Overview of the TREC 2009 Chemical IR Track

Mihai Lupu[1], Florina Piroi[1], Xiangji (Jimmy) Huang[2], Jianhan Zhu[3], John Tait[1]
[1]Information Retrieval Facility (IRF), Vienna, Austria
[2]Information Retrieval Lab, York University, Toronto, Canada
[3]Department of Computer Science, University College London, London, UK
{m.lupu,f.piroi,john.tait}@ir-facility.org, jhuang@yorku.ca, jianhan.zhu@ucl.ac.uk

Abstract

TREC 2009 was the first year of the Chemical IR Track, which focuses on evaluation of search techniques for discovery of digitally stored information on chemical patents and academic journal articles. The track included two tasks: Prior Art (PA) and Technical Survey (TS) tasks. This paper describes how we designed the two tasks and presents the official results of eight participating groups.

1 Introduction

Any evaluation campaign has a set of criteria that generally fall in one of two categories: effectiveness (does the system do what it was designed to be doing?) and efficiency (how fast/reliable/cheap is it?). While in principle these two categories do not conflict, in practice, because human experts have to be involved in the effectiveness category, it is hard to run one experiment that goes both sufficiently deep in the analysis to assess actual effectiveness in real user context and sufficiently large scale to give a clear image of the scalability of the different systems. This is why we divided our track into two sub–tasks.

Technical Survey Task: 18 topics have been kindly provided by chemical patent experts based on their information needs. Participants' systems retrieve a ranked list of documents in response to each topic. In order to alleviate the evaluation work for the experts, and compare ordinary users and experts' views on relevant judgements, we carried out a two step evaluation procedure, where each topic is judged by two graduate students majored in chemistry in the first step, then presented to a patent expert for judgements by taking into account the students' judgements in the second step. This task enables us to understand the pros and cons of the participating systems in finding relevant chemical documents and how effectiveness can be improved.

Prior Art Search Task: The second task asked participating systems to find relevant patents with respect to a set of 1,000 existing patents. The results returned by the systems were not to be manually evaluated, but based on existing citations of the 1,000 patents and their family members. This task also contained a mini–task, where the participants were invited to submit the results to only the first 100 patents in the list. This task was intended to helps us investigate how to design both effective and efficient systems that can retrieve high quality relevant documents for a rather large number of topics.

The track organizers received registrations from 14 research groups from both academia and industry, who were allowed to download the data and topics. Eventually, 8 groups submitted at least one run to at least one of the two tasks. The methods applied vary substantially, from basic IR methods (e.g. vector space models without any pre–processing of the text) to advanced chemistry–specific methods using named entity recognition software and synonyms of chemical substances.

The remainder of the paper is organized as follows. We describe the test collection in Section 2, present Prior Art and Technical Survey task design and results in Section 3 and 4, respectively, and introduce participants' systems in Section 5.

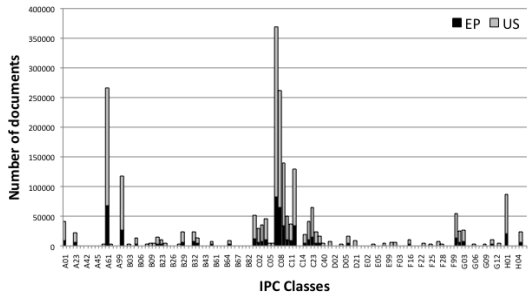


Figure 1: Distribution of files per patent class

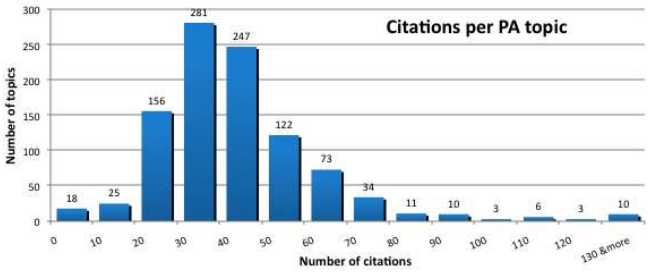


Figure 2: Distribution of the number of citations per topic in the prior art search topic set

2 Test Collection

In 2009, the Information Retrieval Facility (IRF) disposed of 2,648,160 patent files (approximately 112GB in size) from the chemical domain (classified under IPC codes C and A61K) that, after pre-processing, were available to the participants. This collection covers patents in the field until 2007, registered at three major patent offices, i.e., EPO, USPTO and WIPO. The format of the patent files is in XML.

Among these 2.6 million patent files, we distributed 1,185,012 files that contain (claims and (description or abstract)), i.e., that have enough textual information for making them useful in text retrieval. They sum up to 98.22GB of uncompressed data. Among these files, Figure 1 shows their distribution per IPC class. There are two observations to be made: 1. Class C is "Chemistry" 2. Many patent files are classified under more than one class, hence the apparently larger number of files in Figure 1 than the 1,185 million just mentioned.

This year's track has also benefited from chemical scientific articles, kindly provided by the Royal Society of Chemistry (RSC) in the UK. This data set consists of about 59,000 scientific articles, for a total of size of around 3GB, from 31 journals published by the RSC. The format is also in XML, but different schemas have been used for the RSC and patent documents.

3 Prior Art (PA) Task

The PA Task consisted of 1,000 topics which were full–text patent documents (i.e. consisting of at least claims and abstract or description) taken from both the European Patent Office (EPO) and the US Patent Office (USPTO). However, because the number of patents in the collection that came from the USPTO was much larger than the number of patents that came from the EPO, this difference was reflected in the set of topics: only 77 were from the EPO and the rest were from USPTO.

In choosing them, we basically sampled randomly from the collection, trying to optimize two objectives: minimize the number of applicant citations (the US patent applications are notorious for citing large sets of marginally relevant patents) and maximizing the total number of citations.

Participants were also allowed to submit the results for only the first 100 topics, where time and computational resources did not allow for submitting results for all the 1,000 topics. We took these 100 topics to be the first 100 of the 1,000 topics. We had not realized at that time that the order in which the topics had been initially selected from the database, resulted in the EPO topics being first and USPTO second. This resulted in the small PA topic set having a different distribution of topic sources than the large topic set. As we will see, this only has small influence on the systems' performance.

The challenge of the PA task is that the query, i.e., a whole patent document, is typically quite long, and query processing techniques may need to employed. We consider the MRR metric as important for this task since it is desirable to locate the first relevant documents as close as possible to the top rank position.

3.1 Relevance Judgements

Relevance judgements for this topic consisted of citations from the patent document that was used as a topic, as well as citations in family members of the original document.

Here we provide some background knowledge about citations and patent families. Like a research paper, a patent is, at different points in time, associated with other patents (or research papers). Before applying for a patent, the applicant must do a prior art search (as a researcher would include a literature review in a paper submitted to a conference or journal). This list of references are called *Applicant's citations*. Then, upon receiving the application, the patent office will review it and add another set of references, possibly rejecting the patent application. The similarity with academia is again striking: upon receiving a submission, a conference's programme committee would review the paper, suggesting other related work and possibly rejecting the submission for not being sufficiently novel. This set of references, that the patent office adds, are called *Examiner's citations* or *Search report citations*. Finally, in some cases, after a patent is published, a third party (a competing company, for instance) will oppose it, referencing works that neither the applicant, nor the examiner found, but which are very related to the new patent. This is less common in academia, but you can think of it as the situation in which, while presenting the work at the conference, a member of the audience stands up and claims that he had already solved the same problem years ago (hopefully citing some concrete works to prove that). Such references are called *Opposition citations*.

Putting all these references together, our collection of patents can take advantage of the large amount of manual work already done by field experts.

We can further extend this set of citations by looking at patent families. In our collection, we consider "simple" families: patents are related if they are basically the same idea submitted to different patent offices for protection in different jurisdictions[1]. It is a fact that when submitted to a patent office, there is a bias of the examiner towards the collection of that specific patent office, and other documents, published elsewhere, are not cited. We compensate this by considering a patent relevant in one of the following three situations:

1. it is directly cited by the topic patent,
2. it is a family member of a patent directly cited by the topic patent,
3. it is directly cited by a family member of the topic patent.

Given the limited size of the collection and the fact that applicants, examiners and opponents have supposedly carried out rather conclusive searches, we tend to consider that this is a fully judged collection in the Cranfield style and thus most measures returned by `trec_eval` based on the judgements are reliable indicators of systems' performance.

3.2 Results

Eight groups submitted to the short PA task. Among them, five also submitted to the full PA task, i.e., three groups did not submit to the full PA task. Comparing Table 1 and Table 2, the relative standings of the five groups' submitted runs, in terms of MAP, are similar.

Tables 3 and 4 show the results of the participating systems when considering only the EPO or the USPTO topics, respectively. We can see that most runs' performance on the

[1]For a concrete definition of "simple" and "extended" patent families, look at `http://www.epo.org/patents/patent-information/about/families/definitions.html`

Table 1: Results for the best two runs from each team in terms of MAP, and performance on a range of IR metrics for the short PA task.

Team	Run Tag	MAP	b-pref	P_30	Recall_100	NDCG	MRR
Univ of Applied Science Geneva	BiTeM09PAqe_sm	0.1688	0.7432	0.2010	0.4011	0.4357	0.5039
Univ of Applied Science Geneva	BiTeM09PAcba_sm	0.1683	0.7415	0.1990	0.4055	0.4359	0.5013
CMU	CMU09Chmtcdd_sm	0.0894	0.4999	0.1277	0.2610	0.2972	0.3819
Purdue Univ	purduePA09r2	0.0679	0.4236	0.0983	0.2385	0.2476	0.2508
Purdue Univ	purduePA09r1	0.0654	0.4222	0.0893	0.2298	0.2398	0.2468
CMU	CMU09Chmtcd_sm	0.0517	0.3779	0.0810	0.1631	0.2090	0.3002
Univ of Iowa	UIowaS09PA1_sm	0.0485	0.4207	0.0817	0.1888	0.2245	0.2373
DUTIR	DUTIRRun3	0.0204	0.0984	0.0397	0.0517	0.0702	0.0932
DUTIR	DUTIRRun2	0.0203	0.0969	0.0420	0.0508	0.0695	0.0924
York Univ	york09caPA01_sm	0.0180	0.1522	0.0373	0.0552	0.0883	0.1120
York Univ	york09caPA03_sm	0.0160	0.1574	0.0307	0.0480	0.0837	0.0976
Fraunhofer SCAI	SCAI09PAt1a_sm	0.0075	0.4274	0.0037	0.0147	0.1395	0.0403
Fraunhofer SCAI	SCAI09PAf1e_sm	0.0069	0.5444	0.0050	0.0137	0.1686	0.0189
Univ of Iowa	UIowaS09PA3	0.0066	0.1092	0.0133	0.0447	0.0542	0.0490
MSOE	msoe09TSx4ta	0.0022	0.1635	0.0047	0.0172	0.0600	0.0177
MSOE	msoe09TSx5ta	0.0019	0.1365	0.0037	0.0156	0.0496	0.0136

Table 2: Results for the best two runs from each team in terms of MAP, and performance on a range of IR metrics for the full PA task.

Team	Run Tag	MAP	b-pref	P_30	Recall_100	NDCG	MRR
Univ of Applied Science Geneva	BiTeM09PAcba	0.1835	0.6542	0.2567	0.3375	0.4192	0.5328
Univ of Applied Science Geneva	BiTeM09PAqe	0.1823	0.6602	0.2547	0.3335	0.4192	0.5236
CMU	CMU09Chmtcdd	0.0975	0.4570	0.1776	0.2326	0.3091	0.5129
Univ of Iowa	UIowaS09PA1	0.0683	0.4066	0.1329	0.1851	0.2643	0.3864
CMU	CMU09Chmtcd	0.0647	0.3605	0.1261	0.1678	0.2344	0.4004
York Univ	york09caPA01	0.0566	0.3385	0.1109	0.1431	0.2262	0.3361
York Univ	york09caPA03	0.0343	0.1978	0.0748	0.0855	0.1376	0.2337
Fraunhofer SCAI	SCAI09PAf1e	0.0065	0.4004	0.0052	0.0124	0.1471	0.0247
Fraunhofer SCAI	SCAI09PAt1e	0.0060	0.3777	0.0044	0.0113	0.1417	0.0265

USPTO subset is higher than that on the EPO subset. Again, the systems' performance is plotted in Figure 6. We can see that there are considerable differences (we are still to run significance tests on these differences). A particular observation is that run *york09caPA01* seems to be doing much better, according to all measures, for USPTO topics than for EPO topics—this is something worth investigating.

We were also interested in calculating the statistical significance of the difference of the results returned by the runs. We used the randomized test introduced by Smucker in [4] to compute the difference between the MAP results. We compared this approach with the t–test, which was argued in a series of works [4, 5, 2, 1, 6] to be the second most reliable. Figure 3 compares the results of the randomization test and the t–test. We can see that though they seem to correlate to a large extent, the t–test overestimates many p–values. Further investigation results on the groups of runs whose differences are not statistically significant, as well as on the correlation between the two types of the significance tests will be provided in the final version of this report.

Table 3: Results for topics that come only from EP patents

Team	Run Tag	MAP	b-pref	P_30	Recall_100	NDCG	MRR
Univ of Applied Science Geneva	BiTeM09PAcba	0.1404	0.7655	0.1494	0.4075	0.4165	0.4226
Univ of Applied Science Geneva	BiTeM09PAqe	0.1396	0.7668	0.1468	0.4043	0.4170	0.4280
CMU	CMU09Chmtcdd	0.0845	0.4894	0.1030	0.2612	0.2798	0.3357
Univ of Iowa	UIowaS09PA1	0.0468	0.4287	0.0649	0.1960	0.2133	0.2078
CMU	CMU09Chmtcd	0.0426	0.3535	0.0515	0.1521	0.1827	0.2510
York Univ	york09caPA03	0.0082	0.1033	0.0130	0.0287	0.0435	0.0336
Fraunhofer SCAI	SCAI09PAt1a	0.0068	0.4212	0.0026	0.0158	0.1266	0.0307
Fraunhofer SCAI	SCAI09PAf1e	0.0062	0.5597	0.0048	0.0143	0.1602	0.0254
York Univ	york09caPA01	0.0042	0.0632	0.0108	0.0216	0.0289	0.0415

Table 4: Results for topics that come only from US patents

Team	Run Tag	MAP	b-pref	P_30	Recall_100	NDCG	MRR
Univ of Applied Science Geneva	BiTeM09PAcba	0.1871	0.6450	0.2656	0.3316	0.4194	0.5420
Univ of Applied Science Geneva	BiTeM09PAqe	0.1859	0.6513	0.2637	0.3276	0.4194	0.5316
CMU	CMU09Chmtcdd	0.0986	0.4543	0.1838	0.2303	0.3116	0.5277
Univ of Iowa	UIowaS09PA1	0.0701	0.4047	0.1386	0.1842	0.2686	0.4013
CMU	CMU09Chmtcd	0.0665	0.3611	0.1323	0.1691	0.2387	0.4129
York Univ	york09caPA01	0.0609	0.3615	0.1193	0.1532	0.2426	0.3607
York Univ	york09caPA03	0.0365	0.2057	0.0800	0.0902	0.1455	0.2504
Fraunhofer SCAI	SCAI09PAf1e	0.0065	0.3871	0.0052	0.0122	0.1460	0.0246
Fraunhofer SCAI	SCAI09PAt1e	0.0061	0.3722	0.0045	0.0107	0.1427	0.0274

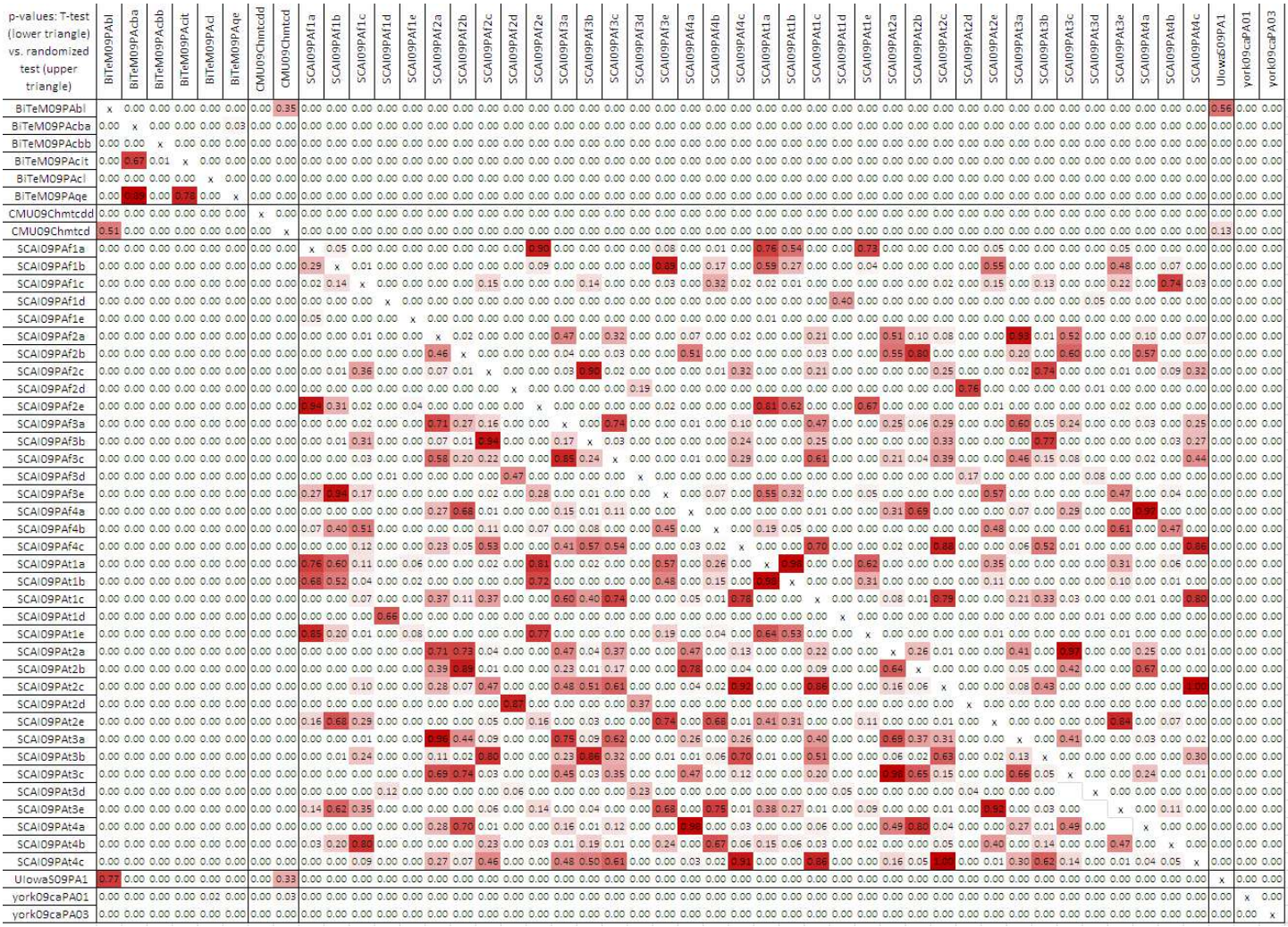
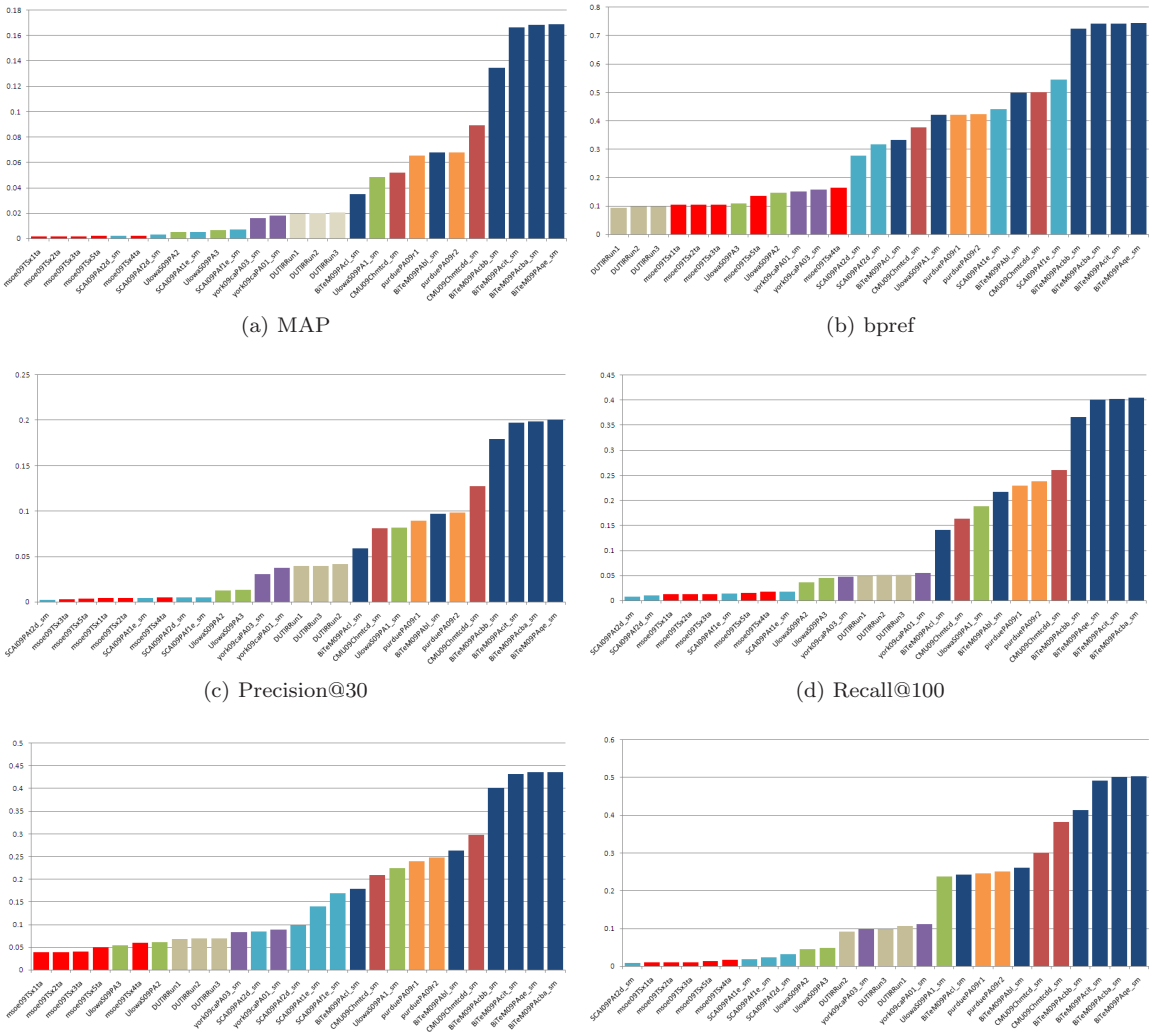


Figure 3: Statistical significance of the differences between the MAP results of the runs submitted for the full PA task, according to the t–test (lower triangle) and the randomization test (upper triangle). Darker cells represent higher p-values.

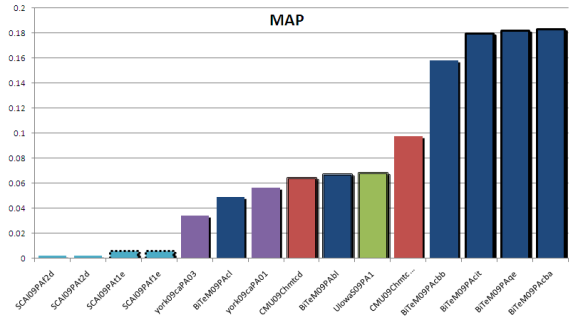


(a) MAP (b) bpref

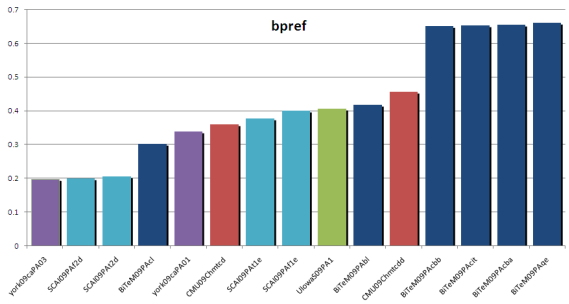
(c) Precision@30 (d) Recall@100

(e) NDCG (f) MRR

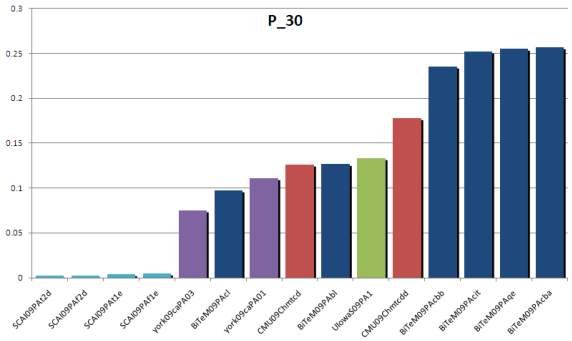
Figure 4: Systems' performance by six IR measures for the small set of patents PA task



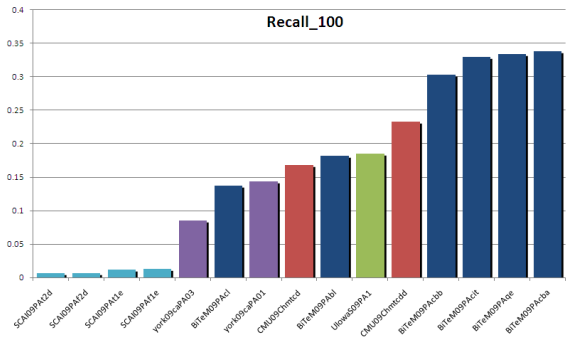
(a) MAP



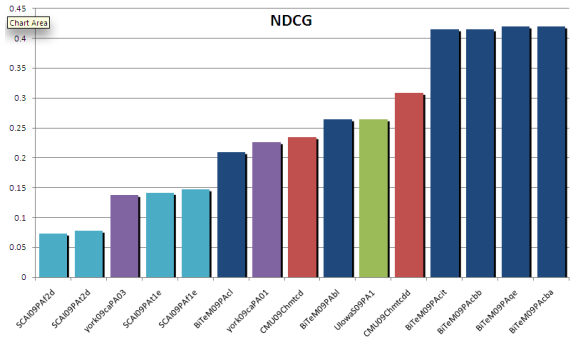
(b) bpref



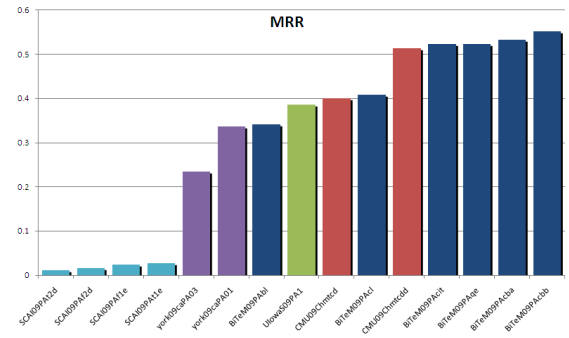
(c) Precision@30



(d) Recall@100

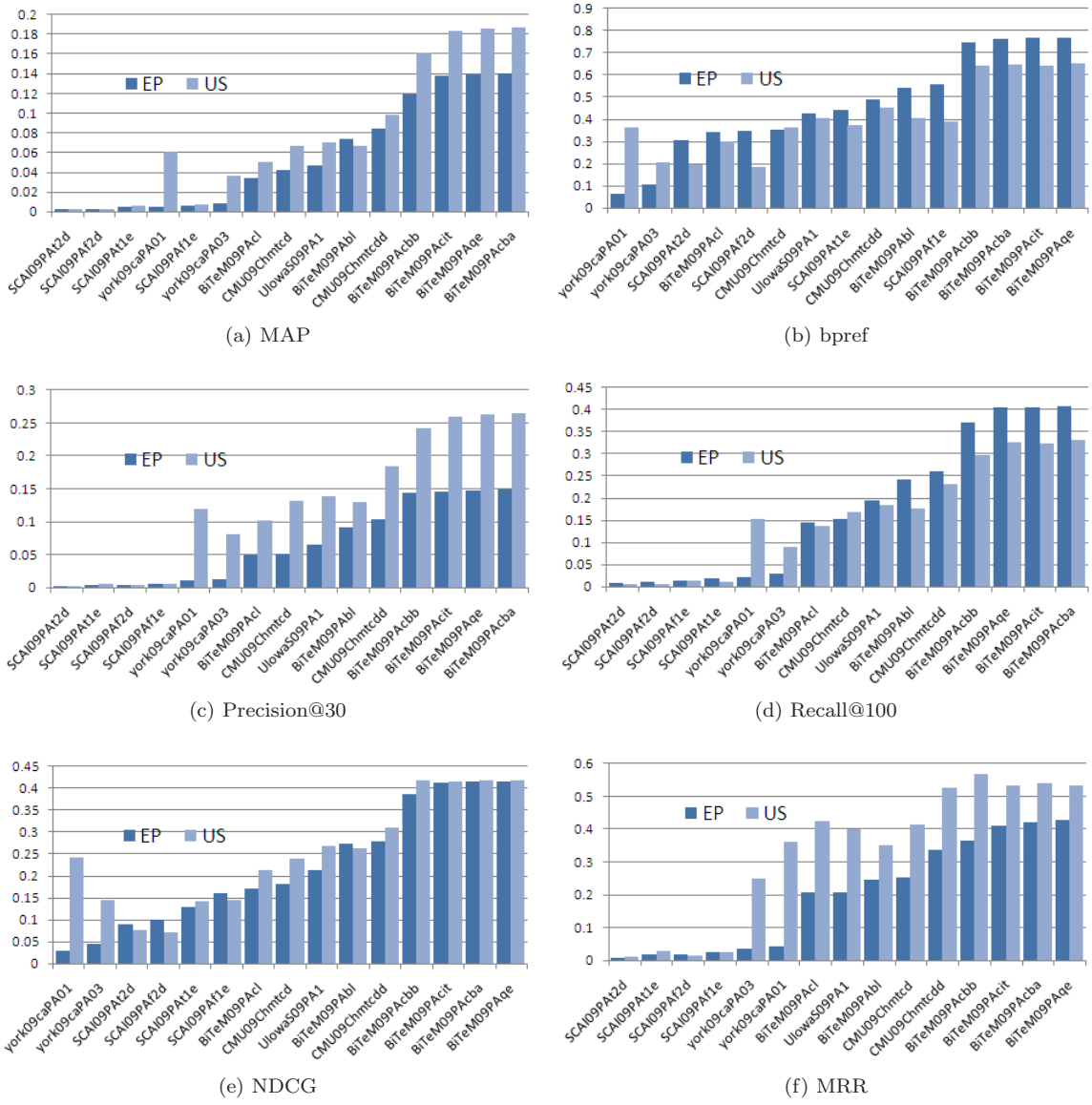


(e) NDCG



(f) MRR

Figure 5: Systems' performance by six IR measures for the full PA task



(a) MAP

(b) bpref

(c) Precision@30

(d) Recall@100

(e) NDCG

(f) MRR

Figure 6: Differences in results between topics coming from the EPO and topics coming from USPTO

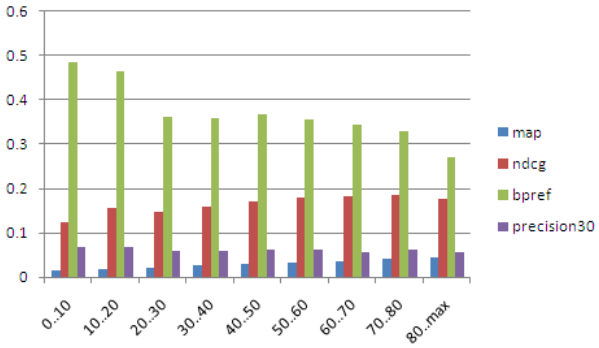


Figure 7: Average performance according to four measures, over the number of relevant documents per topic

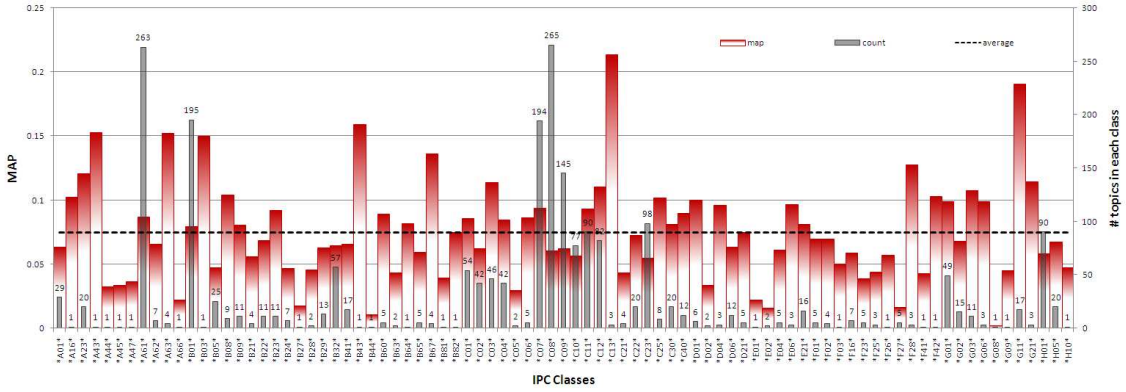


Figure 8: Distribution of MAP values over the different classes of topic patents. The grey bars represent the support of this average (i.e. how many topics were in this class)

3.3 Other observations

There are many aspects to be analysed with respect to this corpus and the measures calculated during this evaluation campaign. We could look at the distribution of these measures across patent classes, number of citations per patent, size of the patent, provenance (e.g. EPO or USPTO), kind of documents used as topics (application vs. granted patent), year of publication. In this section we look at some of them, that we found most interesting. Many others are available on the track's wiki page[2]

Figure 7 shows how measures 'prefer' topics which have more or less relevant documents. As we can see, MAP seems to favour topics with many relevant results, while bpref the converse.

We also looked to see if some types (i.e. classes) of patents are easier to retrieve than others. In Figure 8, we can see for instance a clear difference between two of the most popular chemical classes: C07 (Organic Chemistry) and C08 (Organic Macromolecular Compounds), with C08 topics being on average harder than C07 ones. We can also see that A61 (Medical or Veterinary Science) has performed quite well. C23 (Coating metallic material) seems to perform less than average. Inorganic Chemistry (C01), Glass/Minerals (C03) and Cements/Concrete/Artificial Stone/Ceramic (C04) on the other hand tended to do better. Similar results are visible when looking at the Recall@100 measure (Figure 9

On the other hand, looking at the size of the topics did not seem to provide significant insights. Figure 10 shows that the size of the topics did not have a major impact on the values of the NDCG measure.

Each participating group is encouraged to do this analysis on their own results.

[2]http://wiki.ir-facility.org/

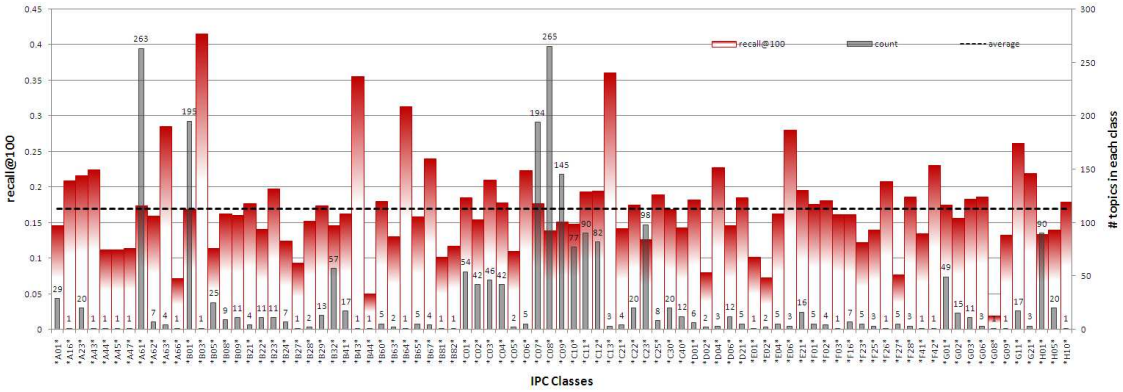


Figure 9: Distribution of Recall@100 values over the different classes of topic patents. The grey bars represent the support of this average (i.e. how many topics were in this class)

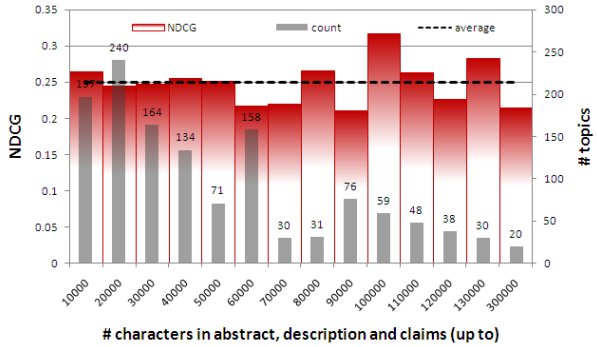


Figure 10: Distribution of NDCG values over the sizes of the topic patents. The grey bars represent the support of this average (i.e. how many topics were in each size interval)

4 Technical Survey (TS) Task

The TS task is similar to a traditional ad hoc retrieval task, however, the challenge is the way to deal with chemical specific problems such as synonyms and abbreviations. Five patent experts have kindly provided topics from their experience as patent searchers. Table 9 in the Appendix shows the 18 topics, which aim to find patents or articles about organic, high molecular weight, pharmaceutical, and inorganic chemistry.

4.1 Sampling and Relevance Judgements

We employed the stratified sampling approach [8], where we sampled 100% documents to depth 10, 30% to depth 30 and 10% to depth 100. On average, 300 documents per topic were sampled for evaluation. In order to save the valuable time of the patent experts, we adopted a two–stage evaluation process. In the first stage, each topic was evaluated by two graduate students major in chemistry independently. Then their evaluation results were merged by setting a document as relevant if at least one of the two students has judged it as relevant. More precisely, the resulting evaluation was taken as the maximum between those of the two students, using the following order: $-1 < -2 < 0 < 1 < 2$, where the numerical values for relevance judgements are -2–"unsure", -1–"unjudged", 0–"not relevant", 1–"relevant", and 2–"highly relevant".

In the second stage, these merged evaluations were presented to the experts for official evaluation. They had the option to view subsets of the pool evaluated by students as: show all documents, documents judged as "relevant"/"highly relevant", or documents judged as "unsure"/"unjudged", respectively.

4.1.1 Statistics of Students' Evaluations

Tables 5 and 6 present the situation after the first phase of manual assessments conducted by graduate students in the chemical field.

Tables 5 shows the degree of disagreements per topic among the assessing students. A *strict* disagreement means that the two students chose two different relevance judgements, a *conflictual* disagreement means that one assessor considered a document as (highly) "relevant" to the topic, while the other considered it as "not relevant". Lastly, a *lenient* disagreement means that one of the assessors judged a document as "unsure" or left it "unjudged". Naturally, the conflictual and lenient disagreements sum up to the number of strict disagreements.

The high percentage of strict disagreements for topic TS–4 reflects the fact that at the time of closing the first evaluation stage one of the assessors did not finish his task. We can hypothesize that a high disagreement rate means a difficult topic, while low disagreement rate indicates an easy topic.

Table 5: Student judgement disagreements

topic	**#docs**	**strict**		**conflict**		**lenient**	
TS–1	325	152	46.77%	98	30.15%	54	16.62%
TS–2	287	87	30.31%	62	21.60%	25	8.71%
TS–3	213	59	27.70%	25	11.74%	34	15.96%
TS–4	350	316	90.29%	7	2.00%	309	88.29%
TS–5	255	12	4.71%	9	3.53%	3	1.18%
TS–6	261	46	17.62%	31	11.88%	15	5.75%
TS–7	318	103	32.39%	69	21.70%	34	10.69%
TS–8	195	81	41.54%	48	24.62%	33	16.92%
TS–9	292	201	68.84%	56	19.18%	145	49.66%
TS–10	283	122	43.11%	48	16.96%	74	26.15%
TS–11	354	108	30.51%	90	25.42%	18	5.08%
TS–12	284	125	44.01%	95	33.45%	30	10.56%
TS–13	327	12	3.67%	9	2.75%	3	0.92%
TS–14	367	61	16.62%	28	7.63%	33	8.99%
topic	**#docs**	**conflict**		**strict**		**lenient**	

Continued on next page

Table 5: Student judgment disagreements, continued

topic	#docs	strict		conflict		lenient	
Ts–15	342	56	16.37%	52	15.20%	4	1.17%
Ts–16	334	117	35.03%	54	16.17%	63	18.86%
Ts–17	357	217	60.78%	215	60.22%	2	0.56%
Ts–18	374	111	29.68%	87	23.26%	24	6.42%
total	**5518**	**1986**	**35.99%**	**1083**	**19.63%**	**903**	**16.36%**

Table 6 shows the percentages of relevant/not relevant documents after the two sets of judgements per topic were merged.

Table 6: Student judgement: results

topic	#docs	highly relevant(2)		relevant(1)		irrelevant(0)		unsure(-2)		unjudged(-1)	
Ts–1	325	81	24.92%	156	48.00%	88	27.08%	0	0.00%	0	0.00%
Ts–2	287	94	32.75%	75	26.13%	117	40.77%	0	0.00%	1	0.35%
Ts–3	213	107	50.23%	30	14.08%	75	35.21%	0	0.00%	1	0.47%
Ts–4	350	0	0.00%	21	6.00%	270	77.14%	0	0.00%	59	16.86%
Ts–5	255	9	3.53%	23	9.02%	223	87.45%	0	0.00%	0	0.00%
Ts–6	261	38	14.56%	22	8.43%	201	77.01%	0	0.00%	0	0.00%
Ts–7	318	84	26.42%	75	23.58%	153	48.11%	4	1.26%	2	0.63%
Ts–8	195	62	31.79%	50	25.64%	83	42.56%	0	0.00%	0	0.00%
Ts–9	292	168	57.53%	104	35.62%	15	5.14%	0	0.00%	5	1.71%
Ts–10	283	89	31.45%	71	25.09%	119	42.05%	0	0.00%	4	1.41%
Ts–11	354	30	8.47%	93	26.27%	231	65.25%	0	0.00%	0	0.00%
Ts–12	284	65	22.89%	75	26.41%	142	50.00%	0	0.00%	2	0.70%
Ts–13	327	13	3.98%	7	2.14%	307	93.88%	0	0.00%	0	0.00%
Ts–14	367	67	18.26%	26	7.08%	274	74.66%	0	0.00%	0	0.00%
Ts–15	342	9	2.63%	53	15.50%	280	81.87%	0	0.00%	0	0.00%
Ts–16	334	101	30.24%	54	16.17%	171	51.20%	0	0.00%	8	2.40%
Ts–17	357	114	31.93%	116	32.49%	121	33.89%	5	1.40%	1	0.28%
Ts–18	374	178	47.59%	70	18.72%	120	32.09%	0	0.00%	6	1.60%
total	**5518**	**1309**	**23.72%**	**1121**	**20.32%**	**2990**	**54.19%**	**9**	**0.16%**	**89**	**1.61%**

4.1.2 Statistics of Experts' Evaluations

Table 7 shows the percentages of relevant/not relevant documents after the second phase of the manual assessments.

Table 7: Expert judgement: results

topic	#docs	highly relevant(2)		relevant(1)		irrelevant(0)		unsure(-2)		unjudged(-1)	
Ts–1	325	81	24.92%	156	48.00%	88	27.08%	0	0%	0	0%
Ts–2	287	96	33.45%	52	18.12%	139	48.43%	0	0%	0	0%
Ts–3	213	107	50.23%	30	14.08%	75	35.21%	1	0.47%	0	0%
Ts–4	350	15	4.29%	28	8.00%	303	86.57%	4	1.14%	0	0%
Ts–5	255	8	3.13%	24	9.41%	223	87.45%	0	0%	0	0%
Ts–6	261	36	13.79%	33	12.64%	192	73.56%	0	0%	0	0%
Ts–7	318	1	0.31%	3	0.94%	314	98.74%	0	0%	0	0%
Ts–8	195	29	14.87%	18	9.23%	143	73.33%	5	2.56%	0	0%
Ts–9	292	26	8.90%	11	3.77%	255	87.33%	0	0%	0	0%
Ts–10	283	89	31.45%	71	25.09%	119	42.05%	4	1.41%	0	0%
Ts–11	354	30	8.47%	93	26.27%	231	65.25%	0	0%	0	0%
Ts–12	284	58	20.49%	74	26.15%	150	53.00%	1	0.35%	0	0%
Ts–13	327	11	3.36%	7	2.14%	309	94.50%	0	0%	0	0%
topic	**#docs**	**highly relevant(2)**		**relevant(1)**		**irrelevant(0)**		**unsure(-2)**		**unjudged(-1)**	

Continued on next page

Table 7: Expert judgment: results, continued

topic	#docs	highly relevant(2)		relevant(1)		irrelevant(0)		unsure(-2)		unjudged(-1)	
Ts–14	367	47	12.81%	16	4.36%	292	79.56%	12	3.27%	0	0%
Ts–15	342	8	2.34%	18	5.26%	315	92.11%	0	0%	1	0.29%
Ts–16	334	7	2.10%	24	7.19%	295	88.32%	8	2.39%	0	0%
Ts–17	357	114	31.93%	116	32.49%	126	35.29%	1	0.28%	0	0%
Ts–18	374	178	47.59%	70	18.72%	125	33.42%	1	0.27%	0	0%
total	**5518**	**941**	**17.06%**	**844**	**15.30%**	**3694**	**66.96%**	**37**	**0.67%**	**1**	**0.02%**

4.1.3 Exceptions

For topics Ts–9 and Ts–16, the expert decided that too many documents had been judged as relevant by the students. After further discussions, we decided to have the respective students re–evaluate the two topics, following additional instructions from the expert. While these evaluations have been done by the students, we include their results in the expert's result evaluations (see section sub4.1.2). This is because the expert assessments of these new judgements do not differ much from the student's judgements (for Ts–9 there is no difference).

Unfortunately, topics Ts–1, Ts–3, Ts–10 were not judged by the patent expert, as the respective person was unavailable at the time of assessments. Therefore, the three greyed out rows in table 7 contain the relevance assessments made by the students.

4.2 Procedure and used measures

We evaluated the results sampled as described in section 4.1 and computed the extended inferred average precision (xinfAP) and the inferred normalized discounted cumulative gain (infNDCG) as described in [8]

4.2.1 xinfAP

In [8], Yilmaz and colleagues extended the infAP measure [7] by taking non–random samples from the pool of documents. We adopted this measure because it appears to estimate AP more accurately than infAP, given the same evaluation effort.

4.2.2 infNDCG

Also based on a stratified sampling approach, infNDCG extends nDCG [3], whose aim is to represent the common view that relevant documents returned higher in the ranked list are more important than similarly relevant documents returned lower in the list.

4.3 Analysis of results

In figure 11 we've put together the results of the extended inferred AP for the 31 runs submitted and for each of the 18 topics. The idea was to get a quick image of which topics were generally not well answered, which ones were answered only by some systems and which ones were answered by all systems.

The first thing that draws our attention is the fact that only one run (SCAI09TSMAN) submitted a fair result for topic Ts–13. Topic 13 refers to the use of tetrahydrocannabinol (Thc) as an anti–tumor agent. We note that the narrative of this topic is quite small and that the only run that performed well used a manual query formulation. We may conclude then that the text of the topic was insufficient for the systems to extract sufficiently meaningful chemical information to be able to correctly retrieve relevant documents, but this is to be further investigated in collaboration with Fraunhofer SCAI.

It is equally interesting to look at situations where one run for a particular participant performed significantly differently from the other runs of the same participant. This happens for BiTeM09po. Particularly, this run seems to do better than its peers on topics Ts–1 and Ts–11. This run used patent families as a basic unit to index instead of a document. It is unclear how this generalizes to scientific articles though.

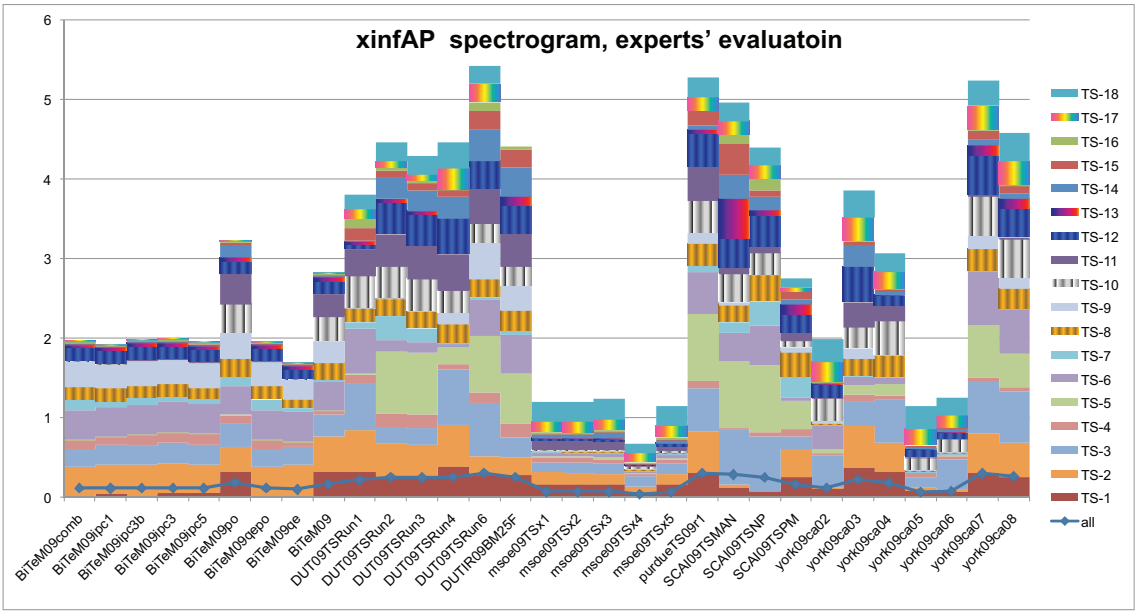


Figure 11: "Spectrogram" of the extended inferred average precision of the runs submitted for the TS tasks, based on the experts' evaluation.

Table 8: Results from the best two runs from each team in terms of xinfAP, and performance in terms of the inferred NDCG. Based on experts' evaluation.

Team	Run	xinfAP	infNDCG
DUT	DUT09TSRun6	0.301352563	0.535624287
Purdue	purdueTS09r1	0.292944735	0.535857751
York Univ	york09ca07	0.290945835	0.495564009
Fraunhofer SCAI	SCAI09TSMAN	0.275340089	0.493232092
York Univ	york09ca08	0.254126663	0.478079741
DUT	DUT09TSRun2	0.24820495	0.488650961
Fraunhofer SCAI	SCAI09TSNP	0.244157701	0.446849598
Univ. of Applied Sciences Geneva	BiTeM09po	0.179119822	0.393264533
Univ. of Applied Sciences Geneva	BiTeM09	0.157118933	0.345235351
Milwakee School of Engineering	msoe09TSx3	0.068977164	0.316426733
Milwakee School of Engineering	msoe09TSx1	0.066913758	0.313434303

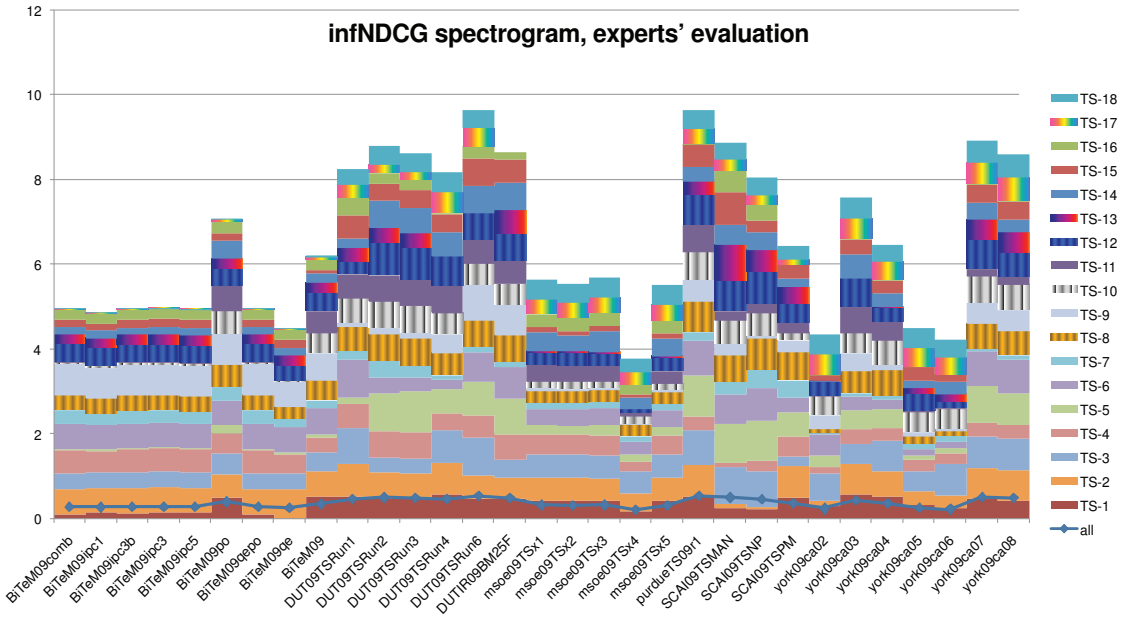


Figure 12: "Spectrogram" of the inferred nDCG of the runs submitted for the TS tasks, based on the experts' evaluation.

5 Approaches

The following are descriptions of the approach taken by different groups. These paragraphs were contributed by participants and are intended to be a road map to their papers in the TREC proceedings. Below each group name is a list of their runs submitted to each task. For the PA task, runs submitted for only 100 topics are identified with *'priorart_sm'*.

5.1 Carnegie Mellon University

Prior Art runs: CMU09Chmtcdd, CMU09Chmtcd
Technical Survey runs: none

The focus is on date filtering and formulating ranking queries. For filtering, different date filtering criteria are tested and reported, the best strategy being filtering out publications with publication date after the latest priority date of the query patent. For ranking, using term occurrence statistics from the whole patent improves over just using terms from title and claims of the query patent. Experiments were conducted easily with the structured retrieval support of the out-of-box Lemur/Indri tool–kit.

5.2 Univ of Applied Science Geneva

Prior Art runs: BiTeM09PAbl, BiTeM09PAcba, BiTeM09PAcbb, BiTeM09PAcit, BiTeM09PAcl, BiTeM09PAqe
Technical Survey runs: BiTeM09qepo, BiTeM09qe, BiTeM09po, BiTeM09ipc5, BiTeM09ipc3b, BiTeM09ipc3, BiTeM09ipc1, BiTeM09comb, BiTeM09

The goal of the first TREC Chemical track was to retrieve documents relevant to a given patent query, within a large collection of patents in chemistry. Regarding this objective, for the Prior Art subtask, our runs performed significantly better that runs submitted by other participating teams. Baseline retrieval methods achieved relatively poor performances (Mean Average Precision = 0.067). Query expansion, driven my chemical named entity recognition resulted in some modest improvement (+2 to 3%). Filtering based on IPC codes did not result in any significant improvement. A re–ranking strategy, based on claims only improved MAP by about 3%. The most effective gain was obtained by using patent citation patterns.

Somehow similar to feed–back but restricted to citations, we used patents cited in the retrieved patents in order to boost the retrieval status value of the baseline run. This strategy led to a remarkable improvement (MAP 0.18, +168 %). Nevertheless, as official topics were sampled from the collection disregarding their creation date, our strategy happened to exploit citations of patents which were patented after the topic itself. From a user perspective, such a setting is questionable. We think that future TREC-CHEM competitions should address this issue by using patents filed as recently as possible.

5.3 Dalian University of Technology

Prior Art runs: DUTIRRun1 (priorart_sm), DUTIRRun2 (priorart_sm), DUTIRRun3 (priorart_sm)
Technology Survey runs: DUT09TSRun[1..6], DUTIR09BM25F

For the technology survey task, our experiment was conducted on two text retrieval models, BM25 and Language Model for IR (LMIR). The first three runs were based on the LMIR model, and we used a combination of the title and narrative of the topic to retrieve the documents. Moreover, we attempted the pseudo-relevance feedback (PRF) method to expand the original query in DUT09TSRun3. The last three runs in this task were mainly focused on structure based retrieval. We used the combination of the title and narrative to do retrieval on different fields of patents. In our experiment, some different weighting functions were used to rank each fields (e.g. title, abstract, description, claims) based on BM25F structure-based weighting model. For prior art task, we focused on formulating the queries from the query patents. We selected 60 terms with the largest probabilities (TF-IDF scores) from different fields (e.g. title, abstract, description, claims) as the original query to retrieval the relevance documents.

5.4 Iowa University

Prior Art runs: UIowaS09PA1,UIowaS09PA2 (priorart_sm), UIowaS09PA3 (priorart_sm)
Technology Survey runs: none

We submitted three separate runs to the Prior Art track. Our first submission eliminated patents with priority dates later than (after) a given query patent's priority date. It also included "unnested" claims so they could stand as independent documents for comparisons. We produced two separate Lucene indexes: one with patent claims alone; the other with the title, description, abstract, and classification portions (TDAC index) and determined the most effective weighting of these summary functions when merging the results into a single list. In our second submission, we used only the primary classification information from the TDAC index to retrieve those patents with matching IPC primary classification codes. Our third submission approximated patent priority dates by patent number and assumed patent numbers reflected a temporal sequence—only those patent numbers lower than the target patent number were included.

5.5 Fraunhofer SCAI

Prior Art runs: SCAI09PA[f,t]1[a..e],SCAI09PA[f,t]2[a..e],SCAI09PA[f,t]3[a..e], SCAI09PA[f,t]4[a..c]
Technical Survey runs: SCAI09TSMAN, SCAI09TSNP, SCAI09TSPM

The core of our framework is an index of 1.2 million chemical patents provided as a dataset by TREC. For technology survey task, we submitted three runs based on simple entity recognition, manual querying and automatically generated noun phrase based querying. It was observed that manual querying outperformed the remaining runs with the best nDCG score of 0.49. For prior art search task, we introduced several new fields into the conventional index that contained normalized noun phrases,biomedical as well as chemical entities. We submitted 36 runs for this task that are based on automatically querying with tokens, noun phrases and entities. The results showed that the token based search performed better than the remaining runs with the best bpref score of 0.40.

5.6 Purdue University

Prior Art runs: purduePA09r1 (priorart_sm), purduePA09r2 (priorart_sm)
Technical Survey runs: purdueTS09r1

For technology survey task, we used a weighted combination of query title and narrative to do retrieval on different fields (e.g. title, abstract, claims, description, whole doc, etc.) of patents and documents in a weighted way. We used synonyms (from PubChem) for chemicals that have been identified, used simple entity recognition to extract information that is later used to increment or decrement weights of some terms and to filter out documents from the ranked list. For prior art search task; we used all title words, and selected sets of terms from the claims, abstract and description fields of query patents (top 30 terms wrt. TF–IDF scores are chosen for each) to construct the queries. From those terms, chemical entities are extracted and synonyms for the identified chemical entities are also included from PubChem. Then structured queries are formed to do retrieval over different fields of documents with different weights. Then some of the retrieved documents are filtered out from the ranked list because of date constraints. If priority dates don't exist, we use publication date as the corresponding dates. Last but not the least, IPC similarities between query patent and its retrieved patents are exploited to re–rank the retrieved documents.

5.7 Milwaukee School of Engineering Team

Prior Art runs: msoe09TSx[1..5]1ta (priorart_sm)
Technical Survey runs: msoe09TSx[1..5]

A distributed information retrieval system based on a dimensional data model. The indexing model supports chemical named entity identification and efficient aggregation of term and entity statistics at multiple levels of patent structure including individual words, sentences, claims, descriptions, abstracts, and titles. Multiple forms of evidence are integrated within a probabilistic graphical model. The system is deployed on 12 Amazon EC2 instances.

5.8 York University

Prior Art runs: york09caPA01, york09caPA03
Technology Survey runs: york09ca0[2..7]

Our chemical experiments mainly focus on addressing three major problems in two chemical information retrieval tasks, technology survey task and prior art task. The three problems are: (1) how to deal with chemical terminology synonyms? (2) how to deal with chemical terminology abbreviation? (3)how to deal with long queries in Prior Art (PA) task? For technology survey, we proposed a chemical terminology expansion algorithm with the professional chemical domain information from two chemical websites, ChemID plus and PubChem. We also introduced an algorithm using the collection's information in prior art task for keyword selection. The Mean Average Precision (MAP) for our TS task run "york09ca07" using *Algorithm* 1 was 0.2519 and for our PA task run "york09caPA01" using *Algorithm* 2 was 0.0566. The evaluation results show that both algorithms are effective for improving retrieval performance.

6 Acknowledgements

The organizers of the TREC Chemistry Track wish to thank

- Emine Yilmaz (Microsoft Research) and Evangelos Kanoulas (North-Eastern University) for their help in the sampling and evaluation of the results of the Technical Survey (TS) task.
- NIST organizers for their constant guide
- Richard Kidd at the Royal Society of Chemistry, for providing the set of scientific articles
- Matrixware Information Services Gmbh, and in particular the Data Services team, for providing the data and helping us understand it
- The TREC CHEM committee members

References

[1] Ben Carterette. On rank correlation and the distance between rankings. In *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 436–443, New York, NY, USA, 2009. ACM.

[2] Ben Carterette and Mark D. Smucker. Hypothesis testing with incomplete relevance judgments. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 643–652, New York, NY, USA, 2007. ACM.

[3] K. Järvelin and J. Kekäläinen. Cumulated Gain-based Evaluation of IR Techniques. *ACM Transactions on Information Systems*, 20(4), 2002.

[4] Mark D. Smucker, James Allan, and Ben Carterette. A comparison of statistical significance tests for information retrieval evaluation. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 623–632, New York, NY, USA, 2007. ACM.

[5] Mark D. Smucker, James Allan, and Ben Carterette. Agreement among statistical significance tests for information retrieval evaluation at varying sample sizes. In *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 630–631, New York, NY, USA, 2009. ACM.

[6] William Webber, Alistair Moffat, and Justin Zobel. Statistical power in retrieval experimentation. In *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management*, pages 571–580, New York, NY, USA, 2008. ACM.

[7] E. Yilmaz and J. A. Aslam. Estimating Average Precision with Incomplete and Imperfect Judgements. In *Proc. of CIKM*, 2006.

[8] Emine Yilmaz, Evangelos Kanoulas, and Javed A. Aslam. A simple and efficient sampling method for estimating AP and NDCG. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 603–610, New York, NY, USA, 2008. ACM.

7 Appendix

Table 9: Topics used in the Technical Survey Track of TREC–CHEM 2009

#	Topic Title	Category	Narrative
1	Curable composition of Organopolysiloxane and its thermal treating	organic, high molecular weight	We are a research group at a university/company and plan starting a new project. For this we need to have knowledge on paints, coating and curable compositions. More specifically we would like to have information about a curable composition made of an organopolysiloxane having a functional group capable of condensation reaction at both molecular termini, plus one hydrophobic and, optionally, one hydrophilic silica. We would also need information on thermally treating such a curable composition.
2	Dipetidyl peptidase-IV inhibitors	pharmaceuticals	We are a new pharmaceutical company that is interested in entering the area of Dipetidyl peptidase-IV inhibitors. This is a relatively new therapeutic area for the treatment of type-2 diabetes but we know that there are compounds already generated by several pharmaceutical companies (including a marketed drug from Merck called Januvia) for this indication. We are interested in discovering the compounds that have been identified so far for inhibiting this enzyme and which companies they are associated with. It would also be useful to determine if there is more than one chemical class of compounds that is used to inhibit this enzyme or if several classes have been identified.
3	methods for controlling molecular weight of polyhydroxyalkanoate	organic, high molecular weight	We are a group of researchers in an university/company. We want to start a project and, before starting, we want more informations about polyhydroxyalkanoate and about methods for controlling molecular weight of polyhydroxyalkanoate constituted of units containing residue of phenyl-, thienyl-, or cyclohexyl-structure in side chain of molecule. Therefore, we are doing an information search to have a better understanding of this area. We are particularly interested in finding methods for controlling the molecular weight of a polyhydroxyalkanoate containing at least one of a 3-hydroxy-?-substituted alkanoic acid unit (1) and a 3-hydroxy-ω-cyclohexylalkanoic acid unit (2) wherein a microorganism is cultivated, in the presence of a hydroxyl group-containing compound, which is capable of producing the polyhydroxyalkanoate containing at least one of the units represented by (1) or (2) from an ω-substituted alkanoic acid (3) or ω-cyclohexylalkanoic acid (4). The 3-hydroxy-ω-substituted alkanoic acid unit (1) should also have a residue having a ring structure of any one selected from the group consisting of a phenyl structure and a thienyl structure. In the presence of plural units, the residues are selected independently for the respective units. The 3-hydroxy-ω-cyclohexylalkanoic acid unit (2) should exhibit a substituent on the cyclohexyl group selected from the group consisting of H atom, CN, NO2, halogen atom, CH3, C2H5, C3H7, CF3, C2F5 and C3F7. In the presence of plural units, the substituents are selected independently for the respective units. The ω-substituted alkanoic acid (3) should also have a residue having a ring structure of any one selected from the group consisting of a phenyl structure and a thienyl structure. In the presence of plural units, the residues are selected independently for the respective units. The ω-cyclohexylalkanoic acid (4) should exhibit a substituent on the cyclohexyl group selected from the group consisting of H atom, CN, NO2, halogen atom, CH3, C2H5, C3H7, CF3, C2F5 and C3F7. In the presence of plural units, the substituents are selected independently for the respective units.
4	Color development baths for Silver Halide	inorganic	We would like to find all documents published after 1997 describing the use of (HADS or hydroxylamine derivatives or di(2-sulphoethyl)hydroxylamine) for color development baths of AgX (Silver Halide) material.
5	(Pregna-4,17-diene-3,16-dione or Guggulsterone or RN:95975-55-6)	organic, high molecular weight	Documents on (Pregna-4,17-diene-3,16-dione or Guggulsterone or RN:95975-55-6) - particularly on preparation
6	Purification of tetrahydrocannabinol	organic, high molecular weight	Please find all documents on purification of tetrahydrocannabinol (the active ingredient of cannibis)
7	Formulations of minitabs Factor Xa inhibitor	pharmaceuticals	Please identify documents with formulations of minitabs, containing a Factor Xa inhibitor
8	Lisuride	pharmaceuticals	Lisuride: transdermal and subcutaneous formulations for Parkinsons and Restless Leg Syndrome
9	formulations of ink-jet inks	formulations	We represent a company that makes ink-jet printing inks for high volume applications. We have started working with hyperbranched polyesteramides and in order to create our own formulations we need to know what other substances are typically combined with these in ink solutions. We would like to know if there are any patent or literature references which describe the use of hyperbranched polyesteramides in ink formulations.
10	catalysts for simultaneous hydrotreating and hydrodewaxing of hydrocarbons	inorganic	We are a group of researchers in an university/company. We want to start a project and, before starting, we want more informations about catalysts for the simultaneous hydrotreating and hydrodewaxing of hydrocarbons.Therefore, we are doing an information search to have a better understanding of this area. We are particularly interested in finding a catalyst for dewaxing a hydrocarbon feedstock containing waxy paraffins under conditions of elevated temperature and pressure. This catalyst comprises a Group VIB metal component on a support comprising a porous refractory oxide in intimate admixture with an essentially Group IIIA metal-free crystalline silica molecular sieve having channels with apertures defined by ten membered rings of oxygen atoms. This silica molecular sieve is a silicalite, having a mean refractive index of 1.39.+-.0.01 and a specific gravity between about 1.65 and 1.80 at 25.degree. C.
11	thermally processable imaging element	organic, low molecular weight	We are interested in finding a thermally processable imaging element, said element comprising a support, a thermographic or photothermographic imaging layer, and an overcoat layer comprising: (A) 50 to 90% by weight of poly(silicic acid) and (B) 10 to 50% by weight of a mixture of: (i) a water-soluble hydroxyl-containing polymer; and (ii) a water-soluble polyvinyl acetal wherein the acetal-group is substutiuted by alkyl - groups only.
12	Diazepam or RN: 439-14-5	Pharmaceuticals	Please identify all documents on diazepam being used as a muscle relaxant (there are other names see ChemID plus or PubChem)
13	tetrahydrocannabinol as an anti-tumour agent	Pharmaceuticals	Patents on use of THC as anti-tumor agent (not anti-cancer pain)
14	Inhalations	formulations	Inhalation formulations containing lactose and magnesium stearate as excipients
15	Betaines for peripheral arterial disease	Pharmaceuticals	Cardiovascular uses of betaines, especially peripheral arterial disease.
16	synthetic routes used to perform Diels-Alder reactions on a multi-gram scale	Reaction Conditions	We are a specialty chemical company that do custom synthesis for different clients in the pharmaceuticals and fine chemicals communities. One of our customers has asked us to synthesize a compound using a Diels-Alder cycloaddition reaction on a multi-gram scale. The procedures they used were done on a much smaller scale and utilize expensive and potentially dangerous compounds. We would like to see a survey of the synthetic routes used to perform Diels-Alder reactions on a multi-gram scale. Specifically we are interested in seeing what solvents are used, what starting materials have been tried and if there have been any catalysts or exotic reaction conditions that have been used to carry out this reaction
17	water in oil in oil	emulsions	When searching for different types of emulsion, there exist a set of predefined phrases that identify particular types. While generally fixed, such phrases may display minor differences that make them hard to find. For instance, to help us, find all documents containing the exact phrase water in oil in oil where the phrase may also have the terms separated by other punctuations than 'space' (e.g. "water-in-oil-in-oil") and where the terms may also be represented by their initial (e.g. W/O/O)
18	water in oil	emulsions	When searching for different types of emulsion, there exist a set of predefined phrases that identify particular types. While generally fixed, such phrases may display minor differences that make them hard to find. For instance, to help us, find all documents containing the exact phrase "water in oil" where the phrase may also have the terms separated by other punctuations than 'space' (e.g. "water-in-oil") and where the terms may also be represented by their initial (e.g. W/O)

Table 10: Results for 5 popular measures for the short PA task

Runs	MAP	b-pref	MRR	P_30	Recall_100	NDCG
BiTeM09PAbl_sm	0.0678	0.4982	0.2609	0.0967	0.2173	0.2630
BiTeM09PAcba_sm	0.1683	0.7415	0.5013	0.1990	0.4055	0.4359
BiTeM09PAcbb_sm	0.1345	0.7241	0.4133	0.1793	0.3666	0.4022
BiTeM09PAcit_sm	0.1663	0.7416	0.4910	0.1970	0.4021	0.4325
BiTeM09PAcl_sm	0.0351	0.3336	0.2425	0.0590	0.1414	0.1785
BiTeM09PAqe_sm	0.1688	0.7432	0.5039	0.2010	0.4011	0.4357
CMU09Chmtcdd_sm	0.0894	0.4999	0.3819	0.1277	0.2610	0.2972
CMU09Chmtcd_sm	0.0517	0.3779	0.3002	0.0810	0.1631	0.2090
DUTIRRun1	0.0195	0.0932	0.1060	0.0397	0.0491	0.0683
DUTIRRun2	0.0203	0.0969	0.0924	0.0420	0.0508	0.0695
DUTIRRun3	0.0204	0.0984	0.0932	0.0397	0.0517	0.0702
msoe09TSx1ta	0.0015	0.1040	0.0103	0.0040	0.0127	0.0395
msoe09TSx2ta	0.0015	0.1040	0.0107	0.0040	0.0127	0.0395
msoe09TSx3ta	0.0015	0.1055	0.0109	0.0030	0.0133	0.0399
msoe09TSx4ta	0.0022	0.1635	0.0177	0.0047	0.0172	0.0600
msoe09TSx5ta	0.0019	0.1365	0.0136	0.0037	0.0156	0.0496
purduePA09r1	0.0654	0.4222	0.2468	0.0893	0.2298	0.2398
purduePA09r2	0.0679	0.4236	0.2508	0.0983	0.2385	0.2476
SCAI09PAf1a_sm	0.0058	0.5164	0.0163	0.0040	0.0116	0.1591
SCAI09PAf1b_sm	0.0056	0.5170	0.0153	0.0043	0.0153	0.1587
SCAI09PAf1c_sm	0.0056	0.4822	0.0201	0.0040	0.0136	0.1500
SCAI09PAf1d_sm	0.0027	0.3067	0.0093	0.0017	0.0091	0.0965
SCAI09PAf1e_sm	0.0069	0.5444	0.0239	0.0050	0.0137	0.1686
SCAI09PAf2a_sm	0.0055	0.5165	0.0149	0.0040	0.0144	0.1556
SCAI09PAf2b_sm	0.0053	0.5115	0.0134	0.0030	0.0102	0.1531
SCAI09PAf2c_sm	0.0056	0.4799	0.0212	0.0037	0.0162	0.1468
SCAI09PAf2d_sm	0.0029	0.3168	0.0317	0.0047	0.0101	0.0980
SCAI09PAf2e_sm	0.0062	0.5540	0.0356	0.0027	0.0112	0.1675
SCAI09PAf3a_sm	0.0055	0.4898	0.0193	0.0027	0.0114	0.1497
SCAI09PAf3b_sm	0.0054	0.5013	0.0217	0.0030	0.0115	0.1529
SCAI09PAf3c_sm	0.0049	0.4587	0.0170	0.0037	0.0103	0.1401
SCAI09PAf3d_sm	0.0027	0.3102	0.0236	0.0023	0.0070	0.0944
SCAI09PAf3e_sm	0.0054	0.5273	0.0088	0.0023	0.0127	0.1595
SCAI09PAf4a_sm	0.0043	0.4524	0.0081	0.0010	0.0073	0.1372
SCAI09PAf4b_sm	0.0056	0.5366	0.0098	0.0033	0.0105	0.1628
SCAI09PAf4c_sm	0.0055	0.4967	0.0235	0.0030	0.0140	0.1521
SCAI09PAt1a_sm	0.0075	0.4274	0.0403	0.0037	0.0147	0.1395
SCAI09PAt1b_sm	0.0053	0.4519	0.0181	0.0043	0.0147	0.1420
SCAI09PAt1c_sm	0.0041	0.3935	0.0152	0.0027	0.0066	0.1236
SCAI09PAt1d_sm	0.0028	0.2660	0.0211	0.0033	0.0136	0.0876
SCAI09PAt1e_sm	0.0053	0.4413	0.0189	0.0040	0.0173	0.1402
SCAI09PAt2a_sm	0.0053	0.4577	0.0176	0.0037	0.0117	0.1415
SCAI09PAt2b_sm	0.0050	0.4586	0.0175	0.0037	0.0176	0.1417
SCAI09PAt2c_sm	0.0043	0.4091	0.0130	0.0030	0.0132	0.1256
SCAI09PAt2d_sm	0.0021	0.2763	0.0085	0.0023	0.0072	0.0853
SCAI09PAt2e_sm	0.0052	0.4877	0.0126	0.0040	0.0124	0.1496
SCAI09PAt3a_sm	0.0042	0.4302	0.0119	0.0017	0.0115	0.1315
SCAI09PAt3b_sm	0.0049	0.4461	0.0272	0.0033	0.0152	0.1385
SCAI09PAt3c_sm	0.0040	0.3947	0.0114	0.0013	0.0114	0.1215
SCAI09PAt3d_sm	0.0026	0.2728	0.0145	0.0040	0.0119	0.0857
SCAI09PAt3e_sm	0.0048	0.4649	0.0145	0.0027	0.0151	0.1433
SCAI09PAt4a_sm	0.0044	0.4048	0.0280	0.0023	0.0074	0.1255
SCAI09PAt4b_sm	0.0049	0.4752	0.0108	0.0020	0.0114	0.1458
SCAI09PAt4c_sm	0.0048	0.4339	0.0265	0.0037	0.0131	0.1346
UIowaS09PA1_sm	0.0485	0.4207	0.2373	0.0817	0.1888	0.2245
UIowaS09PA2	0.0049	0.1457	0.0454	0.0127	0.0368	0.0616
UIowaS09PA3	0.0066	0.1092	0.0490	0.0133	0.0447	0.0542
york09caPA01_sm	0.0180	0.1522	0.1120	0.0373	0.0552	0.0883
york09caPA03_sm	0.0160	0.1574	0.0976	0.0307	0.0480	0.0837

Table 11: Results for 5 popular measures for the full PA task

	MAP	bpref	MRR	P_30	recall_100	ndcg
BiTeM09PAbl	0.0672	0.4180	0.3413	0.1271	0.1818	0.2643
BiTeM09PAcba	0.1835	0.6542	0.5328	0.2567	0.3375	0.4192
BiTeM09PAcbb	0.1581	0.6502	0.5515	0.2348	0.3034	0.4149
BiTeM09PAcit	0.1798	0.6538	0.5226	0.2516	0.3294	0.4148
BiTeM09PAcl	0.0490	0.3018	0.4080	0.0972	0.1375	0.2089
BiTeM09PAqe	0.1823	0.6602	0.5236	0.2547	0.3335	0.4192
CMU09Chmtcdd	0.0975	0.4570	0.5129	0.1776	0.2326	0.3091
CMU09Chmtcd	0.0647	0.3605	0.4004	0.1261	0.1678	0.2344
SCAI09PAf1a	0.0059	0.3777	0.0256	0.0047	0.0109	0.1386
SCAI09PAf1b	0.0057	0.3894	0.0228	0.0047	0.0111	0.1414
SCAI09PAf1c	0.0053	0.3501	0.0246	0.0043	0.0110	0.1285
SCAI09PAf1d	0.0027	0.2137	0.0172	0.0029	0.0075	0.0797
SCAI09PAf1e	0.0065	0.4004	0.0247	0.0052	0.0124	0.1471
SCAI09PAf2a	0.0046	0.3418	0.0188	0.0041	0.0102	0.1232
SCAI09PAf2b	0.0045	0.3344	0.0206	0.0041	0.0092	0.1207
SCAI09PAf2c	0.0051	0.3500	0.0223	0.0041	0.0103	0.1266
SCAI09PAf2d	0.0021	0.1990	0.0158	0.0028	0.0065	0.0734
SCAI09PAf2e	0.0059	0.3925	0.0260	0.0046	0.0114	0.1424
SCAI09PAf3a	0.0047	0.3380	0.0234	0.0040	0.0100	0.1227
SCAI09PAf3b	0.0050	0.3536	0.0239	0.0041	0.0097	0.1288
SCAI09PAf3c	0.0048	0.3367	0.0222	0.0043	0.0097	0.1222
SCAI09PAf3d	0.0022	0.2035	0.0157	0.0026	0.0063	0.0756
SCAI09PAf3e	0.0056	0.3811	0.0227	0.0049	0.0122	0.1388
SCAI09PAf4a	0.0044	0.3181	0.0220	0.0038	0.0093	0.1163
SCAI09PAf4b	0.0055	0.3734	0.0229	0.0050	0.0114	0.1361
SCAI09PAf4c	0.0049	0.3485	0.0228	0.0046	0.0104	0.1271
SCAI09PAt1a	0.0058	0.3601	0.0261	0.0051	0.0117	0.1359
SCAI09PAt1b	0.0058	0.3826	0.0252	0.0053	0.0127	0.1430
SCAI09PAt1c	0.0048	0.3336	0.0215	0.0048	0.0101	0.1255
SCAI09PAt1d	0.0028	0.2138	0.0191	0.0032	0.0076	0.0825
SCAI09PAt1e	0.0060	0.3777	0.0265	0.0044	0.0113	0.1417
SCAI09PAt2a	0.0046	0.3355	0.0195	0.0034	0.0086	0.1238
SCAI09PAt2b	0.0045	0.3314	0.0196	0.0038	0.0101	0.1225
SCAI09PAt2c	0.0049	0.3405	0.0233	0.0041	0.0096	0.1262
SCAI09PAt2d	0.0021	0.2048	0.0118	0.0024	0.0059	0.0772
SCAI09PAt2e	0.0056	0.3775	0.0212	0.0049	0.0109	0.1399
SCAI09PAt3a	0.0047	0.3369	0.0207	0.0039	0.0101	0.1252
SCAI09PAt3b	0.0050	0.3514	0.0227	0.0048	0.0108	0.1311
SCAI09PAt3c	0.0046	0.3290	0.0172	0.0038	0.0086	0.1222
SCAI09PAt3d	0.0024	0.2105	0.0177	0.0028	0.0067	0.0803
SCAI09PAt3e	0.0055	0.3726	0.0245	0.0051	0.0118	0.1389
SCAI09PAt4a	0.0044	0.3166	0.0251	0.0037	0.0094	0.1186
SCAI09PAt4b	0.0054	0.3666	0.0260	0.0045	0.0110	0.1367
SCAI09PAt4c	0.0049	0.3440	0.0254	0.0045	0.0107	0.1284
UIowaS09PA1	0.0683	0.4066	0.3864	0.1329	0.1851	0.2643
york09caPA01	0.0566	0.3385	0.3361	0.1109	0.1431	0.2262
york09caPA03	0.0343	0.1978	0.2337	0.0748	0.0855	0.1376

Table 12: Results for topics that come only from EP patents

	MAP	bpref	MRR	P_30	recall_100	ndcg
BiTeM09PAbl	0.0734	0.5402	0.2462	0.0913	0.2408	0.2729
BiTeM09PAcba	0.1404	0.7655	0.4226	0.1494	0.4075	0.4165
BiTeM09PAcbb	0.1188	0.7451	0.3649	0.1437	0.3711	0.3863
BiTeM09PAcit	0.1381	0.7658	0.4103	0.1455	0.4050	0.4134
BiTeM09PAcl	0.0336	0.3440	0.2078	0.0498	0.1444	0.1702
BiTeM09PAqe	0.1396	0.7668	0.4280	0.1468	0.4043	0.4170
CMU09Chmtcdd	0.0845	0.4894	0.3357	0.1030	0.2612	0.2798
CMU09Chmtcd	0.0426	0.3535	0.2510	0.0515	0.1521	0.1827
SCAI09PAf1a	0.0046	0.5204	0.0077	0.0017	0.0115	0.1465
SCAI09PAf1b	0.0047	0.5354	0.0118	0.0026	0.0156	0.1511
SCAI09PAf1c	0.0049	0.4996	0.0164	0.0026	0.0137	0.1435
SCAI09PAf1d	0.0026	0.3247	0.0063	0.0009	0.0096	0.0958
SCAI09PAf1e	0.0062	0.5597	0.0254	0.0048	0.0143	0.1602
SCAI09PAf2a	0.0051	0.5569	0.0108	0.0022	0.0141	0.1560
SCAI09PAf2b	0.0050	0.5552	0.0083	0.0017	0.0103	0.1548
SCAI09PAf2c	0.0055	0.5211	0.0217	0.0022	0.0172	0.1491
SCAI09PAf2d	0.0027	0.3464	0.0164	0.0039	0.0107	0.0991
SCAI09PAf2e	0.0055	0.5925	0.0211	0.0017	0.0108	0.1658
SCAI09PAf3a	0.0052	0.5256	0.0173	0.0017	0.0105	0.1497
SCAI09PAf3b	0.0047	0.5356	0.0101	0.0022	0.0106	0.1510
SCAI09PAf3c	0.0048	0.4955	0.0161	0.0022	0.0111	0.1421
SCAI09PAf3d	0.0029	0.3426	0.0279	0.0026	0.0075	0.0980
SCAI09PAf3e	0.0051	0.5634	0.0081	0.0017	0.0136	0.1589
SCAI09PAf4a	0.0041	0.4896	0.0069	0.0009	0.0067	0.1383
SCAI09PAf4b	0.0052	0.5712	0.0079	0.0022	0.0107	0.1609
SCAI09PAf4c	0.0054	0.5344	0.0221	0.0017	0.0140	0.1526
SCAI09PAt1a	0.0068	0.4212	0.0307	0.0026	0.0158	0.1266
SCAI09PAt1b	0.0045	0.4661	0.0132	0.0022	0.0130	0.1345
SCAI09PAt1c	0.0034	0.4004	0.0081	0.0013	0.0052	0.1148
SCAI09PAt1d	0.0027	0.2799	0.0095	0.0030	0.0144	0.0862
SCAI09PAt1e	0.0042	0.4432	0.0159	0.0030	0.0186	0.1293
SCAI09PAt2a	0.0052	0.4929	0.0183	0.0039	0.0138	0.1427
SCAI09PAt2b	0.0048	0.4948	0.0124	0.0026	0.0195	0.1424
SCAI09PAt2c	0.0040	0.4370	0.0128	0.0030	0.0152	0.1252
SCAI09PAt2d	0.0022	0.3046	0.0073	0.0017	0.0079	0.0882
SCAI09PAt2e	0.0048	0.5190	0.0114	0.0030	0.0127	0.1479
SCAI09PAt3a	0.0039	0.4575	0.0083	0.0009	0.0119	0.1298
SCAI09PAt3b	0.0043	0.4727	0.0129	0.0035	0.0163	0.1354
SCAI09PAt3c	0.0037	0.4192	0.0099	0.0013	0.0117	0.1201
SCAI09PAt3d	0.0028	0.2999	0.0148	0.0039	0.0136	0.0884
SCAI09PAt3e	0.0043	0.4907	0.0083	0.0017	0.0152	0.1399
SCAI09PAt4a	0.0039	0.4351	0.0192	0.0022	0.0051	0.1245
SCAI09PAt4b	0.0044	0.5010	0.0092	0.0013	0.0113	0.1425
SCAI09PAt4c	0.0043	0.4621	0.0164	0.0030	0.0145	0.1327
UIowaS09PA1	0.0468	0.4287	0.2078	0.0649	0.1960	0.2133
york09caPA01	0.0042	0.0632	0.0415	0.0108	0.0216	0.0289
york09caPA03	0.0082	0.1033	0.0336	0.0130	0.0287	0.0435

Table 13: Results for topics that come only from US patents

	MAP	bpref	MRR	P_30	recall_100	ndcg
BiTeM09PAbl	0.0667	0.4078	0.3492	0.1301	0.1769	0.2635
BiTeM09PAcba	0.1871	0.6450	0.5420	0.2656	0.3316	0.4194
BiTeM09PAcbb	0.1614	0.6423	0.5671	0.2424	0.2978	0.4173
BiTeM09PAcit	0.1833	0.6445	0.5319	0.2604	0.3231	0.4149
BiTeM09PAcl	0.0502	0.2982	0.4247	0.1011	0.1369	0.2121
BiTeM09PAqe	0.1859	0.6513	0.5316	0.2637	0.3276	0.4194
CMU09Chmtcdd	0.0986	0.4543	0.5277	0.1838	0.2303	0.3116
CMU09Chmtcd	0.0665	0.3611	0.4129	0.1323	0.1691	0.2387
SCAI09PAf1a	0.0060	0.3658	0.0271	0.0050	0.0108	0.1379
SCAI09PAf1b	0.0057	0.3772	0.0238	0.0049	0.0107	0.1406
SCAI09PAf1c	0.0053	0.3376	0.0252	0.0044	0.0107	0.1273
SCAI09PAf1d	0.0027	0.2045	0.0181	0.0030	0.0073	0.0783
SCAI09PAf1e	0.0065	0.3871	0.0246	0.0052	0.0122	0.1460
SCAI09PAf2a	0.0046	0.3239	0.0194	0.0043	0.0099	0.1205
SCAI09PAf2b	0.0044	0.3160	0.0216	0.0043	0.0091	0.1178
SCAI09PAf2c	0.0050	0.3357	0.0224	0.0043	0.0097	0.1248
SCAI09PAf2d	0.0020	0.1867	0.0157	0.0027	0.0061	0.0713
SCAI09PAf2e	0.0059	0.3758	0.0264	0.0048	0.0115	0.1405
SCAI09PAf3a	0.0047	0.3224	0.0239	0.0042	0.0100	0.1205
SCAI09PAf3b	0.0051	0.3385	0.0251	0.0043	0.0097	0.1269
SCAI09PAf3c	0.0048	0.3235	0.0227	0.0045	0.0096	0.1205
SCAI09PAf3d	0.0021	0.1919	0.0147	0.0026	0.0062	0.0737
SCAI09PAf3e	0.0057	0.3659	0.0239	0.0052	0.0121	0.1371
SCAI09PAf4a	0.0044	0.3038	0.0233	0.0041	0.0095	0.1145
SCAI09PAf4b	0.0055	0.3569	0.0242	0.0052	0.0114	0.1340
SCAI09PAf4c	0.0049	0.3330	0.0228	0.0048	0.0101	0.1249
SCAI09PAt1a	0.0058	0.3550	0.0257	0.0053	0.0113	0.1367
SCAI09PAt1b	0.0059	0.3756	0.0262	0.0055	0.0127	0.1437
SCAI09PAt1c	0.0050	0.3280	0.0226	0.0051	0.0105	0.1264
SCAI09PAt1d	0.0028	0.2083	0.0199	0.0032	0.0070	0.0821
SCAI09PAt1e	0.0061	0.3722	0.0274	0.0045	0.0107	0.1427
SCAI09PAt2a	0.0045	0.3224	0.0196	0.0034	0.0081	0.1222
SCAI09PAt2b	0.0044	0.3178	0.0203	0.0039	0.0093	0.1208
SCAI09PAt2c	0.0050	0.3324	0.0241	0.0042	0.0092	0.1263
SCAI09PAt2d	0.0021	0.1965	0.0122	0.0025	0.0057	0.0763
SCAI09PAt2e	0.0056	0.3657	0.0220	0.0051	0.0107	0.1393
SCAI09PAt3a	0.0047	0.3268	0.0218	0.0042	0.0099	0.1248
SCAI09PAt3b	0.0051	0.3413	0.0235	0.0049	0.0104	0.1307
SCAI09PAt3c	0.0046	0.3215	0.0178	0.0040	0.0084	0.1224
SCAI09PAt3d	0.0024	0.2030	0.0179	0.0027	0.0061	0.0796
SCAI09PAt3e	0.0056	0.3628	0.0258	0.0054	0.0115	0.1388
SCAI09PAt4a	0.0044	0.3067	0.0256	0.0039	0.0098	0.1182
SCAI09PAt4b	0.0054	0.3554	0.0273	0.0048	0.0109	0.1362
SCAI09PAt4c	0.0049	0.3341	0.0261	0.0047	0.0104	0.1281
UIowaS09PA1	0.0701	0.4047	0.4013	0.1386	0.1842	0.2686
york09caPA01	0.0609	0.3615	0.3607	0.1193	0.1532	0.2426
york09caPA03	0.0365	0.2057	0.2504	0.0800	0.0902	0.1455

Table 14: Results in terms of **xinfAP** for the each TS topic after experts' evaluation

Run	TS-1	TS-10	TS-11	TS-12	TS-13	TS-14	TS-15	TS-16	TS-17	TS-18	TS-2	TS-3	TS-4	TS-5	TS-6	TS-7	TS-8	TS-9
BiTeM09comb	0.016	0.003	0.000	0.155	0.045	0.009	0.010	0.023	0.000	0.000	0.379	0.221	0.108	0.001	0.374	0.132	0.156	0.333
BiTeM09ipc1	0.052	0.007	0.000	0.155	0.044	0.012	0.005	0.022	0.000	0.000	0.359	0.258	0.095	0.001	0.370	0.073	0.162	0.309
BiTeM09ipc3b	0.037	0.006	0.000	0.156	0.048	0.009	0.012	0.022	0.000	0.000	0.376	0.242	0.136	0.001	0.370	0.089	0.152	0.327
BiTeM09ipc3	0.062	0.007	0.000	0.154	0.057	0.009	0.013	0.021	0.000	0.000	0.372	0.256	0.138	0.001	0.370	0.072	0.153	0.306
BiTeM09ipc5	0.062	0.007	0.000	0.152	0.057	0.009	0.013	0.021	0.000	0.000	0.352	0.255	0.137	0.001	0.367	0.064	0.140	0.315
BiTeM09po	0.324	0.360	0.386	0.145	0.054	0.146	0.035	0.023	0.004	0.000	0.314	0.294	0.099	0.008	0.359	0.116	0.235	0.323
BiTeM09qepo	0.019	0.004	0.000	0.156	0.045	0.009	0.009	0.023	0.000	0.000	0.373	0.221	0.115	0.001	0.374	0.132	0.159	0.320
BiTeM09qe	0.005	0.001	0.000	0.115	0.045	0.009	0.012	0.027	0.000	0.000	0.409	0.222	0.073	0.001	0.374	0.042	0.105	0.258
BiTeM09	0.320	0.303	0.283	0.159	0.052	0.026	0.003	0.027	0.003	0.001	0.448	0.280	0.046	0.003	0.350	0.026	0.221	0.276
DUT09TSRun1	0.322	0.388	0.349	0.052	0.034	0.015	0.161	0.115	0.121	0.181	0.532	0.590	0.099	0.009	0.566	0.091	0.176	0.003
DUT09TSRun2	0.257	0.397	0.405	0.394	0.042	0.271	0.094	0.030	0.083	0.244	0.420	0.212	0.174	0.775	0.140	0.303	0.211	0.016
DUT09TSRun3	0.255	0.396	0.418	0.394	0.040	0.254	0.093	0.029	0.077	0.246	0.408	0.212	0.176	0.773	0.130	0.175	0.208	0.014
DUT09TSRun4	0.385	0.278	0.465	0.440	0.000	0.282	0.084	0.003	0.271	0.323	0.527	0.697	0.071	0.201	0.056	0.004	0.238	0.138
DUT09TSRun6	0.281	0.242	0.437	0.349	0.000	0.395	0.233	0.113	0.225	0.231	0.239	0.676	0.123	0.712	0.454	0.032	0.223	0.460
DUTIR09BM25F	0.284	0.238	0.420	0.351	0.108	0.377	0.218	0.044	0.000	0.000	0.223	0.252	0.167	0.630	0.492	0.044	0.250	0.314
msoe09TSx1	0.164	0.014	0.106	0.036	0.000	0.050	0.001	0.025	0.133	0.248	0.156	0.123	0.066	0.009	0.050	0.003	0.018	0.001
msoe09TSx2	0.164	0.018	0.107	0.036	0.000	0.043	0.001	0.025	0.133	0.248	0.149	0.123	0.068	0.010	0.050	0.003	0.019	0.001
msoe09TSx3	0.165	0.013	0.105	0.036	0.000	0.081	0.001	0.025	0.140	0.256	0.131	0.129	0.060	0.027	0.049	0.003	0.019	0.001
msoe09TSx4	0.046	0.036	0.004	0.004	0.000	0.021	0.001	0.013	0.120	0.116	0.088	0.134	0.031	0.008	0.024	0.003	0.029	0.004
msoe09TSx5	0.166	0.013	0.060	0.039	0.003	0.051	0.001	0.025	0.142	0.256	0.135	0.128	0.059	0.007	0.050	0.003	0.019	0.001
purdueTS09r1	0.313	0.401	0.434	0.407	0.054	0.052	0.189	0.001	0.158	0.249	0.518	0.548	0.088	0.843	0.519	0.086	0.268	0.145
SCAI09TSMAN	0.126	0.354	0.077	0.368	0.494	0.304	0.395	0.112	0.170	0.225	0.043	0.697	0.010	0.839	0.354	0.138	0.210	0.040
SCAI09TSNP	0.086	0.273	0.069	0.401	0.063	0.165	0.077	0.153	0.166	0.225	0.007	0.679	0.050	0.839	0.504	0.302	0.328	0.009
SCAI09TSPM	0.252	0.073	0.109	0.221	0.131	0.061	0.103	0.000	0.038	0.120	0.362	0.152	0.096	0.352	0.043	0.258	0.312	0.066
york09ca02	0.106	0.289	0.000	0.179	0.000	0.000	0.013	0.000	0.254	0.286	0.014	0.420	0.022	0.045	0.305	0.001	0.021	0.033
york09ca03	0.377	0.253	0.324	0.444	0.000	0.260	0.057	0.003	0.295	0.341	0.520	0.307	0.085	0.126	0.103	0.014	0.215	0.133
york09ca04	0.318	0.425	0.193	0.134	0.000	0.041	0.034	0.000	0.217	0.236	0.376	0.535	0.054	0.147	0.074	0.014	0.264	0.013
york09ca05	0.089	0.163	0.001	0.104	0.002	0.009	0.034	0.000	0.207	0.283	0.038	0.129	0.050	0.001	0.007	0.002	0.024	0.003
york09ca06	0.073	0.173	0.000	0.072	0.007	0.048	0.006	0.000	0.158	0.230	0.028	0.383	0.041	0.007	0.007	0.003	0.024	0.001
york09ca07	0.316	0.487	0.028	0.485	0.130	0.082	0.109	0.006	0.297	0.321	0.489	0.648	0.055	0.660	0.684	0.000	0.276	0.164
york09ca08	0.256	0.476	0.031	0.352	0.134	0.058	0.102	0.006	0.295	0.354	0.438	0.639	0.048	0.422	0.564	0.002	0.250	0.146
average	0.184	0.197	0.155	0.214	0.055	0.102	0.068	0.030	0.120	0.168	0.297	0.344	0.085	0.240	0.275	0.072	0.164	0.144

Table 15: Results in terms of **infNDCG** for the each TS topic after experts' evaluation

Run	TS-1	TS-10	TS-11	TS-12	TS-13	TS-14	TS-15	TS-16	TS-17	TS-18	TS-2	TS-3	TS-4	TS-5	TS-6	TS-7	TS-8	TS-9
BiTeM09comb	0.113	0.051	0.003	0.411	0.235	0.155	0.179	0.265	0.004	0.007	0.595	0.363	0.548	0.035	0.581	0.340	0.345	0.743
BiTeM09ipc1	0.151	0.064	0.003	0.411	0.231	0.179	0.155	0.259	0.003	0.007	0.585	0.376	0.492	0.035	0.569	0.273	0.349	0.726
BiTeM09ipc3b	0.132	0.059	0.003	0.411	0.240	0.157	0.186	0.261	0.003	0.007	0.597	0.370	0.547	0.035	0.573	0.306	0.347	0.737
BiTeM09ipc3	0.153	0.066	0.003	0.409	0.247	0.159	0.190	0.259	0.003	0.007	0.597	0.374	0.542	0.036	0.569	0.297	0.353	0.720
BiTeM09ipc5	0.153	0.066	0.003	0.412	0.247	0.159	0.190	0.259	0.003	0.007	0.584	0.374	0.540	0.035	0.568	0.283	0.346	0.731
BiTeM09po	0.516	0.513	0.600	0.408	0.247	0.420	0.174	0.264	0.058	0.015	0.546	0.501	0.466	0.195	0.580	0.320	0.515	0.741
BiTeM09qepo	0.116	0.054	0.003	0.412	0.235	0.155	0.173	0.264	0.003	0.007	0.594	0.363	0.550	0.035	0.581	0.340	0.348	0.737
BiTeM09qe	0.067	0.015	0.000	0.354	0.234	0.166	0.199	0.265	0.008	0.011	0.647	0.359	0.463	0.035	0.604	0.199	0.263	0.617
BiTeM09	0.532	0.476	0.523	0.430	0.241	0.203	0.074	0.265	0.052	0.038	0.596	0.446	0.344	0.091	0.598	0.195	0.471	0.639
DUT09TSRun1	0.534	0.589	0.573	0.279	0.343	0.210	0.538	0.426	0.316	0.363	0.761	0.845	0.586	0.141	0.894	0.217	0.555	0.086
DUT09TSRun2	0.483	0.618	0.619	0.752	0.358	0.639	0.409	0.253	0.183	0.449	0.621	0.346	0.623	0.902	0.357	0.403	0.611	0.169
DUT09TSRun3	0.492	0.616	0.625	0.751	0.344	0.608	0.409	0.254	0.174	0.451	0.587	0.345	0.626	0.953	0.322	0.297	0.620	0.156
DUT09TSRun4	0.587	0.494	0.652	0.672	0.000	0.589	0.415	0.028	0.488	0.469	0.731	0.778	0.402	0.567	0.230	0.100	0.523	0.445
DUT09TSRun6	0.473	0.481	0.566	0.628	0.000	0.646	0.643	0.298	0.419	0.432	0.558	0.890	0.527	0.784	0.706	0.118	0.633	0.840
DUTIR09BM25F	0.492	0.475	0.566	0.631	0.550	0.653	0.536	0.180	0.000	0.000	0.498	0.407	0.592	0.859	0.727	0.137	0.622	0.726
msoe09TSx1	0.437	0.145	0.395	0.293	0.041	0.447	0.110	0.316	0.345	0.461	0.551	0.535	0.480	0.211	0.391	0.145	0.275	0.063
msoe09TSx2	0.437	0.153	0.396	0.292	0.041	0.363	0.111	0.316	0.346	0.461	0.553	0.535	0.460	0.214	0.391	0.145	0.276	0.063
msoe09TSx3	0.435	0.141	0.371	0.296	0.042	0.496	0.110	0.316	0.357	0.473	0.528	0.549	0.459	0.249	0.389	0.145	0.276	0.063
msoe09TSx4	0.196	0.163	0.093	0.085	0.000	0.266	0.087	0.222	0.286	0.326	0.421	0.510	0.216	0.193	0.295	0.123	0.259	0.042
msoe09TSx5	0.439	0.141	0.297	0.316	0.035	0.413	0.110	0.316	0.357	0.472	0.540	0.544	0.456	0.204	0.388	0.147	0.277	0.062
purdueTS09r1	0.540	0.634	0.659	0.705	0.314	0.335	0.552	0.014	0.339	0.450	0.738	0.823	0.326	0.952	0.832	0.205	0.697	0.532
SCAI09TSMAN	0.249	0.548	0.216	0.725	0.841	0.458	0.772	0.520	0.255	0.408	0.115	0.873	0.087	0.933	0.672	0.301	0.629	0.278
SCAI09TSNP	0.234	0.533	0.215	0.755	0.512	0.417	0.270	0.391	0.218	0.408	0.038	0.849	0.261	0.933	0.767	0.449	0.735	0.061
SCAI09TSPM	0.515	0.174	0.241	0.439	0.390	0.218	0.314	0.000	0.115	0.331	0.727	0.236	0.475	0.574	0.332	0.417	0.662	0.281
york09ca02	0.321	0.445	0.017	0.329	0.019	0.002	0.143	0.011	0.461	0.471	0.117	0.631	0.169	0.265	0.492	0.032	0.084	0.336
york09ca03	0.574	0.464	0.616	0.676	0.000	0.558	0.360	0.026	0.470	0.499	0.725	0.468	0.350	0.443	0.321	0.090	0.504	0.440
york09ca04	0.532	0.563	0.446	0.354	0.009	0.320	0.300	0.014	0.427	0.406	0.594	0.721	0.299	0.443	0.225	0.090	0.604	0.128
york09ca05	0.327	0.484	0.034	0.393	0.143	0.157	0.326	0.000	0.438	0.486	0.324	0.478	0.271	0.095	0.151	0.125	0.188	0.088
york09ca06	0.260	0.447	0.033	0.159	0.177	0.283	0.170	0.000	0.395	0.438	0.289	0.747	0.253	0.127	0.147	0.148	0.107	0.058
york09ca07	0.489	0.608	0.179	0.700	0.472	0.400	0.414	0.041	0.482	0.519	0.706	0.743	0.334	0.864	0.819	0.057	0.595	0.498
york09ca08	0.443	0.598	0.192	0.553	0.479	0.302	0.415	0.040	0.533	0.557	0.713	0.751	0.315	0.749	0.782	0.118	0.551	0.514
average	0.368	0.351	0.295	0.466	0.234	0.340	0.291	0.205	0.243	0.304	0.541	0.553	0.421	0.393	0.511	0.212	0.439	0.397