Retrieval Efficiency of Normalized Query Expansion

Sofia Stamou and Dimitris Christodoulakis

Computer Engineering and Informatics Department, Patras University,
Research Academic Computer Technology Institute, 26221, Greece
{stamou, dxri}@cti.gr

Abstract. In this paper we experimentally study the impact of normalized query expansion on Web Information Retrieval. In this respect, we have implemented a query expansion module, which firstly normalizes the user submitted queries and subsequently attempts to enrich them with semantically related terms that are obtained from WordNet. Experimental results demonstrate that for certain query types our module has a potential in giving improved search results in terms of relevance, compared to the results retrieved for the same queries by other retrieval methods.

1 Introduction

To support information seekers in overcoming terminological problems when searching for information on the Web, several approaches have been addressed in the literature, the most prominent of which imply the expansion of the issued queries with semantically related terms. In this paper we seek to get an improved insight on how normalized query expansion can effectively cope with vocabulary mismatches, in an attempt to improve retrieval relevance when querying the Web. To challenge that, we built a prototype query expansion module that interacts with normalization techniques, applied to the subjected queries. The query expansion module we introduce explores the semantic information encoded in WordNet and determines which terms are the most suitable to be used for enriching a given query.

We evaluated our system's performance in successfully enriching queries by having humans judge the relevance of the results retrieved in response to a set of queries after these have been expanded by our system, and compare them to the relevance of the results retrieved for the same queries after employing other searching techniques. Our findings indicate that our expansion approach improves retrieval performance for certain query types, compared to the performance of other retrieval techniques. Retrieval improvements are pronounced for long queries (i.e. multiword and phrase queries) due to our system's effectiveness in disambiguating them.

In the remaining of this paper, we describe how our system proceeds in generating expanded queries (Section 2) and we report on our empirical findings (Section 3) that proof the impact of our approach in helping information seekers find alternative wordings for expressing their information needs. We conclude our work (Section 4) with a discussion on our system's contribution in retrieval performance. Before presenting our system, we should note that although we experimented with Greek data, we believe that our approach can be useful to other languages as well.

A. Gelbukh (Ed.): CICLing 2005, LNCS 3406, pp. 604 – 607, 2005.

2 Normalized Query Expansion

A core module in our query expansion system is a normalizer introduced in [3]. Normalization is a computational process, which assisted by morphological lexicons, identifies (query) term inflectional and derivational variants and reduces them to a single canonical form, i.e. lemma. Lemmatized (query) terms pass though our expansion module, which maps them to Greek WordNet (GWN) [1] synsets and retrieves their semantically equivalent terms (i.e. synonyms). Retrieved synonyms are used as supplementary terms for enriching queries, following the approach described below.

A critical factor for the successful expansion of a query requires the prior successful match between the sense of the query and the senses of the GWN synsets that map the given query. For determining the appropriate sense of a query that matches multiple GWN synsets, we rely on the matching synsets' correlation in GWN, represented by a correlation factor that is determined by the synsets' position in the WordNet graph. More specifically, synsets that belong to a lower level in the WordNet hierarchy and that also share a minimal distance from each other are deemed as highly correlated. The *Correlation* of a word pair (w_1, w_2) is formally determined as the product of the words' *Depth* [5] and their conceptual similarity (*Sim*) [4] in WordNet. Depth of a word pair (w_1, w_2) is defined as:

$$DepthScore\ (w_1, w_2) = Depth\ (w_1)^2 \bullet Depth\ (w_2)^2. \tag{1}$$

and words' similarity $Sim(w_1, w_2)$ is determined by the set of WordNet synsets (c) that subsume both w_1 and w_2, in any sense of either word with a probability Pr, i.e.:

$$Sim\ (w_1, w_2) = \max\ c \in subsumers\ (w_1, w_2) \quad [\ -\log \Pr\ (c)\ ]. \tag{2}$$

Finally, the *Correlation* between w_1 and w_2 is defined as:

$$Correlation\ (w_1, w_2) = Depth\ (w_1, w_2)\ \bullet\ Sim\ (w_1, w_2). \tag{3}$$

Using the above formulas, query terms that match multiple GWN synsets are expanded with the synonyms of the synset that has the greatest *Correlation* score to them, of all its correlation scores. While synset selection is straightforward for multiword queries, for which there is enough data to resolve ambiguities, synset selection for single term queries is more complicated, essentially due to the lack of enough query terms that would help disambiguate the query. In selecting a GWN synset for a single term query expansion, we follow the assumption of [2] that terms occurring together in the same document are related to the same theme, and we attempt query sense resolution as follows: the top 10 ranked documents retrieved as relevant to a single term query are merged together forming a sample corpus, from which we extract query co-occurrence data. We first define the optimum window size within which we consider co-occurrence to 15 words, we then merge together these context windows and we assign frequency weights to their terms according to the normalized TF.IDF scheme. Terms with frequency values above a given threshold and the initial query are mapped to GWN synset. Selection of the best matching synset relies on the synsets' *Correlation* scores and takes place as described above.

3 Evaluating Expansion Performance

To evaluate the efficiency of our module in successfully enriching queries, we conducted an experiment in which we compared the relevance of the results retrieved for a set of 18 queries after these have been expanded by our system to the results retrieved for the same queries after employing other searching techniques. Experimental queries were manually selected and they correspond to three query type formulations, i.e. single term, multiword and phrase queries. The other retrieval methods against which we compared our system's efficiency are: keyword-based searching, referred to as "baseline retrieval" and searching with query morphological variants, known as "normalized" retrieval. For our comparison, we relied on the relevance judgments of external evaluators, who analyzed the top 30 retrieved documents for each of the queries across the three searching techniques. Relevance judgments were scored on a 10-point scale, ranging between 0.1 for marginally relevant results and 1 for highly relevant results. At the end of the experiment, we collected human relevance judgments, grouped them in three categories, i.e. relevance of the top-10, top-20 and top-30 retrieved documents respectively, and we calculated the average relevance for each of the categories across the three searching methods (i.e. baseline, normalized and expansion retrieval). Average relevance was defined as the ratio of the sum of relevance scores at the specified ranking points to the total number of the documents considered, and is defined as:

$$R(r) = \left| \sum_{i=1}^{Nd} R_i(r) \right| / \left| Nd \right|. \tag{4}$$

where R(r) is the query average relevance at ranking level (r), Nd is the number of pages considered, and $R_i(r)$ is the relevance at ranking level (r) for query i. Having computed separately average relevance scores for the results retrieved across the different ranking points, we merged together relevance judgments and we computed the overall relevance of each retrieval method by summing the average relevance scores at the specified ranking points and then dividing by the total number of queries. Overall relevance of each retrieval technique is computed as:

$$\sum_{i=1}^{NUM} R_r / NUM. \tag{5}$$

where R(r) is the relevance at ranking level (r), (in our case the top 30 retrieved records) and NUM is the number of queries issued. We assessed the retrieval efficiency of our system by comparing the relevance scores that it delivered at each of the specified rankings, to the relevance of the results returned for the same queries by other searching methods, at the same ranking cut-offs. Obtained results are briefly discussed next. The interested reader can obtain a more elaborate view on the results retrieved for each of the queries across the three searching methods by referring to a longer version of the paper available at [6].

4 Discussion

An analysis of the experimental results indicates that for multiword and phrase queries, our expansion module outperforms keyword-based and morphology-based retrieval. In detail, our findings demonstrate that our system, when supplied with a sufficient number of query terms (i.e. typically 2 or 3), it can effectively resolve the senses of the underlying queries and hence enrich them with their corresponding synonyms found in WordNet. On the other side, the successful expansion of single term queries is mainly dependent on the lexical cohesion of the sample corpus that our system employs for obtaining query co-occurring words. Due to our system's dependence on corpus co-occurrence data for resolving query senses, it becomes evident that upon selection of the wrong co-occurring terms, expansion fails. Expansion failure was implied in our work by the deteriorated relevance scores that expansion gave for some single term queries. Moreover, expansion performance depends on the normalizer's ability to resolve query terms' part-of-speech ambiguities and thus induce them to their correct lemmas. Again, normalization failure leads to the selection of semantically irrelevant terms for query expansion, which results to the retrieval of irrelevant documents among the results. Summarizing, we have found that normalized query expansion has a potential in improving retrieval efficiency when dealing with multi-term queries, whose senses can be effectively inferred by their WordNet semantic correlations. For other query types, our findings suggest that expansion has a promising potential as long as it uses a balanced corpus of topics for disambiguating queries.

References

1. Grigoriadou, M., Kornilakis, H., Galiotou E., Stamou, S., Papakitsos, E.: The Software Infrastructure for the Development and Validation of the Greek WordNet. In Romanian Journal of Information Science and Technology, Vol. 7(1-2) (2004)
2. Mandala, R., Takenobu, T., Tanaka, H.: Combining Multiple Evidence from Different Types of Thesaurus for Query Expansion. In Proceedings of the 22nd ACM SIGIR Conference (1999)
3. Ntoulas, A., Stamou, S., Tzagarakis, M.: Using a WWW Search Engine to Evaluate Normalization Performance for a Highly Inflectional Language. In Proceedings of the ACL Student Workshop, France (2001)
4. Resnik, Ph.: Disambiguating Noun Groupings with Respect to WordNet Senses. In Proceedings of the 3rd Workshop on Very Large Corpora, MIT (1995)
5. Song, Y.I, Han, K.S., Rim, H.C.: A Term Weighting Method Based on Lexical Chain for Automatic Summarization. In Proceedings of the 5th CICLing Conference (2004)
6. Normalized Query Expansion Overview and Results: ftp://150.140.4.154/ftproot/