Diversity in Recommender Systems

Bridging the gap between users and systems

Laurent Candillier[*], Max Chevalier[**], Damien Dudognon[*,**], Josiane Mothe[**,***]

[*]OverBlog	[**]IRIT, UMR5505	[***]IUFM
WikioGroup	University of Toulouse	University of Toulouse
Toulouse, France	Toulouse, France	Toulouse, France
firstname.name@wikiogroup.com	firstname.name@irit.fr	firstname.name@toulouse.iufm.fr

Abstract—Recommender systems aim at automatically providing objects related to user's interests. The angular stone of such systems is a way to identify documents to be recommended. Indeed, the quality of these systems depends on the accuracy of its recommendation selection method. Thus, the selection method should be carefully chosen in order to improve end-user satisfaction. In this paper, we first compare two sets of approaches from the literature to underline that their results are significantly different. We also provide the conclusion of a survey done by thirty four students showing that diversity is considered as important in recommendation lists. Finally, we show that combining existing recommendation selection methods is a good mean to obtain diversity in recommendation lists.

Keywords-Information Retrieval; Recommender System; Document similarity; Diversity

I. INTRODUCTION

Information Retrieval (IR) usually sorts the retrieved documents according to their similarity with the user's query. Doing so, they assume that document relevance can be calculated independently from other documents [10]. As opposed to this assumption, various studies consider that a user may prefer to get documents that treat of various aspects of her information need rather than possibly redundant aspects within documents [6][34]. In a recent user study, in which thirty four M. Sc. students in management studies were questioned, we found that more than 80% of the students prefer a system that provides document diversity. They indicate that they want document diversity in order to get a more complete map of the information.

Document diversity has many applications. It is considered to be one solution to query term ambiguity. Indeed, queries as expressed by users are not enough to disambiguate terms. To answer such queries, IR can provide the user with a range of documents that corresponds to the various term senses [10]. In that case, redundancy can be penalized by lowering the rank of a document that is too similar to a document already ranked.

Diversity can also be useful in the context of social network analysis [36] and recommender systems [21].

Recommender Systems (RS) aim at providing the user with items related to the current browsed item. Relationships between items can correspond to large range of users' interests and should be linked to their interest in document diversity. Indeed, relevance can be evaluated regarding various criteria. Mothe and Sahut [27] consider the following criteria for information relevance:
- Related to the topic;
- Novelty;
- Understandability;
- Type of media and length;
- Completeness and scope.

In an RS, document similarities are used to associate a score with documents to be recommended. Similarity measures are either based on document content or structure, or on document usage considering popularity or collaborative search.

Similarities from the literature include:
- Similarities based on document content: to be similar two documents should share indexing terms. Example of such measures are the cosine measure [29], or semantic measures [11][32];
- Similarities based on document popularity such as the BlogRank [20];
- Collaborative similarity measures: The document score depends on the scores that previous users assigned to it [1];
- Browsing and classification similarities: document similarity is based either on browsing path [12] or considering the categories users made [5];
- Social similarities based on relationships between contents and users [3][26].

The hypothesis of our work is that diversity in similarity measures is a mean to obtain document diversity for RS. Indeed, not only facets of a topic are diverse but also users are, as well as users' expectations. Even if a unique recommendation method is efficient in the majority of the cases, it is useful to consider the other users' points of views. In that case, diversity on content but also other sorts of diversity should be considered in recommendations. This paper aims at showing the importance of diversity and to present a method and results we obtained when combining various methods in an RS in the context of blogs.

The paper is organized as follows: Section 2 presents the related works. In Section 3, we analyze the overlapping rate of the results retrieved by the best systems that participate to various IR tasks. Section 4 presents a user study on a blog platform composed of more than 20 million articles. Section 5 concludes this paper.

II. RELATED WORKS

Users are different and therefore their interests are different. To deal with this variety of interests, IR systems try to diversify the retrieved documents. Doing this, they attempt to maximize the chances of retrieving at least one relevant document for the user [30]. However, it is difficult to define what diversity is: several terms are used in the literature to describe this concept.

The literature distinguishes topicality and topical diversity. Topicality refers to which extent the document may be related to a particular topic [33]. Topical diversity groups the extrinsic diversity and the intrinsic diversity. The former helps to dispel the uncertainty resulting from the ambiguity of the user's need or from the lack of knowledge about her needs [28]. The intrinsic diversity, or novelty, intends to avoid redundancy in the retrieved documents [10]. It allows to present to the user various points of view, an overview of the topic that can only be achieved by considering simultaneously several documents, or even to check the information reliability [28].

To introduce topical diversity, two strategies are generally considered either as a clustering problem, or as a selection method close to the Maximal Marginal Relevance (MMR) proposed in [6].

Among the clustering approaches, He *et al.* [15] use Single Pass Clustering (SPC). Better results are obtained in [4] with the k-means algorithm [23]. Assignment to different clusters is generally done by using Euclidean distance and Cosine measure, eventually weighted by the terms frequency. Meij *et al.* [25] apply a hierarchical clustering algorithm on the top fifty documents retrieved by a language modeling approach. The selection of the documents used to build the result list is based on metrics of quality and stability of clusters. The best result from each cluster is selected.

The clustering step usually takes place after a set of documents has been retrieved to reorder them according to the sub-topics identified by the clusters.

Another way to topically diversify the results is to select documents considering those that already occur in the result. To reduce redundancy in the retrieved document list, MMR [6] or sliding window approaches [19] aim at selecting the documents maximizing the similarity with the query and, at the same time, minimizing the similarity with all the documents already selected. The similarity between the document and the previous selection can differ from the similarity with the query [6].

Several approaches select the documents using indicators or filters to increase the diversity in the results. Kaptein *et al.* [19] employ two types of document filters: a filter, which considers the number of new terms brought by the document to the current results and a link filter, which uses the value-added of new input or output links to select new documents. Furthermore, Ziegler *et al.* [37] propose an intra-list similarity metric to estimate the diversity of the recommended list. This metric uses a taxonomy-based classification.

Finally, some user's needs cannot be simply satisfied by topic-related documents. Serendipity, which aims at bringing to the user attractive and surprising documents she might not have otherwise discovered [16], is an alternative to topical diversity. Alternatively, Lathia *et al.* [21] investigate the case of temporal diversity and Cabanac *et al.* [5] consider organizational similarity.

To be able to evaluate and compare topical diversity oriented approaches, TREC Web 2009 campaign defines a dedicated diversity task. This task is based on the ClueWeb09 dataset, which consists in about 25TB of documents in multiple languages. The set B of the corpus, which we use for our experiments, only focuses on the English-language documents, roughly 50 million documents. The diversity task uses the same 50 queries than the *adhoc* tasks (http://trec.nist.gov).

Clarke *et al.* [9] present the panel of metrics used to estimate and compare the performances of the topical diversity approaches. In our experiments, we only consider the Normalized Discounted Cumulative Gain (α-nDCG) [18] and the Intent Aware Average Precision (MAP-IA) [2].

All these approaches and the available evaluation framework are mainly focused on content and on topical diversity. It seems difficult to develop an offline experimental framework, such the TREC Web diversity task, to evaluate other types of diversity, like serendipity. This task is even more difficult in the case of RS. In this context, a users study becomes necessary [14].

III. EXPERIMENTS

We believe that there is no one single approach that would satisfy all users' needs, but a set of complementary approaches. Considering this, we hypothesize that it is interesting to combine them. To demonstrate this, a first step is to verify that two distinct approaches retrieve different documents in various IR contexts (*adhoc*, *diversity*), even if they aim at the same goal.

For this, we consider several systems, which have been evaluated within the same framework to ensure they are comparable. Another selection criterion is the availability of the evaluation runs. Therefore, we focus on the *adhoc* and *diversity* tasks of the TREC Web 2009 campaign (considering only the set B of the corpus). Moreover, we choose the best systems for each task rather than taking into account all the submitted ones.

The runs comparison is done by computing, for each pair of runs, the overlap, that is to say the number of common documents between the two compared runs. The overlap is computed for the N first documents.

In this section, we first compare the four systems, which have the higher MAP at the *adhoc* task and then the four systems with the higher α-nDCG@10 at the diversity task.

A. Adhoc task experiment

1) Adhoc Task and compared runs

"The goal of the task is to return a ranking of the documents in the collection in order of decreasing probability of relevance. The probability of relevance of a

document is considered independently of other documents that appear before it in the result list." [8].

The performances of the different systems proposed are compared using the Mean Average Precision (MAP) metric. For this experiment, we consider the best run submitted by the four best competitors at the TREC Web 2009 *adhoc* task. Their scores are presented in Table I.

TABLE I.　　TREC WEB 2009 ADHOC TASK RESULTS

Group Id	Run Id	MAP
UDel	udelIndDRSP	0,2202
UMD	UMHOOsd	0,2142
uogTr	uogTrdphCEwP	0,2072
EceUdel	UDWAxBL	0,1999

The approaches evaluated (*Run Id*) for these runs are the following ones:

- *udelIndDRSP*: this run combines the query-likelihood language model with the MRF model of term dependencies and the pseudo relevance feedback with relevance models. It also used a document prior called Trust Domain [7];
- *UDWaxQEWeb*: incorporates the semantic term matching method in the axiomatic retrieval framework [35];
- *UMHOOsd*: uses a model based on the Markov Random Fields (MRF) in a distributed retrieval system [22];
- *uogTrdphCEwP*: uses the DPH weighting model derived from the Divergence From Randomness (DFR) model [24].

2) Results

Figure 1 presents the average overlap and precision for the runs selected in the experiments, considering the fifty queries of the task. We note that when considering only first retrieved documents the overlap is low, in spite of the fact that they are most relevant. For example, if we consider the ten first documents for which the precision reaches its highest value (0.386), the overlap is only 22.4%.
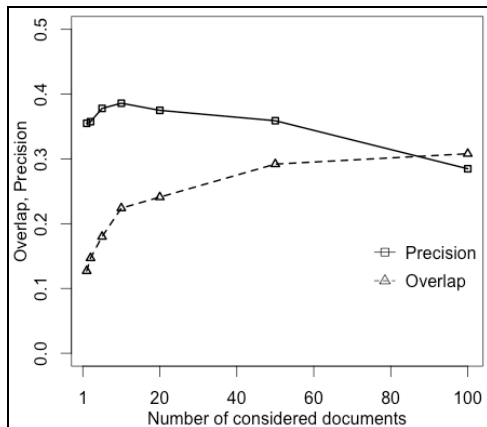

Figure 1.　Average overlap and precision for TREC Web 2009 *adhoc* task

These results demonstrate that for a given query, two distinct systems are unlikely to return the same documents.

In the next experiment, we check if we get similar observations when we focus on approaches designed to diversify the results.

B. Diversity task experiment

1) Diversity Task and compared runs

In this experiment, similarly to the previous experiment, we center on several systems submitted at the TREC Web Diversity Task. All these systems aim at providing users with diversified result lists.

"The goal of the diversity task is to return a ranked list of pages that together provide complete coverage for a query, while avoiding excessive redundancy in the result list. For this task, the probability of relevance of a document is conditioned on the documents that appear before it in the result list" [8]. The queries are similar for the *adhoc* and the *diversity* tasks. Table II presents the scores obtained by the different systems at their best run.

TABLE II.　　TREC WEB 2009 DIVERSITY TASK RESULTS

Group Id	Run Id	α-nDCG@10	MAP-IA@10
Waterloo	Uwgym	0.369	0.144
uogTr	uogTrSYCcsB	0.282	0.132
ICTNET	ICTNETDivR3	0.272	0.095
Amsterdam	UamsDancTFb1	0.257	0.082

For the *diversity* task, we retained the following runs:

- *uwgym*: this run acts as a baseline run for the track. It was generated by submitting the queries to one of the major commercial search engines. The results were filtered to keep only the documents included in the set B of the ClueWeb Collection [8];
- *uogTrSyCcsB*: relies upon the DPH DFR model and expands the queries using the Wikipedia documents retrieved [24];
- *ICTNETDivR3*: applies the k-means algorithm to a set of documents using Euclidean distance or Cosine measure to define the nearest cluster [4];
- *UamsDancTFb1*: uses a sliding window approach, which intends to maximize the similarity with the query and, at the same time, to minimize the similarity with the previous document selection. The selection process is completed by a link filter and a term filter [19].

2) Results

As shown in Figure 2, we get similar behavior than in the previous experiment: the overlap is also very low when we consider the first documents. These observations confirm our hypothesis that distinct approaches produce distinct results, even if they attempt to reach the same goal.

C. Conclusion

Whatever the purpose of the different approaches, whether they intend to introduce diversity in the result set, or they are designed to exactly match the needs expressed, the overlap in returned documents is low. Few documents are retrieved in multiple lists. We note that these observations are especially true when we consider only the first documents, which should theoretically be the most relevant.
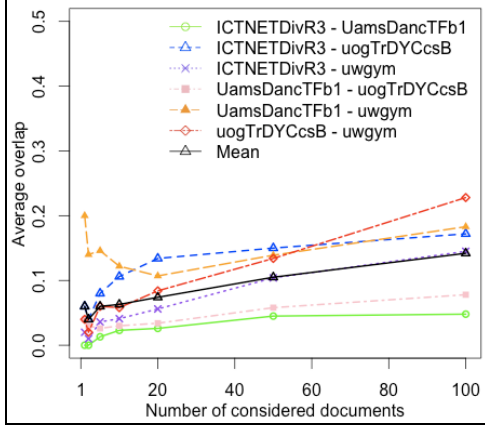
Figure 2. Average overlap for TREC Web 2009 *diversity* task

This leads us to believe that there is no better approach than others. Moreover, it is an indicator that they are complementary. Therefore the choice of the approaches is crucial. Several questions arise:

- How to choose these approaches?
- And how to eventually combine them to maximize the chances of satisfying the user?

IV. USERS STUDY: THE CASE OF OVERBLOG

A. Diversify the recommendations

We conducted an experiment with real users to check some hypotheses about the interest of producing diversified recommendations to users in RS. Among the hypotheses are:

- most of the time, users in an IR process search for focus information (topicality);
- sometimes, users want to enlarge the subject they are interested in (topical diversity);
- some users are in a process of discovery and search for new information (serendipity);
- the interesting links between documents do not only concern their content similarity;
- integrating diversity in an RS process is valuable because it allows to answer additional users' needs.

To check these hypotheses, we recruited thirty four students and asked them to test and compare various RS. The users were first asked to type a query on our search engine (first time imposed, to ensure overlap about the documents they all considered, and then a query of their choice). They had then to choose one document and were then shown two lists of recommended documents:

- one list was based on one of the five systems we designed: *mlt* and *searchsim* use topicality, *kmeans* uses topical diversity, and *topcateg* or *blogart* use serendipity (see system description Section B);
- the other list was our RS, designed by merging the results of those five previous systems (choosing the first document in the result list of each system).

Each list contains five documents, and the users do not know which system it corresponds to. They are then asked to choose which list they find the most relevant, and which one they find the most diversified.

Finally, the two lists were mixed into one, and the users has to assess which documents were relevant according to them.

B. Data and systems

We focused on the French documents of the *OverBlog* platform. The data used represent more than 20 million articles distributed on 1.2 million blogs.

The five systems used to get recommendation lists are:

- *blogart*: returns documents randomly selected in the same blog of the visited document (serendipity);
- *kmeans*: classifies the retrieved documents with the k-means algorithm [23] and build the final result list selecting one document per cluster (topical diversity);
- *mlt*: uses *Apache Solr's MoreLikeThis* module (http://lucene.apache.org/solr) to retrieve similar documents considering all the content of the visited document (topicality);
- *searchsim*: performs a search using a vector-space search engine. The query used is the title of the visited document (topicality);
- *topcateg*: retrieves the most popular documents randomly selected in the same category (from the OverBlog's hierarchy) of the visited document (serendipity).

The use of these systems aims at simulating the various types of diversity (topicality, topical diversity, serendipity) and intents to limit the overlap between the documents they retrieve. To ensure that the systems used in the user study retrieve distinct results, we compute the overlap between each pair, similarly to the previous experiments. We observe the same trends as in the experiments led on the *adhoc* and *diversity* tasks: the overlap is low between the approaches based on content similarities (*mlt*, *searchsim* and *kmeans*) and is null in the case of serendipity (*blogart*, *topcateg*).

C. Results

Table III shows the feedback returned by the user panel concerning the interest of the proposed lists, and their impression on their diversity. For example (4th row), 76.5% of the lists provided by *mlt* have been considered as more relevant than those of *fused*. We can see that the systems perceived as the most relevant are those that focus on topicality. Then the *fused* system is seen as more relevant than its opponent roughly once upon two times on average. We get the same result for *blogart*. That is more surprising, but confirms that users are sometimes interested in links that does not only concern the content of documents. The results obtained for the question "Which one of the following result lists seems the most diversified to you?" is even more surprising, since there are not high differences between the systems. We think this can be explained by the fact that users have difficulties in defining the notion of diversity. We should have probably helped them by clarifying our question.

TABLE III. PERCENTAGE OF USERS WHO CONSIDER THE *SYSTEM* TO BE MORE RELEVANT/DIVERSIFIED THAN THE *FUSED* SYSTEM

System	Relevance	Diversity
blogart	0.447	0.553
kmeans	0.708	0.333
mlt	0.765	0.500
searchsim	0.643	0.429
topcateg	0.154	0.654

Table IV describes for each RS what is the system precision, that is to say the number of documents that have been considered as relevant among the documents retrieved. We check again that the approaches that use content similarities are seen as more relevant. *kmeans*, that proposes topical diversity, has the best results. On the contrary, *topcateg* and *blogart* that search for serendipity, have lower results. As expected, the *fused* system offers a compromise between these different systems.

TABLE IV. PRECISION PER SYSTEM

System	blogart	kmeans	mlt	searchsim	topcateg	fused
Precision	0.147	0.385	0.265	0.307	0.038	0.267

Finally, Table V compares the fused system with the others. It gives the proportion of relevant documents that have been retrieved by each system. For example, when comparing *mlt* to *fused* (4th column), 54.69% of the relevant documents have been retrieved by *mlt* only, 32.81% by fused only and 12.50% only by both. We can thus observe that, even if more relevant documents come from the systems searching for topicality, a significant part of them comes from the fused system. We think that justifies our approach, because more than 20% of relevant documents coming from our system only means that one document among the five proposed is considered as relevant and would not have been returned if using any system alone.

TABLE V. DISTRIBUTION OF THE RELEVANT DOCUMENTS

System: *fused* against	blogart	kmeans	mlt	searchsim	topcateg
Only retrieved by the system	35.00%	52.46%	54.69%	52.43%	8.77%
Only retrieved by *fused* system	65.00%	21.31%	32.81%	38.83%	91.23%
Commons	0.00%	26.23%	12.50%	8.74%	0.00%

D. Conclusion

The *fused* system we propose offers a new framework to combine various RS. The one implemented and tested here does not outperform the others, but that was not our goal. Rather, our idea is to promote diversity, and we have seen with the user experiments that this is a relevant track. Indeed, by diversifying our recommendations, we are able to answer different and additional users' needs, when the other systems focus on the majority needs: most often the content similarity. The systems we tested here for serendipity were quite simple. Nevertheless, the results they returned were considered as relevant by some users, and we think this is an encouraging sign for developing RS since users are interested in various forms of diversity in result lists.

V. CONCLUSIONS

We are all different, and any system that aims at providing tailored results to their users must take this fact into account. Information retrieval and recommender systems have this ambition, especially since users become used to be given personalized tools, and since they need such systems to overcome the huge amount of data they can access. We have also different expectations depending on the context. Indeed, we may search once for focus information, and the next time search for novelty. This behavior has been confirmed by our user study: the majority of the users' interests has been for topicality (*kmeans, mlt, searchsim*), but sometimes, interests for originality (*blogart, topcateg*) emerged. This motivates the design of a system able to handle such heterogeneity of the users' needs.

Our first contribution in this area has been to study the overlap between the documents retrieved by several IR systems of the literature using state-of-the-art datasets. Although those systems are all content similarity-based, we have noted that they are based on different underlying assumptions, and that their overlap was low. This low overlap between the documents retrieved by these systems indicates that there is not a perfect system able to satisfy the diversity of the users' needs, but a set of complementary systems.

The strength of our approach is that it is designed to combine different RS. We could easily add to the framework any other RS that would cover new interests. It seems that some work could especially be done on the approaches that offer serendipity: *blogart* seems to be an interesting one; *topcateg* is to be improved. Other RS fusion approaches have been proposed in the literature. In particular, Shafer *et al.* [31] and Jahrer *et al.* [17] present a "meta RS". However, when they choose to focus on results shared by the different RS used by the meta system, we instead propose to select the best recommendations of each system to ensure diversity. We assume that it is important to give a chance to all possible "points of view" proposed by every retrieved document.

We will direct our future work towards designing an RS architecture promoting the diversity of recommendations. When existing approaches focus on designing methods to force diversity in their results (using clustering or MMR), we choose to consider multiple systems to build the recommendation list and ensure diversity. Moreover, it is important to see that every retrieved document may give rise to a wide range of interests for readers. So, next step for this work is to study the learning mechanism to find the proportion of documents coming from each RS to be fused. That is to say we would learn the main interests that are important for end-users (readers). To do this, our idea is to use an automatic learning process based on the users' feedbacks. We could for example simply initialize the system with equal distribution for each RS, and then increase the proportion of recommendations for the ones that are more often clicked by the users, and decrease the proportion for RS less often considered. Our meta RS would thus be well focused on the users' needs. Considering the results of

the experiments presented in this paper, we could expect a 80% proportion for topicality systems, and 20% for more original systems.

Finally, we will see if these assumptions stand on a real scale experiment using the online blog platform *OverBlog*. Then, we will conduct statistical analysis to study what could be the influence of the document type on those proportions.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering, 17(6), 2005, pp. 734-749

[2] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong, "Diversifying search results", Inter. Conf. on Web Search and Data Mining, 2009, pp. 5-14

[3] L. Ben Jabeur, L. Tamine, and M. Boughanem, "A social model for Literature Access: Towards a weighted social network of authors", Inter. Conf. on Adaptivity, Personalization and Fusion of Heterogeneous Information RIAO, 2010, pp. 32-39

[4] W. Bi, X. Yu, Y. Liu, F. Guan, Z. Peng, H. Xu, and X. Cheng, "ICTNET at Web Track 2009 diversity task", Text REtrieval Conf., 2009

[5] G. Cabanac, M. Chevalier, C. Chrisment, and C. Julien, "An Original Usage-based Metrics for Building a Unified View of Corporate Documents", Inter. Conf. on Database and Expert Systems Applications, 2007, LNCS V. 4653, 2007, pp. 202–212

[6] J. Carbonell and J. Goldstein, "The use of MMR, diversity-based reranking for reordering documents and producing summaries", ACM Conf. on Research and Development in Information Retrieval, 1998, pp. 335-336

[7] P. Chandar, A. Kailasam, D. Muppaneni, L. Thota, and B. Carterette, "Ad Hoc and Diversity Retrieval at the University of Delaware", Text REtrieval Conf., 2009

[8] C. L. A. Clarke, N. Craswell, and I. Soboroff, "Overview of the TREC 2009 Web Track", Text REtrieval Conf., 2009

[9] C. L. A. Clarke, N. Craswell, I. Soboroff, and A. Ashkan, "A comparative analysis of cascade measures for novelty and diversity", ACM Inter. Conf. on Web Search and Data Mining, 2011

[10] C. L. A. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I.n MacKinnon, "Novelty and Diversity in Information Retrieval Evaluation", ACM Conf. on Research and Development in Information Retrieval, 2008, pp. 659-666

[11] D. Dudognon, G. Hubert, J. Marco, J. Mothe, B. Ralalason, J. Thomas, A. Reymonet, H. Maurel, M. Mbarki, P. Laublet, and V. Roux, "DYNAMic ontology for information retrieval", Inter. Conf. on Adaptivity, Personalization and Fusion of Heterogeneous Information RIAO, 2010, pp. 213-215

[12] I. Esslimani, A. Brun, and A. Boyer, "A Collaborative Filtering approach combining Clustering and Navigational based correlations", Web Information Systems and Technologies, 2009, pp. 364-369

[13] E. A Fox and J. A. Shaw, "Combination of multiple searches", Text REtrieval Conf., NIST special publication, 1994, pp. 243–252

[14] C. Hayes, P. Massa, P. Avesani, and P. Cunningham, « An online evaluation framework for recommender systems », Workshop on Personalization and Recommendation in E-Commerce, 2002

[15] J. He, K. Balog, K. Hofmann, E. Meij, M. de Rijke, M. Tsagkias, and W. Weerkamp, "Heuristic Ranking and Diversification of Web Documents", Text REtrieval Conf., 2010

[16] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating Collaborative Filtering Recommender Systems", ACM Trans. Information Systems, 22(1), 2004, pp. 5-53

[17] M. Jahrer, A. Töscher, and R. Legenstein, "Combining predictions for accurate recommender systems", ACM SIGKDD Inter. Conf. on Knowledge Discovery and Data mining, 2010

[18] K. Järvelin, and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques", ACM Transactions on Information Systems, 20(4), 2002, pp. 422–446

[19] R. Kaptein, M. Koolen, and J. Kamps, "Result diversity and entity ranking experiments: Anchors, links, text and Wikipedia", Text REtrieval Conf., 2009

[20] A. Kritikopoulos, M. Sideri, and I. Varlamis, "Blogrank: ranking weblogs based on connectivity and similarity features", Inter. Workshop on Advanced Architectures and Algorithms for Internet Delivery and Applications, 2006, art. 8

[21] N. Lathia, S. Hailes, L. Capra, and X. Amatriain, "Temporal diversity in recommender systems", ACM Conf. on Research and Development in Information Retrieval, 2010, pp. 210-217

[22] J. Lin, D. Metzler, T. Elsayed, and L. Wang, "Of Ivoryand Smurfs: Loxodontan MapReduce Experiments for Web Search", Text REtrieval Conf., 2009

[23] J. MacQueen, "Some methods for classification and analysis of multivariate observations", Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297

[24] R. McCreadie, C. Macdonald, I. Ounis, J. Peng, and R. L. T. Santos, "University of Glasgow at TREC 2009: Experiments with Terrier", Text REtrieval Conf., 2009

[25] E. Meij, J. He, W. Weerkamp, and M. de Rijke, "Topical Diversity and Relevance Feedback", Text REtrieval Conf., 2010

[26] J. Mothe, C. Chrisment, T. Dkaki, B. Dousset, and S. Karouach, "Combining mining and visualization tools to discover the geographic structure of a domain", Computers, Environment and Urban Systems, 30(4), 2006, pp. 460-484

[27] J. Mothe and G. Sahut, "Is a relevant piece of information a valid one? Teaching critical evaluation of online information". Approaches to Teaching and Learning Information Retrieval, N. Efthimiadis, Manuel Fernández Luna, Juan Huete, Andrew MacFarlane (Eds.), 2011 (to appear)

[28] F. Radlinski, P. N. Bennett, B. Carterette, and T. Joachims. "Redundancy, diversity and interdependent document relevance", SIGIR Forum, 43(2), 2009, pp. 46–52

[29] G. Salton and M. J. McGill, "Introduction to Modern Information Retrieval", McGraw-Hill, 1983

[30] R. L. T. Santos, C. Macdonald, and I. Ounis, "Selectively Diversifying Web Search Results", ACM Inter. Conf. on Information and Knowledge Management, 2010

[31] J. B. Schafer, J. A. Konstan, and J. Riedl, "Meta-recommendation systems: user-controlled integration of diverse recommendations", Inter. Conf. on Information And Knowledge Management, 2002, pp. 43-51

[32] Z. Wu, M. Palmer, "Verb semantics and lexical selection", Annual Meeting of the Associations for Computational Linguistics, 1994, pp. 133-138

[33] Y. C. Xu and Z. Chen, "Relevance judgment: What do information users consider beyond topicality", Journal of the American Society for Information Science and Technology, 57(7), 2006, pp. 961–973

[34] X. Yin, J. X. Huang, and Z. Li, "Promoting Ranking Diversity for Biomedical Information Retrieval Using Wikipedia", European Conf. on Information Retrieval, 2010, pp. 495-507

[35] W. Zheng and H. Fang, "Axiomatic Approaches to Information Retrieval - University of Delaware at TREC 2009 Million Query and Web Tracks", Text REtrieval Conf., 2009

[36] X. Zhu, A. B. Goldberg, J. Van Gael, and D. Andrzejewski, "Improving Diversity in Ranking using Absorbing Random Walks", Human Language Technologies - Conf. North American Chapter Assoc. Comp. Linguistics, 2007, pp. 97-104

[37] C. Ziegler, S. McNee, J. A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification", Inter. Conf. on World Wide Web, 2005, pp. 22–32