An e-Librarian Service that Understands Natural Language Questions from Students

Serge Linckels[1,2], Christoph Meinel[1], and Thomas Engel[2]

[1] University of Potsdam, Hasso-Plattner-Institut (HPI) für Softwaresystemtechnik GmbH, D-14440 Potsdam, Germany
{linckels,meinel}@hpi.uni-potsdam.de
[2] University of Luxembourg, Luxembourg International Advanced Studies in Information Technologies (LIASIT), L-1359 Luxembourg, Luxembourg
thomas.engel@uni.lu

Abstract. In this paper we present an *e-librarian service* which is able to retrieve multimedia resources from a knowledge base in a more efficient way than by browsing through an index or by using a simple keyword search. Our premise is that more pertinent results would be retrieved if the e-librarian service had a semantic search engine which understood the sense of the user's query. This requires that the user must be given the means to enter semantics. We explored the approach to allow the user to formulate a complete question in natural language. The background theory was implemented in two different prototypes.

1 Introduction

Our vision is to create an *e-librarian service* which is able to retrieve multimedia resources from a knowledge base in a more efficient way than by browsing through an index or by using a simple keyword search. Our premise is that more pertinent results would be retrieved if the e-librarian service had a *semantic search engine* which understood the sense of the user's query. This requires that the user must be given the means to enter semantics. We explored the approach to allow the user to formulate a complete question in natural langauge (NL). Linguistic relations within the user's NL question and a given context, i.e. an ontology, are used to extract precise semantics and to generate a semantic query. The e-librarian service does not return the answer to the user's question, but it retrieves the most pertinent document(s) in which the user finds the answer to her/his question.

In section 2 we give a brief overview of the technical background of the e-librarian service. In section 3 we present the two prototypes. We conclude with some (dis)advantages and some future work in section 4.

2 Technical Background

Our e-librarian service is an ontology driven expert system about a given domain (e.g. computer history, fractions in mathematics). It is composed of a domain language and a concept taxonomy. As an illustration, in the concept taxonomy for our prototype CHESt (section 3), a document describing the transistor would be placed in the concept "EComponent" (electronic component), which is a hyponym of "Hardware". On the one hand, the more detailed the taxonomy is, the more exact the system can classify the documents. On the other hand, a very detailed taxonomy reduces the tolerance for the user question, so that it must be very well and precisely formulated.

The user can enter his question in natural language (NL), which is then translated into an unambiguous logical form w.r.t. a given ontology. This step is called non-standard inference [5]. Therefor, the expert system is able to find implicit consequences of its explicitly represented knowledge. This means that the content of a document is only of minor importance, but the meaning of the clip as a whole has to be machine readable. The meaning of each clip is described by additional data — called metadata — that are encoded using a specific ontology framework. We use the W3C recommendation *Web Ontology Language, Description Logics* (OWL DL) (`http://www.w3.org/2004/OWL/`) to describe all documents (*resources*) with metadata. We published in [6] how DL can be used in our e-librarian service, where the concept taxonomy of the ontology is translated into an acyclic $\mathcal{ALC}$-concept description. The language $\mathcal{ALC}$ [10] is sufficiently expressive for our purposes. It is in fact a subset of the logics implemented in most "state of the art" DL reasoners; we use *Pellet* [11]. The returned results are logical consequences of the inference rather than of keyword matchings.

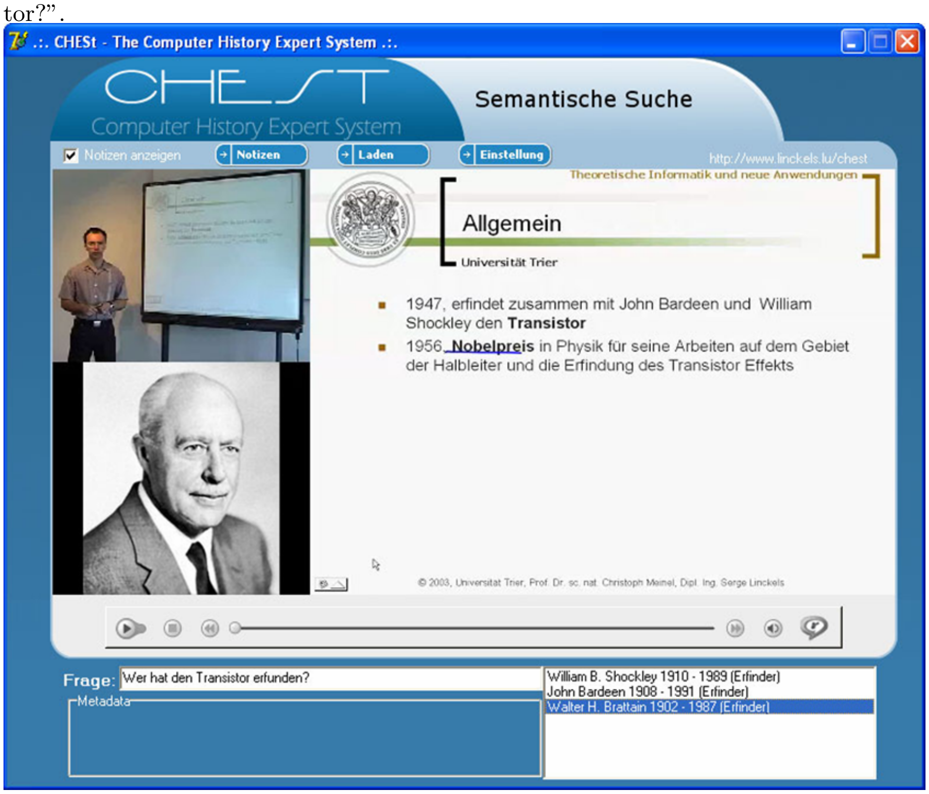
In our e-librarian service, the retrieval itself is a simple extension of the semantic interpretation of the user question. In fact, a given semantic interpretation is used to generate a semantic query, and to logically infer over the knowledge base. A semantic query over a knowledge base $\mathcal{K}$ w.r.t. a domain ontology H, and a query in an $\mathcal{ALC}$ terminology means that there must exist at least one model $\mathcal{I}$ of $\mathcal{K}$ such $(R_q)^{\mathcal{I}} \neq \emptyset$, written $\mathcal{K} \models R_q$. As illustration, lets suppose the NL user question q = "Who invented the transistor?". Below is the according $\mathcal{ALC}$ query expression w.r.t. the CHESt ontology and the variable y, which is the missing part, and which should be the result of the query.

$$\begin{aligned}\mathcal{K}_{CHESt} \models R_q = EComponent(x)\ \wedge\ hasTitle(x, \text{"}transistor\text{"})\\ \wedge\ hasInventor(x, y?)\ \wedge\ Inventor(y?)\end{aligned}$$

3 Implementation

The background theory for our e-librarian service was implemented prototypically in two different educational tools. The multimedia documents (*clips*) were recorded with tele-TASK (`http://www.tele-task.de`). A first prototype is CHESt (*Computer History Expert System*), an e-learning tool where the user can freely formulate his question in NL (`http://www.linckels.lu/chest`). CHESt understands the user's question and returns a precise and short answer in multimedia form. The tool has a knowledge base with 300 multimedia clips that cover the main events in computer history. The result of experiments were published in [7].

Fig. 1. CHESt with a semantic search and the question: "Who invented the transistor?".



A second prototype of our e-librarian service is MatES (*Mathematic Expert System*), which explores a slightly different pedagogical approach. The knowledge base covers the topic of fractions in mathematics taught in schools (7^{th} grade). All clips were recorded, mainly by pupils, w.r.t. the official school programme. The tool has a knowledge base with 115 multimedia clips that cover all questions about fractions for this level.

4 Conclusion

Experiments [7] confirmed that our background theory for an e-librarian service can be implemented in an educational tool. Such a tool can be used as a complement to traditional courses. The presentation of the knowledge in the form of short multimedia clips, and the fast response time of the search engine were strongly appreciated by the students. But we learned that our semantic search mechanism can be improved in several ways, e.g. all queries are built "bottom up", without consideration of neither former queries, nor queries from other users. A collaborative information retrieval approach like the one proposed by [4], or the study by [12], seem interesting solutions to explore.

We understood that it is not an easy task to use search engines as a didactical tool in schools. Firstly, in a free discussion with the students, the problem was often mentioned that the topic (e.g. computer history) was too complicated. Users need training and domain knowledge before they are able to successfully use search engines for that topic. Secondly, when using search engines, the students are relatively free to act as they like, which is quite unusual for most. As confirmed by [1, 8, 3, 9, 2], users need guidance in how to formulate effective queries even if they are free to formulate their question in NL.

References

1. F. Blondel. La recherche d'informations sur internet par des lycéens, analyse et assistance à l'apprentissage. In P. J. Vries E., Pernin J.P., editor, *Hypermédias et Apprentissages 5 : Actes du cinquième colloque*, pages 119–133, 2001.
2. C. Bonk. *Online Teaching in an Online World*. CourseShare.com, `http://www.publicationshare.com/`, 2001.
3. R. Fidel, R. K. Davies, M. H. Douglass, J. K. Holder, C. J. Hopkins, E. J. Kushner, B. K. Miyagishima, and C. D. Toney. A visit to the information mall: Web searching behavior of high school students. *Journal of the American Society for Information Science*, 50(1):24–37, 1999.
4. S. Klink. Improving document transformation techniques with collaborative learned term-based concepts. In A. Dengel, M. Junker, and A. Weisbecker, editors, *Reading and Learning*, volume 2956 of *Lecture Notes in Computer Science*, pages 281–305. Springer, 2004.
5. R. Küsters. *Non-Standard Inferences in Description Logics*, volume 2100 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, 2001. Ph.D. thesis.
6. S. Linckels and C. Meinel. A simple application of description logics for a semantic search engine. In N. Guimarães and P. Isaías, editors, *Proceedings of the IADIS International Conference Applied Computing 2005, Lisbon, Algarve, Portugal, February 22-25, 2005*, volume 2, pages 306–311. IADIS, 2005.
7. S. Linckels, C. Meinel, and T. Engel. Teaching in the cyber-age: Technologies, experiments, and realizations. In *3. Deutsche e-Learning Fachtagung Informatik der Gesellschaft fr Informatik, DelFI 2005, Rostock, Germany, September 13-16, 2005*, 2005.
8. P. Martin. *Web Intelligence*, chapter Knowledge Representation, Sharing and Retrieval on the Web, pages 263–297. Springer-Verlag, Jan 2003.
9. R. Navarro-Prieto, M. Scaife, and Y. Rogers. Cognitive strategies in web searching. In *5th Conference on Human Factors & the Web*, `http://zing.ncsl.nist.gov/hfweb`, 1999.
10. M. Schmidt-Schauß and G. Smolka. Attributive concept descriptions with complements. *Artificial Intelligence*, 48(1):1–26, 1991.
11. E. Sirin and B. Parsia. Pellet: An owl dl reasoner. In V. Haarslev and R. Möller, editors, *Description Logics*, volume 104 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2004.
12. J. D. Wilson, S. A. Cordry, and N. King. Building learning communities with distance learning instruction. *AECT TechTrends*, 48(6):20–22, 2005.