

Sentence Retrieval with LSI and Topic Identification

David Parapar and Álvaro Barreiro

IR Lab, Department of Computer Science,
University of A Coruña, A Coruña, Spain
`dparapar@udc.es`, `barreiro@udc.es`

Abstract. This paper presents two sentence retrieval methods. We adopt the task definition done in the TREC Novelty Track: sentence retrieval consists in the extraction of the relevant sentences for a query from a set of relevant documents for that query. We have compared the performance of the Latent Semantic Indexing (LSI) retrieval model against the performance of a topic identification method, also based on Singular Value Decomposition (SVD) but with a different sentence selection method. We used the TREC Novelty Track collections from years 2002 and 2003 for the evaluation. The results of our experiments show that these techniques, particularly sentence retrieval based on topic identification, are valid alternative approaches to other more ad-hoc methods devised for this task.

1 Introduction and motivation

In this work we understand the task of sentence retrieval in the way defined in the TREC Novelty Track. The Novelty Track was introduced for the first time in the TREC 2002 conference [1] and is composed of two main tasks. The first one is sentence retrieval: starting with a set of relevant documents for a query (topic in the TREC terminology), the system must extract from those documents the relevant sentences for that topic, removing the ones that do not contain significant information or that are related to different topics. The second task starts from the sentences retrieved in the first task or from the relevant sentences selected by human assessors. Taking in account this set, the system must retrieve only the novel sentences, i.e., sentences that contain new information with respect to the previous sentences in the set. In this paper we have focused only in the first task.

Among the applications of sentence retrieval we find query-biased text summarization and the presentation to the users of the most relevant sentences of the documents retrieved in a results list [2]. Furthermore, the novelty task would remove the redundant information in the extracted sentences. Another application could be the construction of question answering systems because query relevant sentences can be useful to obtain the user's query.

The research done for the Novelty Track can be divided in two groups. Some systems try to adapt classical document retrieval techniques to sentence retrieval with a different definition of the parameters of interest. For example,

the classical vector space model can be adapted with a new definition of term frequency, inverse document frequency and document length. On the contrary, some participants work with techniques specifically developed for related tasks like summarization or passage retrieval. Anyway, every retrieval model based on matching of query and sentence terms will have difficulties with the short length of sentences. For this reason most systems use pseudo-relevance feedback [3] and query and/or document expansion [4].

Despite of the research effort done, the effectiveness in the sentence retrieval task still can be improved. Our idea was to test Latent Semantic Indexing (LSI) because it had not been used before in this task and because it can lead to more general and less ad-hoc solutions and because only a small set of documents has to be analyzed in this task. Therefore, we expect effective and efficient solutions.

LSI [5, 6] is a retrieval model based on dimensionality reduction. An initial space of terms and documents is reduced to represent concepts instead of terms. With this transformation LSI claims to remove the noise produced for the variability in the use of terms, extracting the underlying semantic concepts in a document collection. Most retrieval models are based on the number of query and document matching terms with different weighting schemes. Therefore, for the sentence retrieval task, LSI can be appropriate because the query-sentence similarity measure is not obtained in the initial space of terms and sentences. In the TREC Novelty sentence retrieval task researchers are given a set of relevant documents for each query. This homogeneous set can facilitate the extraction of the latent semantic structure. LSI can be more effective than query expansion since this strategy introduces noise in the queries trying to increase term matching, and the expansion based on co-occurrence is going to be more difficult in the case of the sentences. Meanwhile LSI reduces the noise at the cost of a possible lost of information in the final reduced space. To achieve a good balance between noise filtering and information loss, the selection of the dimensions in the reduced space is crucial.

In this paper we also present another retrieval strategy based on topic identification. This strategy was devised starting from the Gong and Liu [7] summarization method. It uses Singular Value Decomposition (SVD) and topic identification for developing a generic summarization method for single documents. This process consists of two steps. First, the main topics of a document are identified from the SVD of the document text. Next, a sentence is selected for each main topic and added to the summary. In our case we need to produce a query-relevant set of sentences from a set of relevant documents for the query. Instead of identifying the main topics of a single document, we first identify the main topics of the query. In this last usage a topic is an aspect of the query and a query can be about several topics. Please note that in some sections we will use topic with the meaning of "TREC topic". Hereinafter the context is enough to distinguish between the two meanings of the word topic. The sentence selection process is also different. In the case of a generic summary of a single document, we intend to maximize the coverage of the document's main content by selecting one sentence for each main topic. On the other hand, to address the sentence

retrieval task, it is necessary to maximize the number of relevant sentences retrieved for each topic of the query. After the identification of the query topics, we retrieve the set of relevant sentences for the topics.

We advance here that sentence retrieval based on topic identification outperforms the method based on LSI retrieval. In addition, it is competitive compared with other more specific techniques.

The rest of the paper is organized as follows. In the next section we introduce the LSI model and explain the alternative retrieval method proposed for sentence retrieval. In section 3 the experimental setting is presented. In section 4 we show and analyze the results. In section 5 different considerations and future research lines are introduced. The paper ends with the conclusions section.

2 LSI and Topic Identification

2.1 LSI: Model

The LSI model [5, 6] is an extension of the the vector space model for information retrieval based on a dimensionality reduction technique. LSI claims to capture the latent semantic structure in the documents and represent them in function of basic concepts and ideas instead of terms. This allows us to deal with the polysemy and synonymy problems.

The first stage in LSI is the Singular Value Decomposition (SVD) in which a matrix of terms by documents, obtained from the document collection, is decomposed into three matrices and then truncated to a reduced space. In our case we have a matrix A of terms by sentences (instead of documents) in which the value of cell ij is associated with the apparitions of term i in sentence j weighted with local and global weights. This matrix can be decomposed in the following way:

$$A_{t \times s} = T_{t \times r} \Sigma_{r \times r} S_{r \times s}^t \quad (1)$$

where,

- t: number of terms
- s: number of sentences
- r: rank of A
- T: matrix of left singular vectors
- S: matrix of right singular vectors
- Σ : diagonal matrix of singular values

The Σ matrix is a diagonal matrix of singular values in decreasing order; the singular values are the positive square roots of the eigenvalues of the matrix $A \times A^t$, where A^t is the transposed matrix of A. Each singular value represents a dimension of the space, dimensions with higher singular values are more important in this space. The rows of the T matrix are the term vectors and the rows of the $S_{s \times r}$ are the sentence vectors. To avoid the noise in term usage and to capture the latent structure in a document collection the SVD is truncated

to a reduced number of dimensions k . With this truncation, the k higher singular values are kept with the correspondent dimensions in the $T_{t \times k}$ and $S_{s \times k}$ matrices:

$$\hat{A}_{t \times s} = T_{t \times k} \Sigma_{k \times k} S_{k \times s}^t \quad (2)$$

where \hat{A} is the closest matrix of rank k to A in terms of $\| \cdot \|_2$ and $\| \cdot \|_F$ norms. In the sentence matrix $S_{s \times k}$ each sentence vector has k components. The usual interpretation of this transformation is that each component now represents a concept extracted from some relationship between terms and sentences in the original space.

In this point it is worth to revisit the example given in one of the pioneer LSI works. In the example presented in [5], after a truncated SVD with $k = 2$ of an original space of 18 terms and 14 documents, it is clear that the documents pertaining to a certain topic are clustered above the x-axis which is associated with the first dimension of the final reduced space, while documents pertaining to another topic are clustered near the lower y-axis which is associated with the second dimension of the final reduced space.

2.2 Retrieval in LSI

Once we have this decomposition and truncation for an information retrieval system we must do the same transformation with the queries that arrive to the system. We need to project the queries to the same reduced space using the following formula:

$$\hat{q} = q^t T_{t \times k} \Sigma_{k \times k}^{-1} \quad (3)$$

With this transformation the query vector \hat{q} is just like a pseudo-sentence vector that can be compared with the rows of $S_{s \times k}$ in a retrieval task. Actually, in order to make this comparison the rows of $S_{s \times k}$ and \hat{q} are scaled multiplying them by Σ . Note that this scaling only produces a stretched version of the reduced space $S_{s \times k}$ giving more importance to the dimensions with higher singular values. Finally, query and sentences can be compared using the cosine or dot product. If the similarity measure is larger than a threshold thr_{LSI} then the sentence is considered relevant.

2.3 Retrieval based on topic identification

First we will give a brief description of the method and then we will provide a pseudo-code style description. Our intuition here is that a sentence contains a very low number of different concepts or topics, i.e., in most cases a sentence is only about one or two things. Considering that in the reduced space each dimension represents a concept, we can consider only a few dimensions of the sentence vector to determine its relevance. We considered that the most important topics in a query are those associated with the largest magnitude components, because

they are the most discriminative in the space. So we take these dimensions as the representatives of the query and we select the sentences with the largest values for those dimensions.

In this method the query \hat{q} and the sentences, rows of matrix $S_{s \times k}$, are not scaled with Σ because this method is specifically devised for operating in the reduced space. The retrieval algorithm uses the $S_{s \times k}$ matrix and the projected query \hat{q} as follows:

1. For a projected query $\hat{q} = (\hat{q}_1, \hat{q}_2, \dots, \hat{q}_k)$ the n components with the largest absolute values are chosen.
2. Let \hat{q}_i be any of the selected components of \hat{q} obtained in the previous step. Let $S_i = (s_{1i}, s_{2i}, \dots, s_{si})$ be the column i of matrix $S_{s \times k}$. Each element s_{ji} represents the weight of the component i in the sentence j (row j of $S_{s \times k}$). Let thr_{TI} be a positive threshold. For each \hat{q}_i selected in 1:
 - For each $j:1..s$:
 - If \hat{q}_i and s_{ji} are positive and $s_{ji} > thr_{TI}$ the sentence j is selected.
 - If \hat{q}_i and s_{ji} are negative and $|s_{ji}| > thr_{TI}$ the sentence j is selected.
3. The union of sentences obtained in the previous step is returned.

Let us remark that in the inner loop of step 2 we only select those sentences that have components with values larger than the threshold and the same sign than the query component because we suppose that queries and sentences pertaining to a certain topic must be located in the same side of the axis associated to that dimension. It is possible to find an optimal number of retrieved sentences by changing the value of the threshold thr_{TI} introduced in the algorithm.

Now we emphasize the differences with the algorithm exposed in [7] where the goal is to produce a generic summary of a document. First, in the algorithm presented here the goal is to obtain the set of relevant sentences to a query. Second, in [7] the first step obtains the singular values associated with the main concepts of a single document; in our algorithm the main concepts in the query are obtained in the first step. Third, in [7] for each important singular value, a sentence is selected trying to maximize the coverage of the document's main content; in the problem addressed in this paper we must cover the query topics but for each query topic we must retrieve not only a representative sentence but the set of relevant sentences.

The evaluation of the sentence retrieval task in the TREC Novelty track is set-based. The method for retrieval of sentences based on LSI described in section 2.1 produces a ranking. Therefore, it was necessary to define a similarity threshold to obtain a set. The method based on topic identification directly produces a set. For other tasks the algorithm based on topic identification could be modified to produce a ranking.

3 Experimental settings

We used the collections of the TREC Novelty Track from years 2002 and 2003 for the evaluation of the two retrieval methods.

The 2002 collection consists of 49 topics selected from the ad-hoc task of TREC in the range from 300 to 450 [1]. For these topics a set of 25 relevant documents was provided. In fact, 25 is the upper limit, as for some topics the number of documents is lower. There are two relevance judgments for the relevant sentences in this track corresponding to the minimum and maximum assessor. The minimum assessor, the one with the lower number of relevant sentences, was taken as the official for this task. The average percentage of relevant sentences is very low, 2.54%.

The 2003 collection consists of 50 topics specifically created for the 2003 track [8]. Two kinds of topics were created: *events*, for news articles, and *opinions* for articles about controversial subjects. For each topic, 25 relevant documents were selected from the AQUAINT collection, composed of articles from New York Times News Service, Associated Press and Xinhua News Service. The authors of the topics, primary assessors, made the relevance judgments of their own topics and a secondary group of assessors provided a second set of judgments. The percentage of relevant sentences from the primary assessors, that was taken as the official for this task, was much higher than in 2002, 41.13%.

In both collections the sentences are previously identified in the documents and tagged separately to allow a quick processing of the documents. This separation is based in the points of the original text so the division in sentences is not part of the difficulties of the task.

We tested different stemmers; the best results were obtained with the Krovetz stemmer. The stop words were removed. All the experiments presented in the next section used the Krovetz stemmer and the same stop list.

In this task the relevant sentences only can be found in the set of 25 documents associated with each TREC topic. For this reason, the matrix A of equation (1) was generated independently for each set of relevant documents associated with each TREC topic in both methods. This indexing scenario represents a real situation in which a standard IR system retrieves the top-ranked documents and these documents are analyzed to retrieve relevant sentences.

Retrieval was done with the programs provided by the LSI software package of Telcordia Technologies [9]. We used this software with the SVDLIBC [10] library to perform the SVD transformation.

The evaluation measures are the set-based precision, set-based recall, and the F measure for each topic and the average for the 50 topics. The metric taken as reference was the F measure given by the formula:

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

This measure gives the same importance to precision and recall, so the goal of retrieval is a good trade-off between precision and recall.

4 Results

Different weighting schemes provided by the LSI software package of Telcordia were tested for local and global weights. For the local weights the best weight

was the binary one. Anyway, if the log or raw term frequency option is used, the difference in the result obtained is small, since the information of the number of times a term appears in a sentence is usually not so important. The global weight idf^2 , which is the square of the logarithmic idf, also allows the best results in both methods and collections. For the LSI method the similarity measure chosen for the two collections was the dot product.

In figure 1 the best experiments for the 2002 collections with both methods are shown. For the 2002 collection the first method results were bad. The best results were obtained using 20 factors ($k = 20$) in the truncation of the SVD, although there are similar results with k varying in the range from 15 to 25. With the dot product as the similarity measure we determined the value of the threshold THR_{LSI} that produces the best F value. The higher F value was 0.06 with THR_{LSI} in the range between 38 and 50. The F value is very low compared with the best runs submitted to the TREC track. Moreover, the groups participating in the track had only four topics for training while we used the full set of relevance judgments to adjust the used parameters. In fact, a random retrieval of sentences produces an F value of 0.04, so we can say that the LSI method is not able to work properly in this collection.

In the same collection the topic identification method improves substantially the performance. The best range for the number of factors is the same than in the LSI method, with the best value for $k = 15$. We used the same weighting scheme. The used threshold type is different: a minimum value for all of the components was fixed and only sentences with absolute values higher than that threshold were retrieved. The best results are obtained when the similarity threshold THR_{TI} has a value of 0.05 and the parameter n for the number of query components has a value of 1. In all experiments an increment in the number of query components produced a small decrement in the performance. The best F obtained was 0.141 which supposes a big improvement with respect to the LSI results. But these results still do not reach the ones obtained by the best runs submitted to the track with an F value around 0.23. The F obtained with the second human judgments (maximum assessor) is 0.371.

In the evaluation of the two methods we only established the values of k, THR_{LSI}, THR_{TI} and n mentioned above to obtain the best results. Most systems participating in the TREC Novelty track applied specific techniques to improve performance: specific analysis of the queries or documents to be used in query or document expansion, query expansion with linguistic resources, pseudo-relevance feedback techniques, features extraction from the TREC topics or document clustering before sentence retrieval. For these reasons, it is difficult to compare our results with the results obtained for the runs submitted to the TREC tracks. The participants did not have the relevance judgments to find the best parameters. Actually, in the 2002 track four training topics were provided. These topics were used for some participants to determine, for example, the percentage of sentences that should be retrieved. In the 2003 track no training topics were provided. However, since participants in the competition were allowed to send five runs, we presented here the comparison with the best results of TREC.

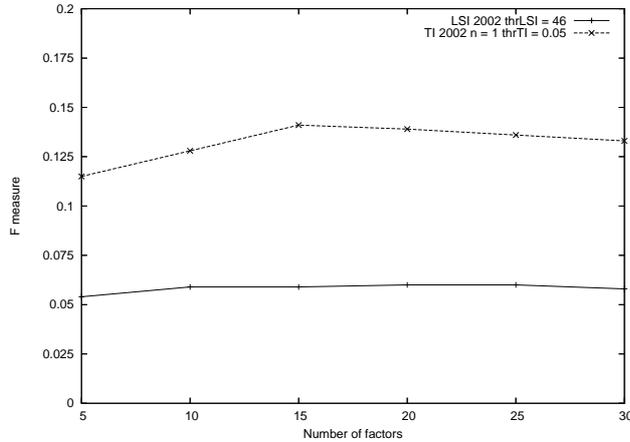


Fig. 1. F measure for the best runs in function of the number of factors (k), 2002 collection

Besides that, we presented the comparison of our results with the results of the secondary assessor, therefore this comparison gives an idea of the behavior of our methods with respect to a human performing the task.

The results of the LSI method with the 2003 collection were good. Parameter tuning has to be different for this collection so different experiments were repeated again to determine the best parameters. The optimal number of factors turned out to be $k = 5$. The reason of this could be that the 2003 collection is a bit smaller or more homogeneous than the 2002 collection. With the same weighting scheme and similarity measure, the value of the threshold thr_{LSI} that achieves the best results is also different, in this case because the relevance judgments are very different and the percentage of relevant sentences is higher. As expected, the value of the threshold is lower than in the case of the 2002 collection and the range for the best results is between 16 and 20.

The best F with $k = 5$ factors and a threshold thr_{LSI} of 16 is 0.593. This value is similar to those submitted for the best runs participating in the track. Indeed, the F value of the secondary assessor is 0.58. Our value is better in this case because our recall is very high 0.90, while the recall of the human assessor is 0.67. In terms of precision we obtain 0.5 while the human obtains 0.69. Results with better precision are obtained increasing the threshold value. With these increased values the F value decreases because the recall decreases faster than the precision increases. For example, for a value of $thr_{LSI} = 50$, a recall of 0.62 similar to the human is obtained, but the precision has a value of 0.55 and the F value is 0.533. It is a good F value close to the human assessor F value and more similar in the trade-off precision/recall.

Although the difference is not so large as in the 2002 collection, the topic identification method also performs better than LSI for the 2003 collection. The optimal number of factors is the same that for the LSI method ($k = 5$). Compared

with the topic identification algorithm in the 2002 collection the threshold thr_{TI} is also lower. The highest F value is 0.640 and it is obtained with $thr_{TI} = 0.02$. This value is even better than the value provided by the secondary assessor and than the best runs submitted to the track, but it is obtained with a high recall (0.89) and a precision of 0.58; with $thr_{TI} = 0.03$ we obtain a recall of 0.73 (the secondary assessor obtains 0.67) and a precision of 0.62 (the secondary assessor obtains 0.69) and an F value of 0.603 which is still higher than the human assessor that obtains a 0.58. In figure 2 the changes of the best results in function of k for the 2003 collection with both methods are shown.

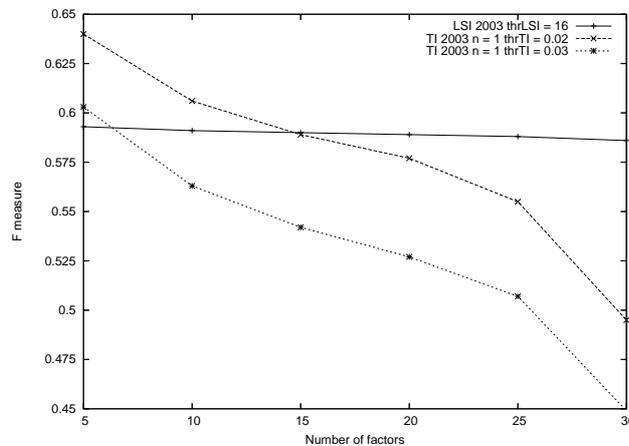


Fig. 2. F measure for the best runs in function of the number of factors (k), 2003 collection

In both collections, after selecting the best k , the topic identification method performs better than the LSI method. This is especially important in the 2002 collection where LSI simply does not work. An important fact is that in both methods the optimal number of factors is the same for the same collection. But, in the case of the topic identification method, the variation in the number of factors has a major influence in the performance of the system while in the LSI method the influence of the number of factors is quite soft. This variation can be seen in figures 1 and 2 but especially in the graphic from the 2003 collection where the performance of topic identification is lower than LSI for $k > 20$. So a previous statistical analysis to determine what number of factors is the best for a concrete collection will be more important in the topic identification than in the LSI algorithm.

5 Discussion and future work

In [11] different experiments with real relevance feedback and LSI were made. These experiments in several small collections achieved a big improvement in performance. For example, the use of the first relevant document retrieved as the new query is enough to improve the results in all the collections tested. We think that the use of pseudo-relevance feedback techniques can also improve the performance of LSI and, therefore, of our topic identification method in sentence retrieval. In fact, the best precision values obtained in the topic identification experiments can help to make the pseudo-relevance feedback more effective in this method.

The results presented here reinforce the hypothesis that each dimension in the reduced space of the truncated SVD represents a topic of the original documents. The algorithm proposed in [7] starts from the hypothesis that a pattern in the word usage in the original document space can be associated with a topic. LSI claims that it captures the variability in the word usage, reducing terms with similar usage pattern to the same dimension in the reduced space. If these associated terms represent a topic we can use each dimension in the reduced space as the representative of a concept in a document collection. Starting from these ideas Gong & Liu devised the algorithm for generic text summarization.

Following that research line, in our LSI sentence retrieval method each singular value is supposed to represent the importance of those topics, therefore the highest singular values represent the most important topics in the indexed text. In our topic identification approach the retrieval is not directly driven by the singular values. We projected the query in the truncated SVD space and the retrieval of relevant sentences is driven by the most important query components in this reduce space. Therefore, the good results of the method support the hypothesis that the algorithm finds query topics and can retrieve sentences about these topics.

The topic identification method always gets the best result using only one component of the query ($n = 1$). The main reason is that the best results are obtained with a small number of factors. Probably most queries are very focused in a single theme, so the use of one component can be the optimal. We made an individual query analysis, maintaining the same number of factors, and we observed that for some queries better results can be obtained for $n > 1$. This is an expected result because some queries can be about different related themes, in this case the selection of the same number of components as main themes is better. Obviously, this opens a research line whose goal is to produce a topic identification method in which the n parameter could be established for each individual query. In fact this research line is related with some recent work in predicting query performance [12] [13]. In these works the objective is to predict the query difficulty. Weighting functions or query expansion parameters can be changed depending upon the prediction of query difficulty. In our case we are interested in prediction of query topicality but we believe that some of the predictors of query difficulty can be useful for query topicality. It is also interesting to mention another recent work analyzing the reasons of the failure

of queries in different systems [14]: in many cases the query failure comes from the difficulty in determining the query topicality.

As we introduced in the first section, the sentence retrieval task can be seen like a type of summarization. Although the evaluated methods are presented to perform a query-relevant retrieval, they could be adapted to perform generic summaries using, for example, the full document to summarize instead of the query. If the summarization requires text modification, sentence retrieval can be simply the first step of the process, previous to the generation of a new summary.

Another application of sentence retrieval is the presentation of retrieval results. The traditional presentation strategies of web search engines usually include the title, the URL and the terms of the search in the context in which they appear. Some experiments with users [2] examined the effects of the presentation of relevant sentences instead of the documents surrogates. The main goal of using top-ranking sentences is the presentation of content to the users, encouraging users interaction and reducing the number of decisions that the users take to decide the relevance of a document and read it. Indeed, experiments of implicit feedback were developed changing the order of the sentences presented in function of the information extracted implicitly from the user. The experience of the users was positive with respect to these experiments, so the task of sentence retrieval has promising applications.

The novelty collections from the years 2002 and 2003 are very different. Due to the variability in the results obtained and the shifting of the parameters optimized to get these results we plan to repeat the evaluation for the 2004 Novelty track [15] collection. In the 2004 collection the major change is the inclusion of irrelevant documents into the documents sets associated with each topic, actually the irrelevant documents are close matches to the relevant ones, and not random irrelevant documents. After the evaluation with this collection we plan to research about the stability of these techniques given the variability in the input data.

6 Conclusions

Two methods were tested for the task of sentence retrieval. The LSI-based method had not been used before in sentence retrieval. The results of this method are very different for the two collections employed in the evaluation. For the 2002 TREC Novelty track collection the method has a performance only slightly better than a random retrieval, while in the 2003 TREC Novelty track collection the performance is competitive with the best systems tested with this collection.

The second method used is a new method based on topic identification. This method obtains better results than the previous one in the two evaluated collections. In the 2002 collection the performance is not as good as the performance of the best systems participating in the track. For the 2003 collection the results are among the best participant systems and the second human assessor, at least in terms of the F measure.

Most groups participating in the TREC Novelty track used techniques like query expansion or relevance feedback to address the difficulty of matching query and sentence terms. We have proposed two generic methods with a formal background getting good results and still being able to improve the performance.

7 Acknowledgments

The work reported here was co-funded by "Ministerio de Educación y Ciencia" and FEDER funds under research projects TIC2002-00947 and TIN2005-08521-CO2-02.

References

1. Harman, D.: Overview of the TREC 2002 Novelty Track. NIST Special Publication 500-251: The Eleventh Text REtrieval Conference. (2002) 17–28
2. White, R. W., Jose, J. M., Ruthven, I.: Using top-ranking sentences to facilitate effective information access. *JASIST*. **56**(10) (2005) 1113–1125
3. Larkey, L. S., Allan, J., Connell, M. E., Bolivar, A., Wade, C.: UMass at TREC 2002: Cross Language and Novelty Tracks. NIST Special Publication 500-251: The Eleventh Text REtrieval Conference. (2002) 721–732
4. Zhang, M., Song, R., Lin, C., Ma, S., Jiang, Z., Jin, Y., Liu, Y., Zhao, L.: THU TREC 2002: Novelty Track Experiments. NIST Special Publication 500-251: The Eleventh Text REtrieval Conference. (2002) 591–595
5. Berry, M. W., Dumais, S. T., Letsche, T. A.: Computational Methods for Intelligent Information Access. *Supercomputing '95: Proceedings of the 1995 ACM/IEEE conference on Supercomputing*. (1995)
6. Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., Harshman, R. A.: Indexing by Latent Semantic Analysis. *JASIS*. **41**(6) (1990) 391–407
7. Gong, Y., Liu, X.: Generic text summarization using relevance measure and latent semantic analysis. *SIGIR '01, Proceedings of the 24th Annual International ACM SIGIR conference on Research and Development in Information Retrieval*. (2001) 19–25
8. Soboroff, I., Harman, D.: Overview of the TREC 2003 Novelty Track. NIST Special Publication 500-255: The Twelfth Text REtrieval Conference. (2003) 38–53
9. Telcordia Technologies: LSI Software Home. <http://lsi.research.telcordia.com>
10. Rohde, D.: SVDLIBC. <http://tedlab.mit.edu/~dr/SVDLIBC>
11. Dumais, S.: Enhancing Performance in Latent Semantic Indexing. TM-ARH-017527 Technical Report, Bellcore. (1990)
12. Cronen-Townsend, S., Zhou, Y., Croft, W. B.: Predicting Query Performance. *SIGIR '02, Proceedings of the 25th Annual International ACM SIGIR conference on Research and Development in Information Retrieval*. (2002) 299–306
13. He, B., Ounis, I.: Inferring Query Performance Using Pre-retrieval Predictors. *SPIRE '04, Proceedings of the 11th International Conference String Processing and Information Retrieval, LNCS 3246*. (2004) 43–54
14. Buckley, C.: Why current IR engines fail. *SIGIR '04, Proceedings of the 27th Annual International ACM SIGIR conference on Research and Development in Information Retrieval*. (2004) 584–585
15. Soboroff, I.: Overview of the TREC 2004 Novelty Track. NIST Special Publication 500-261: The Thirteenth Text REtrieval Conference Proceedings. (2004)