First Results of a Semantic Web Technologies Evaluation

York Sure [1] and Victor Iosif [2,3]

[1] Institute AIFB, University of Karlsruhe, D-76128 Karlsruhe, Germany
Tel.: +49 721 608 6592
sure@aifb.uni-karlsruhe.de

[2] EnerSearch AB, Malmö, Sweden
[3] Lund University Business School, Sweden
victor@enersearch.se

Abstract

How do we measure ontology based search tools? How can we know if they perform better than the traditional methods for finding information on the Internet? We believe that one of the best way to answer this is to carry out a case study including a carefully designed field experiment where the tools are tested. In this paper we discuss the design considerations and set-up of filed experiments with Semantic Web tools, and we also give some result from our own field studies that was carried out within the EU IST-1999-10132 On-To-Knowledge project.

1. Introduction

Numerous tools and applications of semantic web technologies are already available and the number is growing fast (among others [DAML], [OntoWeb] and [OTK]). However, the Semantic Web technologies are still in pre-mature phase, and numerous important questions still need answers. We will address in this paper aspects of the following questions, that need at least to be discussed before industry is willing to adopt semantic web technologies: How do we measure ontology based search tools? How can we know if they perform better than the traditional methods for finding information on the Internet? How can we measure ontology based search tools? We believe that one of the best way to answer this is to carry out a case study including a carefully designed field experiment where Semantic Web tools are tested and compared to traditional approaches. We will discuss in this paper the set-up of a field experiment with Semantic Web tools, and we give some first results from our own field study that was carried out within the EU IST-1999-10132 On-To-Knowledge project [OTK].

2. Context of the experiment

The *On-To-Knowledge* project (cf. [OTK] or [Davies2002]) builds an ontology-based tool environment to speed up knowledge management, dealing with large numbers of heterogeneous, distributed, and semi-structured documents typically found in large company intranets and the World Wide Web. The project's target results are: (1) a toolset for semantic information processing and user access; (2) OIL, an ontology-based inference layer on top of the World Wide Web; (3) an associated methodology and validation by industrial case studies.

Ontologies [Fensel2001] are a core element of the project and a key enabling technology for the semantic web in general. They need to interweave human understanding of symbols with their machine processability. More recently, the concept of ontology is gaining tremendous ground in fields, such as intelligent information integration, cooperative information systems, information retrieval, electronic commerce, and knowledge management. The reason ontologies are becoming so popular is largely due to what they promise: *a shared and common understanding of a domain that can be communicated between people and application systems*.

EnerSearch is an industrial research consortium focused on IT and energy. Its aim is to create and disseminate knowledge on how the use of advanced IT will impact the energy utility sector, particularly in view of the view of the fact that this industry branch is being liberalized across Europe. EnerSearch has a structure that is very different from a traditional research company. Research projects are carried out by a varied and changing group of researchers spread over different countries (Sweden, US, Netherlands, Germany, France). Many of them, although funded for their work, are not even employees of EnerSearch. Thus, for its knowledge creation function EnerSearch is organized as a virtual research organization. Due to this wide geographical spread, EnerSearch has the character of a virtual organization also from the knowledge distribution point of view. In addition, for the general public interest it maintains a website (www.enersearch.se) where it publishes many of its research results as papers, reports and books. Thus, dissemination of produced knowledge on IT and energy is a key function for EnerSearch. Within the On-To-Knowledge project, EnerSearch investigates whether Semantic Web methods and tools might be helpful to improve on this function, especially focused on its web information provisioning.

3. Hypotheses to be explored

In sum, in designing Semantic Web experiments different design dimensions are of importance: variations in information modes, in target user groups, and in individual information-processing styles. Any experiment must be based on one or more clearly formulated hypotheses that can be verified or falsified, for example by empirical-statistical methods. A possible list of testable hypotheses regarding Semantic Web-based information seeking is:

1. Users will be able to complete information-finding tasks in less time using the ontology-based semantic access tools than with the current mainstream keyword-based free text search.
2. Users will make fewer mistakes during a search task using the ontology-based semantic access tools than with the current mainstream keyword-based free text search.
3. The reduction in completion and number of mistakes will be more noticeable for less experienced users.
4. The reduction in time will also be more noticeable for users lacking a detailed knowledge of the underlying technical system implementation.
5. The ontology-based semantic access tools will be perceived as more useful than free text search by different types of persons for a broad range of domains and knowledge-acquisition scenarios.
6. The effort for developing and maintaining the ontology and information structure will not significantly exceed the effort to develop and maintaining the free text approach. In a field experiment, hypotheses such as these are tested for their

significance, and it is investigated how their validity varies with different information modes, target user groups, and individual information-processing styles.

4. Setting up the Experiment

The involvement of knowledge users in the experiment from the beginning is important because of the interaction between the users and tasks on the one hand, and their pre-knowledge, or lack thereof, of domain and/or systems on the other hand. We identified several types of target users for the tests through conducting a set of pre-trial interviews. As a result, the evaluation experiment includes three different types of interest groups, as outlined previously. One group consists of staff members from four different shareholder companies (the companies involved in the case study are: Sydkraft AB, Sweden, Iberdrola from Spain, Electricidade de Portugal, and ECN, The Netherlands). A second group consists of researchers from different scientific fields, several having at some time participated in EnerSearch projects. The third and final group intends to represent more or less a general outside audience and consists of students (studying at the department of software engineering and computer science at the Blekinge Institute of Technology in Sweden). Finding information on a personal basis is important for all of these three groups, but for various reasons (that well reflect those found in geographically spread virtual organizations) they are generally limited in their time to invest for searching knowledge. The majority of the test users are familiar with the EnerSearch web and have used it before. There were of course also those who had never heard of the EnerSearch web but were introduced to and instructed on how the EnerSearch web is functioning. We included in the Appendix A the list of 30 questions that each user should answer. We divided the total of 45 test users into 6 groups. We then mixed up the order of the questions for each group as well as the tool that each user group should use for answering each question, i.e. each group had 3 blocks – one block per available tool – with 10 questions each.

The test persons will use three different tools: the two ontology-based semantic tools **QuizRDF** (developed by BT Exact Technologies, GB) and **Spectacle** (developed by Aidministrator, NL) on the one hand and the free text search tool **EnerSEARCHer** on the other hand. The On-To-Knowledge tools QuizRDF and Spectacle have the following advantages compared to a free text search tool like the EnerSEARCHER: (i) **QuizRDF** has the advantage that the user can start with simple queries consisting of only small number of search terms in order to get a picture of what kind of information is available in the EnerSearch knowledge base. There is a continuum here from the common keyword-based search to different levels of semantic search. (ii) The browsing tool **Spectacle** presents information according to the inherent structuring that is offered by an underlying domain ontology. This gives a valuable semantic context for the user not available through standard information retrieval tools. The information is presented in such a way that the path that leads to the information adds to the users' understanding of the semantic role of the information. Each concept chosen is ”surrounded” by other ontologically similar concepts. Semantic generalizations and specializations (in different dimensions) are also offered to the user when browsing for information.

In Appendix B we illustrate the technical setting by giving screenshots of the participating tools and the GUI that was used for guiding users through the questionnaire.

5. Preliminary Results

So far we only evaluated a small amount of data from the experiment, i.e. from seven test users, that partially answered our online questionnaire (cf. Appendix A). We are far from

having statistically significant results available, but we are able to illustrate some first impressions from the evaluation of our data. Each log entry contains the following items: question number, answer (text), name of the user, time duration for answering the question, the usergroup to which the user belongs, the tool used for getting the answer and how easy the user found to answer this question with the particular tool (on a scale from 1-"easy" to 5-"hard" plus 6-"I give up"). Until now we received 177 log entries from 7 test users. In our first analysis we concentrated on getting an impression for the hypothesis 1 and 2 (cf. Section 3.3).

Figure 1 shows the calculated results for answering the question: "How relatively often did users give (W)rong, (R)ight or (N)o answers with each tool?". The figure shows the following preliminary results: For EnerSEARCHER, 23,19% of the questions answered in total (with EnerSEARCHer) were wrongly answered, 37,68% were answered right and in 39,19% of the cases the user gave up, thus resulting in having no answer at all for the question. For RDF-quiz, 10,20% of the questions were answered wrong, 57,14% were answered right and in 32,65% the user gave up. For Spectacle, 23,73% of the questions were answered wrong, 40,68% were answered right and in 35,59% cases the user gave up.

Thus, as a first result, our hypothesis 2, "Users will make fewer mistakes during a search task using the ontology-based semantic access tools than with the current mainstream keyword-based free text search", is supported by this result.

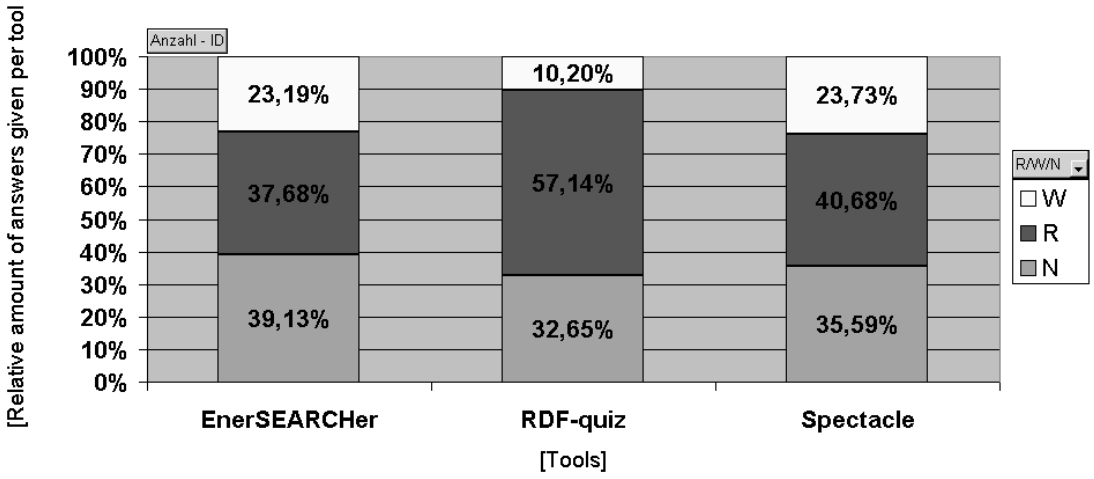

Figure 1: How relatively often did users give (W)rong, (R)ight or (N)o answers with each tool?

Figure 2 shows the calculated results for answering the question: "What relative average amount of time needed users for (W)rong, (R)ight or (N)o answering of one single question?". We highlight the most relevant detail of this figure (the reader might use the figure for further interpretations): To answer a question right, users needed in average the shortest amount of time with RDF-quiz (25,77%), followed by EnerSEARCHer (34,71%) and Spectacle (39,52%).

Thus, as a second result, our hypothesis 1, "Users will be able to complete information-finding tasks in less time using the ontology-based semantic access tools than with the current mainstream keyword-based free text search", is partially supported by this result.

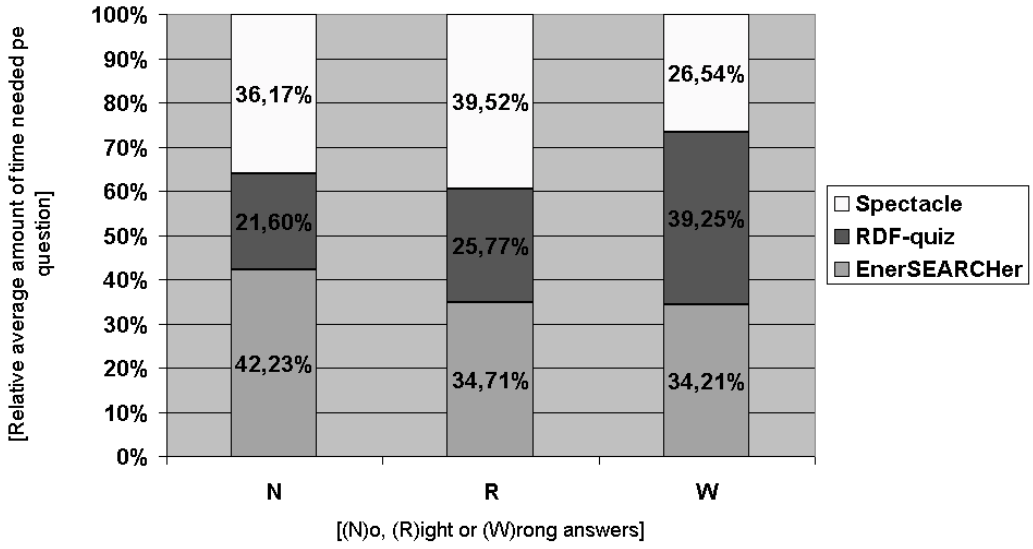

Figure 2: What relative average amount of time needed users for (W)rong, (R)ight or (N)o answering of one single question?

A careful look at the data revealed, that for some few entries in the logfile exist large durations for the answering of the questions, that are far beyond the "typical time" needed for finding an answer. These few entries have a significant influence on the average duration for answering questions and they exist for every given tool. We will investigate further to clarify whether users simply did other things then answering the questions during that time or what the reasons were for these long durations.

6. Conclusion

To demonstrate the real value of Semantic Web we need to carry out field experiments. We have outlined a number of hypotheses that we try to find answer to in our case study. We also have described what kind of variables have to be taken into account, how data collection, evaluation, experiment procedure and system design can be done, and we have sketched the importance of the information processing. First results indicate that (i) our hypothesis 2, "Users will make fewer mistakes during a search task using the ontology-based semantic access tools than with the current mainstream keyword-based free text search", is supported by our results, and (ii) our hypothesis 1, "Users will be able to complete information-finding tasks in less time using the ontology-based semantic access tools than with the current mainstream keyword-based free text search", is partially supported by our result.

So far we only evaluated a small amount of data, i.e. from seven test users, and only investigated in proving hypothesis 1 and 2. However, the experiment is an ongoing task and we are aiming at finishing the experiment by September. The ongoing tasks include the collection and complete evaluation of the data from the remaining test users (total of 45) as well as e.g. post-trials with the users to get feedback on the experiment as a whole. We will investigate which cognitive style users prefer and how that correlates with their usage of the tools, and, last but not least, how the clustering in the different user groups influence the results of our experiment. In a future version of this paper we will include these results.

Acknowledgements. Part of this work is financed by the European Commission in the EU IST-1999-10132 project On-To-Knowledge. We thank all project partners for their friendly collaboration and fruitful discussion. In particular, we thank Prof. Hans Akkermans (Free

University Amsterdam), Prof. Rikard Larsson (Lund University Business School), Peter Mika (Free University Amsterdam) and Henrik Bylund (Webmaster EnerSearch) who contributed essential parts of this work presented.

References

[Berners-Lee2001] T. Berners-Lee, J. Hendler, and O. Lassila: The Semantic Web, *Scientific American*, May 2001.

[DAML] US DARPA project "DARPA Agent Markup Language (DAML)", cf. http://www.daml.org.

[Davies2002] J. Davies, D. Fensel, F. van Harmelen (eds.). *On-To-Knowledge: Semantic Web enabled Knowledge Management.*, Wiley, to appear 2002.

[Driver1993] M.Driver, K.R. Brousseau, and P.L. Hunsaker. The Dynamic Decision Maker. Jossey-Bass, San Fransisco, CA, 1993.

[Fensel2001] D. Fensel: *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag, Berlin, 2001.

[OntoWeb] EU IST-2000-29243 thematic network "OntoWeb: Ontology-based Information Exchange for Knowledge Management and Electronic Commerce", cf. http://www.ontoweb.org.

[OTK] EU IST-1999-10132 project "On-To-Knowledge: Content-driven knowledge management tools through evolving ontologies", cf. http://www.ontoknowledge.org. In particular, the deliverables with descriptions of the tools and the EnerSearch case study can be found at: http://www.ontoknowledge.org/del.shtml.

Appendix

A. List of questions

1. Name a Knowledge Management methodology?
2. How does load management effects pricing?
3. How large energy savings did the multi-agent system simulation of building control indicate?
4. When is it reasonable to adapt to changing communication conditions of a channel?
5. How does energy prices in Sweden compare to prices in Europe?
6. How does the liberation of an energy market influence the effectiveness of DSM measures?
7. How can you model the communication between intelligent agents on a society level?
8. How can spot pricing affect a distributor?
9. Name an organization that works with preparing protocol and equipment standards?
10. Who are the owners of EnerSearch?
11. What will happen to the price levels of electricity in Europe the next few years?
12. In what project has Claes Badenschneider been active?
13. Which are the three main mechanisms for a proper utility strategy according to Brousseau et al.?
14. What is the name of the building in Ronneby where a lot of field tests have been performed?
15. When did Hans Ottosson design and implement his first load management system?
16. To whom should a registration for ISPLC 2001 be sent?
17. What are the case studies in the
18. What are the two technology-related problem areas where improvements are needed with respect to PLT?
19. To what e-mail address should general questions to EnerSearch be sent?
20. How can agents communicate with devices in an intelligent building?
21. What is life cycle cost and how to calculate it?
22. Which approach could be used to improve utility -customer relationship?
23. How does agent load management benefits the customer?
24. What styles of decision making could be distinguished?
25. What are some of the aspects to consider when assessing the costs of agent based load management?
26. What approaches are used to describe business processes?
27. What are some of the weaknesses of the existing tools used to manage documents?
28. How is Information Technology changing the way organizations operate?
29. What is the name for an interorganizational relationship in which independent organizations share their resouces, knowledge, costs and risks in order to produce a product?
30. Where can we find information on virtual organizations?

B. Illustration of the Technical Setting

The following Figures B.1-B.3 show screenshots of all three tools during a user trial including the GUI that was used to guide users through the questionnaire during the experiment. The tools are all accessible with a typical browser.

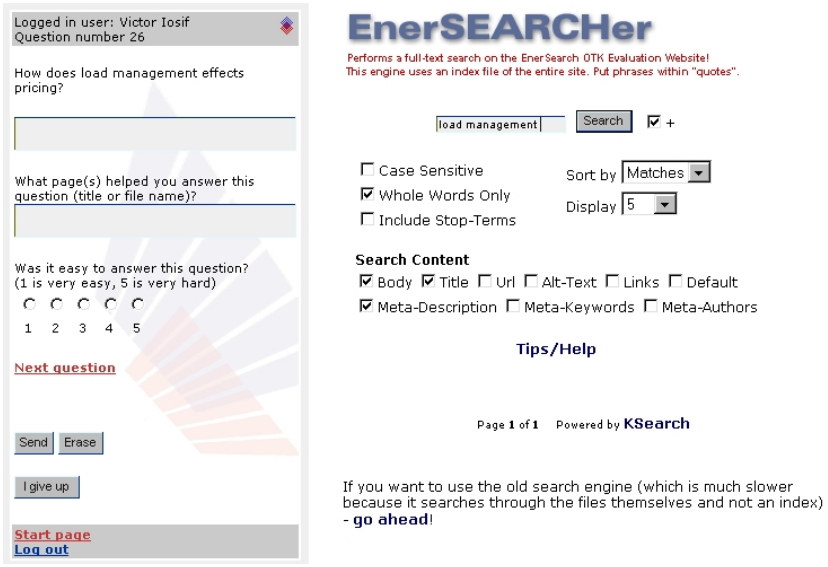

Figure B.1: EnerSEARCHer (free text)

Figure B.1 shows the setting during the experiment for the EnerSEARCHer. The screen is splitted into two parts, i.e. frames. (i) On the left side you see a frame that is used to guide users through the questionnaire during the experiment. After a user is logged into the system, (s)he gets presented question by question in the upper part and can type in the answers. Along with each answer a user should also note how easy (s)he found to answer a particular question (on a scale from 1-"easy" to 5-"very hard"). If a user could not find any answer (s)he could push the "I give up button". Each user had to answer 30 questions, i.e. 10 with each tool. As mentioned in the paper, the questions were mixed up for different user groups. Which leads us to the second part, the tools themselves. (ii) On the right side the currently active tool for answering a question (EnerSEARCHer, QuizRDF or Spectacle) was presented to the user, i.e. the tool a user had to use for answering a question was given in this frame. Here you see the GUI of EnerSEARCHer, where you can easily recognize a typical query interface for keyword based search engines.

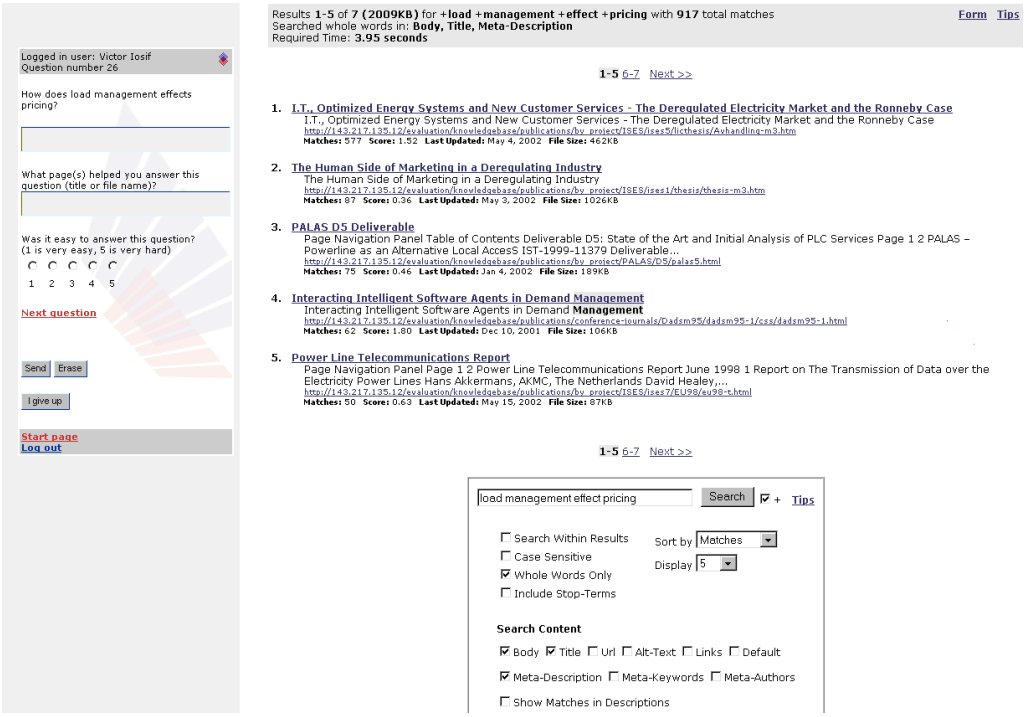

Figure B.2: EnerSEARCHer results

In our first example screenshot Figure B.1 the user "Victor Iosif" was currently trying to answer the question 26 "How does load management effects pricing?". He had to answer the question with EnerSEARCHer. E.g. other users were given QuizRDF or Spectacle. Figure B.2 shows how EnerSEARCHer presents a result list corresponding to a query to the user (in a typical keyword based search engine like manner).

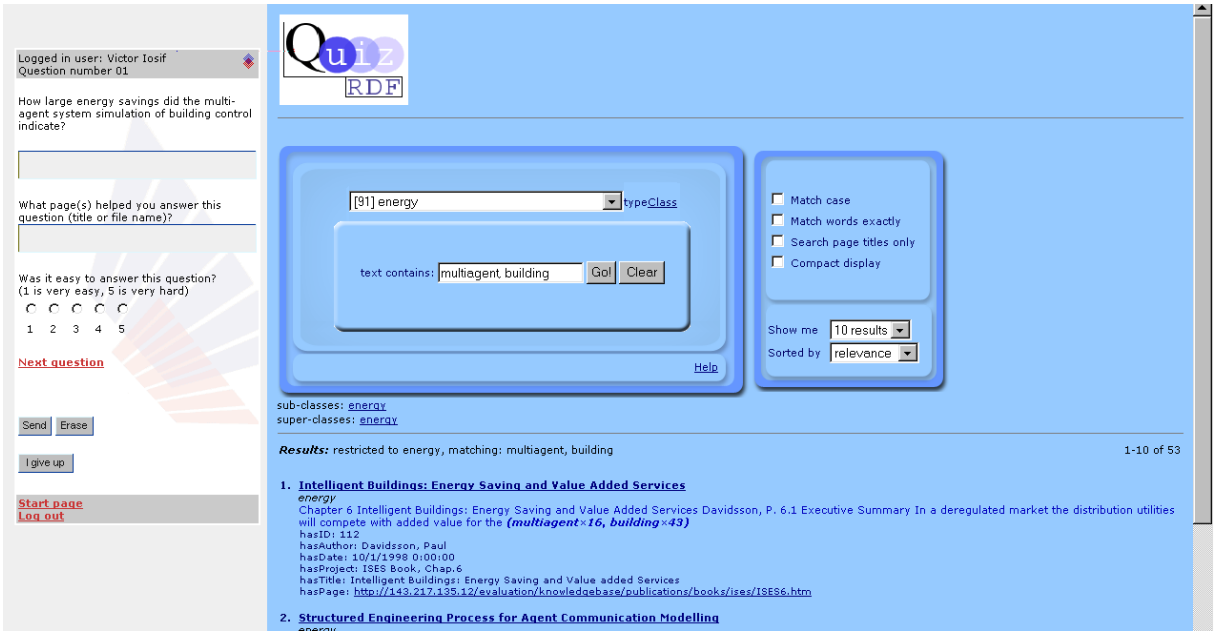

Figure B.3: QuizRDF (ontology based)

Figure B.3 depicts the query interface of QuizRDF. The user can start with entering keywords that (s)he is looking for - similar to a typical keyword based search like EnerSEARCHer. Additionally, (s)he can choose a concept (or "Class") from a predefined ontology in the upper text box of QuizRDF. QuizRDF then returns documents that are relevant to the chosen concept and/or the given keywords. In the lower part a typical search result is presented. Additionally to the name of a document, QuizRDF gives a short summary of a found document and metadata according to the predefined ontology.

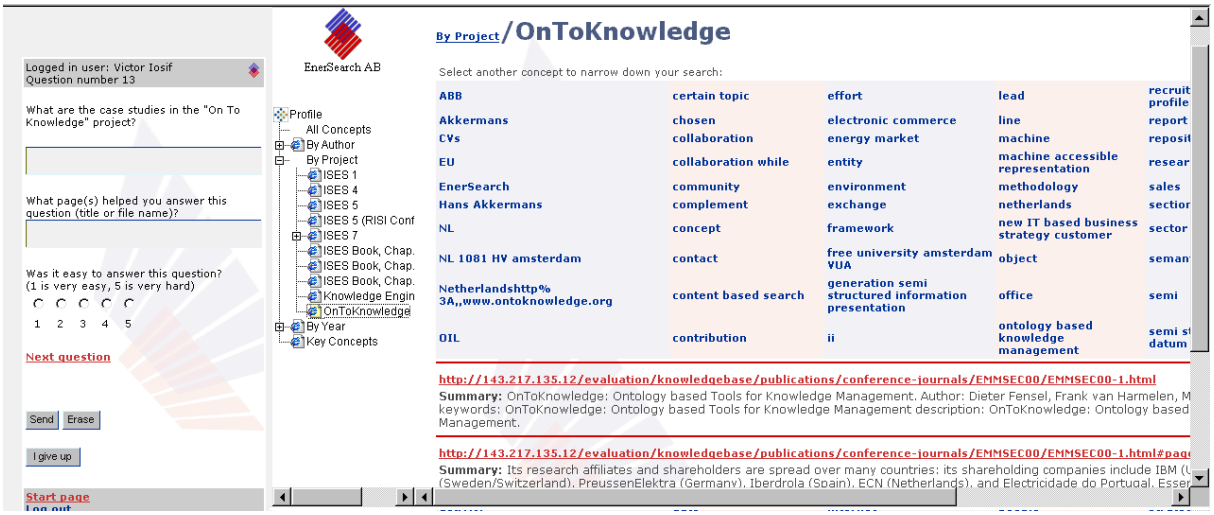

Figure B.4: Spectacle (ontology based)

Figure B.4 illustrates the interface of Spectacle. On the left side you see an explorer-tree like browsing structure which is according to the underlying ontology. Spectacle offers the user some predefined views on the ontological structure, e.g. to start a search by looking for "Autors", "Projects" or "Year". These concepts were identified as an intuitive starting point for the search. Known Projects are shown as childs of the "Projects"-Node, e.g. the On-To-Knowledge project. On the right side of the interface you see on the upper side the currently chosen navigational path, e.g. "By Project/OnToKnowledge". Below the user gets presented the context of a chosen concept, i.e. related concepts. Users may follow these links to naroow down their navigational search. At the bottom you see a result list indicating the relevant documents for the currently chosen navigational path in the ontology.

Further information regarding the tools as well as the case study and the On-To-Knowledge project in general can be found at [OTK].