Using SKOS vocabularies for improving Web Search

Bernhard Haslhofer
University of Vienna
Computer Science
haslhoferb@acm.org

Flávio Martins
Universidade Nova de Lisboa
Fac. Ciências e Tecnologia
flaviomartins@acm.org

João Magalhães
Universidade Nova de Lisboa
Fac. Ciências e Tecnologia
jm.magalhaes@fct.unl.pt

ABSTRACT

Knowledge organization systems such as thesauri or taxonomies are increasingly being expressed using the Simple Knowledge Organization System (SKOS) and published as structured data on the Web. Search engines can exploit these vocabularies and improve search by expanding terms at query or document indexing time. We propose a SKOS-based term expansion and scoring technique that leverages labels and semantic relationships of SKOS concept definitions. We also implemented this technique for Apache Lucene and Solr. Experiments with the Medical Subject Headings vocabulary and an early evaluation with Library of Congress Subject Headings indicated gains in precision when using SKOS-based expansion compared to pseudo relevance feedback and no expansion. Our findings are important for publishers and consumer of Web vocabularies who want to use them for improving search over Web documents.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: [retrieval models, query formulation, search process]; H.3.1 [**Content Analysis and Indexing**]: [thesauri]

General Terms

Algorithms, Experimentation, Performance

Keywords

Search, Query Expansion, Thesaurus, SKOS, Linked Data

1. INTRODUCTION

The Simple Knowledge Organization System (SKOS) [9] has evolved as the de-facto standard for expressing and publishing knowledge organization systems such as thesauri, classification schemes, and taxonomies as structured data on the Web. A recent study [8] has shown that hundreds of published SKOS vocabularies can now be found online, including examples such as the United Nations AGROVOC thesaurus[1], the Library of Congress subject headings [11], the New York Times authoritative news vocabularies[2], or the UK Archival Thesaurus (UKAT)[3].

Web vocabularies are typically domain-specific and are maintained by expert users, some of them over a long term. Often they were created with the intent to facilitate information organization and retrieval within organizations and their closed information systems. Since these vocabularies are now being expressed in SKOS and published on the Web, they could easily be used by search engines to improve search and retrieval within and across organizational boundaries.

Term expansion is one possible retrieval technique that can benefit from public accessibility of structured SKOS vocabularies. Applied at query-time usually (*query expansion*), it deals with the problem that real-world concepts are referred to using different terms and expands terms by labels associated with SKOS concept definitions (*label-based expansion*). An information retrieval system can help users and also automatically refine their queries by exploiting the semantic relationships between terms. If, for instance, the system is aware that the concept "Weapons" is semantically narrower than "Military Equipment" or that "Armaments" and "Arms" are different terms denoting the same concept (synonyms) for "Weapons", it could expand query terms to improve retrieval effectiveness.

While query expansion is a well-known technique [3] and many modern IR systems already support it using thesauri such as WordNet, existing approaches do not yet take into account existing Web-based vocabularies and are not tailored to the specifics of SKOS, which distinguishes between the notion of URI identifiable *concepts* and associated *terms*. We propose a more general term expansion technique that supports *label-* and *URI-based* term expansion at query or document indexing-time. Our contributions are:

- A SKOS-based term expansion and scoring technique.
- An open source implementation of this technique for Apache Lucene and Solr.

We also report results obtained from evaluating our approach for query expansion using the MeSH and LCSH vocabularies. They show that SKOS vocabularies can improve search. Our implementation is publicly available at `https://github.com/behas/lucene-skos`.

[1] `http://aims.fao.org/standards/agrovoc/about`
[2] `http://data.nytimes.com/`
[3] `http://www.ukat.org.uk/`

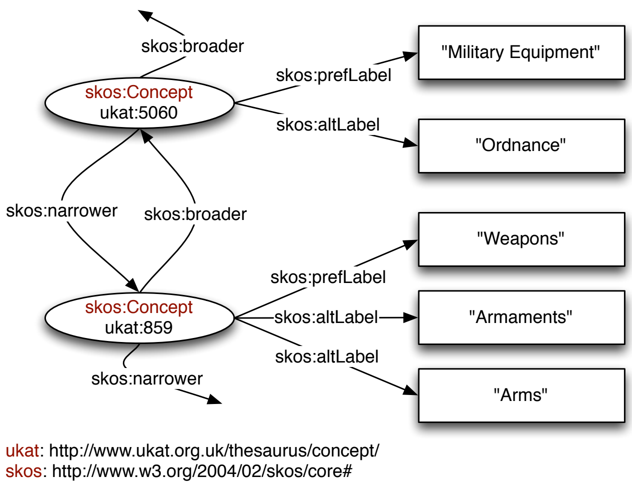

Figure 1: SKOS Example—UKAT Thesaurus.

2. BACKGROUND

We now provide a brief introduction to SKOS and then discuss related work in the area of term expansion.

2.1 SKOS

The Simple Knowledge Organization System (SKOS) [10] is a model for expressing structured vocabularies (classification schemes, thesauri, taxonomies, etc.) on the Web. Figure 1 shows two example concept definitions taken from the UKAT thesaurus: one defines the concept "Weapons" and two synonyms ("Armaments, Arms"), the other represents the concept "Military Equipment" and provides a single synonym definition "Ordnance". Since "Military Equipment" is semantically broader than "Weapons" and vice-versa, the example also includes broader and narrower relationships between these concepts.

Concepts are a fundamental SKOS feature. Each has a unique URI identification and, optionally, a set of properties declaring the preferred indexing term (skos:prefLabel), alternative or synonym terms (skos:altLabel), and semantic relationships (e.g., skos:broader, skos:narrower). The use of broader and narrower relationships between different concepts creates a taxonomy-like structure of conceptual resources. For further details on available SKOS model features, we refer to its online documentation[4]

SKOS is an application of the Resource Description Framework (RDF) and allows vocabularies to be published as dereferenceable resources on the Web, which makes them easy to retrieve and reuse in applications. It plays an important role in ongoing Linked Data activities [4] where vocabularies are published on the Web and connected to other vocabularies or documents via concept URI links. They can easily be consumed by other applications by dereferencing vocabulary URIs, or, in many cases, by downloading vocabulary dumps.

2.2 Term Expansion

Most existing work focuses on label-based expansion at query time (query expansion). Therefore, we proceed to focus on that technique while discussing related work.

Typical query expansion techniques [3] involve taking advantage of statistical correlation, synonyms, knowledge of the morphology of words, and the use of dictionaries to improve the quality of search results. Voorhees [14] found that automatic expansion using synonym sets from WordNet can degrade performance, while hand picking concepts improves short or ambiguous queries. Hersh et al. [5] described OHSUMED, a medical test collection with judgments, based on journals from the National Library of Medicine (NLM). Later in [6], Hersh et al. expanded all child terms in this test collection because journals in OHSUMED are indexed using the narrowest indexing terms. They also found that simply adding the new expansions to the query is a simple yet effective approach.

Broader techniques have been proposed since to cover more formal descriptions of domain-specific thesauri. This field has been quite active as is described by Bhogal et al. [2] who reviewed several ontology-based query expansion techniques. Trying to balance between simple thesauri and ontology representations, the Simple Knowledge Organization System (SKOS) [9] is a standard model for sharing and linking controlled vocabularies (thesauri, taxonomies, classification systems, etc.) on the Web. It defines a set of classes and properties to represent vocabularies as Web graphs. The fundamental building block of a SKOS vocabulary is the *Concept*. Each *Concept* has an unique URI identifier and associated preferred indexing terms, alternative terms, and semantic relations. Thus, SKOS provides a mechanism to fully encode domain-specific thesauri.

When expanding a query term, the relationships among terms also play a key role in the process. Bai et al. [1] confirmed the importance of term-relationships expansion, and also proposed to expand queries based on the relationship between a set of terms and a single term. Thus, SKOS also provides multiple types of relationships to better assist in the query-expansion process: the *skos:altLabel* property is used to declare additional synonyms, abbreviations and acronyms for the concept. The use of broader-narrower relationships between concepts creates a hierarchically organized graph of conceptual resources. In addition, the property *skos:related* can be used to assert a non-hierarchical, associative relationships between two concepts.

More recent approaches, have mastered several techniques for integrating thesaurus-like resources in query expansion. Lin and Demner-Fushman [7] and Zhou et al. [15] investigated the creation of custom similarities to score concept matches with more emphasis than word matches. Zhou et al. used the MeSH thesaurus with one hierarchical level of relationships and synonyms. In our case, we provide support for multiple levels (MeSH SKOS has 12 hierarchical levels) and multiple relation categories. Theobald et al. [12] developed a system that performs expansion only when thematic similarity is above a threshold, but found that it implies a high execution cost and hard-tuning of parameters.

Besides label-based expansion at query time, our approach also supports URI-based expansion. This is important when documents contain explicit links to SKOS concepts, which is frequently the case in the Linked Data Web. If documents contain links to SKOS concepts, expansion should be performed at indexing-time. Although this is possible with our technique, our experiments deal with query-time term expansion only due to lack of benchmarks for indexing-time evaluation.

[4]SKOS Primer: `http://www.w3.org/TR/skos-primer/`

Title	Spearhead
Description	Roman iron spearhead. The spearhead was attached to one end of a wooden shaft...
Subject	Weapons

Figure 2: Example Metadata Description with label-based subject property.

3. SKOS-BASED TERM EXPANSION

A search engine can load one or more SKOS vocabularies from the Web and expand terms in search queries or documents based on known concept definitions and semantic relationships. Before we discuss possible expansion techniques, consider Figure 2, which shows an example metadata description (document) about an image of a roman iron spearhead with a label-based subject property "Weapons".

If a user, who is searching for images of weapons from the period of the roman empire, enters the terms "roman spearhead" or "roman weapon" the described image will be found because of matching terms in the query and the document. However, if the user enters a query "roman arms", the image will be ranked lower because of the lexical mismatch between the terms "arms" and "weapons".

Web vocabularies, such as the UK Archival Thesaurus, formalize knowledge about concepts in a certain domain and define, for instance, that the term "arms" and the term "weapons" are synonyms denoting the same real-world concept. A search engine can use that kind of knowledge and expand queries or documents with additional terms, causing additional documents to match queries, which would not match previously.

In the following, we describe two SKOS-aware expansion techniques: *label-based* and *URI-based* expansion. Subsequently, we describe how SKOS relationship types can be considered in *scoring* models and an *implementation* of these techniques for the widely used full-text search engine library Apache Lucene.

3.1 Label-based Expansion

Term expansion requires SKOS vocabularies to be loaded by the search engine. This can be performed either by dereferencing the URIs of concept definitions and following links to related concepts (crawling) or downloading and expanding packaged vocabularies (dump load) from a given URI.

Label-based expansion can be applied on tokenized queries and documents before any stemming or lemmatization is performed. For each token the SKOS expander looks up matching concepts in a system-internal vocabulary representation, which could be an inverted index having SKOS labels in the dictionary and concept URIs in the posting list. Labels of matching concepts (*expansion terms*) are then added to the internal query or document representation.

Each expansion term carries an attribute indicating the type of SKOS property (e.g., skos:altLabel) that caused the expansion. This allows domain experts to fine-tune the influence of expanded terms on scoring of query results according to the needs in their information retrieval scenario. Figure 3 shows how the query "roman weapons" can be expanded and normalized in order to match terms in the document shown in Figure 2.

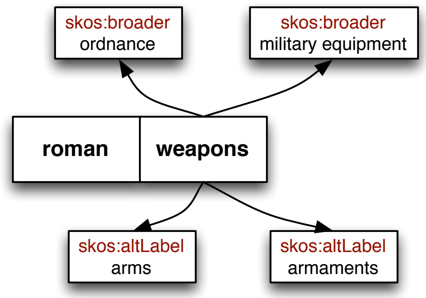

Figure 3: SKOS Label-based (Query) Expansion example

Title	Spearhead
Description	Roman iron spearhead. The spearhead was attached to one end of a wooden shaft...
Subject	`http://www.ukat.org.uk/thesaurus/concept/859`

Figure 4: Example Metadata Description with URI-based subject property.

Label-based expansion is typically implemented at query time and also known as *query expansion*, which is a widely implemented technique in modern information retrieval systems. This has the advantage that term expansion is performed on the fly without affecting the underlying index size and index statistics. If vocabularies are small, they can be loaded into memory to minimize concept lookup effort at query time. However, in certain cases, e.g., when vocabularies are large, it is desirable to perform label-based expansion at document indexing-time. This increases the size of the index, but might reduce time costs at query time.

3.2 URI-based Expansion

SKOS distinguishes between the notion of *concepts* and *terms* and thereby, different from other models, reduces semantic ambiguities. As a consequence, documents might feature explicit links to SKOS concept URIs instead of using label-based terms. Figure 4 shows the same metadata description as in Figure 2 but links to the SKOS concept representing "Weapons" in the subject property. At document indexing time, the SKOS expander looks up term definitions and semantically related concepts and expands URIs the same way it expands labels. Figure 5 shows how the URI `http://www.ukat.org.uk/thesaurus/concept/859` is expanded by SKOS definition.

This technique is useful when indexed documents are structured, as it is in the case with metadata descriptions, and when it is known that certain properties (e.g., subject) carry SKOS URIs instead of labels, which is frequently the case in Linked Data sources. Search engines can then be configured to perform URI-based expansion on these properties, and potentially label-based expansion on others.

Technically, this technique could also be applied at query-time but since users typically use natural language terms instead of URIs in their queries, it is mostly used at document indexing-time.

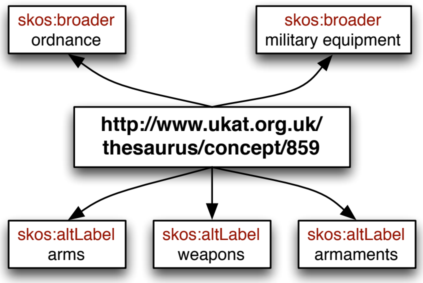

Figure 5: SKOS URI-based (Query) Expansion example

3.3 Scoring

Term expansion can lead to a large number of terms added to the original query or document. While this should generally increase retrieval effectiveness, it might also cause *query drift*, which means that an expanded query does not reflect the user's initial information need anymore. Instead of applying expensive pruning methods, our proposed scoring model leverages the explicit declaration of expansion types in term representations. This allows the configuration of the influence of certain SKOS definitions on the final score.

Our approach works with regular text retrieval functions, which are minimally modified to enable scoring based on expansion types. The following function shows our scoring method as applied to the *tf–idf* retrieval model. The score of a query in relation to a given document can be succinctly conveyed as follows, assuming that query terms are expanded:

$$Score(q,d) = coord_{q,d} \cdot \left(\sum_{t \in q} (tf_{t,d} \cdot idf_t) + \sum_{c \in expansions(q)} (tf_{c,d} \cdot idf_c \cdot boost_{c_{type}}) \right)$$

$expansions(q)$ denotes the set of expansions found for query q, $coord_{q,d}$ is a score factor calculated based on the number of terms matching the document and $boost_{c_{type}}$ is a boost factor that varies with the expansion type of c.

With the $coord_{q,d}$ factor, we ensure that a document with more matching query terms will score higher. The score for documents that match more alternative terms, broader, narrower or related concepts resulting from the expansion of the original query will also be higher.

Since not all query terms are equally important, especially when they were not explicitly entered by the user, as it is the case with query expansion, original query terms and terms resulting from expansions have to be weighted differently. This can be controlled by the boost factor $boost_{c_{type}}$. It balances the contributions of the original and expanded query terms to the final score and allows flexible weighting schemes based on expansion types. One could, for instance, weight *skos:related* expansion types less than synonym definitions (*skos:altLabel*). An optimal setup is a combination of boost values, one for each SKOS expansion type.

Similar scoring strategies can be applied for URI-based term expansion and other retrieval models. We implemented expansion-type aware scoring for BM25 also but omit the description of this model since it is analogous.

3.4 Implementation

We implemented SKOS-based term expansion in an *analyzer* module for Apache Lucene[5] 3.x and Solr[6] 3.x and made it available as open source at `https://github.com/behas/lucene-skos`. Analyzers in Lucene tokenize field text into its most fundamental indexed representation, terms. They can be combined to perform any number of operations on text, which could include text normalization, stop word removal, stemming, and lemmatization.

Our analyzer implementation expands terms by matching SKOS concept definitions and keeps track of expansion types. it is configurable w.r.t the used vocabulary and expansion strategies applied on given document fields. One could for instance, use label-based expansion for textual fields and URI-based expansion for fields linking to SKOS URIs, similar to the example in Figure 4.

4. EVALUATION

We evaluated our SKOS-based term expansion approach using two dataset bundles: (i) the modified **OHSUMED** dataset from TREC-9 in combination with the Medical Subject Headings Vocabulary (**MeSH**), and (ii) a set of metadata records and queries from the Library of Congress (**LoC**) in combination with the Library of Congress Subject Headings (**LCSH**). Since these datasets contain label-based indexing terms, we focused our experiments on label-based expansion at query-time. We normalized queries and SKOS vocabularies (stop word removal, lowercase filtering, whitespace and punctuation removal) and expanded each query term t by collecting concepts that have t in any of their labels (*skos:prefLabel*, *skos:altLabel*, *skos:hiddenLabel*). We then added these terms to the initial query and applied predefined boost weights. The boost weights selected were the ones that maximized performance metrics. We performed the retrieval experiments for various boost values in intervals of 0.05 to find the best result for each metric.

We compared the retrieval effectiveness of our proposed approach (**SKOS**) against two baselines:

- **No term expansion (NoExp)**: queries are executed over documents (metadata descriptions) without SKOS-based term expansion.
- **Pseudo Relevance Feedback (PRF)**: an initial search is performed with the original query and terms are selected from the top k retrieved documents. A new query is then resubmitted with the original query terms and the expansion terms collected from the k-retrieved documents. This expands the query with in-corpus terms which co-occur with the query terms.

We believe that these baselines are significant because they are two real-world alternatives to relying on domain knowledge represented in the SKOS: systems can either apply no expansion (NoExp) or expand based on statistical characteristics of the indexed data.

For each dataset bundle, we measured the overall performance of our approach by averaging performance over the

[5]`http://lucene.apache.org/core/`
[6]`http://lucene.apache.org/solr/`

	LTC TF-IDF						BM25					
	P@1	P@3	P@10	nDCG@1	nDCG@3	nDCG@10	P@1	P@3	P@10	nDCG@1	nDCG@3	nDCG@10
NoExp	0.333	0.302	0.260	0.407	0.376	0.356	0.381	0.344	0.265	0.450	0.414	0.374
PRF	0.322	0.282	**0.302**	0.379	0.360	**0.393**	0.377	0.317	0.275	0.443	0.397	0.369
SKOS	**0.419**	**0.366**	0.276	**0.484**	**0.429**	0.379	**0.500**	**0.366**	**0.282**	**0.548**	**0.435**	**0.397**

Table 1: Precision and nDCG results on the OHSUMED dataset.

query set provided with each dataset (63 OHSUMED/MESH, 10 LoC/LCSH). We focused on two measures, precision at rank n (P@n) for both dataset bundles and nDCG at rank n (nDCG@n) for the OHSUMED/MESH bundle, which provides ordinal relevance judgments. Two retrieval models were used for ranking documents: LTC TF-IDF and BM25.

4.1 Dataset Characteristics

The **OHSUMED/MeSH** bundle comprises a subset of the PubMed database with about 350K metadata records (documents) from 270 journals from 1987 to 1991. Each record contains fields, including title, authors and abstract. In addition, each document contains indexing terms from Medical Subject Headings. We use the OHSU set, a selection of 63 queries from the original 101 in OHSUMED. The collection includes absolute relevance judgments with three levels given by human assessors to documents.

Medical Subject Headings (MeSH) is a controlled vocabulary managed by the U.S. National Library of Medicine (NLM) used to index millions of articles in the PubMed database. Physicians and medical librarians can use the terms in this vocabulary to search for relevant documents. It is structured in 12 levels establishing thesaurus-like hierarchic relationships and is easily mapped to SKOS [13].

The **LoC/LCSH** bundle consists of 8,905 MODS metadata records about historical maps hosted by the Library of Congress Map Collections[7]. Each metadata record has fields like title, person/role (e.g., creator, cartographer), place, related items, physical location in library, identifier, etc. It also contains several subject fields with indexing terms taken from several controlled vocabularies. We obtained all queries over this collection for the years 2009 a 2010. Since this experiment is still in an early stage, we selected 10 representative queries each delivering more than 10 results without query expansion. We used Mechanical Turk to obtain three relevance judgments for each query/document pair.

The Library of Congress Subject Headings (LCSH) have been actively maintained since 1898 and is used in the LoC catalog but also internationally, often in translation. It includes personal and corporate name headings as well as geographical terms.

4.2 Results and Discussion

Table 1 summarizes the results from our experiment on the OHSUMED/MeSH bundle, in which we compared the performance of SKOS-based expansion to no expansion and pseudo-relevance feedback using two different retrieval models (LTC TF-IDF and BM25). It shows that our approach exceeded precision of NoExp and PRF baselines at rank 1 and 3 over all retrieval models. Only in the case of $tf\text{–}idf$ PRF yielded higher precision and nDCG at rank 10.

SKOS-expansion adds specific and curated knowledge to the query, while PRF widens the query with terms that are only statistically correlated, therefore this is an expectable outcome. In our case, the PRF expansion method uses the 25 most common terms in the first 5 results for a query, without knowledge about the relevancy of each term added, apart from a document collection statistical point of view. PRF might lead to better recall i.e., return more relevant documents, but hurt precison at top ranks. On the other hand, with SKOS-expansion the terms used in expansion are always from a SKOS vocabulary and not from the corpus. Therefore, we are already only expanding using relevant terms, since the terms' existence in the vocabulary is a good hint that the term is important.

	LTC TF-IDF		
	P@1	P@3	P@10
NoExp	0.500	0.500	0.370
PRF	0.300	0.433	**0.430**
SKOS	**0.600**	**0.533**	0.370

Table 2: Precision with LTC TF-IDF results on the Library of Congress dataset.

Table 2 shows the results we obtained from early experiments with the Library of Congress Subject Headings. The results are similar to the results with OHSUMED. The PRF method is best at 10, but SKOS-expansion is more effective ranking the top positions.

In both experiments we observed that SKOS-expansion provides the best results for the top ranks. The PRF method can be competitive for longer result lists because it provides better retrieval performance with scoring models such as LTC TF-IDF. However, this greater retrieval performance of PRF at 10 has a cost: the loss of precision at the very top ranks. This means that, while the overall number of relevant documents in the top 10 ranks might be larger with PRF, the top ranked documents for a number of queries are not relevant. This problem with PRF is very noticeable in both datasets when measuring precision at rank 1, where PRF considerably hurts performance in contrast to NoExp. On the other hand, SKOS-expansion and PRF can be combined. We can take advantage of the improved precision at the top ranks of the SKOS-expansion method to choose the best top documents to use for PRF, for example.

5. CONCLUSIONS

Knowledge organization systems expressed in SKOS are available as structured data on the Web and can be used to improve Web search. In this paper, we presented a SKOS-based term expansion technique that supports both label- and URI-based term expansion either at query or document indexing-time. It keeps track of SKOS expansion types and exploits that information for scoring. Our experiments, which were conducted on two distinct dataset bundles, have

[7] http://memory.loc.gov/ammem/gmdhtml/gmdhome.html

indicated gains in retrieval effectiveness compared to settings without expansion or pseudo-relevance feedback.

Potential adopters can easily apply the proposed technique by loading our analyzer library implementation with Apache Lucene or Solr, configuring a SKOS vocabulary to be used, and defining SKOS expansion for selected Lucene Document fields. Overall, we believe that this is a simple and pragmatic solution for using human curated knowledge organization systems expressed in SKOS with search engines.

Our immediate next step will focus on a more thorough evaluation using the LoC/LCSH dataset, which requires an extended gold standard. We also seek to expand our analysis by collecting document corpora, queries, and vocabularies from other domains. While our implementation is already operational, it could be improved by adding vocabulary loaders (e.g., crawlers), and additional configurable scoring models. We also believe that our technique could be expanded beyond the scope of SKOS and applied for general structured Web data, as long as concepts are identified by URIs and described by a set of labels.

6. ACKNOWLEDGMENTS

This work was supported by the QSearch project (23133) co-financed by Agência de Inovação Portugal. This work was also supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme (PIOF-GA-2009-252206).

7. REFERENCES

[1] J. Bai, D. Song, P. Bruza, J.-Y. Nie, and G. Cao. Query expansion using term relationships in language models for information retrieval. In *SIGIR'05*, pages 688–695. ACM, 2005.

[2] J. Bhogal, A. Macfarlane, and P. Smith. A review of ontology based query expansion. *Information Processing & Management*, 43(4):866–886, 2007.

[3] C. Carpineto and G. Romano. A survey of automatic query expansion in information retrieval. *ACM Computing Surveys (CSUR)*, 44(1):1, 2012.

[4] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011.

[5] W. Hersh, C. Buckley, T. J. Leone, and D. Hickam. OHSUMED: an interactive retrieval evaluation and new large test collection for research. In *SIGIR '94*, pages 192–201, Aug. 1994.

[6] W. Hersh, S. Price, and L. Donohoe. Assessing thesaurus-based query expansion using the UMLS Metathesaurus. In *Proceedings of the AMIA Symposium*, page 344, Jan. 2000.

[7] J. Lin and D. Demner-Fushman. The role of knowledge in conceptual retrieval. In *SIGIR '06*, pages 99–106, Aug. 2006.

[8] N. A. A. Manaf, S. Bechhofer, and R. Stevens. The current state of skos vocabularies on the web. In *ESWC*, pages 270–284, 2012.

[9] A. Miles and S. Bechhofer. SKOS Simple Knowledge Organization System Reference. Recommendation, W3C, 2009.

[10] A. Miles, B. Matthews, M. Wilson, and D. Brickley. SKOS Core: Simple knowledge organisation for the Web, Dec. 2005.

[11] E. Summers, A. Isaac, C. Redding, and D. Krech. Lcsh, skos and linked data. *arXiv preprint arXiv:0805.2855*, 2008.

[12] M. Theobald, R. Schenkel, and G. Weikum. Efficient and self-tuning incremental query expansion for top-k query processing. In *SIGIR '05*, pages 242–249, Aug. 2005.

[13] M. Van Assem, V. Malaisé, A. Miles, and G. Schreiber. A Method to Convert Thesauri to SKOS. *The Semantic Web: Research and Applications*, pages 95–109, 2006.

[14] E. M. Voorhees. Query expansion using lexical-semantic relations. In *SIGIR '94*, pages 61–69, Aug. 1994.

[15] W. Zhou, C. Yu, N. Smalheiser, V. Torvik, and J. Hong. Knowledge-intensive conceptual retrieval and passage extraction of biomedical literature. In *SIGIR '07*, pages 655–662, July 2007.