

The Xtrieval Framework at CLEF 2008: Domain-Specific Track

Jens Kürsten, Thomas Wilhelm and Maximilian Eibl
Chemnitz University of Technology
Faculty of Computer Science, Dept. Computer Science and Media
09107 Chemnitz, Germany
[jens.kuersten | thomas.wilhelm | maximilian.eibl] at cs.tu-chemnitz.de

Abstract

This article describes our participation at the *Domain-Specific track*. We used the *Xtrieval* framework [2], [3] for the preparation and execution of the experiments. The translation of the topics for the cross-lingual experiments was realized with a plug-in to access the Google AJAX language API². This year, we submitted 20 experiments in total. In all our experiments we applied a standard top-k pseudo-relevance feedback algorithm. Also, all of our submissions were merged experiments, where multiple stemming approaches for each language were combined to improve retrieval performance. The evaluation of the experiments showed that the combination of stemming methods works very well. Translating the topics for the bilingual experiments deteriorated the retrieval effectiveness only between 8 and 15 percent in comparison to our best monolingual experiments.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval

Keywords

Evaluation, Cross-Language Information Retrieval, Domain-Specific Retrieval

1 Introduction and Outline

The *Xtrieval* framework [2],[3] was used to prepare and execute this years domain-specific text retrieval experiments. The core retrieval functionality is provided by Apache Lucene¹. For the *Domain-Specific track* three different corpora mainly with sociological content in German, English and Russian were employed. We conducted monolingual experiments on each of the collections and also submitted experiments for the bilingual and multilingual subtasks. For the translation of the topics the Google AJAX language API² was accessed through a JSON³ programming interface. We also used the provided bilingual thesauri to find out how much they can help in translating domain-specific terms.

The remainder of the paper is organized as follows. Section 2 describes the general setup of our system. The individual configurations and the results of our submitted experiments are presented in section 3. In sections 4 and 5 we summarize the results and sum up our observations.

¹<http://lucene.apache.org>

²<http://code.google.com/apis/ajaxlanguage/documentation>

³<http://json.org>

2 Experimental Setup

The approach we used for this years participation is mainly based on the following ideas. At first we combine several stemming methods for each language in the retrieval stage. The combination of the results was done by our implementation of the *Z-Score* operator [4]. We compared standard retrieval experiments with query expansion based on the provided domain-specific thesauri to investigate their impact in terms of retrieval effectiveness.

3 Configurations and Results

The detailed setup of our experiments and the results of the evaluation are presented in the following subsections.

3.1 Monolingual Experiments

We submitted 5 monolingual experiments in total, 2 for the English and the German subtasks and 1 for the Russian subtask. For all experiments a language-specific stopword list was applied⁴. We used different stemmers for each language: Porter⁵ and Krovetz [1] for English, Snowball⁵ and a n-gram variant decompounding stemmer⁶ for German as well as an Java implementation of a stemmer⁴ for Russian. For two experiments the provided thesauri were used for query expansion (tqe) and in all experiments a standard pseudo-relevance feedback algorithm for top-k documents was used. In table 1, the retrieval performance of our experiments is presented in terms of mean average precision (map) and the absolute rank of the experiment in the evaluation.

Table 1: Experimental Results for the monolingual subtask

<i>id</i>	<i>lang</i>	<i>tqe</i>	<i>map</i>	<i>rank</i>
cut_merged	DE	no	0.4367	3/10
cut_merged_thes	DE	yes	0.4359	4/10
cut_merged	EN	no	0.3891	1/12
cut_merged_thes	EN	yes	0.3869	2/12
cut_merged	RU	no	0.0955	9/9

Our experiments on the German and English collections had very strong overall performance. In contrast to that our experiment on the Russian collection performed very bad. It is also obvious that the thesaurus based query expansion did not improve the retrieval performance, but at least it did not significantly deteriorate the effectiveness.

3.2 Bilingual Experiments

We submitted 12 experiments in total for the bilingual subtask, i.e. 4 experiments were submitted for each target language collection. We compared the translation from different source languages and the performance of pure topic translation with combined translation. For the combined translation we used the pure topic translation and tried to improve the translation with the help of the bilingual thesauri, i.e. for every term occurring in the bilingual thesauri we added its provided translation to the topic. Again, we used a standard pseudo-relevance feedback algorithm to improve retrieval effectiveness. In Table 2 we compare each of the bilingual experiments with respect to the performance of the corresponding monolingual experiment.

Probably due to the quality of Google’s translation service and the strong performance of our monolingual runs the retrieval effectiveness of our bilingual experiments is also very good. Surprisingly one of our bilingual experiments on the Russian target collection performed best, although our monolingual experiment had the worst overall performance. This is thought to be due to the smaller number of submissions for the bilingual

⁴<http://members.unine.ch/jacques.savoy/clef/index.html>

⁵<http://snowball.tartarus.org>

⁶<http://www-user.tu-chemnitz.de/wags/cv/clr.pdf>

Table 2: Experimental Results for the bilingual subtask

<i>id</i>	<i>lang</i>	<i>tqe</i>	<i>map</i>	<i>rank</i>
cut_merged	DE	no	0.4367	3/10
cut_merged_em2de	EN→DE	no	0.3702 (-15.23%)	1/12
cut_merged_en2de_thes	EN→DE	yes	0.3554 (-18.62%)	2/12
cut_merged_ru2de	RU→DE	no	0.3244 (-25.72%)	3/12
cut_merged_ru2de_thes	RU→DE	yes	0.2843 (-34.90%)	4/12
cut_merged	EN	no	0.3891	1/12
cut_merged_ru2en	RU→EN	no	0.3385 (-13.00%)	1/9
cut_merged_de2en	DE→EN	no	0.3363 (-13.57%)	2/9
cut_merged_ru2en_thes	RU→EN	yes	0.3276 (-15.81%)	3/9
cut_merged_de2en_thes	DE→EN	yes	0.3135 (-19.43%)	4/9
cut_merged	RU	no	0.0955	9/9
cut_merged_en2ru	EN→RU	no	0.0882 (-07.64%)	1/8
cut_merged_de2ru	DE→RU	no	0.0681 (-28.69%)	3/8
cut_merged_en2ru_thes	EN→RU	yes	0.0597 (-37.49%)	5/8
cut_merged_de2ru_thes	DE→RU	yes	0.0499 (-47.75%)	8/8

subtask, which can also be seen in the spread of the ranks of our bilingual Russian experiments. Again, the translation supported by the provided thesauri did not improve the retrieval effectiveness, but with the exception of one experiment (cut_merged_ru2de_thes) it did not deteriorate the performance significantly.

3.3 Multilingual Experiments

For the participation at the multilingual subtask 3 experiments were submitted. Topics in all three languages were used, with one language as source for one experiment. All three target collections were queried for each multilingual experiment. The results of the evaluation are shown in table 3.

Table 3: Experimental Results for the multilingual subtask

<i>id</i>	<i>lang</i>	<i>map</i>	<i>rank</i>
cut_merged_de2x	DE→X	0.2816	1/9
cut_merged_en2x	EN→X	0.2751	2/9
cut_merged_ru2x	RU→X	0.2357	3/9

The retrieval performance of our multilingual experiments was very good, especially in comparison to the experimental results of the years before^{7,8,9}. We assume this to be due to Google’s translation service on the one hand but also to the result list fusion algorithm of the *Xtrieval* framework. It is obvious that the performance is almost equal for the experiments, where we used the German and English topics. The experiment with the Russian topic has a small decline in retrieval effectiveness.

4 Result Analysis - Summary

The following list provides a summary of the analysis of our retrieval experiments for the *Domain-Specific track* at CLEF 2008:

- *Monolingual*: The performance of our monolingual experiments was very good for the German and English collections and worse for the Russian collection. Interestingly, the retrieval effectiveness could not be improved by utilizing the provided domain-specific thesauri for query expansion.

⁷http://www.clef-campaign.org/2005/working_notes/workingnotes2005/appendix_a.pdf - p. 61

⁸http://www.clef-campaign.org/2006/working_notes/workingnotes2006/Appendix_Domain_Specific.pdf - p. 63

⁹http://www.clef-campaign.org/2007/working_notes/AppendixC.pdf - p. 206

- *Bilingual*: Probably due to the used translation service our bilingual experiments performed very well and achieved the best results on each target collection. Astonishingly, we could not improve the retrieval performance by using the provided bilingual thesauri.
- *Multilingual*: Again, mainly due to the quality of the translation and the result list combination capabilities of the *Xtrieval* framework we achieved very impressive results in term of retrieval effectiveness. There was no significant difference between the experiments with English and German topics.

5 Conclusion

This year, we achieved very good retrieval performance in almost all subtasks of the *Domain-Specific track*. Since our main research focus shifted to Multimedia Information Retrieval there were no interesting contributions to retrieval community in this work, except for the fact that combining different stemming approaches helped to improve retrieval performance. Another important observation in all our experiments for this years CLEF campaign was that the translation service provided by Google seems to be extremely superior to any other approach or system. This should motivate the cross-language community to investigate and improve their current approaches.

Acknowledgments

We would like to thank Jaques Savoy and his co-workers for providing numerous resources for language processing. Also, we would like to thank Giorgio M. di Nunzio and Nicola Ferro for developing and operating the DIRECT system¹⁰.

This work was partially accomplished in conjunction with the project *sachsMedia*, which is funded by the *Entrepreneurial Regions*¹¹ program of the German Federal Ministry of Education and Research.

References

- [1] Robert Krovetz. Viewing morphology as an inference process. In *SIGIR '93: Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 191–202, New York, NY, USA, 1993. ACM.
- [2] Jens Kürsten, Thomas Wilhelm, and Maximilian Eibl. The xtrieval framework at clef 2007: Domain-specific track. In C. Peters, V. Jijkoun, Th. Mandl, H. Müller, D.W. Oard, A. Peñas, V. Petras, and D. Santos, editors, *LNCS - Advances in Multilingual and Multimodal Information Retrieval*, volume 5152, Berlin, 2008. Springer Verlag.
- [3] Jens Kürsten, Thomas Wilhelm, and Maximilian Eibl. Extensible retrieval and evaluation framework: Xtrieval. *LWA 2008: Lernen - Wissen - Adaption, Würzburg, October 2008, Workshop Proceedings*, October 2008, to appear.
- [4] Jaques Savoy. Data fusion for effective european monolingual information retrieval. *Working Notes for the CLEF 2004 Workshop, Bath, UK*.

¹⁰<http://direct.dei.unipd.it>

¹¹The Innovation Initiative for the New German Federal States