Question Answering on Top of the BT Digital Library

Philipp Cimiano, Peter Haase, York Sure, Johanna Völker, Yimin Wang Institute AIFB University of Karlsruhe 76131 Karlsruhe, Germany +49 721 608 6592

{pci,pha,ysu,jvo,ywa}@aifb.unikarlsruhe.de

ABSTRACT

In this poster we present an approach to query answering over knowledge resources that makes use of different ontology management components within an application scenario of the BT Digital Library. The novelty of the approach lies in the combination of different semantic technologies providing a clear benefit for the application scenario considered.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries – user issues, system issues.

I.2.1 [Artificial Intelligence]: Applications and Expert Systems – *Natural Language Interfaces*.

I.2.7 [Artificial Intelligence]: Natural Language Processing – *Language parsing and understanding.*

General Terms

Design, Experimentation, Human Factors, Languages, Theory

Keywords

Question Answering, Web Ontologies, Ontology Learning, Natural Language Processing.

1. INTRODUCTION

Enhancing the knowledge access to the Digital Library of the British Telecom is the goal of one of the case studies in the EU IST integrated project SE Knowledge Technologies (SEKT).

In current interfaces to Digital Libraries, users pose keywordbased queries to perform document retrieval. However, these keywords to do not directly represent the semantics of the information need of the user.

We have implemented an approach that allows the user to perform structured natural language queries against the information contained in the Digital Library. The semantics of the information and the user queries is defined by an underlying ontology. Further, in order to allow structured queries against the initially unstructured content of the library, we rely on ontology learning techniques to make both the structure and the semantics of the content explicit

As a result, users are able to ask queries such as "Who wrote a document which talks about network protocols?", i.e. queries that (1) allow to relate different knowledge sources (bibliographic metadata and concepts from the unstructured content), (2) do not

only allow to return documents, but structured answers to the query. Figure 1 shows a screenshot of the web browser-based knowledge portal to the BT Digital Library, displaying the result of such a structured natural language query.

Figure 1	1	Screenshot	of	the	вт	Digital	Library
I Igui C .	-	Dereensnot	•••	une	~	Digital	Libiary

	dit Favoritan Extras ?		
	👔 📓 🟠 🔎 Suchen 🤺 Favoriten 🥹 😭 • 🎄 🗔 et 8080/btdemo/SPARQL Tquery = Which + document + talks + .	the second se	ww_meta=*85earch=Ask+ 🛩 🚺 Wechseln zu - Links * 🕴
ame INTAZ INT Ted	ey Services IIT Ralp		Search BT . er Directory O for
вт 😥	BT Library		
and a second	Al Areas 🕱 Au	t the library	
Ibrary Links bout Us cronyms If Patents Yormation Spaces	Which document talks about which concept 7 31 answer(s) relieved.		
oumai List Xher Resources	WebC4V based open source collaborative development anononment	ineliwork protocol	
What is New These are party course for your visat the forduring and the second second second second New Books Dealer That the Head the second	Knowledge management and the framing of information: a contribution to ORIMS practice and pedagogy	decision maker	
	Encodedge management and the training of information: a contribution to ORIME practice and pedagogy	utralegy	
	Knowledge management and the framing of internation, a contribution to OR/MS practice and pedagoop	reite	
	Providedge management and the framing of information: a contribution to ORMAII analysis and percepty	(BRIVES	
	Roowledge management and the framing of information, a contribution to ORIMI practice and periapoly	anation	
	Knowledge management and the franking of information, a contribution to ORIMI sharehold and addepage	retation	
	Proviledge management and the framing of information: a contribution to CRMM practice and pedhopy	problem descration	
	Providege management and the training of internation: a contribution to ORME practice and pertaiopy		
	Providing management and the framing of information, a contribution to Offstell practice and pediagoge	builting	
	Knowledge management and the transing of information, a contribution to ORIME practice and pedagogy	bese	
	Roowledge management and the framing of information: a contribution to OR/MII practice and pertagopy	maght	
	Poweradge management and the transmig structureation: a contribution to ORMII practice and pedagoge	situation	
	Roowladge management and the framing of internation: a contribution to ORIMI practice and pediapoly	affect	
	Providege management and the haming of information, a contribution to ORARS practice and pedagogy	-	
	King-Martige management and the Staming of information: a control during of information and pedagogy	Berry	

Figure 2, shows the conceptual architecture of the application, which we briefly explain in the following.

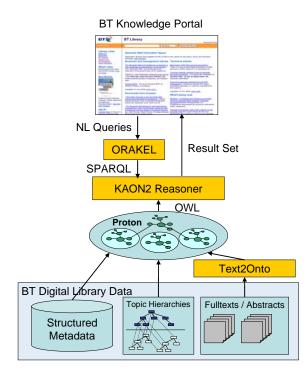
2. INTEGRATING HETEROGENEOUS KNOWLEDGE SOURCES

As shown in the bottom of Figure 2, the knowledge sources of the BT Digital Library comprise databases with bibliographic metadata, topic hierarchies, such as INSPEC [6], but also unstructured sources such as fulltext documents with different formats. All these heterogeneous knowledge sources are integrated into a common ontology, which is based on Proton [5]. While the structured information sources are integrated using a mapping of the underlying structures to the ontology, we obtain structured ontologies from the unstructured sources with the help of Text2Onto [2].

The aim of Text2Onto is to support developers in the ontology construction process by applying text mining techniques. Ontologies automatically generated with Text2Onto can be exported to a number of formats, among these the Web Ontology Language OWL. We can easily perform some user-oriented actions like querying and managing to both structured and unstructured heterogeneous knowledge source, after constructing the ontology using Text2Onto. According to our experiences, the ontologies constructed by Text2Onto are usable per se and furthermore represent a basis which the ontology engineering process can build on.

The integration of heterogeneous knowledge at the semantic level then allows to apply OWL reasoners such as KAON2 to answer the queries and provide an additional level of integration via reasoning.

Figure 2 Conceptual Architecture of the Application



3. ONTOLOGY MANAGEMENT AND QUERY ANSWERING WITH KAON2

The integrated ontology is managed by the KAON2 ontology management system, which is also the component responsible for the actual query answering. We here rely on SPARQL as the query language which is currently supported by KAON2.

In our system, we are using the Proton ontology as the knowledge base. Proton is the SEKT-specific domain ontology which the BT digital library data is based on. The library data are mainly captured from databases and stored as OWL instances, so that the system can apply SPARQL query to the data.

Figure 2 shows that our system, besides importing the Proton ontology as well as library data captured from the data base, also includes information automatically generated by Text2Onto. The KAON2 reaonser handles the subsequent operations to manage the ontology and answer the queries. At last the result set is

Copyright is held by IW3C2. WWW 2006, May 22–26, 2006, Edinburgh, UK. processed and sent back to be displayed by the BT query webpage portal.

4. NATURAL LANGUAGE INTERFACE

ORAKEL [1] is a natural language interface which translates natural language queries to structured queries. This translation relies on a lexicon for the underlying Proton ontology, which specifies the possible lexical representations of the ontology elements in the user queries. ORAKEL generates the lexicon partially automatically from the underlying ontology. The lexicon can be refined manually with appropriate tool support

From the user's view, they are able to directly interact with BT digital library portal, by accessing the library data with natural language questions, which are translated into SPARQL queries by a component called ORAKEL. The underlying mechanism however is hidden from the users – the only thing user need to do is to input the query just as their normal questions and then get the result from the portal in a moment.

From the view of usability and human factor engineering, this interface has the big advantage of bringing the user out of the game of guessing and trying the keywords in the entry of the webpage portal. Obviously, most people have the experience of struggling with the keywords of the query, especially when their searching target is uncertain. This interface enables users to query the data by the relations among them without knowing any keyword included in the data.

5. CONCLUSION

We have presented an approach that combines different ontology management, learning and reasoning techniques in order to allow question answering in the BT Digital Library. The users are able to perform structured natural language queries against a variety of knowledge sources in an integrated manner with a well-defined semantics provided by the underlying ontology. The novelty of our system lies in the combination of different tools for natural language question interpretation, ontology learning, query answering as well as reasoning.

6. ACKNOWLEDGMENTS

The work reported here has been partially financed by the EU projects IST -2003-506826 SEKT, IST- -2003-507483, DIP and IST- 2001-34038 DOT.KOM.

7. REFERENCES

- P. Cimiano. ORAKEL: A Natural Language Interface to an F-Logic Knowledge Base. In Proceedings of NLDB'04, Salford, UK, June 2004.
- [2] P. Cimiano, J. Völker Text2Onto A Framework for Ontology Learning and Data-driven Change Discovery. In Proceedings of NLDB'05, June 2005.
- [3] M. Lytras et al. Digital libraries in the knowledge era: Knowledge management and Semantic Web technologies. Journal of Library Management, Vol. 26 Issue 4/5 P. 170 – 175, May 2005.
- [4] <u>http://kaon2.semanticweb.org</u>
- [5] <u>http://proton.semanticweb.org/</u>
- [6] http://www.iee.org/publish/inspec/