SemSearch: Refining Semantic Search

Victoria Uren, Yuangui Lei, and Enrico Motta

Knowledge Media Institute, The Open University,
Milton Keynes, MK7 6AA, UK
{y.lei,e.motta,v.s.uren}@ open.ac.uk

Abstract. We demonstrate results presentation and query refinement functions of the SemSearch engine for semantic web portals and intranets.

Keywords: Semantic search, query refinement, semantic intranet.

1 Introduction

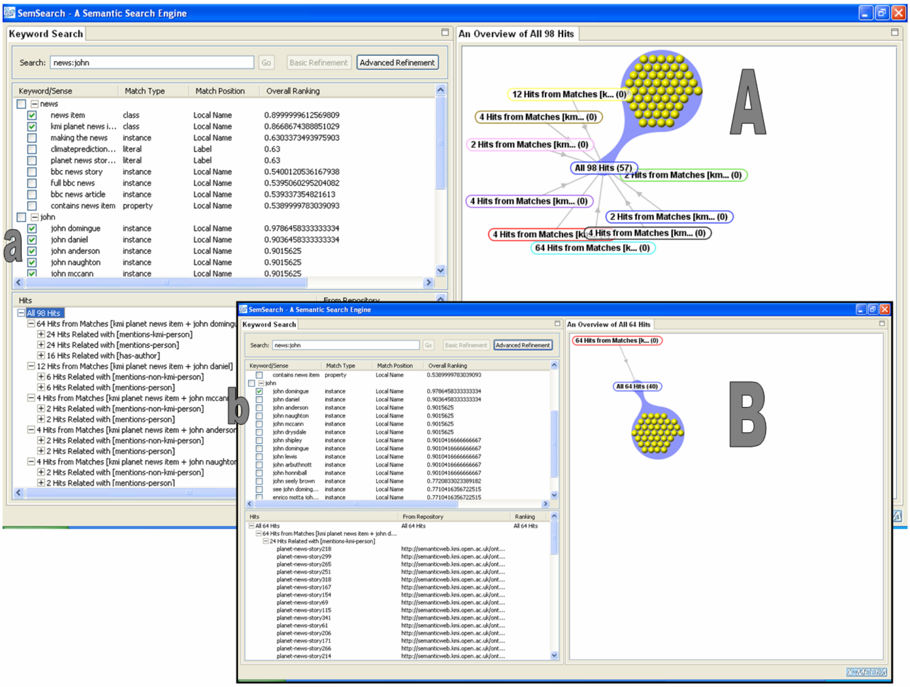
SemSearch is a search engine for RDF knowledge bases [1] [2]. The driving factor in its design is to make the formulation of semantic queries straightforward for users who may not know the details of the ontology underlying the knowledge base. To achieve this, it has a query translation engine which takes keyword input and translates it into formal semantic queries.

The issue of the usability of semantic search systems is being addressed actively. In a recent review [3], we identified four main query modes: keyword, form, visual and natural language systems. We have taken keyword systems as the stepping off point because we believe they are the most familiar to users. This requires a "translation" process that converts keyword input to formal, semantic queries. Our approach to automatic query formulation from keywords is closest to those proposed in [4] and [5]. However, in many cases, users need to refine their searches to get closer to the results they want, a fact widely acknowledged by IR researchers, e.g. [6] and [7]. To achieve this, they need both clear presentations of the results they have so far and mechanisms to constrict, or alternatively broaden, their search. Our hypothesis is that different query modes come into their own at different stages of the search refinement cycle. In this demo, we focus on recent developments in the results presentation and query refinement facilities of SemSearch that exploit several query formulation modes.

SemSearch is intended for two kinds of scenario. The first is the relatively familiar scenario of semantic web portals; the example in figure 1 is taken from the KMi Semantic Web Portal[1]. The second scenario could be called "semantic intranets", i.e., intranets in which organizations employ semantic web technology to give access to heterogeneous resources which may be scattered across an organization's servers on different sites. Both scenarios would be expected to employ a relatively small number of known and trusted ontologies. However, in the case of semantic intranets the volume of resources covered may be very large scale.

[1] http://semanticweb.kmi.open.ac.uk:8080/ksw/index.html

S. Bechhofer et al.(Eds.): ESWC 2008, LNCS 5021, pp. 874–878, 2008.


SemSearch - A Semantic Search Engine
Keyword Search
An Overview of All 98 Hits
12 Hits from Matches [k... (0)
4 Hits from Matches [km... (0)
2 Hits from Matches [km... (0)
All 98 Hits (57)
2 Hits from Matches [km... (0)
4 Hits from Matches [km... (0)
2 Hits from Matches [km... (0)
4 Hits from Matches [k
4 Hits from Matches [km... (0)
64 Hits from Matches [k... (0)
A
a
b
An Overview of All 64 Hits
64 Hits from Matches [k... (0)
All 64 Hits (40)
B
Go
Basic Refinement
Advanced Refinement

Search: news:john

	Keyword/Sense	Match Type	Match Position	Overall Ranking
□	news			
☑	news item	class	Local Name	0.8999999612569809
☑	kmi planet news i…	class	Local Name	0.8668674388851029
□	making the news	instance	Local Name	0.6303373493975903
□	climateprediction…	literal	Label	0.63
□	planet news stor…	literal	Label	0.63
□	bbc news story	instance	Local Name	0.5400120536167938
□	full bbc news	instance	Local Name	0.5395060295204082
□	bbc news article	instance	Local Name	0.539337354821613
□	contains news item	property	Local Name	0.5389999783039093
□	john			
☑	john domingue	instance	Local Name	0.9786458333333334
☑	john daniel	instance	Local Name	0.9036458333333334
☑	john anderson	instance	Local Name	0.9015625
☑	john naughton	instance	Local Name	0.9015625
☑	john mccann	instance	Local Name	0.9015625

Hits

- All 98 Hits
  - 64 Hits from Matches [kmi planet news item + john domingu
    - 24 Hits Related with [mentions-kmi-person]
    - 24 Hits Related with [mentions-person]
    - 16 Hits Related with [has-author]
  - 12 Hits from Matches [kmi planet news item + john daniel]
    - 6 Hits Related with [mentions-non-kmi-person]
    - 6 Hits Related with [mentions-person]
  - 4 Hits from Matches [kmi planet news item + john mccann
    - 2 Hits Related with [mentions-non-kmi-person]
    - 2 Hits Related with [mentions-person]
  - 4 Hits from Matches [kmi planet news item + john anderson
    - 2 Hits Related with [mentions-non-kmi-person]
    - 2 Hits Related with [mentions-person]
  - 4 Hits from Matches [kmi planet news item + john naughton
    - 2 Hits Related with [mentions-non-kmi-person]
    - 2 Hits Related with [mentions-person]

Search: news:john

	Keyword/Sense	Match Type	Match Position	Overall Ranking
□	contains news item	property	Local Name	0.5389999783039093
□	john			
☑	john domingue	instance	Local Name	0.9786458333333334
□	john daniel	instance	Local Name	0.9036458333333334
□	john anderson	instance	Local Name	0.9015625
□	john naughton	instance	Local Name	0.9015625
□	john mccann	instance	Local Name	0.9015625
□	john drysdale	instance	Local Name	0.9015625
□	john shipley	instance	Local Name	0.9010416666666667
□	john domingue	instance	Local Name	0.9010416666666667
□	john lewis	instance	Local Name	0.9010416666666667
□	john arbuthnott	instance	Local Name	0.9010416666666667
□	john honniball	instance	Local Name	0.9010416666666667
□	john seely brown	instance	Local Name	0.7720833023389182
□	see john doming…	instance	Local Name	0.7710416356722515
□	enrico motta ioh…	instance	Local Name	0.7710416356722515

Hits	From Repository	Ranking
All 64 Hits	All 64 Hits	All 64 Hits
64 Hits from Matches [kmi planet news item + john d…		
24 Hits Related with [mentions-kmi-person]		
planet-news-story218	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story299	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story265	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story251	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story318	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story167	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story154	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story69	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story115	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story341	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story61	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story206	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story171	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story266	http://semanticweb.kmi.open.ac.uk/ont…	
planet-news-story214	http://semanticweb.kmi.open.ac.uk/ont…	

Fig. 1. Results presentation in SemSearch for the search "news: john" showing: a) the list of entities automatically selected by SemSearch, A) ClusterMap results visualisation for the search, at a glance it is messy, too many Johns have been selected, b) the form interface is used to reduce the list to just john-domingue, B) with a clear change to the visualisation the user can judge whether the search refinement has had the desired effect.

2 System Details

Implementation. SemSearch is a Java application implemented using the Eclipse[2] rich client platform. The query translation function is also implemented as a web service that can be built into web portals. Lucene[3] is used to make an index of the textual parts of the RDF knowledge base. This is exploited by the query interpretation engine, which generates formal SeRQL queries that are sent to the Sesame[4] query engine. Results are visualized using the Aduna Cluster Map[5] tool.

Query Translation. The query translation engine finds out the semantic meanings of the keywords specified in user queries and composes multiple senses into formal queries. This process is described in detail elsewhere [1],[2]. Here, we provide a summary to assist understanding of other parts of the description.

[2] Eclipse http://www.eclipse.org/

[3] Lucene http://lucene.apache.org/

[4] Sesame http://www.openrdf.org/

[5] Cluster Map http://www.aduna-software.com/technologies/clustermap/overview.view

From the semantic point of view, a keyword entered by the user may match i) general concepts (e.g., the keyword "news" which matches the concept news-item), ii) semantic relations between concepts, (e.g. the keyword "author" matches the relation has-author), or iii) instance entities (e.g., the keyword "Yuangui" which matches the instance Yuangui-Lei). The system exploits the Lucene text search to match the keyword against the indexes of local names, labels, and short literal values built with Lucene. This process can produce a number of matches for each input keyword. These matches must be assembled into SeRQL queries. In the example shown in Figure 1, the keyword "news" has eight possible matches and the keyword "phd students" has three, giving twenty four possible pairings for which queries need to be constructed. SemSearch ranks the queries and initially searches the high ranking ones. In the example, "news" matches two classes, which are selected, ignoring the instances. But "john" matches only instances, ten of which rank high enough to be selected. The current ranking process is described in detail elsewhere [1].

Presenting Results. Semantic search systems provide access both to the RDF data itself, as triples, and to documents with RDF annotations. Summary results are particularly helpful at the search refinement stage when the user needs to determine whether and in what way to modify their query. The most basic summary result presentation is the ranked listing. SemSearch provides such a listing using the same ranking method it applies for selecting queries. The RDF results can also be summarized using Cluster Map (shown on the right in figure 1). This generates a graphical representation that shows how many occurrences of each entity have been found and whether they are linked to other entities found in the search. The resulting visualizations give rapid insight into which entities dominate the results. In particular, it can indicate that "something is wrong" much faster than scanning a results listing or reading through documents. In our example, it indicates that too many instances have been included in searches. Finally, the user may need to see the original documents from which RDF annotations were derived. Again, a ranking is provided to let them see the documents that best represent each entity first.

By examining the results presented in these ways the user can determine whether they have found a satisfactory answer to their query. If they are not satisfied, the presentations should provide them with the clues they need to proceed to the next step: query refinement.

Query Refinement. One way a user can refine their SemSearch query is by changing the selection of matching entities using a simple form. The close matches are presented as a list, with the entities automatically selected by SemSearch ticked (shown in figure 1 in the top left panel of the interface). By selecting and deselecting terms in the list a different selection of semantic queries can be run. Motivations for this kind of refinement include removing spurious matches, which contain the same string as the query but are not interesting (the extra "john" instances in our example), and including matches that were not selected but look promising to the user.

Another means of query refinement tackles the case in which a user's initial query has been pitched at the wrong conceptual level and they need to narrow down or broaden the search scope. This is a case where the hierarchical organization of ontologies gives semantic search systems a real advantage over pure text search in which the user has to keep guessing at appropriate new keywords until they get the results they want. A standard approach is to use an ontology browser. However, we wished to avoid this in SemSearch because 1) it can be more information than the user requires, 2) we plan to extend SemSearch for use with multiple ontologies and so need an approach that will scale up. Therefore, we have developed a visual query formulation function in which the user is shown only the immediate neighbourhood of an entity as an interactive graph. Working from this fragment the user can navigate up or down through the hierarchy to find the right level.

3 Work in Progress

We are continually improving SemSearch and adding new functionality. One open issue currently under investigation is semantic ranking. This topic is being actively researched by the semantic web community. Methods are being developed for ranking whole ontologies or RDF/OWL documents, e.g. [8], [9] and [10], for associations between semantic entities, e.g., [11] and [12], and for query results, e.g. [13]. For SemSearch, we are presently investigating alternative algorithms for ranking matches to semantic entities and combinations of matches in formal queries generated by the query translation engine. Currently, the matches are ranked using an algorithm that exploits the similarity, the domain context, and the query context factor. Improved rankings are also needed for annotations found as results, and we are investigating methods for this.

We have demonstrated how SemSearch makes keyword, form and visual search modes available at different stages of the query refinement process. Future work will extend the use of multiple, complementary search modes demonstrated in the current prototype.

Acknowledgement

This work funded by the X-Media project (www.x-media-project.org) sponsored by the European Commission as part of the Information Society Technologies (IST) programme under EC grant number IST-FP6-026978.

References

1. Lei, Y., Uren, V., Motta, E.: SemSearch: a search engine for the semantic web. In: Staab, S., Svátek, V. (eds.) EKAW 2006. LNCS (LNAI), vol. 4248, pp. 238–245. Springer, Heidelberg (2006)
2. Lei, Y., Lopez, V., Motta, E., Uren, V.: An Infrastructure for Building Semantic Web Portals. Journal of Web Engineering 6(4), 283–308 (2007)
3. Uren, V., Lei, Y., Lopez, V., Liu, H., Motta, E., Giordanino, M.: The usability of semantic search tools: a review. Knowledge Engineering Review 22, 361–377 (2007)
4. Tran, T., Cimiano, P., Rudolph, S., Studer, R.: Ontology-Based Interpretation of Keywords for Semantic Search. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ISWC 2007. LNCS, vol. 4825, pp. 523–536. Springer, Heidelberg (2007)

5. Zhou, Q., Wang, C., Xiong, M., Wang, H., Yu, Y.: SPARK: Adapting Keyword Query to Semantic Search. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ISWC 2007. LNCS, vol. 4825, pp. 694–707. Springer, Heidelberg (2007)
6. Belkin, N.J., Cool, C., Kelly, D., Lin, S.-J., Park, S.Y., Perez-Carballo, J., Sikora, C.: Iterative exploration, design and evaluation of support for query reformulation in interactive information retrieval. Information Processing and Management 37(3), 403–434 (2005)
7. White, R.W., Kules, B., Bederson, B.: Exploratory Search Interfaces: Categorization, Clustering and Beyond. Report on the XSI 2005 Workshop at the Human-Computer Interaction Laboratory, University of Maryland. ACM SIGIR Forum 39(2), 52–56 (2005)
8. Ding, L., Pan, R., Finin, T., Joshi, A., Peng, Y., Kolari, P.: Finding and ranking knowledge on the semantic web. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729, pp. 156–170. Springer, Heidelberg (2005)
9. Alani, H., Brewster, C.: Ontology Ranking based on the analysis of concept structures. In: Proceedings of the 3rd International Conference on Knowledge Capture 2005 (K-CAP 2005), pp. 51–58. ACM Press, New York (2005)
10. Hogan, A., Harth, A., Decker, S.: ReConRank: A Scalable Ranking Method for Semantic Web Data with Context. In: 2nd International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2006), Athens, GA, USA, November 5 (2006)
11. Aleman-Meza, B., Halaschek-Wiener, C., Arpinar, I.B., Ramakrishnan, C., Sheth, A.P.: Ranking Complex Relationships on the Semantic Web. IEEE Internet Computing 9(3), 37–44 (2005)
12. Anyanwu, K., Maduko, A., Sheth, A.P.: SemRank: ranking complex relationship search results on the semantic web. In: Proceedings of the 14th international conference on World Wide Web, pp. 117–127. ACM Press, New York (2005)
13. Stojanovic, N., Maedche, A., Staab, S., Studer, R., Sure, Y.: SEAL: a framework for developing SEmantic PortALs. In: Proceedings of the 1st International Conference on Knowledge Capture (K-CAP 2001), pp. 155–162. ACM Press, New York (2001)