Domain Specific Retrieval Experiments with MIMOR at the University of Hildesheim

René Hackl, Ralph Kölle, Thomas Mandl, Christa Womser-Hacker

University of Hildesheim, Information Science, Marienburger Platz 22
D-31141 Hildesheim, Germany
`{koelle, mandl, womser}@rz.uni-hildesheim.de`

Abstract. For our first participation in CLEF we chose the domain specific GIRT corpus. We implemented the adaptive fusion model MIMOR (Multiple Indexing and Method-Object Relations) which is based on relevance feedback. The linear combination of several retrieval engines was optimized. As a basic retrieval engine, IRF from NIST was employed. The results are promising. For several topics, our runs achieved a performance above the average. The optimization based on last years' topics and relevance judgements proved to be a fruitful strategy.

1 Introduction

For the CLEF 2002 campaign we tested an adaptive fusion system based on the MIMOR model [Womser-Hacker 1997; Mandl & Womser-Hacker 2001b]. The experiments are done fully automatic and deal with the domain specific corpus of social science documents. We choose GIRT for our first participation data because we are especially interested in the challenges of the domain specific task and because it allows mono-lingual retrieval. As basic retrieval engine for our fusion system, we used the IRF package from NIST.

2 Fusion in Information Retrieval

The basic idea behind fusion is to delegate a task to different algorithms and consider all the results returned. These single results are then combined into one final result. This approach is especially promising, when the single results are very different.

As investigations on the outcome of the TREC conference have shown, the results of information retrieval systems performing similarly well are often different. This means, the systems find the same percentage of relevant documents, however, the overlap between their ranked lists is sometimes low [Womser-Hacker 1997].

An overview of fusion methods in information retrieval is given by [McCabe et al. 1999]. Research concentrates on issues like which methods can be combined, how the retrieval status values of two or more systems can be treated mathematically and which features of collections indicate that a fusion might lead to positive results.

Different retrieval methods can be defined according to various parameters. One possibility is using different indexing approaches, like word and phrase indexing.

The values are combined statically by taking the sum, the minimum or the maximum of the results from the individual systems [Mandl & Womser-Hacker 2001a]. Linear combinations assign a weight to each method which determines its influence on the final result. These weights may be improved for example by heuristic optimization or learning methods [Vogt & Cottrell 1998].

In experimental systems, the methods to be fused are applied to the same collection. However, fusion has been applied to collections without overlap as well. A corpus can be split into artificial sub-sets which are treated by a retrieval system. In such a case, the goal of the fusion can be regarded as an attempt to derive knowledge about which collection leads to good results. For internet meta search engines, fusion usually means elimination of documents returned by at least two search engines.

3 MIMOR

MIMOR (Multiple Indexing and Method-Object Relations) represents a learning approach to the fusion task which is based on results of information retrieval research which show that the overlap between different

systems is often small [Womser-Hacker 1997]. On the other hand, relevance feedback is a very promising strategy for improving retrieval quality. As a consequence, the linear combination of different results is optimized through learning from relevance feedback. MIMOR represents an information retrieval system managing poly-representation of queries and documents by selecting appropriate methods for indexing and matching [Mandl & Womser-Hacker 2001b]. By learning from user feedback on the relevance of documents, the model adapts itself by assigning weights to the different basic retrieval engines. Therefore, MIMOR follows a long term learning strategy in which the relevance assessments are not just used for the current query. MIMOR is not limited to text documents but open to other data types such as structured data and multimedia objects. Learning is implemented as a delta rule. More complex learning schemes should be evaluated [Joachims 1998; Drucker et al. 1999, Iyer et al. 2000].

4 CLEF Retrieval Experiments with the GIRT Corpus

The GIRT data is part of a digital library for the social sciences. The domain specific task is described in [Kluck & Gey 2000].

4.1 Tools

Our main tool was the IRF retrieval package available from NIST[1]. IRF is an object oriented retrieval framework which is programmed in JAVA. The sample IR application proved to meet our retrieval needs sufficiently, which means ranked results, setting of parameters etc. Thus, only few changes were necessary.

In addition we used the Lucene package for topic and data preprocessing. It is written in JAVA, too, and it can be obtained from the Jakarta Apache Project's homepage[2]. Lucene comprises both a built-in German analyzer and a German stemmer apart from plenty other features. Both may easily be altered as desired. The indexing application could be run without major changes, so we slightly modified the filter settings and provided a different, more thorough stop word list[3].
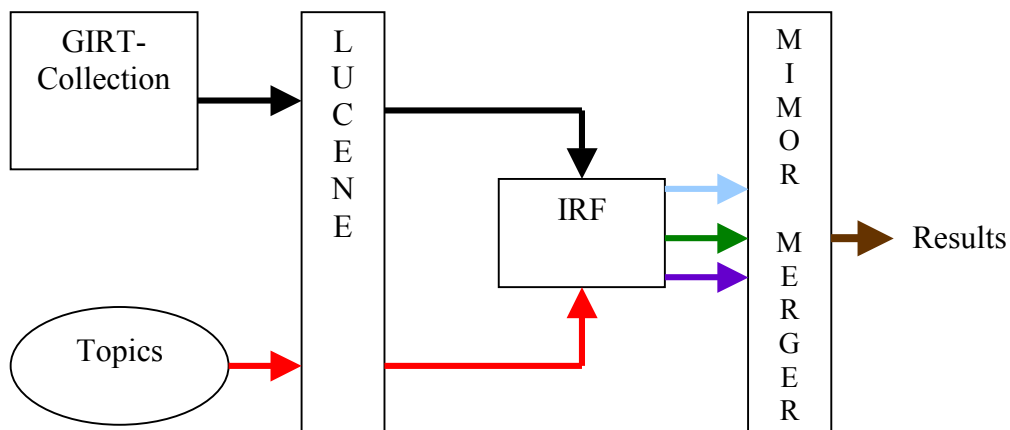

Diagramm 1. Sequence of operations

4.2 Processing steps

The current MIMOR implementation of the University of Hildesheim could not be employed for all processing steps. Therefore the necessary in-between actions were carried out by AWK[4] scripts. We had some formatting

---

[1] http://www.itl.nist.gov/iaui/894.02/projects/irf/irf.html
[2] http://jakarta.apache.org/lucene/docs/index.html
[3] http://www.unine.ch/info/clef/
[4] http://members.cox.net/dos/txtfrmt.htm

done and cut off all information except for but document number, title and text. The most important function however was to simulate MIMOR's merging algorithm (cf. diagramm 1). Several scripts were necessary to statically simulate this functionality.

Three different indexing parameter settings of the IRF – *inverse document frequency* (weight title:1, weight text:1), *inverse document frequency* (weight title:3, weight text:1), *keyword* (weight title:1, weight text:1) - represented the individual retrieval systems.

We also ran the topics from last year with the same setup. Subsequently, we compared the results obtained with the relevant documents assessed for last years campaign. Based on these results we globally optimized the fusion by assigning a higher weight to the system with better results.

For comparison we submitted two runs:

- Run with equal weights for both individual systems

- Run with optimized weights derived from last years results

4.3 Results

As table 1 shows, the average precision for both runs is almost equal. Speaking in absolute figures, run UHi02r2 (where optimized fusion was applied) produced ten more relevant documents.

Table 1. Result overview

Run	Retrieved	Relevant	Rel_ret	Avg. prec.
UHi02r1	23.751	961	387	0.1093
UHi02r2	23.751	961	397	0.1097

Query 67 – where out of five possible relevant documents none was retrieved – is considered an outlier and is left out for the following operations.

Taking the difference between the two precision scores as a method to compare, the fused result was better on nine, equal on two and worse on twelve occasions. However the sum of differences

$$\sum_{i=51}^{75} \text{prec}(r1) \quad - \quad \sum_{i=51}^{75} \text{prec}(r2)$$

produces a slight advantage in favor of the optimized fusion: -0.0089 which is less than one percent and results in a fourth decimal point mean deviation of 0.000387 for each query.

Twenty out of 23 topics (=86.96%) could be found within a 0.0060 range and still nine of 23 (=39.13%) were within a 0.0010 range. However, these deviations are not necessarily statistically significant.

The precision at five documents is at 0.2667 respectively 0.2750, which we consider not too bad for our first participation. Some of our results were above the average of all participants, others were below. The analysis of the particular documents will presumably reveal IRF's strengths and weaknesses.

To summarize, we gained valuable experience from the CLEF experiments. The lack of genuinely different information retrieval engines makes a more sophisticated evaluation of fusion versus non fusion unavailing as can be seen from the overall rather narrow interval scales. Consequently, high priority will be given to the integration of further information retrieval engines in the MIMOR architecture. We expect that a learning process over some cycles will eventually lead to an advantageous system performance.

5 Outlook

Several improvements need to be implemented for next years participation in the GIRT experiments:
- MIMOR needs to be further automated

- The optimization process needs to be refined
- More basic retrieval systems need to be integrated

Furthermore, we plan to further extend our work on the CLEF challenge and participate also in the multi-lingual track with our fusion system.

References

Drucker, H.; Wu, D.; Vapnik, V. (1999): Support Vector Machines for Spam Categorization. In: IEEE Trans. on Neural Networks , vol 10 (5). pp. 1048-1054.

Joachims, T. (1998): Text Categorization with Support Vector Machines: Learning with Many Relevant Features. In: European Conference on Machine Learning (ECML). pp. 137-142.

Kluck, M.; Gey, F. (2000): The Domain-Specific Task of CLEF - Specific Evaluation Strategies in Cross-Language Information Retrieval. In: Peters, Carol (ed.): Cross-Language Information Retrieval and Evaluation. Workshop of the Cross-Language Information Evaluation Forum (CLEF 2000) Lisbon, Portugal, Sept.21-22, 2000. Springer [LNCS 2069] pp. 48-56.

Iyer, R.; Lewis, D.; Schapire, R.; Singer, Y.; Singhal, A. (2000): Boosting for Document Routing. In: Ninth ACM Conference on Information and Knowledge Management (CIKM).

Mandl, T.; Womser-Hacker, C. (2001a): Fusion Approaches for Mappings Between Heterogeneous Ontologies. In: Constantopoulos, Panos; Sølvberg, Ingeborg (eds.): Research and Advanced Technology for Digital Libraries: 5th European Conference (ECDL 2001) Darmstadt Sept. 4-8 2001. Springer [LNCS 2163]. pp. 83-94.

Mandl, T.; Womser-Hacker, C. (2001b): Probability Based Clustering for Document and User Properties. In: Ojala, T. (ed.): Infotech Oulo International Workshop on Information Retrieval (IR 2001). Oulo, Finnland. Sept 19-21 2001. pp. 100-107.

McCabe, M. Catherine; Chowdhury, A.; Grossmann, D.; Frieder, O. (1999): A Unified Framework for Fusion of Information Retrieval Approaches. In Eigth ACM Conference on Information and Knowledge Management (CIKM). pp. 330-334.

Vogt, C.; Cottrell, G. (1998): Predicting the Performance of Linearly Combined IR Systems. In: 21th Annual Intl ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '98). pp. 190-196.

Womser-Hacker, C. (1997): Das MIMOR-Modell. Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval. Habilitationsschrift. Universität Regensburg, Informations-wissenschaft.