Ontology-Based Query Expansion Widget for Information Retrieval

Jouni Tuominen, Tomi Kauppinen, Kim Viljanen, and Eero Hyvönen

Semantic Computing Research Group (SeCo)
Helsinki University of Technology (TKK) and University of Helsinki
http://www.seco.tkk.fi/
firstname.lastname@tkk.fi

Abstract. In this paper we present an ontology-based query expansion widget which utilizes the ontologies published in the ONKI Ontology Service. The widget can be integrated into a web page, e.g. a search system of a museum catalogue, enhancing the page by providing a query expansion functionality. We have tested the system with general, domain-specific and spatio-temporal ontologies.

1 Introduction

In information retrieval systems the relevancy of search results depends on the user's ability to represent her information needs in a query [1]. If the vocabularies used by the user and the system are not the same ones, or if the shared vocabulary is used in different levels of specificity, the search results are usually poor. Query expansion has been proposed to solve these issues and to improve information retrieval by expanding the query with terms related to the original query terms. Query expansion can be based on corpus, e.g. analyzing co-occurences of terms, or on knowledge models, such as thesauri [2] or ontologies [1]. Methods based on knowledge models are especially useful in cases of short, incomplete query expressions with few terms found in the search index [1, 2].

We have implemented a web widget providing query expansion functionality to web-based systems as an easily integrable service with no need to change the underlying system. The widget uses ontologies to expand the query terms with semantically related concepts. The widget extends the previously developed ONKI Selector widget, which is used for selecting concepts especially for annotation purposes [3].

The user does not have to be familiar with the ontologies used in content annotations by utilizing the autocompletion search feature of the widget, as the system suggests matching concepts as the user is writing the query string. Also, to help the user to disambiguate concepts the ONKI Ontology Browsers [4] can be used to get a better understanding of the semantics of the concepts, e.g. by providing a concept hierarchy visualization.

The query expansion widget supports Semantic web and legacy systems[1], i.e. either the concept URIs or the concept labels can be used in queries. In

[1] By legacy systems we mean systems that do not use URIs as identifiers.

legacy systems cross-language search can be performed, if the used ontology contains concept labels in several languages. In addition to the widget, the query expansion service can also be utilized via JavaScript and Web Service APIs. The query expansion widget and the APIs are available for public use as part of the ONKI Ontology Service[2] [4]. The JavaScript code needed for integrating the widget into a search system can be generated by using the ONKI Widget Generator[3].

The contribution of this paper is to present an approach to perform query expansion in systems cost-effectively, not to evaluate how the chosen query expansion methods improve information retrieval in the systems.

2 Ontologies used for Query Expansion

The ONKI query expansion widget can be used with any ontology published in the ONKI Ontology Service. The service contains some 60 ontologies at the time of writing. Users are encouraged to submit their own ontologies to be published in the service by using the Your ONKI Service[4]. In the following, we describe how we have used different types of ontologies for query expansion.

2.1 Query Expansion with General and Domain-specific Ontologies

For expanding general and domain-specific concepts in queries we have used The Finnish Collaborative Holistic Ontology KOKO[5] which consists of The Finnish General Upper Ontology YSO [5] and several domain-specific ontologies expanding it. To improve poor search results caused by using vocabularies in different levels of specificity in queries and in the search index we have used the transitive is-a relation (*rdfs:subClassOf*[6]) for expanding the query concepts with their subclasses. So for example, when selecting a query concept *publications*, the query is expanded with concepts *magazines*, *books*, *reports* and so on.

Using other relations in addition or instead of the is-a relation in query expansion might be beneficial. When considering general associative relations, caution should be exercised as their use in query expansion can lead to uncontrolled expansion of result sets, and thus to potential loss in precision [6, 7]. In case of a legacy system (not handling URIs, using labels instead) the use of alternative labels of concepts (synonyms) may improve the search. The relations used in the query expansion of an ontology can be configured when publishing the ontology in the ONKI Ontology Service.

[2] `http://www.yso.fi/`
[3] `http://www.yso.fi/onkiselector/`
[4] `http://www.yso.fi/upload/`
[5] `http://www.seco.tkk.fi/ontologies/koko/`
[6] Defined in the RDFS Recommendation, `http://www.w3.org/TR/rdf-schema/`

2.2 Query Expansion with the Spatio-temporal Ontology SAPO

A spatial query can explicitly contain spatial terms (e.g. Helsinki) and spatial relations (e.g. near), but implicitly it can include even more spatial terms that could be used in query expansion [8]. For example, in a query "museums near Helsinki" not only Helsinki is a relevant spatial term, but also its neighboring municipalities. Spatial terms – i.e. geographical places – do not exist just in space but also in time [9, 10]. This is especially true for museum collections where objects have references to places from different times. This sets a requirement to utilize also relations between historical places and more contemporary places in query expansion. To provide these mappings we used a spatio-temporal ontology SAPO (The Finnish Spatio-temporal Ontology) [11].

In SAPO regional overlap mappings are expressed as depicted in Figure 1, where example Turtle RDF[7] statements[8] express that the region of the latest temporal part of place *sapo:Joensuu* — i.e. the one valid from the beginning of year 2009 — overlaps the region of the temporal part of *sapo:Eno* of years 1871–2008. The temporal part of the place simply means the place during a certain time-period such that different temporal parts might have different extensions (i.e. borders) [11].

```
sapo:Joensuu(2009-)
    sapo:begin
            "2009-01-01" ;
    sapo:overlaps
            sapo:Eno(1871-2008) ,
            sapo:Pyhaselka(1925-2008) ,
            sapo:Joensuu(2005-2008) .
```

Fig. 1. Overlap mappings between temporal parts of places.

```
sapo:Joensuu
     sapo:unionof
             sapo:Joensuu(1848-1953) ,
             sapo:Joensuu(1954-2004) ,
             sapo:Joensuu(2005-2008) ,
             sapo:Joensuu(2009-) ;
     sapo:overlapsAtSomeTime
             sapo:Eno ,
             sapo:Pyhaselka ,
             sapo:Tuupovaara ,
             sapo:Pielisensuu ,
             sapo:Kiihtelysvaara .
```

Fig. 2. A place is a union of its temporal parts. Moreover, places may have overlapped other places *at some time*.

For example, the place *sapo:Joensuu* is a union of four temporal parts, defined in the example depicted in Figure 2. However, annotations of items likely utilize places rather than their temporal parts. For this reason the model uses property *sapo:overlapsAtSomeTime* to explicate that e.g. a place *sapo:Joensuu* has — at some point in the history — overlapped together five different places (*sapo:Eno* and four others). In other words, e.g. at least one temporal part of *sapo:Joensuu* has overlapped at least one temporal part of *sapo:Eno*. We have used this more generic property *sapo:overlapsAtSomeTime* between places for query expansion.

[7] `http://www.dajobe.org/2004/01/turtle/`

[8] The example uses the following prefix - sapo: http://www.yso.fi/onto/sapo/

3 A Use Case of the Query Expansion Widget

We have created a demonstration search interface[9] consisting of the original Kantapuu.fi search form[10] and integrated ONKI widgets for query expansion. Kantapuu.fi is a web user interface for browsing and searching for collections of Finnish museums of forestry, using simple matching algorithm of free text query terms with the item index terms. The ontologies used in the query expansion are the same ones as used in annotation of the items[11], namely The Finnish General Upper Ontology YSO, Ontology for Museum Domain MAO[12] and Agfiforest Ontology AFO[13]. For expanding geographical places the Finnish Spatio-temporal Ontology SAPO is used.

When a desired query concept is selected from the results of the autocompletion search of the widget or by using the ONKI Ontology Browser, the concept is expanded. The resulting query expression is the disjunction of the original query concept and the concepts expanding it, formed using the Boolean operation OR. The query expression is placed into a hidden input field, which is sent to the original Kantapuu.fi search page when the HTML form is submitted.

An example query is depicted in Figure 3, where the user is interested in old publications from place Joensuu. User has used the autocompletion feature of the widget to input to the *keywords* field a query term "publicat", which has been autocompleted to the concept *publications*, which has been further expanded to its subclasses (their Finnish labels). Similarly, the place *Joensuu* has been added to the field *place of usage* and expanded with the places it overlaps.

The result set of the search contains four items, from which two are magazines used in place Eno and the rest two are cabinets for books used in place Joensuu. Without using the query expansion the result set would have been empty, as the place *Eno* and the concept *books* were not in the original query.

4 Discussion

When implementing the demonstration search interface for the Kantapuu.fi system with ONKI widgets we faced some challenges. If a query concept has lots of subconcepts, the expanded query string may become inconveniently long, as the concept URIs/labels of the subconcepts are added to the query. This may cause problems because the used HTTP server, database system or other software components may set limits to the length of the query string. With lengthy queries the system may not function properly or the response times of the system may increase.

---

[9] `http://www.yso.fi/kantapuu-qe/`

[10] `http://www.kantapuu.fi/`, follow the navigation link "Kuvahaku".

[11] To be precise, the ontologies are based on thesauri that have been used in annotation of the items.

[12] `http://www.seco.tkk.fi/ontologies/mao/`

[13] `http://www.seco.tkk.fi/ontologies/afo/`

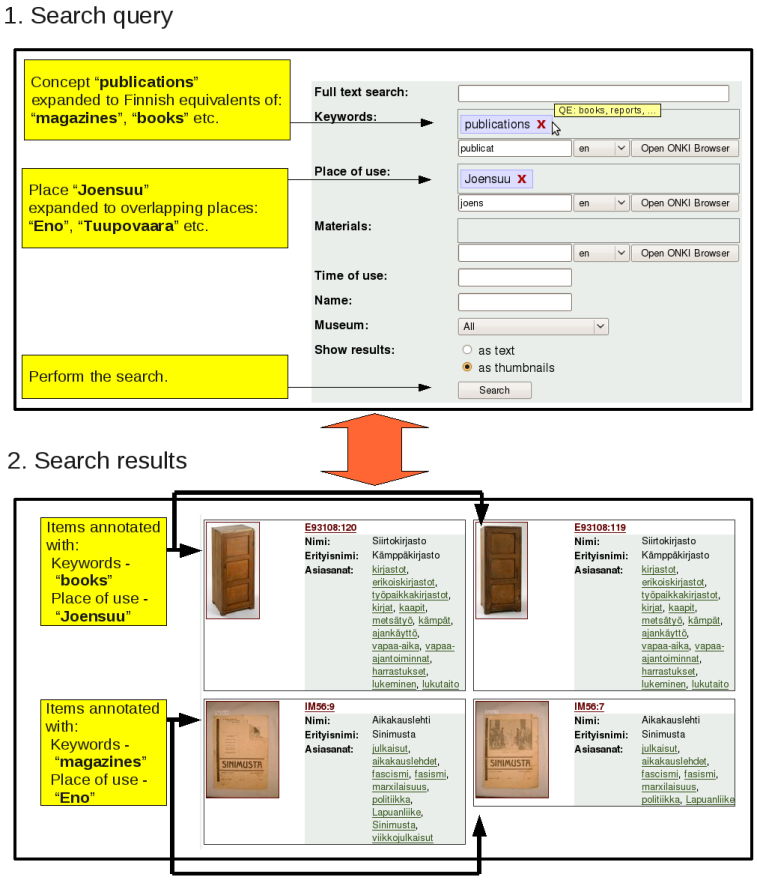

Fig. 3. Kantapuu.fi system with integrated ONKI widgets.

Future work includes user testing for finding out if users consider the query expansion of the concepts and places useful. Also, systematic evaluation of the search systems used would be essential to find out if the query expansion improves the information retrieval, and specifically which semantic relations improve the results the most. The user interface of the query expansion widget needs further developing, e.g., the user should be able to select/unselect the suggested query expansion concepts.

Acknowledgements

We thank Ville Komulainen for his work on the original ONKI server and Leena Paaskoski and Leila Issakainen for cooperation on integrating the ONKI query expansion widgets into the Kantapuu.fi system. This work has been partially

funded by Lusto The Finnish Forest Museum[14] and partially by the IST funded EU project SMARTMUSEUM[15] (FP7-216923). The work is a part of the National Semantic Web Ontology project in Finland[16] (FinnONTO) and its follow-up project Semantic Web 2.0[17] (FinnONTO 2.0, 2008-2010), funded mainly by the National Technology and Innovation Agency (Tekes) and a consortium of 38 private, public and non-governmental organizations.

References

1. Voorhees, E.M.: Query expansion using lexical-semantic relations. In: Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, Dublin, Ireland (July 3-6 1994) 61–69
2. Wang, Y.C., Vandendorpe, J., Evens, M.: Relational thesauri in information retrieval. Journal of the American Society for Information Science **36**(1) (1985) 15–27
3. Viljanen, K., Tuominen, J., Hyvönen, E.: Publishing and using ontologies as mashup services. In: Proceedings of the 4th Workshop on Scripting for the Semantic Web (SFSW 2008), 5th European Semantic Web Conference 2008 (ESWC 2008), Tenerife, Spain (June 1-5 2008)
4. Viljanen, K., Tuominen, J., Hyvönen, E.: Ontology libraries for production use: The Finnish ontology library service ONKI. In: Proceedings of the 6th European Semantic Web Conference (ESWC 2009). (May 31 - June 4 2009)
5. Hyvönen, E., Viljanen, K., Tuominen, J., Seppälä, K.: Building a national semantic web ontology and ontology service infrastructure—the FinnONTO approach. In: Proceedings of the 5th European Semantic Web Conference (ESWC 2008). (June 1-5 2008)
6. Tudhope, D., Alani, H., Jones, C.: Augmenting thesaurus relationships: Possibilities for retrieval. Journal of Digital Information **1**(8) (2001)
7. Hollink, L., Schreiber, G., Wielinga, B.: Patterns of semantic relations to improve image content search. Journal of Web Semantics **5**(3) (2007) 195–203
8. Fu, G., Jones, C.B., Abdelmoty, A.I.: Ontology-based spatial query expansion in information retrieval. In: In Lecture Notes in Computer Science, Volume 3761, On the Move to Meaningful Internet Systems: ODBASE 2005. (2005) 1466–1482
9. Kauppinen, T., Hyvönen, E.: Modeling and reasoning about changes in ontology time series. In Kishore, R., Ramesh, R., Sharman, R., eds.: Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems. Integrated Series in Information Systems, New York, NY, Springer-Verlag, New York (NY) (January 15 2007) 319–338
10. Jones, C., Abdelmoty, A., Fu, G.: Maintaining ontologies for geographical information retrieval on the web. Volume 2888., Sicily, Italy, Springer Verlag (November 2003) 934–951
11. Kauppinen, T., Väätäinen, J., Hyvönen, E.: Creating and using geospatial ontology time series in a semantic cultural heritage portal. In: S. Bechhofer et al.(Eds.): Proceedings of the 5th European Semantic Web Conference 2008 ESWC 2008, LNCS 5021, Tenerife, Spain. (June 1-5 2008) 110–123

[14] `http://www.lusto.fi`
[15] `http://smartmuseum.eu/`
[16] `http://www.seco.tkk.fi/projects/finnonto/`
[17] `http://www.seco.tkk.fi/projects/sw20/`