University of Padua at TREC 2013: Federated Web Search Track

Emanuele Di Buccio, Ivano Masiero, and Massimo Melucci

Department of Information Engineering, University of Padua, Italy
{dibuccio,masieroi,melo}@dei.unipd.it

Abstract. This paper reports on the participation of the University of Padua to the TREC 2013 Federated Web Search track. The objective was the experimental investigation in Federated Web Search setting of TWF·IRF, which is a recursive weighting scheme for resource selection. The experimental results show that the TWF component, that is peculiar of this scheme, is sufficient to obtain an effective search engine ranking in terms of NDCG@20 when compared with the baseline and the runs of other track participants.

1 Introduction

This paper reports on the participation of the Information Management System (IMS) Research Group of the University of Padua to the TREC 2013 Federated Web Search track (FedWeb13).[1] The participation to the FedWeb13 track aimed at the investigation of the effectiveness of the TWF·IRF weighting framework when adopted in Federated Web search setting.

The weighting framework was originally introduced in [3] to address the problem of resource selection in Hybrid Hierarchical Peer-To-Peer (P2P) Networks. In unstructured P2P networks each participating node can act both as client and server — in an Information Retrieval (IR) perspective it can both submit a query to the other participating nodes and act as a search server, providing the most relevant documents in its local collection in response to a given query. In a hybrid unstructured network there are two types of nodes, i.e. peers and super-peers. Yet, a peer has to update and transfer the data structures which summarizes its own document collection to the super-peers. A query is sent from a peer to the super-peers and then it is routed from a super-peer to the other super-peers on the basis of the summaries stored in each super-peer. While all the peers are involved when routing the query in an unstructured network, only the super-peers are involved in routing in a hybrid unstructured network. When routing the query a super-peer ranks both the other super-peers and the peers by expected recall.

The TWF·IRF addresses the problem of informative resource ranking in architectures with an arbitrary number of resource levels. In the FedWeb13 track setting there are three resource levels: (i) document, (ii) search engines, and (iii)

[1] The identifier adopted in TREC for our research group is `UPD`.

set of search engines. In particular, there is a single set of 157 search engines and the objective of the resource selection task is to rank them according to (their predictive capability for) a given query. Even if search engines use many features that in peer-to-peer settings[2] could not be available, the participation to the track would give some insights on the TWF·IRF effectiveness in ranking peers in a group when considering a completely uncooperative environment. Indeed, the task is the same as that a super-peer needs to perform in order to select the most promising peers in its group to which the query should be forwarded; moreover, the summaries stored in the "super engine", i.e. the broker, are the results of query-based sampling performed on the considered search engines since the index of the distinct search engines cannot be accessed — for this reason their are considered "uncooperative".

2 A Recursive Weighting Scheme

As for past works in Distributed IR [2, 4], the specific approach adopted in this work is to describe the informative resources at the diverse levels (document, search engines) in terms of document descriptors, e.g. terms. Therefore, a search engine is described as a set of document descriptors, specifically the distinct descriptors appearing in the documents stored in it. The innovative contribution of our approach consists in the way of computing weights.

The weight of a descriptor in a resource is constituted of two components: TWF and IRF. The Inverse Resource Frequency (IRF) is a generalization of the Inverse Document Frequency (IDF) for the higher resource levels. Generalizations of the IDF were proposed in [2] to rank collections (Inverse Collection Frequency, ICF) and in [4] to rank peers (Inverse Peer Frequency, IPF). The IRF extends this idea for an arbitrary resource level:

$$irf_t^{(z)} = \log N^{(z)}/n_t^{(z)} \tag{1}$$

where t denotes the term, $N^{(z)}$ is the number of resources at level z contained by the resource at level $z+1$ and $n_t^{(z)}$ is the number of those resources that are indexed by t. ICF and IPF are instances of the IRF weight at level 2. In the FedWeb13 settings there are three resource levels: (1) document, (2) search engines and (3) the set of search engines.

Unlike IRF, Term Weighted Frequency (TWF) is peculiar of this scheme. The weight of a descriptor t in a resource i at level z is

$$w_{i,t}^{(z)} = twf_{i,t}^{(z)} \cdot irf_t^{(z)}, \tag{2}$$

where

$$twf_{i,t}^{(z)} = \sum_{r \in R_i^z} twf_{i,t}^{(z-1)} \cdot irf_t^{(z-1)} \tag{3}$$

[2] We consider the case where each participating peer provides search functionalities to access their local collection, e.g. part of the personal documents of a user.

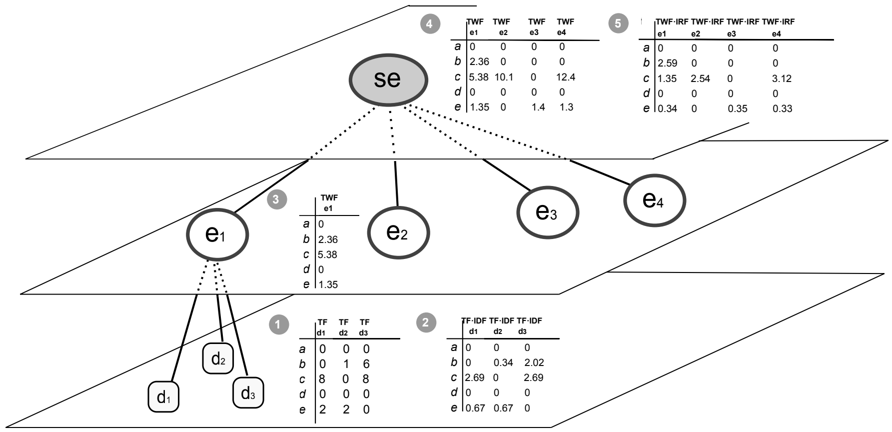

Fig. 1. Computation of the weights for the informative resources at the diverse levels.

and R_i^z denotes the sets of resources in the ith resource at level z. For a given query q, resources at level z can be ranked according to $\sum_{t \in q} w_{i,t}^{(z)}$. Ax example with four resources at level 2 and five descriptors is depicted in Figure 1.

3 Experiments

3.1 Experimental Methodology

The experimental methodology consists of two tasks:

– **resource selection**: given a set of search engines S, a set of queries $Q_{\mathcal{T}}$ and a set of sample documents obtained by query-based sampling performed on each of the search engines, the goal of this task was to return a ranked list of search engines for each query in $Q_{\mathcal{T}}$, where the search engines should be ranked according to their capability to satisfy the user information need expressed by the query;
– **result merging**: given a query and the top 10 results for each of the search engines in S, the goal of this task was to merge the results into a single ranked list.

3.2 Test Collection and Effectiveness Measures

The research questions described in Section 3.1 were addressed using the FedWeb13 test collection. This collection is constituted of a list of 157 search engines[3] and a set of sample search results obtained by performing query-based

[3] The list of search engines is available at the following url:
`http://snipdex.org/datasets/fedweb2013/FW13-engines.txt`

sampling on those search engines. A set $Q_{\mathcal{S}}$ of 2000 queries was adopted to perform the sampling. For each search engine and for each query in the given query set, the top 10 results were retrieved – both snippets and landing documents. Half of the queries in $Q_{\mathcal{S}}$ was obtained using the *Zips method*, which exploits "single term queries taken evenly from the binned term distribution in ClueWeb09, where terms were binned on a log-scale of their document frequency (df) to ensure that there are queries from the complete frequency distribution." [8]. The other half of the queries was built by randomly selecting terms from the sample documents collected from the search engine.

A set of 200 queries, $Q_{\mathcal{T}}$, were provided by the track organizers to address the two research tasks described in Section 3.1.

The evaluation for the two tasks was performed on a subset of 50 queries among those in $Q_{\mathcal{T}}$. The effectiveness measures adopted for the *resource selection* task were NDCG@20 and ERR@20. The Normalized Discounted Cumulative Gain (NDCG) [6] version adopted in the experiments is that proposed in [1]. The relevance of a search engine was computed by using the graded precision [7] on the top 10.[4] The effectiveness measures adopted for the *result merging* task were NDCG and P@10.

3.3 Parsing and Indexing

The indexing module of our system relies on an XML parser written in Java for extracting the document fields from the sample searches and the sample documents in the test collection, and on the Apache Lucene library. The sample documents in the FedWeb13 Test Collection were indexed by creating a distinct index for each of the 157 search engines. These indexes are *document-level* indexes. Each (Lucene) document in a document-level index is constituted of four fields: link, title, description, and the content of the document associated to the sample search result. For each field, the document-level index stores information on the frequency of the descriptors in each document and in the collection, as well as their positions in each document.

Starting from these indexes, a search engine-level index was built. The set of descriptors in this index is the union of all the distinct descriptors in the distinct document-level indexes associated to the search engines. As for the document-level index, in the search engine-level index a list of posting is associated to each descriptor. Each posting stores information on the identifier of the search engine, the number of documents in the search engine were the descriptor appears, and the TWF of the descriptor. In the specific Lucene-based implementation adopted, the TWF weight was stored in the payload that can be associated to each term; the weight value was approximated and stored as a float.[5]

[4] Details are provided in the FedWeb13 Track web page: `https://sites.google.com/site/trecfedweb/`

[5] Single-precision 32-bit IEEE 754 floating point

3.4 Resource Selection

The runs submitted to the FedWeb13 track exploited the TWF component of the weighting scheme described in Section 2. This score was adopted to rank search engines in the resource selection task. The score of a search engine for a query q was computed as

$$\sum_{t \in q} twf_{i,t}^{(2)} \tag{4}$$

where $twf_{i,t}^{(2)} = \sum_{d_j \in D_i} twf_{i,t}^{(1)} \cdot irf_t^{(1)}$ and D_i denotes the sets of documents in the ith search engine, $twf_{j,t}^{(1)} = tf(t, j)$ is the term frequency of term t in the document d_j. The IRF at the document level was implemented as:

$$irf_t^{(1)} = \log\Big(1 + \frac{N^{(1)} - n_t^{(1)} + 0.5}{n_t^{(1)} + 0.5}\Big) \tag{5}$$

We did not exploit the IRF component of the weighting scheme at level two in order to isolate the effect of the TWF component.

Two runs were submitted for the resource selection task: `UPDFW13sh` and `UPDFW13mu`. In the `UPDFW13sh` run, the query was built by performing an OR among the terms appearing in the query.[6] The ranked list of search engines that constitute the `UPDFW13mu` run was obtained by appending three ranked lists:

- L_1: the list of search engines ranked by their TWF weight with regard to the query, and using the AND boolean constraint among the occurrence of the distinct terms in the query[7];
- L_2: the list of search engines that did not belong to L_1 and ranked by their TWF weight with regard to the query by using the OR boolean constraint among the occurrence of the distinct terms in the query;
- L_3: the list of search engine that did not belong to L_1 and L_2, ranked by their identifier — the identifier associated to the search engine in the test collection.

The final ranked list of search engines was obtained by appending L_2 to L_1, and then L_3 to the fusion of the first two lists.

Results for the resource selection task are reported in Table 1 and Table 2; *sh* denotes `UPDFW13sh` and *mu* denotes `UPDFW13mu`. When considering per topic values, a row reports *mu* and *sh* values for the considered measure, the best and the median values over all the systems for that topic; last row reports the average value of the best, the median, the *mu* and the *sh* runs.

[6] The Lucene query was a BooleanQuery constituted of PayloadTermQuery connected by SHOULD clause.

[7] The Lucene query was a BooleanQuery constituted of PayloadTermQuery connected by MUST clause.

3.5 Result Merging

Two runs were submitted in the result merging task. Both runs were obtained by performing result merging in a round robin fashion. When exploiting round robin for merging, the first document in the merged list is the first document in the first selected search engine, the second document is the first document in the second selected search engine, and so on. In the `UPDFW13rrsh` run the search engine ranking was that obtained in the `UPDFW13sh` run submitted in the resource selection task; the `UPDFW13rrmu` run exploited the ranking obtained in the `UPDFW13mu` run.

Results for the result merging task are reported in Table 3 and Table 4; *sh* denotes `UPDFW13rrsh` and *mu* denotes `UPDFW13rrmu`. When considering per topic values, a row reports *mu* and *sh* values for the considered measure, the best and the median values over all the systems for that topic; last row reports the average value of the best, the median, the *mu* and the *sh* runs.

4 Final Remarks

This paper reported on the evaluation of TWF·IRF in the TREC2013 Federated Web Search track. The results obtained in the resource selection task show the effectiveness of TWF·IRF in ranking search engines when their description exploits the full content of the indexed documents obtained by query-based sampling. Indeed, the `UPDFW13mu` run obtained the highest NDCG@20 value among the other runs; moreover the run effectiveness is comparable with the baseline provided by the track organizers [5].

These promising results suggest further investigations on the variables that can affect TWF·IRF effectiveness in Federated Web Search. Examples of variables are the document “field” adopted to extract resource descriptors – e.g. the use of the snippet instead of the full content of the documents – or the sampling strategy adopted to gather documents for resource description – e.g. the FedWeb12 test collection samples were obtained by random, top and zipf sampling. Moreover, the experiments reported in this paper exploit only the TWF component; future investigation will be focused on the effect of the IRF component on the resource selection effectiveness in order to investigate if the TWF component is “sufficient” or both TWF and IRF are “necessary”.

Acknowledgment

This work has been supported by the project “Un’inchiesta grammaticale sui dialetti italiani: ricerca sul campo, gestione dei dati, analisi linguistica” (Bando FIRB – Futuro in ricerca 2008).

References

1. C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning*, ICML '05, pages 89–96, New York, NY, USA, 2005. ACM.
2. J. P. Callan, Z. Lu, and W. B. Croft. Searching distributed collections with inference networks. In *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '95, pages 21–28, New York, NY, USA, 1995. ACM.
3. R. Castiglion and M. Melucci. An evaluation of a recursive weighing scheme for information retrieval in peer-to-peer networks. In *Proceedings of the 2005 ACM workshop on Information retrieval in peer-to-peer networks*, P2PIR '05, pages 9–16, New York, NY, USA, 2005. ACM.
4. F. M. Cuenca-Acuna and T. Nguyen. Text-Based Content Search and Retrieval in Ad-hoc P2P Communities. In E. Gregori, L. Cherkasova, G. Cugola, F. Panzieri, and G. Picco, editors, *Web Engineering and Peer-to-Peer Computing*, volume 2376 of *Lecture Notes in Computer Science*, pages 220–234. Springer Berlin Heidelberg, 2002.
5. T. Demeester, D. Trieschnigg, D. Nguyen, and D. Hiemstra. Overview of the TREC 2013 Federated Web Search Track. In *Proceedings of the Twenty-Second Text REtrieval Conference (TREC 2013)*. National Institute of Standards and Technology (NIST), 2014.
6. K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems*, 20(4):422–446, Oct. 2002.
7. J. Kekäläinen and K. Järvelin. Using graded relevance assessments in ir evaluation. *Journal of the American Society for Information Science and Technology*, 53(13):1120–1129, 2002.
8. D. Nguyen, T. Demeester, D. Trieschnigg, and D. Hiemstra. Federated search in the wild: the combined power of over a hundred search engines. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, CIKM '12, pages 1874–1878, New York, NY, USA, 2012. ACM.

Table 1. Comparison in terms of NDCG@20 among the resource selections runs. The percentage of increment has been omitted when the median value is 0; those entries are marked with **

Topic	Best	Median	mu	$\Delta_{mu-best}$	$\Delta_{mu-median}$	sh	$\Delta_{sh-best}$	$\Delta_{sh-median}$
7001	0.301	0.000	0.301	0.0	**	0.270	-10.2	**
7003	0.345	0.000	0.345	0.0	> 1000	0.335	-2.8	> 1000
7004	0.398	0.155	0.000	-100.0	-100.0	0.303	-23.8	95.6
7007	0.429	0.000	0.429	0.0	**	0.386	-10.0	**
7009	0.988	0.348	0.359	-63.6	3.3	0.336	-66.0	-3.3
7018	0.396	0.000	0.278	-29.8	**	0.049	-87.6	**
7025	0.387	0.231	0.309	-20.1	34.0	0.387	0.0	67.8
7030	0.971	0.238	0.967	-0.4	305.9	0.500	-48.5	109.9
7033	0.498	0.000	0.000	-100.0	-100.0	0.257	-48.5	> 1000
7034	0.434	0.018	0.015	-96.6	-16.4	0.015	-96.6	-18.3
7039	0.545	0.000	0.545	0.0	> 1000	0.051	-90.6	> 1000
7040	0.493	0.012	0.435	-11.8	> 1000	0.340	-31.1	> 1000
7042	0.231	0.035	0.062	-73.0	80.7	0.034	-85.1	-0.4
7046	0.318	0.005	0.318	0.0	> 1000	0.318	0.0	> 1000
7047	0.356	0.000	0.000	-100.0	0	0.000	-100.0	0
7056	0.997	0.001	0.000	-100.0	-92.7	0.000	-100.0	-100.0
7067	0.239	0.004	0.239	0.0	> 1000	0.016	-93.4	336.9
7068	0.618	0.172	0.618	0.0	258.2	0.425	-31.2	146.4
7069	0.631	0.266	0.500	-20.7	88.4	0.431	-31.7	62.3
7075	1.000	0.465	0.631	-36.9	35.6	0.431	-56.9	-7.4
7076	0.148	0.004	0.000	-100.0	-100.0	0.056	-62.0	> 1000
7080	0.432	0.100	0.272	-36.9	171.2	0.000	-100.0	-100.0
7084	0.390	0.007	0.390	0.0	> 1000	0.390	0.0	> 1000
7087	0.680	0.110	0.107	-84.3	-2.6	0.113	-83.4	2.6
7089	0.999	0.000	0.000	-100.0	0	0.000	-100.0	0
7090	0.684	0.147	0.423	-38.2	188.1	0.173	-74.7	18.1
7094	0.298	0.001	0.290	-2.9	> 1000	0.290	-2.9	> 1000
7096	0.291	0.119	0.291	0.0	144.2	0.272	-6.5	128.4
7097	0.469	0.115	0.469	0.0	309.0	0.432	-8.0	276.3
7099	0.387	0.000	0.000	-100.0	0	0.315	-18.5	**
7103	0.335	0.000	0.294	-12.2	> 1000	0.294	-12.2	> 1000
7109	0.877	0.351	0.350	-60.1	-0.2	0.204	-76.7	-41.7
7115	0.300	0.011	0.300	0.0	> 1000	0.300	0.0	> 1000
7124	0.430	0.228	0.256	-40.5	12.5	0.256	-40.5	12.5
7127	0.468	0.193	0.378	-19.3	95.4	0.378	-19.3	95.4
7129	0.457	0.041	0.331	-27.6	702.4	0.331	-27.6	702.4
7132	0.326	0.000	0.309	-5.3	> 1000	0.309	-5.3	> 1000
7145	0.439	0.082	0.439	0.0	434.4	0.439	0.0	434.4
7209	0.241	0.000	0.000	-100.0	0.0	0.000	-100.0	-25.0
7258	0.309	0.008	0.017	-94.6	103.1	0.016	-94.9	91.7
7348	0.985	0.010	0.985	0.0	> 1000	0.280	-71.6	> 1000
7404	1.000	0.387	0.387	-61.3	0.0	0.387	-61.3	0.0
7406	0.301	0.000	0.279	-7.3	> 1000	0.279	-7.3	> 1000
7407	0.856	0.290	0.294	-65.7	1.2	0.265	-69.0	-8.5
7415	0.349	0.119	0.295	-15.5	147.5	0.335	-4.2	180.6
7465	0.387	0.232	0.387	0.0	67.1	0.270	-30.1	16.7
7485	0.485	0.012	0.012	-97.5	0.0	0.012	-97.5	0.0
7504	0.630	0.002	0.334	-47.0	> 1000	0.334	-47.0	> 1000
7505	0.977	0.374	0.416	-57.4	11.3	0.416	-57.4	11.3
7506	0.387	0.120	0.301	-22.2	150.7	0.301	-22.2	150.7
amean	0.299	0.141	0.299	0.0	112.1	0.247	-17.5	74.9

Table 2. Comparison in terms of ERR@20 among the resource selections runs. The percentage of increment has been omitted when the median value is 0; those entries are marked with **

Topic	Best	Median	mu	$\Delta_{mu-best}$	$\Delta_{mu-median}$	sh	$\Delta_{sh-best}$	$\Delta_{sh-median}$
7001	0.00011	0.00000	0.00011	0.0	**	0.00008	-27.3	**
7003	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7004	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7007	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7009	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7018	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7025	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7030	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7033	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7034	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7039	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7040	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7042	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7046	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7047	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7056	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7067	0.00006	0.00000	0.00006	0.0	**	0.00001	-83.3	**
7068	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7069	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7075	1.00000	0.29167	0.50000	-50.0	71.4	0.25000	-75.0	-14.3
7076	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7080	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7084	0.00020	0.00000	0.00020	0.0	**	0.00020	0.0	**
7087	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7089	1.00000	0.00000	0.00000	-100.0	0.0	0.00000	-100.0	0.0
7090	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7094	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7096	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7097	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7099	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7103	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7109	0.00004	0.00001	0.00001	-75.0	0.0	0.00000	-100.0	-100.0
7115	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7124	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7127	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7129	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7132	0.00033	0.00000	0.00013	-60.6	**	0.00013	-60.6	**
7145	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7209	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7258	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7348	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7404	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7406	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7407	0.00781	0.00100	0.00114	-85.4	14.0	0.00089	-88.6	-11.0
7415	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7465	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7485	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7504	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7505	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
7506	0.00000	0.00000	0.00000	0.0	0.0	0.00000	0.0	0.0
amean	0.02003	0.00835	0.01003	-49.9	20.1	0.00503	-74.9	-39.8

Table 3. Comparison in terms of NDCG among the result merging runs.

Topic	Best	Median	*mu*	$\Delta_{mu-best}$	$\Delta_{mu-median}$	*sh*	$\Delta_{sh-best}$	$\Delta_{sh-median}$
7001	0.606	0.544	0.603	-0.5	10.7	0.603	-0.4	10.9
7003	0.729	0.702	0.729	-0.1	3.9	0.714	-2.1	1.8
7004	0.674	0.338	0.383	-43.1	13.3	0.338	-49.8	0.0
7007	0.606	0.540	0.606	0.0	12.1	0.604	-0.4	11.7
7009	0.563	0.387	0.387	-31.2	0.0	0.382	-32.0	-1.2
7018	0.613	0.507	0.465	-24.2	-8.4	0.467	-23.8	-7.9
7025	0.682	0.543	0.522	-23.6	-4.0	0.517	-24.2	-4.8
7030	0.733	0.480	0.511	-30.3	6.5	0.441	-39.8	-8.1
7033	0.482	0.405	0.387	-19.8	-4.6	0.393	-18.4	-3.0
7034	0.484	0.443	0.443	-8.4	0.0	0.447	-7.6	0.9
7039	0.421	0.343	0.343	-18.5	0.0	0.343	-18.5	0.1
7040	0.542	0.356	0.536	-1.2	50.5	0.542	0.0	52.4
7042	0.654	0.587	0.571	-12.7	-2.6	0.495	-24.3	-15.6
7046	0.383	0.315	0.299	-21.9	-5.0	0.301	-21.3	-4.3
7047	0.575	0.434	0.464	-19.3	6.9	0.434	-24.5	0.0
7056	0.609	0.492	0.503	-17.3	2.3	0.000	-100.0	-100.0
7067	0.624	0.538	0.538	-13.9	0.0	0.538	-13.9	0.0
7068	0.587	0.451	0.511	-12.9	13.3	0.513	-12.5	13.8
7069	0.654	0.614	0.608	-7.0	-0.9	0.577	-11.7	-5.9
7075	0.786	0.629	0.635	-19.2	0.9	0.552	-29.7	-12.3
7076	0.629	0.543	0.579	-7.9	6.6	0.358	-43.1	-34.1
7080	0.600	0.376	0.304	-49.4	-19.3	0.306	-49.0	-18.7
7084	0.774	0.628	0.574	-25.8	-8.6	0.566	-26.8	-9.9
7087	0.601	0.433	0.601	0.0	38.9	0.593	-1.3	37.1
7089	0.723	0.556	0.552	-23.6	-0.7	0.576	-20.3	3.6
7090	0.499	0.422	0.460	-7.7	9.1	0.417	-16.3	-1.1
7094	0.584	0.430	0.475	-18.7	10.5	0.449	-23.2	4.3
7096	0.625	0.575	0.471	-24.7	-18.1	0.466	-25.4	-18.9
7097	0.508	0.446	0.468	-7.8	5.0	0.449	-11.5	0.8
7099	0.599	0.375	0.398	-33.5	6.2	0.420	-29.8	12.2
7103	0.593	0.531	0.532	-10.3	0.1	0.529	-10.6	-0.3
7109	0.608	0.479	0.467	-23.2	-2.6	0.462	-24.0	-3.5
7115	0.705	0.621	0.655	-7.0	5.5	0.643	-8.7	3.5
7124	0.640	0.446	0.467	-27.1	4.6	0.420	-34.3	-5.8
7127	0.540	0.372	0.370	-31.5	-0.5	0.371	-31.3	-0.2
7129	0.497	0.398	0.398	-19.8	0.0	0.385	-22.5	-3.4
7132	0.857	0.504	0.504	-41.2	0.0	0.500	-41.6	-0.7
7145	0.646	0.567	0.538	-16.8	-5.2	0.505	-21.9	-11.0
7209	0.673	0.602	0.616	-8.4	2.4	0.602	-10.5	0.0
7258	0.462	0.381	0.409	-11.5	7.2	0.406	-12.0	6.6
7348	0.495	0.326	0.345	-30.3	5.7	0.250	-49.4	-23.3
7404	0.677	0.636	0.675	-0.3	6.3	0.662	-2.3	4.2
7406	0.597	0.463	0.450	-24.7	-2.9	0.446	-25.3	-3.7
7407	0.796	0.726	0.764	-4.0	5.2	0.731	-8.2	0.6
7415	0.683	0.565	0.511	-25.3	-9.7	0.474	-30.7	-16.2
7465	0.445	0.352	0.445	0.0	26.5	0.422	-5.2	19.8
7485	0.492	0.364	0.332	-32.6	-8.7	0.334	-32.1	-8.1
7504	0.499	0.424	0.370	-25.8	-12.7	0.375	-24.8	-11.6
7505	0.693	0.606	0.585	-15.7	-3.5	0.584	-15.7	-3.5
7506	0.587	0.477	0.477	-18.8	0.0	0.463	-21.1	-2.8
all	0.544	0.469	0.497	-8.7	5.9	0.467	-14.0	-0.3

Table 4. Comparison in terms of P@10 among the result merging runs. The percentage of increment has been omitted when the median value is 0; those entries are marked with **

Topic	Best	Median	mu	$\Delta_{mu-best}$	$\Delta_{mu-median}$	sh	$\Delta_{sh-best}$	$\Delta_{sh-median}$
7001	0.8	0.5	0.5	-37.5	0.0	0.5	-37.5	0.0
7003	1.0	0.8	0.9	-10.0	12.5	0.8	-20.0	0.0
7004	0.4	0.2	0.2	-50.0	0.0	0.2	-50.0	0.0
7007	0.7	0.5	0.7	0.0	40.0	0.7	0.0	40.0
7009	0.5	0.1	0.1	-80.0	0.0	0.1	-80.0	0.0
7018	0.6	0.4	0.3	-50.0	-25.0	0.3	-50.0	-25.0
7025	0.9	0.5	0.0	-100.0	-100.0	0.0	-100.0	-100.0
7030	0.6	0.3	0.1	-83.3	-66.7	0.2	-66.7	-33.3
7033	0.3	0.1	0.3	0.0	200.0	0.3	0.0	200.0
7034	0.3	0.2	0.3	0.0	50.0	0.3	0.0	50.0
7039	0.5	0.1	0.1	-80.0	0.0	0.1	-80.0	0.0
7040	0.3	0.2	0.2	-33.3	0.0	0.2	-33.3	0.0
7042	0.8	0.3	0.1	-87.5	-66.7	0.2	-75.0	-33.3
7046	0.1	0.0	0.0	-100.0	0.0	0.0	-100.0	0.0
7047	0.4	0.1	0.1	-75.0	0.0	0.1	-75.0	0.0
7056	0.5	0.4	0.2	-60.0	-50.0	0.0	-100.0	-100.0
7067	0.7	0.3	0.3	-57.1	0.0	0.3	-57.1	0.0
7068	0.6	0.3	0.3	-50.0	0.0	0.3	-50.0	0.0
7069	0.8	0.6	0.5	-37.5	-16.7	0.5	-37.5	-16.7
7075	0.9	0.4	0.3	-66.7	-25.0	0.2	-77.8	-50.0
7076	0.6	0.3	0.3	-50.0	0.0	0.3	-50.0	0.0
7080	0.4	0.1	0.0	-100.0	-100.0	0.0	-100.0	-100.0
7084	0.9	0.4	0.2	-77.8	-50.0	0.2	-77.8	-50.0
7087	0.6	0.3	0.3	-50.0	0.0	0.3	-50.0	0.0
7089	0.6	0.2	0.1	-83.3	-50.0	0.3	-50.0	50.0
7090	0.4	0.1	0.2	-50.0	100.0	0.2	-50.0	100.0
7094	0.5	0.3	0.3	-40.0	0.0	0.3	-40.0	0.0
7096	0.7	0.5	0.3	-57.1	-40.0	0.3	-57.1	-40.0
7097	0.4	0.0	0.3	-25.0	**	0.2	-50.0	**
7099	0.7	0.3	0.1	-85.7	-66.7	0.3	-57.1	0.0
7103	0.7	0.4	0.4	-42.9	0.0	0.4	-42.9	0.0
7109	0.6	0.2	0.2	-66.7	0.0	0.2	-66.7	0.0
7115	0.8	0.5	0.4	-50.0	-20.0	0.4	-50.0	-20.0
7124	0.7	0.3	0.3	-57.1	0.0	0.3	-57.1	0.0
7127	0.4	0.3	0.3	-25.0	0.0	0.3	-25.0	0.0
7129	0.5	0.2	0.1	-80.0	-50.0	0.1	-80.0	-50.0
7132	0.9	0.1	0.1	-88.9	0.0	0.1	-88.9	0.0
7145	0.6	0.3	0.3	-50.0	0.0	0.3	-50.0	0.0
7209	0.8	0.5	0.4	-50.0	-20.0	0.4	-50.0	-20.0
7258	0.3	0.1	0.1	-66.7	0.0	0.1	-66.7	0.0
7348	0.8	0.1	0.0	-100.0	-100.0	0.0	-100.0	-100.0
7404	0.8	0.5	0.5	-37.5	0.0	0.5	-37.5	0.0
7406	0.7	0.2	0.0	-100.0	-100.0	0.0	-100.0	-100.0
7407	1.0	0.8	0.8	-20.0	0.0	0.8	-20.0	0.0
7415	0.9	0.6	0.4	-55.6	-33.3	0.4	-55.6	-33.3
7465	0.3	0.0	0.3	0.0	**	0.2	-33.3	**
7485	0.4	0.1	0.1	-75.0	0.0	0.1	-75.0	0.0
7504	0.5	0.1	0.1	-80.0	0.0	0.1	-80.0	0.0
7505	0.6	0.3	0.1	-83.3	-66.7	0.1	-83.3	-66.7
7506	0.3	0.2	0.2	-33.3	0.0	0.2	-33.3	0.0
all	0.414	0.318	0.254	-38.6	-20.1	0.254	-38.6	-20.1