OPAS : An Ontology-based Library Help Desk Service

Olli Alm, Eero Hyvönen and Antti Vehviläinen
Semantic Computing Research Group (SeCo)
Helsinki University of Technology (TKK), Laboratory of Media Technology
University of Helsinki, Department of Computer Science
firstname.lastname@tkk.fi, http://www.seco.tkk.fi/

Abstract

This paper argues that ontological knowledge technologies can be utilized in authoring question-answer services. To ease the content indexer's work, we propose a system that provides 1) semi-automatic semantic indexing for annotating question-answer pairs, 2) case-based reasoning techniques for finding similar questions and 3) integration of external information sources via ontologies. A real life ontology-based question-answer application OPAS is presented as a proof of concept.

1 Introduction

We consider question-answer (QA) services where human experts provide answers to written questions. Examples of such services include help desks and the popular Frequently Asked Questions (FAQ) lists. QA services share some characteristics: 1) An archive of existing questions and answers is usually available for browsing and searching. 2) The QA pairs are often annotated with *index terms* for information retrieval purposes. 3) Questions are frequently repeated.

This paper shows how semantic web technologies can be utilized in QA systems from the content indexer's viewpoint. The ideas are presented by describing a semantic QA service OPAS[1][5]. OPAS is based on the existing *Ask a librarian* service[2] offered nationally in Finland by the editors of the Libraries.fi portal. Here the clients can send questions to a virtual librarian via email, and a librarian of the service provides an answer within three working days. All QA pairs are indexed by the librarians using the YSA thesaurus[3] of some 23,000 common Finnish terms. The dataset at the moment consists of over 20,000 QA pairs. A keyword-based search service is available on the web for both end-users and indexers to use. In this paper, we have focused on three aspects to enhance content indexing: 1) how to use semi-automatic semantic indexing to help in choosing appropriate index terms for QA pairs, 2) how to apply case-based reasoning (CBR) for finding existing similar QA-pairs for a new submitted question and 3) how to utilize relevant information services to support answer authoring.

2 Semi-Automatic Semantic Indexing

Two major problems of the current service were identified from the indexer's viewpoint: 1) Choosing the appropriate index terms for a question-answer pair is often time consuming and difficult. 2) There are different conventions used in indexing by different people, which makes the content unbalanced. For example, one librarian may use a few general terms to describe an answer, whereas another uses a large number of more detailed terms.

To address these problems semiautomatic indexing is employed in OPAS. We created an ontology-based information extraction tool POKA[4] for textual data, and integrated it with OPAS. POKA provides the QA indexer with a list of possible index terms as ontological references, and the indexer chooses which terms she wants to use. The basis for indexing with *common noun terms*, such as *dog*, *astronomy*, or *child*, is the General Finnish Upper Ontology, YSO[5] that is an ontologized version of the YSA thesaurus used originally in the service. YSO contains over 20,000 Finnish concepts organized into 10 major subsumption hierarchies. Along with the common concepts, POKA extracts places of a place ontology and utilizes a rule-based person name recognizer. Persons and places act as additional information for the YSO terms. An author may add free indexing terms (*Lassie*) not found in the ontology.

For each free term, an instance of the corresponding YSO-class (e.g., *city, person, dog*) has to be selected. Free indexing terms with the same name can be distinguished with different URIs and with an additional comment.

If the input text is long, POKA yields a considerable number of possible index terms. For that reason it is useful to order the terms according to their likely relevance. In our case, we use the idea [3] of searching for *semantic clusters* from the term set and conclude, that these terms are more relevant than semantically isolated terms. For example terms *doctor*, *sickness* and *medication* form a semantic cluster. For common noun terms we use the concept relations defined in YSO to identify these clusters.

[1] http://www.seco.tkk.fi/applications/opas/

[2] http://www.kirjastot.fi/tietopalvelu

[3] http://vesa.lib.helsinki.fi

[4] http://www.seco.tkk.fi/applications/poka/

[5] see http://www.seco.tkk.fi/ontologies/yso/

3 Retrieval of similar QA-cases

Case-based reasoning (CBR) [1] is a problem solving paradigm in artificial intelligence where new problems are solved based on previously experienced similar problems. OPAS contains a CBR component that automatically searches for similar question-answer pairs based on the terms that POKA has extracted from the question text.

The weighted index term list discussed above is used as the basis for the previous QA search with the following modifications: 1) The terms that the indexer has selected are given a substantially higher weight since their relevance has been confirmed by the indexer. 2) The extracted places, persons and free indexing terms are given a higher weight due to their specificity. Previous QA-cases provide essential background information for the answer composition. If the new question is analogous with the previous one, the existing answer can be linked to the new question and the redundancy of the QA-set is avoided.

4 Integrating Information Sources

Nearly all the librarians of the *Ask a librarian* service use the reference library with real books to find useful resources. To ease the access to the relevant information sources, we integrated two essential sources to the system: 1) the Helsinki City Library collections and 2) the link library maintained by the editors of the Libraries.fi. Both sources are classified with the Helsinki City Library Classification System (HCLCS) [6].

An ontology for a library classification system was created for OPAS, and then the HCLCS was converted into this ontologized form. The basis for the classification ontology is Simple Knowledge Organisation System (SKOS)[7] and the conversion was made following the guidelines given in [4]. In addition to class hierarchies the HCLCS contains index terms, and each of these terms has got a relation to a library class. For example the term *Treatment of alcoholics* has got a relation to the library class 371.71 *Alcohol policy*.

Index terms in the HCLCS contain also *views*. For example the term *pieces of art* (”Teokset”) embodies different viewpoints such as bibliographies and art collections. Each of these viewpoints is related to a library class. These relations between index terms and library classes are used to search for material that could be relevant to the answer. A librarian can browse the HCLCS to find information for answering and also to suggest further reading for the client.

Figure 1 depicts the view that the librarian sees when she has decided to answer a question. The end-user has submitted a question about Arto Paasilinna’s life and his books (see *The user’s question* box). On the right from the question text are the found concepts. There are two common noun concepts “teokset” (writings) and “esitelmät” (plays). POKA has also identified the person name “Arto Paasilinna”. Below the question text, there are *the authoring components* for finding similar questions, books and link library links.

5 Evaluation, Discussion and Further work

OPAS is a working prototype for assisting librarians in a help desk service. It is currently under commercial development. Semantic Web technologies are harnessed to provide 1) the indexing vocabulary (YSO), 2) the similarity measure for the dataset (semantic cliques) and 3) the linkage to external information sources (HCLCS). Because of the complex and open-domain nature of the questions, the indexing is supervised by the librarians. The focus of our work is to provide efficient integration of the help desk service and the relevant background information.

Other approaches and methods can be used for retrieving similar QA-pairs. Conventional TF-IDF indexing, using for example the Java search engine Lucene[8], could yield sufficient results when searching for similar questions. However these methods don’t take into account the semantics of the text, and we want to be able to utilize the semantic relations defined in the common upper ontology YSO.

As for semantic authoring, David Aumuller [2] presents a technique to semantically author Wiki pages. The technique is not just for adding annotations to the pages but also for editing the text. His ideas could be applied in authoring the answers.

This work is a part of the National Semantic Web Ontology Project in Finland (FinnONTO) [9] funded mainly by the Finnish Funding Agency for Technology and Innovation (Tekes).

References

[1] Agnar Aamodt and Enric Plaza. Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Commun.*, 7(1):39–59, 1994.

[2] D. Aumueller. Semantic authoring and retrieval within a wiki, Aug 2005. Demo paper, 2nd European Semantic Web Conference 2005 (ESWC2005).

[3] Guus Schreiber Luit Gazendam, Veronique Malaisé and Hennie Brugman. Deriving semantic annotations of an audiovisual program from contextual texts. In *Semantic Web Annotation of Multimedia (SWAMM’06) workshop*, 2006. http://www.cs.vu.nl/ guus/papers/Gazendam06a.pdf.

[4] Mark van Assem, Maarten R. Menken, Guus Schreiber, Jan Wielemaker, and Bob Wielinga. A method for converting thesauri to rdf/owl. In *Third International Semantic Web Conference ISWC 2004*, volume 3298, 2004.

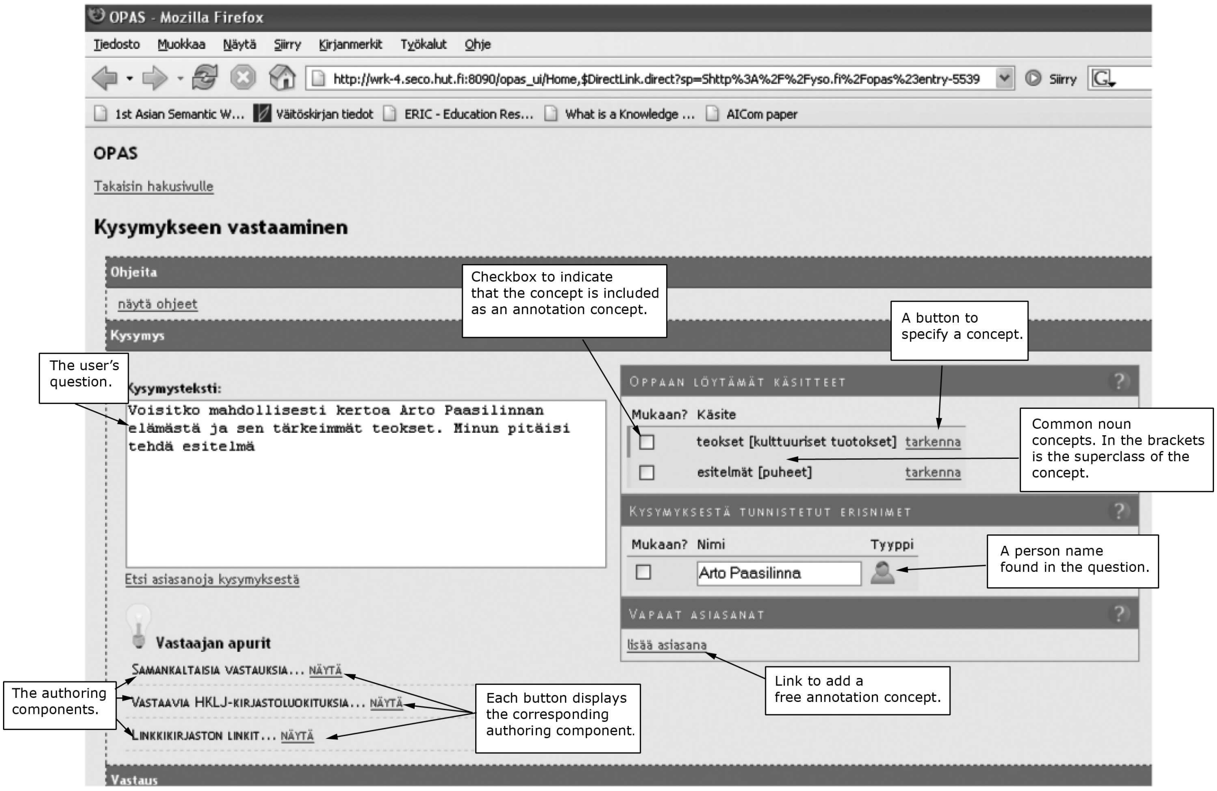
[5] A. Vehvilainen, E. Hyvonen, and O. Alm. A Semi-Automatic Semantic Annotation and Authoring Tool for a Library Help Desk Service. *Proceedings of the first Semantic Authoring and Annotation Workshop, ISWC-2006, Athens, GA, USA*, 2006.

[6] http://hklj.kirjastot.fi/

[7] http://www.w3.org/2004/02/skos/

[8] http://lucene.apache.org

[9] http://www.seco.tkk.fi/projects/finnonto


OPAS - Mozilla Firefox
Tiedosto Muokkaa Näytä Siirry Kirjanmerkit Työkalut Ohje
Siirry
1st Asian Semantic W... Väitöskirjan tiedot ERIC - Education Res... What is a Knowledge ... AICom paper
Takaisin hakusivulle
Ohjeita
näytä ohjeet
Kysymys
Etsi asiasanoja kysymyksestä
Vastaajan apurit
Samankaltaisia vastauksia... näytä
Vastaavia HKLJ-kirjastoluokituksia... näytä
Linkkikirjaston linkit... näytä
Oppaan löytämät käsitteet
Mukaan? Käsite
tarkenna
Kysymyksestä tunnistetut erisnimet
Mukaan? Nimi Tyyppi
Vapaat asiasanat
lisää asiasana
Vastaus

http://wrk-4.seco.hut.fi:8090/opas_ui/Home,$DirectLink.direct?sp=Shttp%3A%2F%2Fyso.fi%2Fopas%23entry-5539

OPAS

Kysymykseen vastaaminen

Kysymysteksti:

Voisitko mahdollisesti kertoa Arto Paasilinnan elämästä ja sen tärkeimmät teokset. Minun pitäisi tehdä esitelmä

☐ teokset [kulttuuriset tuotokset] tarkenna

☐ esitelmät [puheet] tarkenna

☐ Arto Paasilinna

Checkbox to indicate that the concept is included as an annotation concept.

A button to specify a concept.

The user's question.

Common noun concepts. In the brackets is the superclass of the concept.

A person name found in the question.

Link to add a free annotation concept.

The authoring components.

Each button displays the corresponding authoring component.

Figure 1: Question text, concepts found by POKA and similar questions in OPAS UI.