In: Constantopoulos, Panos; Sølvberg, Ingeborg (eds.):

Research and Advanced Technology for Digital Libraries:

5th European Conference (ECDL 2001) Darmstadt 4.-8.9.2001.

Berlin et al.: Springer [Lecture Notes in Computer Science 2163]. pp. 83-94.

Fusion Approaches for Mappings between Heterogeneous Ontologies

Thomas Mandl, Christa Womser-Hacker

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

Abstract. Ordering principles of digital libraries expressed in ontologies may be highly heterogeneous even within a domain and especially over different cultures. Automatic methods for mappings between different ontologies are necessary to ensure successful retrieval of information stored in virtual digital libraries. Text categorization has discussed learning methods to map between full text terms and thesaurus descriptors. This article reports some experiments for the mapping between different ontologies and shows further that fusion methods which have been successfully applied to ad-hoc information retrieval can also be employed for text categorization.

1 Introduction

Ontologies are organized collections of concepts. Their structure expresses a certain view on the world or the domain. A considerable number of ontologies exists in many domains. This especially applies for different countries and cultures. Many ordering systems have concepts in common, however, because they arose within a certain context in response to special demands they may organize them differently. Although this situation is quite natural and not likely to change, it complicates or even prevents the successful communication between communities using different ontologies. The same is true for the exchange of documents between different groups within one digital library. The worldwide exchange of documents involves ontologies from different cultures which are sometimes organized extremely different. Due to the ever growing amount of knowledge available electronically, automated solutions need to be found for these mapping problems.

Text categorization between ontologies or terminologies seems to be the most appropriate technology for this task. Mostly, text categorization assign documents to predefined categories based on a full text analysis [27]. The text is indexed with standard information retrieval methods and represented by weights assigned to words or terms based on their frequency of occurrence. These terms can be regarded as features. In the same manner, terms from a controlled vocabulary like an ontology can serve as features for a text. Thus, the task for text categorization based on full text terms is equivalent to text categorization based on descriptors from ontologies. There are also applications for the second case [4].

Similarity thesauri for cross-lingual retrieval can also be viewed as an additional type of text categorization. In this case, the mapping leads from one set of free text terms to another set of free text terms [20]. Other applications of text categorization include recommender systems [18] and spam detection [6].

This article reviews some learning methods applied to text categorization in the following chapter. Chapter 3 shows how fusion methods are applied to ad-hoc retrieval. Chapter 4 discusses text categorization experiments for real world social science texts and applies fusion algorithms to these text categorization tasks.

2 Learning Methods in Text Categorization

Different learning methods have been applied to text categorization. Most often, statistical association measures like Naive Bayes map between pairs of terms. These learning algorithms derive the knowledge from examples provided as training data and do not rely on further human contribution.

In recent years neural networks and support vector machines have been employed as well [8, 11]. An overview is provided by Aas and Eikvil [1]. Many machine learning methods are available as JAVA code in the WEKA package for data mining [23].

Neural networks and in particular the multi-layer backpropagation network are learning schemes well suited for text categorization tasks [16]. Backpropagation networks are powerful learning systems which can learn complex functions such as non linear separable problems [10]. Neural networks can approximate intuitive expert knowledge which cannot be formalized in a rule based system. Calculations within a neural network are local. Each unit gathers its input from incoming connections and calculates its own activation and output. Signals travel along connections and are modified according to the connection strength. A typical backpropagation network consists of an input and an output layer and of one or more hidden layers. The input vector is propagated into the network which calculates the output [10].

Crestani and Rijsbergen present a backpropagation network for a mapping between different queries in which the representation schemes are equivalent. The same architecture can be modified for general transformations between heterogeneous representation schemes. A query was transformed into queries which achieved better results. Improved queries for training were found using relevance feedback. The transformed queries achieved similar retrieval quality compared to the original query [5]. The process can be seen as a query extension.

Other neural networks have been applied to text categorization tasks as well, e.g. a Hopfield style network to map between two thesauri. In one setting, descriptors from parts of the *ACM Computing Review Classification System* and the *Library of Congress Subject Headings* are represented as neurons in a Hopfield network [4]. However, the learning capabilities of Hopfield networks are restricted to unsupervised learning, therefore the knowledge in the examples typically available in text categorization is not fully exploited. The connections strengths are set according to user defined clusters which are also represented as neurons [4].

A similar approach is presented by Lee and Dubin [15]. Their neural network consists of separated layers for two thesauri. In this application, 4120 terms from the *Astrophysical Journal* and 2305 terms from the *NASA* were mapped. A layer in between the term layers representing the common 15,000 documents defines the transformation. The connection strengths are set according to the co-occurrences of terms in the documents.

3 Fusion Methods

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. Obviously, this approach is especially promising, when the single results are very different.

As investigations on the results 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 results is sometimes low [24].

For retrieval tasks a fusion system needs to integrate different probabilities for the relevance of a document whereas in text categorization fusion means the integration of various probabilities for the assignment of a term from an ontology.

3.1 Committee Machines in Machine Learning

In machine learning, the combination of various supervised learning algorithms for approximation, prognosis or classification is often called a committee machine. This metaphor suggests that these algorithms try to work like a committee of human experts who need to reach a decision based on their individual opinions. The following architectures are used [10]:

- Static methods: the input of single experts does not influence the fusion
    - Ensemble averaging: the output of different experts is combined linearly
    - Boosting: a weak learning algorithm is improved by retraining badly classified examples with another learning algorithm

- Dynamic methods: the input from the experts governs the integration process
    - Mixture of experts: the output of several experts is combined non-linearly
    - Hierarchical mixture of experts: the combination system is organized in a hierarchical manner

All these fusion methods do not need to be based on completely different experts. The results of a neural net usually depends on its random initialization. Thus, different initialization states of the same neural net can be considered as different experts. The same applies to other learning methods with different parameter settings.

3.2 Fusion in Information Retrieval

An overview of fusion methods in information retrieval is given by McCabe et al. [17]. Research concentrates on issues like which methods can be combined, how the retrieval status values of two or more systems can be combined and on 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 [2]. Lee defines the single retrieval methods for his fusion approach according to their different term weighting schemes [14]. He investigates with cosine normalization which takes the varying length of documents into account, maximum normalization based on the maximal term weights and no normalization. Another parameter is the similarity function used [17].

The values are combined statically by taking the sum, the minimum or the maximum of the results from the individual systems [7]. Linear combinations assign a weight to each method which determines its influence on the final result. These weights may be optimized for example by heuristic optimization [21] or learning methods [22, 25].

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 collection can be split into artificial sub-collections which are treated by an retrieval system [22]. 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 often means elimination of documents returned by at least two search engines.

3.3 Fusion and Learning

The ideas from information retrieval and machine learning are combined in the algorithm RankBoost, which applies boosting to ranked result lists [13].

MIMOR represents a learning approach to the fusion task which is based on results of information retrieval research which shows that the overlap between different systems is often small [24]. 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 integrates an information retrieval system managing poly-representation of queries and documents by selecting appropriate methods for indexing and matching. By learning from user feedback on the relevance of documents, the model adapts itself by assigning weights to the different retrieval methods. That way, 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. A formal model of MIMOR has been developed [25]. The conclusion of this paper will relate the MIMOR model to the fusion problem in text categorization.

4 Text Categorization Experiments

This chapter reviews experiments for text categorization for two corpora with a statistical approach and a neural network and analyzes the results.

The data used is part of a digital library for the social sciences [12] and some of it is part of the German collection for the cross language evaluation forum (CLEF), where cross lingual retrieval methods are tested in a way similar to the TREC evaluation studies [19].

The institutions providing the data are the Social Science Information Center, Bonn (IZ) and the library of the University of Cologne with its special collection for the social sciences (USB). Both institutions rely on intellectual indexing. The IZ represents the documents according to two representation schemes:

- the thesaurus: a collection of some 22.000 keywords and synonym relations
- the classification: a hierarchy of scientific disciplines from the perspective of the social sciences containing 157 entries

That way, the documents from the IZ can be used for text categorization. In this paper, a mapping from the thesaurus to the classification is presented.

The USB uses a different thesaurus for content analysis. Studies have shown, that there is a significant overlap between the collections and that therefore, many documents are indexed twice [12]. For these documents, text categorization tasks from the USB representations to the IZ schemes have been implemented. Within a digital library, these automatic transformations can support the user in dealing with the semantic heterogeneity. Terms defined differently by the information providers may cause low retrieval quality for users only familiar with one thesaurus, because they will find documents with unexpected usage of these terms. Users familiar with one representation scheme do not have to understand the other and may still access the documents provided only by one information provider. For such documents, an automatic text categorization system could implement a value added service and derive the relevant terms for the target thesaurus [12].

4.1 Categorization Tasks

As a first task, a mapping from the IZ thesaurus to the IZ classification was chosen. For this experiment, a subset of the IZ databases was selected which is part of CLEF and contains some 12.900 documents. In average each document is assigned 13 descriptors from the thesaurus and 2.3 classification entries. Especially for the number of thesaurus descriptors assigned to each document, the variance is high.

Because the subset did not contain all keywords and classification entries, the experiment was not set up as a direct transformation between thesaurus and classification. Only 5555 out of 22.000 thesaurus entries occurred in the 12.900 documents. Terms appearing at least four times were selected, which led to a representation scheme of 3800 terms. The data set contained 142 of all 157 classification categories. They were further accumulated intellectually to form 70 categories. That way, the task was to map a document described by a term vector of 3.800 elements into 70 categories. The training set contained 12,000 documents and the test set some 900.

The second task took advantage of the overlap between IZ and USB. Some 15,200 documents in common were identified. The second task is a mapping from the USB thesaurus with 3000 terms occurring in the set to 100 categories of the IZ thesaurus. The training set contained 13,000 and the test set some 2,200 documents. The IZ had assigned 11 thesaurus terms per document in average and the USB 2.9 terms.

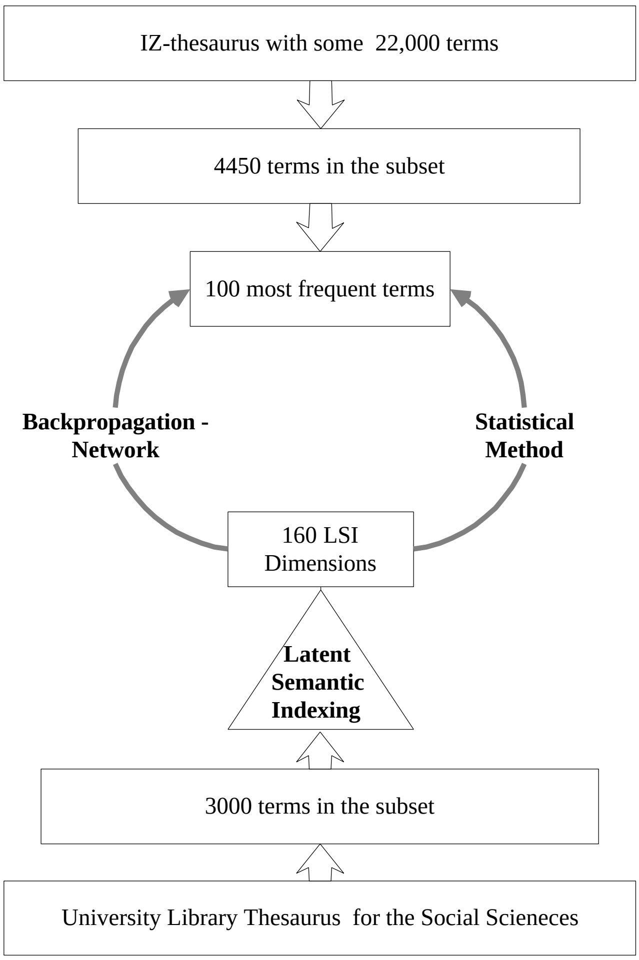

Fig. 1. Second transformation task.

The input spaces of the two tasks underwent a dimensionality reduction using latent semantic indexing [3]. A simple association based statistical method was employed as a first categorization method. It compares the frequency of a co-occurrence of term and category to the frequencies of the occurrence of term and category. As a second categorization method, a neural backpropagation network was applied to both tasks. Figure 1 shows the reduction of features in the second task.

4.2 Evaluation Method and Results

The performance of text categorization is commonly measured with recall and precision. The well known formulas used in information retrieval evaluation can be applied to text categorization. Relevant documents are replaced by relevant categories or correct assignments. In this case, correct assignments means the reproduction of the human classification. The identical formulas suggest that the measures express the same as in information retrieval evaluation. However, recall and precision in text categorization mean something different.

The number of relevant documents is usually determined considering the information need of a user and his relevance judgements. Often domain experts evaluate the documents, nevertheless, the basis for this assessment is still an information need. On the other hand, the assignment of relevant categories by a text categorization method merely shows how well an algorithm can approximate the categorization defined by a human. This categorization is carried out independently from any information need. It is unclear, how this relates to the success of a retrieval process. This needs to be considered when measuring the quality of text categorization methods with recall and precision.

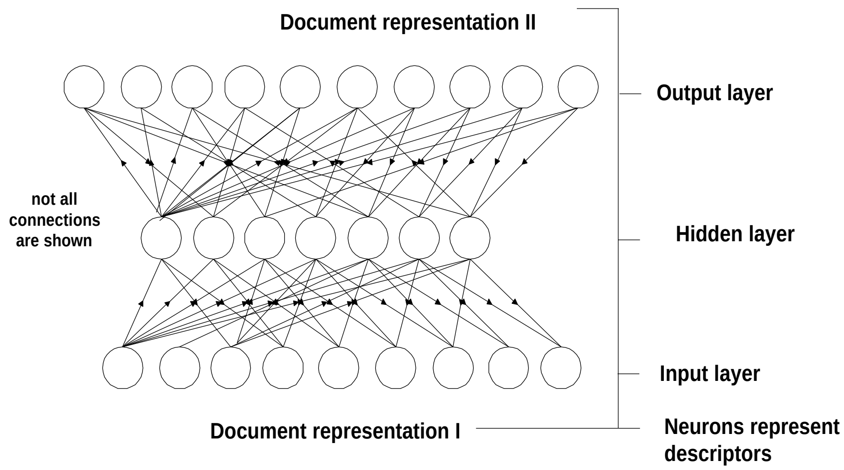

Fig. 2. Typical architecture of a backpropagation network for text categorization

For the first task, both methods performed almost equally well, while the overlap between the results was rather low. Therefore, a fusion of the results was considered in order to improve categorization. For the second task, the neural network performed better than the statistical method. This shows, that the performance depends much on the collection under consideration.

Table 1. Result overview

Task	Documents	Input dimensions	Target dimensions	Result
IZ-Thes. to IZ-Classific.	12,965	5555	70	Almost equal quality
USB-Thes. to IZ-Thes.	15,000	3000	100	Neural network much better

In order to estimate whether fusion methods could lead to a positive result, the overlap of the results for the first task were analyzed The analysis was done for the test set of the categorization task. The similarity between the corresponding results from both text categorization methods was analyzed in two ways. First, the correlation between the ranked lists was calculated. Secondly, the overlap of the result sets was measured, after considering a threshold in the ranked list.

The correlation between two ranked list may be calculated with rank correlation coefficients like the Spearman coefficient [9]. The Spearman correlation measure were calculated for each pair of result sets. The average was then calculated as the correlation between the results. The average correlation reaches only –0.05, which means that according to this coefficient, the results do not correlate and are very different.

Additionally, the overlap between result sets was determined. The ranked lists were cut off after n categories with the highest probability assigned to them by the categorization method. The number of common categories in both sets and the number of relevant or correct categories according to the test data was counted. Table 2 gives an overview of the overlap calculated.

Table 2. Overlap measures for result sets

	Categories found by both methods	in %	Relevant categories found by both methods	in %
First 20 categories	0.27	5.43	0.14	0.71
First 100 categories	4.92	4.92	0.83	0.83

The overlap of the results of the two methods for the categorization task is rather small. For all categories, it reaches some 5% for both thresholds and lies below 1% for the relevant categories only. The quality of the results is almost identical for all recall levels as table 3 shows. That means both methods find approximately the same number of relevant categories, but they find different ones.

5 Fusion Experiments for Text Categorization

The categorization results of both methods were fused using static approaches suggested in [7]. Since the results of the neural network depends much on the random initialization, two differently initialized networks were used for the fusion. Sum, minimum and maximum of both methods were tested.

For the first task, the fusion did not lead to a consistent improvement which would be reflected by the average precision. Only for some recall levels, the best fusion experiments led to an improvement compared to the best single categorization method. The improvement reaches up to 5 % as table 3 shows.

Table 3. Fusion results for the first categorization task

Recall	Neural network I	Neural network II	Statistics	Best Fusion Results:	Fusion stat + NN1 SUM	Fusion stat + NN1 MIN	Maximum improvement in %
0.1	0.306	0.306	0.313		0.321	0.298	+ 2.56
0.2	0.258	0.258	0.262		0.259	0.256	-
0.3	0.213	0.213	0.217		0.223	0.229	+ 5.53
0.4	0.197	0.197	0.201		0.199	0.205	+ 1.99
0.5	0.183	0.183	0.187		0.177	0.177	-
0.6	0.161	0.161	0.166		0.157	0.157	-
0.7	0.136	0.136	0.140		0.143	0.144	+ 2.86
0.8	0.119	0.119	0.123		0.118	0.121	-
0.9	0.102	0.102	0.106		0.092	0.097	-
Average Precision	**0.186**	**0.186**	**0.191**		**0.188**	**0.187**	

For the second task, the fusion leads to an improvement compared to the best single result. For this task, the difference between the single results are significant with the statistic approach performing much weaker than the neural network. However, this weak result still contributed to an improvement of 3 % to the maximum fusion method and of 0.4 % to the sum fusion method. The results of the fusion for the second task are shown in table 4.

Table 4. Fusion results for the second categorization task

Recall	Neural network	Statistic approach	Fusion: MAX	Fusion: SUM	Fusion: MIN
0.1	0.515	0.059	0.525	0.503	0.052
0.2	0.407	0.05	0.418	0.405	0.047
0.3	0.334	0.047	0.351	0.342	0.044
0.4	0.277	0.044	0.287	0.284	0.042
0.5	0.206	0.041	0.211	0.211	0.04
0.6	0.151	0.039	0.154	0.153	0.039
0.7	0.107	0.038	0.11	0.108	0.038
0.8	0.075	0.037	0.077	0.075	0.037
0.9	0.055	0.037	0.055	0.055	0.036
Average Precision	**0.236**	**0.044**	**0.243**	**0.237**	**0.042**

The experiments show that fusion can lead to improvements in text categorization quality. The results differ very much between different corpora. Therefore, careful tests need to be carried out for each domain. The choice of individual and fusion methods has a great impact on the result as well. Furthermore, other fusion methods need to be tested.

6 Conclusion

Different representation schemes express different views of the world or the domain. Varied approaches to the concepts of a domain may result in drastically different ontologies used for intellectual or automatic indexing. Concepts may be ordered differently and the hierarchical organization may be incomparable. Text categorization can be viewed as a transformation between heterogeneous representations for objects stored in digital libraries and there will be a growing demand for applications based on this technology.

The article reviews two text categorization experiments for collections from the social sciences using a neural network and a statistical approach. As the overlap between the results of the two methods is rather small, fusion for text categorization is introduced as a strategy to improve the overall result. For one of the two tasks, the average precision could be increased by 3 %.

Further research should investigate the effect of partitioning the learning space and the application of different learning methods for each partition. The partition can group features of clusters of documents. For example, an optimal learning scheme can be identified for a mapping from a group of terms to a another group of terms in the target ontology. The MIMOR-model provides a framework for such an approach. MIMOR can also be used to optimize a linear fusion based on user feedback [26].

References

1. Aas, K, Eikvil, L. (1999): Text Categorization: a Survey. Report No. 941, June, 1999 Norwegian Computing Center. http://citeseer.nj.nec.com/aas99text.html
2. Bartell, B, Cottrell, G, Belew, R (1994): Automatic Combination of Multiple Retrieval Systems. In 17th Annual Intl ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ´94).
3. Berry, M, Dumais, S., Letsche, T. (1995): Computational Methods for Intelligent Information Access. In ACM Supercomputing '95. San Diego, CA. pp. 1-38.
4. Chen, H. (1995): Machine Learning for Information Rerieval: Neural Networks, Symbolic Learning, and Genetic Algorithms. In Journal of the American Society for Information Science. JASIS 46 (3). pp. 194-216.
5. Crestani, F., Rijsbergen, K. van (1997): A Model for Adaptive Information Retrieval. In Journal of Intelligent Information Systems 8 (1). S. 29-56.
6. 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.
7. Fox, E., Shaw, J (1994): Combination of Multiple Searches. In Harman, D. (ed.): The Second Text Retrieval Conference. NIST Special Publication 500-215.
8. Goller, C., Löning, J., Will, T., Wolff, W. (2000): Automatic Document Classification: A thorough Evaluation of Various Methods. In Knorz, G., Kuhlen, R. (eds.): Informationskompetenz - Basiskompetenz in der Informationsgesellschaft. Proceedings 7. Intl Symposium für Informationswissenschaft. (ISI 2000) pp. 145-162.
9. Hartung, J. (1984): Lehr- und Handbuch der angewandten Statistik. München, Wien.
10. Haykin, S. (1999): Neural Networks: A Comprehensive Foundation.
11. Joachims, T. (1998): Text Categorization with Support Vector Machines: Learning with Many Relevant Features. In European Conference on Machine Learning (ECML). pp. 137-142.
12. Kluck, M., Krause, J., Müller, M., Schmiede, R., Wenzel, H., Winkler, S., Meier, W. (2000): Virtuelle Fachbibliothek Sozialwissenschaften. IZ technical report 19, Informationszentrum Sozialwissenschaften, Bonn. http://www.bonn.iz-soz.de/publications/series/working-papers
13. 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).
14. Lee, J. H. (1995): Combining Multiple Evidence from Different Properties of Weighting Schemes. In: 18th Annual Intl ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 95). pp. 180-188.
15. Lee, J., Dubin, D. (1999): Context-Sensitive Mapping with a Spreading Activation Network. In 22nd Annual Intl ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99). pp. 198-205.
16. Mandl, T. (1998): Vague Transformations in Information Retrieval. In Zimmermann, H., Schramm, V. (Eds.): Knowledge Management und Kommunikationssysteme: Workflow Management, Multimedia, Knowledge Transfer. 6. Intl Symposium für Informationswissenschaft. (ISI '98). pp. 312 – 325.
17. McCabe, M. Catherine, Chowdhury, A., Grossmann, D., Frieder, O. (1999): A Unified Framework for Fusion of Information Retrieval Appproaches. In Eigth ACM Conference on Information and Knowledge Management (CIKM). pp. 330-334.
18. Mooney, R., Roy, L. (2000): Content Based Book Recommending Using Learning for Text Categorization. In Fifth ACM Conference on Digital Libraries.
19. Peters, C. (ed.) (2001): Cross-Language Information Retrieval and Evaluation: Proceedings of the CLEF 2000 Workshop. Springer [LNCS 2069]
20. Sheridan, P., Ballerini, J.P. (1996): Experiments in Multilingual Information Retrieval using the SPIDER System. In 19th Annual Intl ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '96). pp. 58-65.

21. 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.
22. Voorhees, E., Gupta, N., Johnon-Laird, B. (1995): Learning Collection Fusion Strategies. In 18th Annual Intl ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '95). pp. 172-179.
23. Witten, I., Frank, E. (2000): Data Mining: Practical Machine Learning Tools and Techniques with JAVA Implementations.
24. Womser-Hacker, C. (1997): Das MIMOR-Modell. Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval. Habilitationsschrift. Universität Regensburg, Informationswissenschaft.
25. Womser-Hacker, C., Mandl, T. (1999): Adapting Meta Information Retrieval to User Preferences and Document Features. In Bullinger, H-J., Ziegler, J. (eds.): Human-Computer Interaction: Communication, Cooperation and Application Design. HCI International ´99, Munich, Germany, 22-27. August 1999. pp. 604-608.
26. Womser-Hacker, C., Mandl, T. (2000): Ein adaptives Information Retrieval Modell für Digitale Bibliotheken. In Knorz, G., Kuhlen, R. (eds.): Informationskompetenz - Basiskompetenz in der Informationsgesellschaft. Proceedings 7. Intl Symposium für Informationswissenschaft. (ISI 2000) pp. 1-16.
27. Yang, Y., Liu, X. (1999): A re-examination of text categorization methods. In 22nd Intl ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 42-49.