ONTOLOGY-BASED ADAPTIVE QUERY REFINEMENT

Lefteris Kozanidis, Paraskevi Tzekou, Nikos Zotos, Sofia Stamou and Dimitris Christodoulakis
Computer Engineering and Informatics Department, Patras University, 26500, Greece

Keywords: Web Search, Query Refinement, Ontology, User Modeling, Retrieval Relevance.

Abstract: Query refinement is the process of providing Web information seekers with alternative wordings for expressing their information needs. Although alternative query formulations may contribute to the improvement of retrieval results, nevertheless their realization by Web users is intrinsically limited in that alternative query wordings convey explicit information about neither their degree nor their type of correlation to the user-issued queries. Moreover, alternative query formulations are determined based on the semantics of the issued query alone and they do not consider anything about the search intentions of the user issuing that query. In this paper, we introduce a novel query refinement technique which uses a lexical ontology for identifying alternative query formulations that are both informative of the user's interests and related to the user selected queries. The most innovative feature of our technique is the visualization of the alternative query wordings in a graphical representation form, which conveys explicit information about the refined queries correlation to the user issued requests and which allows the user select which terms to participate in the refinement process. Experimental results demonstrate that our method has a significant potential in improving the user search experience.

1 INTRODUCTION

Most of the Web information seekers typically go to their preferred search engine; submit a query that expresses their information needs and receive a list of results that somehow correlate to the information sought. Although searching the Web is a straightforward process, users sometimes encounter difficulties in finding the desired information, basically because their self-selected queries fail to communicate the user information needs in a precise and comprehensible by the engine manner.

To overcome the above difficulty, a significant number of researchers have studied ways for assisting Web users in the query selection process. Most of the work conducted in this respect attempts to refine the user-issued queries by expanding them with semantically similar terms. Query expansion has been an active field of research for quite a long time and operates upon the availability of basic linguistic infrastructure. In particular, most query expansion approaches rely on the use of thesauri or other types of semantic resources from which they leverage alternative wordings for refining the user typed queries. Alternative query formulations are typically in the form of a list of keywords that are returned to the user together with the search results. Upon display of alternative query formulations, the user might, either employ any of the suggested terms and append them to his initial query (therefore enabling expansion), or he might ignore the suggestions altogether and either start a new search session or terminate his search.

Alternative query formulations, although they might be useful in assisting the users find what they are looking for, nevertheless they are intrinsically problematic in that they do not convey any explicit information about their correlation to the user-issued query. As such, they do not help the user realize on the one hand the inadequacy of his self-selected query terms and, on the other, the suitability of the system selected terms in capturing the user's search intention. Another limitation in most of the existing query expansion techniques is that they attempt query refinement based on the query terms alone and without considering anything about the interests of the user issuing that query. In other words, query expansion has been insofar perceived as a query improvement technique that offers the same alternative formulations for a given query regardless of the varying search intentions that are hidden behind that query.

Kozanidis L., Tzekou P., Zotos N., Stamou S. and Christodoulakis D. (2007).
ONTOLOGY-BASED ADAPTIVE QUERY REFINEMENT.
In *Proceedings of the Third International Conference on Web Information Systems and Technologies - Web Interfaces and Applications*, pages 43-50
DOI: 10.5220/0001267300430050

In this paper, we address the above challenges and we propose a novel query refinement technique that provides alternative query formulations in a way that is both informative of and tailored to the user specific information needs. In particular, our technique relies on the semantics of the pages that match the user typed query for identifying a set of keywords to improve that query. Afterwards, it employs a lexical ontology for measuring the semantic correlation between the pages' selected keywords and the terms in the user-issued query. Pages' keywords that are semantically related to the initial query terms constitute the candidate wordings for improving the query. Identified alternative queries are further explored in the ontology in order to determine the semantic correlation that they exhibit to each other. Terms being strongly correlated in the ontology are selected as alternative wordings for refining the query. A refined query comprises a set of semantically similar terms which are interconnected and displayed to the user in the form of a query graph, as illustrated in Figure 1.

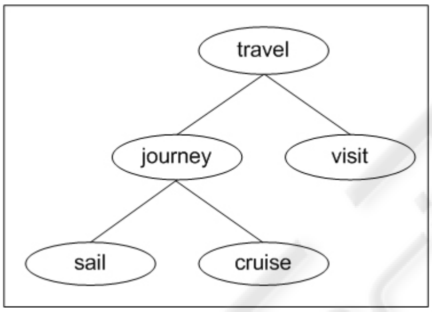

Figure 1: A query graph example.

By selecting alternative query wordings from the matching pages' keywords we ensure that the user's initial search intention is not neglected in the query refinement process. Moreover, by exploring the lexical ontology for picking the terms to refine a query, we guarantee on the one hand the semantic correlation between the user and the system selected terms and on the other that all alternative query terms relate to a common broad concept in the ontology, resolving thus sense ambiguities. Finally, by displaying alternative query formulation in a graphical hierarchical structure we ensure the informativeness of the refined queries.

To evaluate the effectiveness of our ontology-based query refinement technique in a practical setting we carried out a user survey where we studied how accurate our technique is in identifying alternative query formulations that are both comprehensible and useful to the search engine users. Obtained results demonstrate that our method has a significant potential in improving the user search experience.

The remainder of this paper is organized as follows: in Section 2 we outline the core infrastructure of our query refinement technique and we describe in detail the query selection process. In Section 3, we present our experimental study and we discuss experimental results. In Section 4, we review related work and we conclude the paper in Section 5.

2 ONTOLOGY-BASED QUERY REFINEMENT

Most of the queries issued to Web search engines are natural language queries. Although natural language queries are suitable for representing the user information needs, nevertheless they oftentimes fail to retrieve the desired information. This is essentially due to the variety in the vocabulary between the user typed queries and the indexed documents. One way to overcome vocabulary mismatches is to expand the issued queries with semantically similar terms. In this paper, we propose the use of a lexical ontology for selecting alternative query wordings as well as for structuring these alternative terms in a way that is perceptible by humans. The main intuition in our query refinement technique is that humans realize the suitability of alternative query wordings in terms of their explicit correlation to both their interests and self-selected queries.

At a high level, our query refinement method proceeds as follows. A user goes to a search engine and starts a query session representing his information needs. For every query participating in a user's search trace, we collect the retrieved pages and we process them in order to extract a set of keywords that represent the pages' semantic content. We then employ the pages' important keywords and we map them to the lexical elements in the WordNet ontology (Fellbaum, 1998). Query matching keywords that are strongly correlated to each other are deemed as candidate terms for reformulating the query. We then select among the candidate query formulations those terms that are specialized concepts to any of the queries in the user's search trace and we present them to the user in a graphical form.

For editing the refined query graph, we set at the root node the query that has the broader sense among all the user issued queries that participate in a search session. Children nodes represent the set of query terms and selected keywords that are subordinates of the root concept in the ontology. Relations between words are represented as links which are labelled to denote the type of semantic relation that holds among terms. The user can interactively improve his initial query by clicking on any of the graph's nodes. Clicking on a node that has a link to any of the user issued queries implies that the generated refined query is a Boolean "and" query that contains the user defined terms, expanded with the system selected keywords. Alternatively, clicking on a node that has no direct link to any of the user issued queries implies that the generated refined query is a Boolean "or" query that contains any of the system selected terms.

In the following paragraph, we present in detail the process that our model follows for selecting the candidate terms for reformulating a query. In Section 2.2, we describe the way in which our method explores selected terms for building the refined query graph. Finally, in Section 2.2.1, we present how refined query graphs visualize the alternative query formulations so as to assist the user clarify the correlation between his self-defined and system-selected queries. By doing so, we enable the user interests contribute in the query refinement process.

2.1 Query Terms Selection Process

In selecting alternative query formulations for improving the user issued queries, our refinement technique operates on the content of the users' past click data in order to identify the terms that are relevant to both the user interests and the issued query semantics. More specifically, in our method we rely on the queries previously issued by a user and we explore the content of the pages visited for those queries in order to determine a set of candidate terms for describing both the user interests and the user queries semantics. Those candidate terms are used for refining that user's subsequent queries.

In particular, given a set of queries issued by a user during a search session and given also the set of pages visited for those queries, we employ the vector space model and we pre-process pages in order to extract from the pages' main body (i.e. text) a set of candidate terms to participate in the query refinement process. As candidate terms we consider only nouns and proper nouns because these convey most of the semantic information in texts (Gliozzo et al., 2004). We then apply the TF*IDF weighting scheme (Salton and Buckley, 1988) for measuring the candidate terms' importance in the user's past click-through data. Finally, we select those terms whose importance values are within the top 10% of all the term values considered, and we further explore them in the WordNet lexical ontology as described in the following section.

Before delving into the details of the query formulation process, we point out that by selecting the alternative query wordings from the content of the pages that have been previously visited by a user; we ensure that only user relevant terms participate in the refinement process. Moreover, by leveraging a lexical ontology for identifying the semantic correlation between alternative query formulations and the user interests, we guarantee that refined queries are both informative of the user's interests and related to his search intentions as we will show next.

2.2 Refined Query Formulation

So far we have described the process that our model follows for selecting a set of candidate terms to refine a query, based on the content terms in the pages previously visited by a user. We now turn our discussion on how our model explores the lexical ontology for identifying the semantic correlation between these candidate terms and the user issued queries. Selected terms that exhibit a strong correlation to the user queries are the ones that our model suggests as alternative query formulations to the user. Figure 2, illustrates the query refinement process that our model follows. In practice, a refined query contains the set of terms from a user's past click history that are both informative of the user's search interests and correlated to the user's search requests.

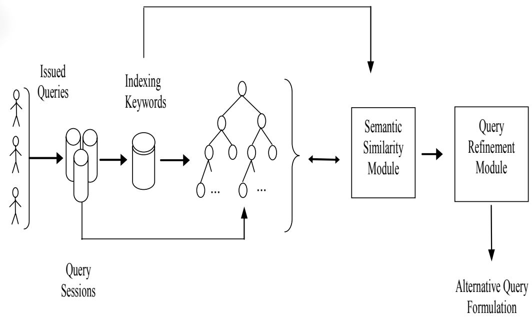

Figure 2: The query refinement process.

In refining a query issued by a user, our model relies on the terms selected from that user's click-through data and proceeds as follows. Selected terms together with all the queries participating in the user's search trace, from which these terms originate, are mapped to the ontology's nodes. Thereafter, our system measures the relevance score of every selected term against each of the queries in the search trace. Relevance score is determined by the semantic similarity measure, introduced in (Resnik, 2005) which is established on the hypothesis that the more information two concepts share in common, the more similar they are. The information shared by two concepts is indicated by the information content of their most specific common subsumer. Formally the semantic similarity between two words, w_1 and w_2, connected in the ontology via a relation r is given by:

$$sim_r(w_1, w_2) = -\log P\left(mscs(w_1, w_2) \right) \quad (1)$$

The measure of the most specific common subsumer (*mscs*) depends on: (i) the length of the shortest path from the root to the most specific common subsumer of w_1 and w_2 and (ii) the density of concepts on this path. Based on the semantic similarity values between the query terms and the words in a user's click history we determine the set of terms that our system suggests for refining a given query of the user.

In picking the terms to refine a user typed query among the set of candidate words, our model operates on a twofold criterion. On the one hand, selected terms should constitute specialized concepts of any of the query concepts in a user's search trace, i.e. the type of relation r in the above equation should be set to hyponymy. On the other hand, selected terms should be strongly correlated to any of the queries in the user's search trace, i.e. the value of their correlation should be among the top 10% of all their correlation values to the queries in the user's search trace.

By enabling only hyponyms of the user submitted queries participate in the refinement process, we ensure that the system suggested terms are both relevant to the user queries and useful in helping the user clarify his vague information needs. Moreover, by allowing from all the candidate query hyponyms only the closest hyponyms (i.e. those with a high correlation value) participate in the refinement process, we ensure that the system suggested terms are not overly specific and as such that they will retrieve information that is both useful and relevant to the user's needs.

Having identified the selection process of the terms that our system suggests to the user for reformulating his query, we now describe the visualization of the refined query. Refined query visualization pertains to the structuring of the alternative query formulations in a query graph. A refined query graph sets a word at each of its nodes and links nodes in a hierarchical structure so as to enable the user's navigation in the system selected terms.

2.2.1 Visualizing Refined Queries

For editing the refined query graph, we set at the root node the query that has the broader sense among all the user issued queries that participate in a search session. In selecting the term to be set at the root of the query graph, we rely on WordNet and we measure the semantic distance that every query in a search trace has from a top level concept. We then pick the query term of the minimum distance from a top node to denote the root concept of the generated query graph. For instance, consider that the following queries participate in a user's search session: *subway*, *railway*, *transportation,* and *car*. Mapping the above queries to WordNet hierarchies and estimating their semantic distance from WordNet's top nodes, results into the identification of the query term *transportation* as the query with the broader sense among all queries in the considered session. In the above example, our model would set the term *transportation* at the root node of the refined query graph.

Having selected the root term of the query graph, we proceed with the structuring of the graphs' children nodes. Children nodes represent the remaining terms in the user's session as well as the terms that our system has selected from the visited pages. For structuring the children concepts under the graph's root, we again explore the semantic relations encoded among the above terms in WordNet and we pick those terms that are interconnected through a specialization link for editing the children concepts of the refined query graph.

At the end of this process our system builds a graph of the terms that are suggested to the user as alternative formulations of his queries. The resulting graph is a hierarchy of concepts that sets at its root the most general term that a user has selected for describing his vague information need. The lower concepts in the hierarchy are a combination of the user and the system selected terms that convey information about more specific concepts which clarify the user's information need.

Figure 3 illustrates the refined query graph generation process. In our example we assume four queries in a search session, i.e. q_1, …q_4 and 7 keywords identified by our system as candidate query reformulations, i.e. k_1, …, k_7. From all the queries in the session, we denote q_3 to be the query of the broader sense. Moreover, we use WordNet ontology for identifying the semantic relations that hold between both query terms and selected keywords. Thereafter, following the approach described above, our system initializes a query graph by setting the query of the broader sense at the root node and proceeds with the children nodes as these are determined on the basis of their specialization (is-a) links in WordNet hierarchies. Query terms and pages' selected keywords that are specializations of the root term are added to the query graph and presented to the user, as shown in the Figure below.

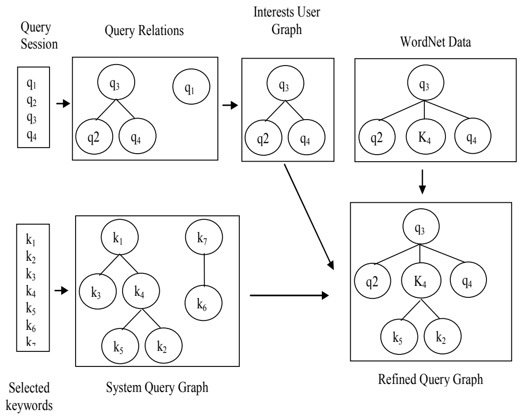

Figure 3: The refined query graph generation process.

The user can interactively improve his initial query by clicking on any of the graph's nodes. Clicking on a node that has a link to any of the user issued queries implies that the generated refined query is a Boolean "and" query that contains the user defined terms, expanded with the system selected keywords. Alternatively, clicking on a node that has no direct link to any of the user issued queries implies that the generated refined query is a Boolean "or" query that contains any of the system selected terms.

Although the proposed query graph has weights on its nodes and links, indicating the degree of semantic similarity between the respective concepts, these values are currently not visible, nor are they editable. We defer the description of both their visualization and editing for a future study.

3 EXPERIMENTAL SETUP

To evaluate the effectiveness of our query refinement technique in providing users with alternative query formulations that are both informative and representative of the user's particular search needs, we experimentally studied the effect that our refined queries have on both retrieval performance and the users' perception of the refined queries correlation to their search intentions.

For our study, we launched a prototype search engine of 500K pages and we implemented a prototype query refinement system using a Pentium R4 server, which is an RDBMS SQL database. WordNet hierarchies were stored in the database and accessed on a demand basis. Moreover, the TF*IDF values of the indexed pages' terms were pre-computed at the index level and stored separately. Similarity scores between the user queries and the pages' selected keywords were computed dynamically and stored in a secondary similarity index. The execution time of our query refinement module is proportional to the number of pages considered for alternative query wordings selection. To minimize our system's complexity, we computed the candidate terms TF*IDF values offline.

Having launched our system, we relied on the query sessions and the clickthrough data of six experienced Web users that we contacted for evaluating the effectiveness of our technique. In particular, we used all the queries participating in the considered query sessions as well as the pages visited for each of those queries as our experimental dataset.

Following the process described above, we processed the pages visited for each of our experimental queries and we selected a number of keywords for refining each of the above queries. System selected keywords together with their corresponding user issued queries were mapped to the WordNet ontology's nodes and following the steps presented in Section 2.2, our system generated a refined query graph for each of the queries examined. Table 1 summarizes some statistics on our experimental data.

Table 1: Statistics on the experimental dataset.

# of sessions	6
# of queries	57
avg. # of queries/user	9.5
avg. # of visited pages /query	5.3
avg. # of selected keywords / refined query	3.8
avg. # of nodes/ refined query graph	5.2

We then presented the generated refined query graphs to the respective participants in our survey and asked them to execute the improved queries by clicking on any of the system selected terms in the query graph and examine the first 10 pages returned for each of their queries. In this respect, we asked our subjects to evaluate how useful and therefore intuitive the refined queries are, by indicating on a 5-point scale (i) how accurate the refined queries are in capturing their search intentions and (ii) how effective the refined queries are in retrieving the desired information. With respect to this last criterion we asked our participants to compare the retrieval performance for each of their queries before and after refinement.

In judging the user ratings, we perceive accuracy of a refined query to be indicative of our technique's ability in considering the user interests in the refinement process. On the other hand, we perceive effectiveness of a refined query to be indicative of our technique's potential in improving the relevance of retrieved results. Obtained results are presented and discussed in the following section.

3.1 Experimental Results

We begin our evaluation, by discussing the effectiveness that our query refinement technique has on retrieval performance. In this respect, we explore the ratings that our users gave to the relevance of the top 10 results retrieved for each of their queries, with and without refinement. We then computed, for each of our subjects an average rating, which indicates that user's overall perception of our technique's efficiency in improving relevance of retrieved results.

In Figure 4 we aggregate the average ratings by participants to show the overall effectiveness of our query refinement technique in improving the user search experience. The x-axis represents our participants and the y-axis indicates the average ratings that each of our subjects gave to the relevance of the retrieved results. Average ratings are given on a 5-point scale, with values ranging between 0 and 0.5, with 0.5 indicating that all the pages considered (i.e. the first ten pages returned for each query) are highly relevant to the issued query. In the Figure, for each participant the first column represents the average ratings for retrieval performance based on the user issued queries alone, whereas the second column represents the average ratings for retrieval performance based on the refined queries, suggested by our system.

Note that, in our experiments, refined queries are Boolean "or" queries since we asked our participants to pick any of the system suggested terms for improving their self-selected queries. By doing so, we ensure that our evaluation reflects the true efficiency of our system in identifying alternative query formulations that are valuable to Web users. Nevertheless, we plan to evaluate the efficiency of Boolean "and" refined queries in a forthcoming study.

Obtained results demonstrate the potential that our query refinement technique has in improving retrieval performance. Specifically, we observe that all our subjects deemed the first ten retrieved results to be more relevant to their queries after these are refined, compared to the relevance of the results returned for the same non-refined queries. Therefore, we claim that our system is efficient in selecting alternative query formulations that are relevant to the user typed queries and being such they contribute to the retrieval of results that are highly relevant to the user information needs.

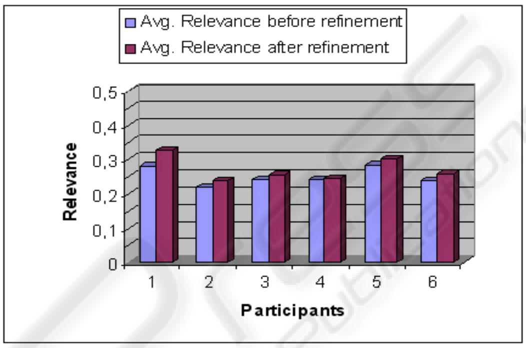

Figure 4: Average relevance of the top ten retrieved pages.

To further support our claim, we measured the user's perception of the refined queries' accuracy in capturing their search intentions. Figure 5, illustrates the user ratings for the refined queries accuracy. The x-axis represents our participants and the y-axis indicates the average ratings that each of our subjects gave to the accuracy of the system selected terms in representing their search intentions. Average ratings are again given on a 5-point scale; with values ranking from 0 to 0.5; with 0 indicating that the system selected keywords are inaccurate in representing the user intentions and 0.5 indicating that the system selected keywords perfectly represent the user intentions.

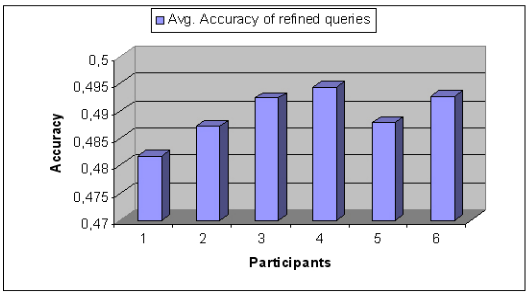

Figure 5: Accuracy of the refined queries in capturing user search intentions.

Obtained results confirm the validity of our assumption that relying on the semantics of the query matching pages for selecting the terms to reformulate a query, results to a refined query that is both relevant to the user typed request and the user search intention. A detailed analysis of the obtained results indicates that all the subjects in our study deemed the alternative query formulations suggested by our system, as highly relevant to their search intentions. Therefore, we claim that our approach has a promising potential in assisting Web users select queries that are expressive of their underlying search intentions.

Although experimental results demonstrate the potential that our query refinement technique has in improving retrieval performance and ultimately the user search experience, nevertheless our study is so far preliminary and involves a small number of both users (i.e. 6) and queries (i.e. 57). We are currently investigating the effectiveness that our refinement technique has on a larger pool of both users and queries. Moreover, we are examining ways of improving our system's effectiveness by incorporating a spell-checker at the query processing module, which relies on both the user's previous queries and the ontology's terms to correct any spelling mistakes that might appear in the user typed queries.

4 RELATED WORK

There has been much work on query refinement, aiming at dealing with vocabulary mismatches in the course of Information Retrieval (IR). Previous studies, address the automatic expansion of queries by using co-occurrence data (Jones and Barber, 1971), syntactic context (Grefenstette, 1992) or relevance information (Smeaton and van Rijsbergen, 1993). For identifying semantically related terms, many types of thesauri have been employed, ranging from hand-crafted (Vossen, 1998), to co-occurrence-based (Chen et al., 1995), (Crouch, 1990), (Qui and Frei, 1993) and head-modifier based thesauri (Grefenstette, 1992) (Jing and Croft, 1994). Aside from automatic query expansion, semi-automatic techniques have been proposed, such as the relevance feedback analysis (Harman, 1992), where the related terms come from user-identified relevant documents, or the local feedback analysis (Xu and Croft, 1996), where the top N retrieved documents are used for finding query-related terms.

Although early IR studies address the problem of query expansion within the limited scope of small text collections, the Web's evolution introduced significant challenges in improving IR efficiency. For the effective expansion of queries in the context of Web searching, many approaches have been addressed. Some of these rely on past queries to improve automatic query expansion. For instance, the work of (Fitzpatrick and Dent, 1997) uses the results of the users' past queries to formulate affinity pools, out of which the terms employed for expansion are selected. This technique has demonstrated an improvement of 15% on TREC-5 collection. More recently, the work reported in (Billerbeck et al., 2003), evaluates a query association technique to expand the TREC-10 Web track (Hawking and Craswell, 2001) queries. This approach concerns associating queries to a document if they share a high statistical similarity with the document. Experimental results showed that query expansion based on associations yields 18%-20% retrieval improvement compared to an optimal conventional expansion approach. Finally, other efforts i.e. (Khan et al., 2004) (Celik and Elci, 2006), concentrate on utilizing conceptual ontologies to find conceptually related terms and thus improve IR effectiveness.

The approaches summarized here, touch upon aspects related to our work and as such we perceive our approach to be complementary to other query refinement techniques. However, what makes our method different from existing techniques is that we assist the user realize the underlying correlation between his self-selected queries and the refined query wordings that are suggested by the system. To achieve that, we visualize refined queries in the form of a lexical graph and we enable the user interact with this suggested query graph by clicking on the terms he wishes to employ in his refined search. This way, our query refinement technique is not only effective in identifying alternative query formulations but it is also adaptive in the sense that it allows a user select different reformulations for the same query, depending on the specific search needs that he has every time he issues a query.

5 CONCLUDING REMARKS

In this paper we have discussed the query refinement problem and we have introduced a novel query refinement technique which uses a lexical ontology for selecting a set of semantically related terms for reformulating a query.

In particular, we proposed the investigation of a user's previous searches and the query relevant documents' semantics for selecting a set of terms that are both informative of the user's search intentions and semantically related to the user issued queries. These terms are then employed by our query refinement module which computes their semantic similarities in WordNet ontology and, based on both their similarity values and their semantic relation types; it determines which terms to participate in the refined query. Refined query terms are organized in a hierarchical structure, the so-called *refined query graph*, which sets at its nodes the refined query terms and links them together, enabling the user navigate from the most general to the most specific terms suggested by the system.

The preliminary experimental evaluation of our technique demonstrates that our query refinement method has a significant potential in improving the user search experience. In particular, experimental results indicate that users perceive the refined queries that our system suggests, to be highly informative and highly relevant to their search intentions. As such, we argue that our method has a promising potential in assisting Web users issue queries that describe their information needs in an accurate and comprehensive manner.

Although, further experimentation is needed before we deploy our technique to a practical setting, nevertheless be believe that our approach can pave the ground for more elaborate approaches in the query refinement process, especially when it comes to the users' interaction with query refinement services. However, one issue that our method leaves open is how to handle cases where a user's search profile gets contaminated from searches that reflect temporary rather than persisting information interests. We defer this study for a future work, since it requires a significant body of research on how users search the Web.

ACKNOWLEDGEMENTS

The work reported here is partially supported by the Greek Secretariat of Research and Technology (GSRT) under a PENED Grant awarded to the first author. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the GSRT.

REFERENCES

Billerbeck, B., Scholer, F., Williams, H.E., Zobel, J., 2003. Query Expansion Using Associated Queries. In *Proceedings of the ACM CIKM International Conference on Information and Knowledge Management*, New Orleans, Louisiana, USA.

Celik, D., Elci, A. 2006. Discovering and Scoring of Semantic Web Services based on Client Requirement(s) through a Semantic Search Agent. In *Proceedings of the 30th Annual International Computer Software and Applications Conference*, Vol. II, IEEE Computer Society Press, pp. 273-278.

Chen, H., Schatz, B., Yim, T., Fye, D., 1995. Automatic Thesaurus Generation for an Electronic Community System. In *ASIS Journal*, Vol. 46(3) pp. 175-193.

Crouch, C., 1990. An Approach to the Automatic Constructions of Global Thesauri. In *Information Processing and Management*, Vol. 26(5) pp. 139-147.

Fellbaum, Ch., 1998. *WordNet: an Electronic Lexical Database,* MIT Press.

Fitzpatrick, L., Dent, M., 1997. Automatic Feedback Using Past Queries: Social Searching? In *Proceedings of the 20th ACM-SIGIR Conference*, pp. 306-313.

Gliozzo, A., Strapparava, C., Dagan, I., 2004. Unsupervised and Supervised Exploitation of Semantic Domains in Lexical Disambiguation. In *Computer Speech and Language*, 18(3) pp. 275-299.

Grefenstette, G., 1992. Use of Syntactic Context to Produce Term Association Lists for Text Retrieval. In *Proceedings of the 15th ACM SIGIR Conference*.

Harman, D., 1992. Relevance Feedback Revisited. In *Proceedings of the 15th ACM SIGIR Conference*.

Hawking, D., Craswell, N., 2001. Overview of the TREC-2001 Web Track. In: Voorhees, E., Harman, D.K. (eds.): The *Tenth Retrieval Conference. NIST Special Publication* pp. 500-250, Washington D.C.

Jing, Y., Croft, B., 1994. An Association Thesaurus for Information Retrieval. In *RIAO Conference*.

Khan, L., McLeod, D., Hovy, E., 2004. Retrieval Effectiveness of an Ontology-Based Model for Information Selection. In *VLDB Journal*, Vol. (13) pp. 71-85.

Qui, Y., Frei, H.P., 1993. Concept Based Query Expansion. In *Proceedings of the 16th ACM SIGIR Conference*.

Resnik, Ph., 2005. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In *Proceedings of the 14th Intl. Joint Conference on Artificial Intelligence*, pp. 448-453.

Salton, G., Buckley, C., 1998. Term Weighting Approaches in Automatic Text Retrieval. In *Information Processing and Management*, Vol. 24(5) pp. 513-523.

Smeaton, A.F., van Rijsbergen, C.J., 1993. The Retrieval Effects on Query Expansion on a Feedback Document Retrieval System. In *Computer Journal*, Vol. 26(3) pp. 239-246.

Spark Jones, K., Barber, E.B., 1971. What Makes an Automatic Keyword Classification. In *ASIS Journal*, Vol. (18) pp. 166-175.

Vossen, P., 1998. *EuroWordNet: a Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers.

Xu, J., Croft, B., 1996. Query Expansion Using Local and Global Document Analysis. In *Proceedings of the 15th ACM SIGIR Conference*

.