Swoogle: A Search and Metadata Engine for the Semantic Web. In *Proc. of the 13thACM Conf. on Information and Knowledge Management*, 2004.

[Esmaili and Abolhassani, 2006] K. S. Esmaili and H. Abolhassani. A categorization scheme for semantic web search engines. In *4th ACS/IEEE Int. Conf. on Computer Systems and Applications (AICCSA-06)*, 2006.

[Guha *et al.*, 2003] R. Guha, R. McCool, and E. Miller. Semantic search. In *WWW '03: Proc. of the 12th int. conf. on World Wide Web*, 2003.

[Heflin and Hendler, 2000] J. Heflin and J. Hendler. Searching the web with SHOE. In *Artificial Intelligence for Web Search. Papers from the AAAI Workshop. WS-00-01.*, 2000.

[Jiang and Tan, 2006] X. Jiang and A.-H. Tan. Ontosearch: A full-text search engine for the semantic web. In *Proc. of the 21st National Conf. on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conf.*, 2006.

[Kiryakov *et al.*, 2004] A. Kiryakov, B. Popov, I. Terziev, D. Manov, and D. Ognyanoff. Semantic annotation, indexing, and retrieval. *Journal of Web Semantics: Science, Services and Agents on the World Wide We*, 2:49–79, 2004.

[Pan *et al.*, 2006] J. Z. Pan, E. Thomas, and D. Sleeman. Ontosearch2: Searching and querying web ontologies. In *Proc. of the IADIS Int. Conf. WWW/Internet 2006*, 2006.

[Patel *et al.*, 2003] C. Patel, K. Supekar, Y. Lee, and E. K. Park. Ontokhoj: a semantic web portal for ontology searching, ranking and classification. In *WIDM '03: Proc. of the 5th ACM int. workshop on Web information and data management*, 2003.

[Priebe *et al.*, 2004] T. Priebe, Ch. Schlager, and G. Pernul. A search engine for rdf metadata. In *DEXA '04: Proc. of the Database and Expert Systems Applications, 15th Int. Workshop on (DEXA'04)*, 2004.

[Rocha *et al.*, 2004] C. Rocha, D. Schwabe, and M. Poggi de Aragão. A hybrid approach for searching in the semantic web. In *Proc. of the 13th int. conf. on World Wide Web (WWW)*, 2004.

[Sauerman *et al.*, 2005] L. Sauerman, A. Bernardi, and A. Dengel. Overview and outlook on the semantic desktop. In *Proc. of the 1st Workshop on The Semantic Desktop at the ISWC 2005 Conf.*, 2005.

[Shah *et al.*, 2002] U. Shah, T. Finin, and A. Joshi. Information retrieval on the semantic web. In *CIKM '02: Proc. of the 11th int. conf. on Information and knowledge management*, 2002.

[Sheth *et al.*, 2002] A. Sheth, C. Bertram, D. Avant, B. Hammond, K. Kochut, and Y. Warke. Managing semantic content for the web. *IEEE Internet Computing*, 6:80–87, 2002.

[Song *et al.*, 2005] J. Song, W. Zhang, W. Xiao, G. Li, and Z. Xu. Ontology-based information retrieval model for the semantic web. In *2005 IEEE Int. Conf. on e-Technology, e-Commerce, and e-Services (EEE)*, 2005.

[Stojanovic *et al.*, 2001] N. Stojanovic, A. Maedche, S. Staab, R. Studer, and Y. Sure. Seal: a framework for developing semantic portals. In *Proc. of the 1st Int. Conf. on Knowledge Capture (K-CAP)*, 2001.

[Stojanovic *et al.*, 2003] N. Stojanovic, R. Studer, and L. Stojanovic. An approach for the ranking of query results in the semantic web. In *Int. Semantic Web Conf.*, 2003.

[Tomassen, 2006] S. L. Tomassen. Research on ontology-driven information retrieval. In *OTM Workshops (2)*, 2006.

[Zhang *et al.*, 2005] L. Zhang, Y. Yu, J. Zhou, Ch. Lin, and Y. Yang. An enhanced model for searching in semantic portals. In *WWW '05: Proc. of the 14th int. conf. on World Wide Web*, 2005.

	W	D	CBQ	TBQ	KBS	IRS	DR	IR	KR
[Heflin and Hendler, 2000]	X		X		X		X		SHOE
[Stojanovic *et al.*, 2001]	(X)[a]		X		X			X	F-Logic
[Davies *et al.*, 2002]		X	(X)[b]	X	X	X		X	RDF & RDFS
[Shah *et al.*, 2002]		X	X[c]	X	[d]	X		X	DAML+OIL
[Sheth *et al.*, 2002]		X	X	[e]	X		X		unknown
[Aleman-Meza *et al.*, 2003]	[f]		X		X			X	RDF
[Guha *et al.*, 2003]	X		(X)[g]		X		X		RDF
[Stojanovic *et al.*, 2003]	(X)[h]		X		X			X	F-Logic
[Bamba and Mukherjea, 2004]		X	X		X			X	RDF & RDFS
[Kiryakov *et al.*, 2004]		X	X	X	X	X	X[i]	X	RDFS
[Priebe *et al.*, 2004]	(X)[j]		X		X			X	RDF
[Rocha *et al.*, 2004]	(X)[k]			X	X	X		X	unknown[l]
[Bangyong *et al.*, 2005]	[m]			X	X	X		X	unknown
[Song *et al.*, 2005]	[n]		X			X		X	none
[Zhang *et al.*, 2005]	(X)[o]		X	(X)[p]	X	X		X	fuzzy DL $\mathcal{ALC}$
[Corby *et al.*, 2006]		X	X		X			X	RDF, RDFS, OWL Lite
[Choudhury and Phon-Amnuaisuk, 2006]		X		X	X	X		X	unknown
[Jiang and Tan, 2006]		X		X	X	X		X	taxonomy[q]
[Chirita *et al.*, 2006]		X		X		X		X	RDF
[Castells *et al.*, 2007]		X	X		X	X		X	OWL

Table 1: Characteristics of approaches to search for documents or ontological elements

[a] search on semantic portal
[b] filtering by type and search in property values
[c] RDF-triples with wildcards
[d] inference in knowledge base to expand set of triples indexed with text
[e] not designed for full-text search, but could be done
[f] search in RDF graphs
[g] nodes in RDF graph
[h] search on semantic portal
[i] entities can be retrieved from knowledge base
[j] search on knowledge portal
[k] search on website
[l] „can be easily mapped to RDF"
[m] model only, no system
[n] model only, no system
[o] search on semantic portal
[p] term based queries are modeled as instances to concepts in knowledge base
[q] ACM Computing Classification System

Search Engine	**Input**	**Crawler**	**Storage**	**Index**	**KR**	**API**	**Online**
OntoKhoj	FE	Yes	-	Yes	RDF(S), DAML+OIL, OWL	-	No
[Biddulph, 2004]	FE	Yes	Yes	No	RDF(S), OWL	Yes	No
Swoogle	K	Yes	No	Yes	RDF(S), DAML+OIL, OWL	Yes	Yes
OntoSelect	F / K	Yes	-	-	RDF(S), DAML, OWL	No	Yes
OntoSearch2	K[a] / FS	No	Yes	No	OWL	Yes	Yes
Watson	K	Yes	Yes	Yes	RDF(S), DAML+OIL, OWL	Yes	Yes

Table 2: Overview of ontology search engines

[a] It seems that results are only available for queries containing a single keyword