Overview of Patent Retrieval Task at NTCIR-5

Atsushi Fujii*, Makoto Iwayama†, Noriko Kando‡

*Graduate School of Library, Information and Media Studies
University of Tsukuba
1-2 Kasuga, Tsukuba, 305-8550, Japan
fujii@slis.tsukuba.ac.jp

†Hitachi, Ltd.
1-280 Higashi-Kougakubo, Kokubunji, 185-8601, Japan
iwayama@crl.hitachi.co.jp
/ Tokyo Institute of Technology

‡ National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, 101-8430, Japan
kando@nii.ac.jp

Abstract

In the Fifth NTCIR Workshop, we organized the Patent Retrieval Task and performed three subtasks; Document Retrieval, Passage Retrieval, and Classification. This paper describes the Document Retrieval Subtask and Passage Retrieval Subtask, both of which were intended for patent-to-patent invalidity search task. We show the evaluation results of the groups participating in those subtasks.

Keywords: *Patent retrieval, Passage retrieval, Classification, NTCIR*

1 Introduction

In the Third NTCIR Workshop (NTCIR-3), the authors of this paper organized the Patent Retrieval Task [5, 6]. This was the first serious effort to produce a test collection for evaluating patent retrieval systems.

The process of patent retrieval differs depending on the purpose of retrieval. In NTCIR-3, the "technology survey" task was performed, in which patents were used as technical publications.

Given a success in NTCIR-3, the authors also performed the Patent Retrieval Task in the Fourth NTCIR Workshop (NTCIR-4) focusing on the "invalidity search" and "patent map generation" subtasks [1, 2, 3].

In NTCIR-4, a number of issues remained open questions. First, in the invalidity search subtask, the number of relevant documents was small and the evaluation result was perhaps less reliable compared with the conventional ad-hoc retrieval tasks. The same problem was identified in the question answering task and has been resolved by increasing the number of questions [8]. Second, although the passage retrieval subtask was planned, the evaluation was not performed due to schedule problems. Third, in the patent map generation subtask, a method for quantitative evaluation was not established.

In view of the above problems, we organized the Patent Retrieval Task at the Fifth NTCIR Workshop (NTCIR-5) and performed the following three subtasks:

- Document Retrieval Subtask

  The invalidity search as in NTCIR-4 was performed, but the numbers of search topics and target documents were increased.

- Passage Retrieval Subtask

  In a document retrieved by a topic for invalidity purposes, the passages were sorted according to the relevance to the topic.

- Classification Subtask

  Classifying patent applications has promise to improve the quality of the patent map generation task. Additionally, the document classification can automatically be evaluated using the patent classification system. In our case, a multi-dimensional classification system called "F-term (File Forming Term)" was used.

After the NTCIR-5 Workshop meeting, the test collections for these subtasks will be available to the public for research purposes[1].

This paper describes the Document Retrieval Subtask and the Passage Retrieval Subtask. The Classification Subtask is described by Iwayama et al. [4].

[1] http://www.slis.tsukuba.ac.jp/~fujii/ntcir5/cfp-en.html

2 Document Retrieval Subtask

2.1 Overview

The purpose of the invalidity search is to find one or more patents that can invalidate the demand in an existing claim. This is a patent-to-patent associative retrieval task. In real world, the invalidity search is usually performed by examiners in a government patent office and searchers of the intellectual property division in private companies.

The Document Retrieval Subtask was performed as follows. First, the task organizers (i.e., the authors of this paper) provided each participating group with a document collection and search topics.

Second, each group submitted retrieval results obtained by the topics. Each group was allowed to use more than one retrieval method and submit multiple retrieval results. In a single retrieval result, up to the top 1000 retrieved documents must be sorted by the relevance score.

Finally, the organizers evaluated the submitted results using relevant documents. The evaluation results were sent to each group, who was also encouraged to analyze the results of their methods and report the obtained knowledge at the workshop meeting.

2.2 Document Sets

In NTCIR-4, the document collection consisted of five years of unexamined Japanese patent applications published in 1993–1997. In NTCIR-5, the document collection consists of ten years of unexamined Japanese patent applications published in 1993–2002. The number of documents in the collection is approximately 3.5 M.

The English patent abstracts, which are human translations of the Japanese Patent Abstracts published in 1993–2002, were also provided to train English-to-Japanese cross-language IR (CLIR) systems. We initially planned a CLIR patent retrieval subtask. However, because search topics were not completed before the formal run, the CLIR subtask was not performed.

2.3 Search Topics

Each search topic is a Japanese patent application rejected by the Japanese Patent Office (JPO). For each topic, one or more citations (i.e., prior art) were identified by examiners of the JPO to invalidate the demand in the topic patent.

To increase the number of topics, we decreased the cost required for producing search topics and relevance judgements as much as possible. We automatically extracted patent applications rejected by the JPO and the citations used for the rejection. For this purpose, we used the citation information in the "*Seirihyoujunka* (Standardized)" Data, which had been produced from the master database in the JPO. We used only the citations as relevant or partially relevant documents and did not perform relevance judgement by human assessors.

An application used as a search topic must satisfy all of the following four criteria.

- the application was not used as a search topic for NTCIR-4.
- the texts of both the application and the corresponding citation are included in the document collection.
- the citation had been published before the application was filed because to invalidate the invention in a topic patent, relevant documents must be prior art.
- the application does not claim the priority; otherwise it is difficult to identify the filing date of the application automatically.

From applications satisfying the above criteria, we selected 1200 applications as search topics. Although the number of topics was determined with no particular reason, we intended to produce more than one thousand topics.

Because candidates of relevant documents for an application can be limited by the filing date of the application, for old applications the retrieval of relevant documents can be a trivial task. Thus, we selected search topics from recently filed applications and produced a pool of topics. This process was repeated until the number of topics in the pool becomes 1200.

However, we tried to avoid selecting applications for the same or similar invention. Applications rejected by the same citation usually claim the same or similar invention. Thus, we did not select applications that were rejected by citations used to reject applications in the pool.

The number of citations per application was also important for the selection of topics. If an application was rejected by a single citation, the decision was made confidently. However, as the number of citations per application increases, the confidence of the decision decreases. As described in Section 2.7, the citation used to reject an application was regarded as relevant, whereas a set of citations used to reject an application was regarded as partially relevant.

Of the 1200 topics, we selected approximately 600 applications rejected by a single citation. However, we did not select an application rejected by more than five citations. Because there is no theory to determine these values, we determined them arbitrary.

The citation information we used did not include the information as to which claim was the target of the rejection. Thus, for each application in the pool

we systematically extracted the first claim, which is usually the target.

Each search topic file includes a number of additional SGML-style tags. The claim used as the target of invalidation is specified by <CLAIM>. The date of filing is specified by <FDATE> and only the patents published before this date can potentially be relevant. Figure 1 shows an example topic claim translated into English.

```
<TOPIC>
<NUM>1048</NUM>
<LANG>EN</LANG>
<FDATE>19950629</FDATE>
<CLAIM>A milk-derived calcium-containing
composition comprising an inorganic salt mainly
composed of calcium obtained by baking a
milk-derived prepared matter containing milk
casein-bonding calcium and/or colloidal calcium.
</CLAIM>
</TOPIC>
```

Figure 1. Fragment of search topic.

During the formal run, we found eleven inappropriate topics. For most of these topics, the automatic method failed to extract the first claim correctly, because the layout of applications is not strictly standardized and can vary depending the applicant. As a result, we used the remaining 1189 topics for evaluation purposes.

2.4 Training Data

For training purposes, the organizers provided participating groups with a collection of search reports produced by professional patent search intermediaries. The source data set was reference data for examiners at the JPO. We used 1000 reports produced in 2001–2003.

Each of the reports includes the ID of a patent application under examination, the ID(s) of one or more candidate documents which can potentially be used to invalidate the target application, and boolean queries used to search for the candidate documents. All IDs were standardized as in our document collection.

Each candidate document is annotated with either "relevant" or "partially relevant". However, because the reports were produced independent of NTCIR, the criteria for relevance judgement is not the same as performed in Section 2.7.

2.5 Baseline Document Retrieval System

To participate in the Document Retrieval Subtask, each participating group was required to develop an entire retrieval system and perform a number of processes, such as query processing and indexing for the 3.5 M patent applications.

To facilitate a partial participation, the organizers developed a baseline document retrieval system and provided participating groups by means of remote access. Thus, a group that developed only query processing was able to participate in the Document Retrieval Subtask. In addition, by sharing the document retrieval system, we can facilitate a glass-box comparative evaluation.

The baseline retrieval system uses Okapi BM25 [7] to compute the relevance score between a query and each document in the collection. In addition, non-textual constraints, such as the International Patent Classification (IPC) codes and publication date, can be used to reduce the number of retrieved documents. For indexing purposes, ChaSen[2] was used to extract words as index terms. Character bigrams were also extracted as index terms. In the formal run, two groups (IFLAB and TRL) used the baseline system.

2.6 Submissions

Each participating group was allowed to submit one or more retrieval results, in which at least one result must be obtained using only the <CLAIM> and <FDATE> fields. For the remaining results, any information in a topic file, such as IPC codes, can be used.

2.7 Evaluation Method

The relevance degree of the citation with respect to a topic was determined based on the following two ranks:

- the citation used to reject an application was regarded as a "relevant document (A)" because the decision was made confidently,
- a citation used to reject an application with another citation was regarded as a "partially relevant document (B)", because each citation is partially related to the claim in the application.

We used Mean Average Precision (MAP), which has commonly been used in past IR literature, to evaluate the submitted runs for the Document Retrieval Subtask.

2.8 Policy for Resource Usage

Because the citations provided by the examiners of the JPO are available to the public, participating groups can obtain relevant documents for the topics before the formal run.

[2] http://chasen.aist-nara.ac.jp/

However, participating groups were not allowed to use the citations corresponding to a topic application for training purposes. This was not applied if their system had already been trained by a large number of citations and it was not easy to remove information, such as statistics, related to specific citations.

Except for the above cases, participating groups were allowed to use any resources for the task according to the "reasonable use policy".

3 Passage Retrieval Subtask

3.1 Task Description

In the Document Retrieval Subtask, we performed the invalidity search task, in which the first claim in a patent application was used to search for similar patent documents. However, because patent documents are long, it is effective to indicate important fragments (i.e., passages) in a relevant document.

The purpose of the Passage Retrieval Subtask is to sort all passages in a relevant document according to the degree to which a passage provides grounds to judge if the document is relevant.

We used the 41 search topics and the 378 relevant documents produced for the dry run and the formal run of the NTCIR-4 Patent Retrieval Task. We call those relevant documents "target documents". The search topics for NTCIR-4 were used to determine criteria as to how the passages in a target document should be sorted.

The passages in each target document were standardized by the official tool provided by the organizers. In Japanese patent applications, paragraphs are identified and annotated with the specific tags by applicants. Because we used these paragraphs as passages, the passage identification process was fully automated.

A high rank should be given to the passages that provide sufficient grounds to judge if a target document is relevant with respect to the search topic. In other words, using a target document as a collection consisting of multiple passages, a search topic was used to search the collection for relevant passages and sort these passages. Passage retrieval results were submitted on a document-by-document basis, instead of on a topic-by-topic basis.

Each participating group was allowed to submit one or more retrieval results, in which at least one result must be obtained using only the <CLAIM> field.

Of the 378 target documents, 356 documents were used for evaluation purposes. For the remaining documents, passages judged as grounds include figures and can not be used to evaluate text retrieval systems. The number of passages per target document was 47. We did not limit the number of passages that can be submitted for a single target document.

3.2 Evaluation Method

Relevant passages were determined based on the following criteria.

- If a single passage can be grounds to judge the target document as relevant or partially relevant, this passage was judged as relevant.
- If a group of passages can be grounds to judge the target document as relevant or partially relevant, this passage group was judged as relevant.

Human assessors exhaustively identified all relevant passages and passage groups. This process was performed during NTCIR-4. We asked 12 members of the Intellectual Property Information Search Committee in the Japan Intellectual Property Association (JIPA) for this task. Each JIPA member belongs to the intellectual property division in the company he or she works for, and they are all experts in patent searching.

A relevant passage group is equally informative as a single relevant passage. We newly introduce the concept of "combinational relevance". This concept provides a salient contrast to the conventional IR evaluation method, in which all relevant items (documents or passages) are independently important and thus combinations of relevant documents are not considered. Thus, MAP cannot be used for the evaluation using the combinational relevance.

We calculate the evaluation score for each run as the rank at which a user obtains sufficient grounds to judge the target document as relevant or partially relevant. To obtain sufficient grounds, a user must read a relevant passage or all passages in a relevant passage group. To calculate the final score, the ranks are averaged over all target documents. In other words, given a list of passages, we calculate the evaluation score as an expected search length at which a user satisfies their information need.

In this paper, we call this score "Combinational Relevance Score (CRS)". However, because this name was determined by the organizers after the formal run, in NTCIR-5 proceedings CRS is called differently depending on the participating group.

4 Evaluation in Document Retrieval Subtask

4.1 General Comparisons

In the formal run of the Document Retrieval Subtask, ten groups participated and the total number of submitted runs was 84. Details of the participating groups are described in Appendix A. Seven groups submitted mandatory runs, each of which was obtained using only the <CLAIM> and <FDATE> fields

in the topics. The remaining three groups used other fields for all of their submissions.

In NTCIR-4, a small number of topics were produced and we used the citations provided by the JPO and the documents judged by human assessors, as relevant or partially relevant documents. In NTCIR-5, a large number of topics were used and only citations were used as relevant or partially relevant documents. In summary, we produced different types of test collections for the invalidity search task.

To compare the results obtained with NTCIR-4 and NTCIR-5 test collections, each participating group was requested to submit the results for the 34 topics, which were used for the formal run in NTCIR-4, and the 1189 topics described in Section 2.3.

Table 1 shows MAP of mandatory runs for different conditions, in which we show only the top result for each participating group. The columns "NTC-X-A" and "NTC-X-B" show the MAP values obtained with the topics for NTCIR-X, which is either of NTCIR-4 or NTCIR-5.

The columns "NTC-4-A" and "NTC-5-A" show the MAP values for which only the relevant documents were used as the correct answers. The columns "NTC-4-B" and "NTC-5-B" show the MAP values for which both the relevant and partially relevant documents were used as the correct answers. For each condition, the MAP values are sorted in descending order. The number of topics used for each condition is shown at the bottom of Table 1.

By comparing the cases of NTC-X-A and NTC-X-B, the relative superiority between groups in MAP did not change whether or not the partially relevant documents were used as the answers. However, by comparing the cases of NTC-4-X and NTC-5-X, the relative superiority between HTC and RDNDC and the relative superiority between ricoh and IFLAB were changed. The ranks of the remaining groups did not changed depending on the condition.

The same tendency was observed in the main task of the NTCIR-4 Patent Retrieval Task [1]. RDNDC achieved the best MAP among the participating groups, specifically when we used the topics for which only the citations provided by the JPO as the answers.

The top four groups in Table 1 participated in the NTCIR-4 Patent Retrieval Task. It is possible that their systems have been well-trained for invalidity search purposes.

Table 2, which uses the same notation as Table 1, shows the MAP values for mandatory and optional runs for the different conditions. Looking at Table 2, one can see that IFLAB achieved the best MAP in NTC-5-X. While IFLAB1 used only the <CLAIM> and <FDATE> fields in the topics, IFLAB5 also used the IPC and retrieved only such documents that were assigned with the same IPC subclass as the topic application. IPC subclasses are the top three levels, such as B27N, G01R, and H01L. Among the participating groups, TRL used both applicant and IPC information for filtering purposes.

It is natural that an application and the citation used to reject the application share IPC codes. It is also often the case that an application and its counterpart citation were filed by the same applicant. We investigated how often these cases happen in our test collection. This investigation is useful to evaluate the effectiveness of methods which do not use application and IPC information.

4.2 Analysis by Topic Categories

In view of the discussion in Section 4.1, we classified the topics for NTCIR-5 based on the applicant and IPC and compared the MAP of submitted runs for each topic category. Tables 3–6 show the results. The column "ALL", which shows the MAP values obtained with all topics, is the same as in the columns NTC-5-A/B of Tables 1 and 2. In Tables 3–6, the best MAP in each column is underlined.

In the column "Applicant", the column "Same" denotes the case where an application and its counterpart citation were filed by the same applicant and the column "Diff" denotes where it is not the case. The same notation is used for the columns "Same" and "Diff" for "IPC". However, because more than one IPC codes can be assigned to a single application, we considered that two documents share the IPC if these documents share at least one subclass. In each application, the applicant and IPC were automatically identified with the INID 71 and 51, respectively.

In the column "IPC section", the columns "A–H" denote the sections of IPC, each of which corresponds to a technology field as in Figure 2. If a topic application is assigned with more than one IPC section, the topic is classified into multiple columns.

In Tables 5 and 6, more than one answer (i.e., relevant or partially relevant documents) for a single topic were allowed to be classified into different conditions. Thus, the total number of topics across the columns can be greater than the number topics in the column ALL.

By comparing the MAP values of Same and Diff in either of Applicant or IPC, one can see that for each run the MAP for Same is significantly greater than the MAP for Diff. This tendency was observed across Tables 3–6. This suggests that to evaluate contributions of methods which do not use applicant and IPC information, the cases of Diff need to be further investigated. It is also suggested that the applicant and IPC information have promise to estimate the difficulty of a topic in the invalidity search.

By comparing different IPC sections, there was no significant differences of MAP in Tables 3–6. However, the number of topics significantly differs depend-

Table 1. MAP for mandatory runs in Document Retrieval Subtask.

NTC-4-A		NTC-4-B		NTC-5-A		NTC-5-B	
Run ID	MAP	Run ID	MAP	Run ID	MAP	Run ID	MAP
HTC10	.3048	HTC10	.2506	RDNDC505	.1949	RDNDC505	.1619
RDNDC501	.2672	RDNDC501	.2369	HTC12	.1944	HTC12	.1573
ricoh3	.2444	ricoh2	.2035	IFLAB1	.1916	IFLAB1	.1539
IFLAB1	.2137	IFLAB1	.1615	ricoh3	.1766	ricoh3	.1447
kle-patent1	.1445	kle-patent1	.1573	kle-patent1	.0786	kle-patent1	.0757
JSPAT2	.1083	JSPAT2	.0772	JSPAT1	.0683	JSPAT1	.0548
TUT-K1	.0989	TUT-K1	.0768	TUT-K1	.0348	TUT-K1	.0283
# of Topics	31	# of Topics	34	# of Topics	619	# of Topic	1189

Table 2. MAP for mandatory and optional runs in Document Retrieval Subtask.

NTC-4-A		NTC-4-B		NTC-5-A		NTC-5-B	
Run ID	MAP	Run ID	MAP	Run ID	MAP	Run ID	MAP
HTC10	.3048	HTC10	.2506	IFLAB5	.2107	IFLAB5	.1684
RDNDC516	.2681	RDNDC501	.2369	RDNDC517	.1999	RDNDC517	.1654
ricoh3	.2444	fj002-02	.2166	HTC12	.1944	HTC12	.1573
fj002-05	.2297	ricoh2	.2035	fj002-07	.1912	fj002-07	.1566
IFLAB1	.2137	IFLAB3	.1823	ricoh3	.1766	ricoh3	.1447
BOLA3	.1655	kle-patent1	.1573	BOLA2	.1675	BOLA2	.1396
kle-patent1	.1445	BOLA3	.1489	TRL1	.0849	kle-patent1	.0757
TUT-K2	.1316	TRL12	.1113	kle-patent1	.0786	TRL1	.0675
JSPAT2	.1083	TUT-K2	.0852	JSPAT1	.0683	JSPAT1	.0548
TRL9	.1060	JSPAT2	.0772	TUT-K2	.0505	TUT-K2	.0413
# of Topics	31	# of Topics	34	# of Topics	619	# of Topic	1189

ing on the IPC section.

In Figure 2, the column "Frequency" shows the frequency of each section assigned to documents in our collection. While the column "Token" shows the frequency of occurrences, in the column "Type" the frequency of each section in a document was counted as one even if the same section appeared in the document more than once. The distributions of IPC sections in Tables 3–6 roughly reflect the distributions in Figure 2. Thus, the selection of search topics is not biased with respect to the IPC section.

5 Evaluation in Passage Retrieval Subtask

In the formal run of the Passage Retrieval Subtask, four groups participated and the total number of submitted runs was 33. All groups submitted mandatory runs, each of which was obtained using only the <CLAIM> field in the topics.

We used CRS (see Section 3.2) to evaluate each run. However, because CRS is a new evaluation measure, it is important to investigate the similarity and difference between CRS and an existing measure.

For this purpose, we used MAP as a comparison. Although as explained in Section 3.2 MAP cannot be used with a group of relevant passages, we decomposed each group into passages and recast the relevant passages as follows:

- relevant (A): a single passage that can be grounds independently
- partially relevant (B): each passage comprising a group that can be grounds

Table 7 shows the CRS and MAP of up to the top two runs for each group, in which "X" denotes the relevance level of target documents with respect to the NTCIR-4 search topics. While in "X=A" only relevant target documents were used for evaluation purposes, in "X=B" partially relevant target documents were also used for evaluation purposes. The number of topics used for each condition is also shown at the bottom of Table 7.

In Table 7, "Y" denotes the relevance level of passages, which should not be confused with the relevance level of target documents. While in "Y=A" only relevant passages were used to calculate MAP,

Table 3. MAP for topic categories: NTC-5-A mandatory runs.

		Applicant		IPC		IPC section							
RunID	ALL	Same	Diff	Same	Diff	A	B	C	D	E	F	G	H
RDNDC505	.1949	.5038	.1318	.2030	.0916	.1869	.1637	.2041	.2995	.2599	.3122	.1283	.1520
HTC12	.1944	.4386	.1445	.2033	.0816	.1928	.1615	.2375	.4118	.2483	.3342	.1123	.1323
IFLAB1	.1916	.4808	.1326	.2020	.0596	.2052	.1416	.2384	.4283	.2528	.3048	.1329	.1348
ricoh3	.1766	.4077	.1290	.1898	.0000	.1818	.1512	.1844	.3361	.2351	.2251	.1087	.1246
kle-patent1	.0786	.1329	.0675	.0785	.0800	.1124	.0589	.1060	.2090	.0622	.0732	.0572	.0445
JSPAT1	.0683	.2020	.0410	.0729	.0093	.0781	.0644	.0382	.1013	.1040	.0906	.0398	.0413
TUT-K1	.0348	.0957	.0224	.0374	.0025	.0541	.0259	.0314	.0772	.0443	.0353	.0155	.0228
# of Topics	619	105	514	574	45	109	126	60	9	39	46	249	185

Table 4. MAP for topic categories: NTC-5-A mandatory and optional runs.

		Applicant		IPC		IPC section							
RunID	ALL	Same	Diff	Same	Diff	A	B	C	D	E	F	G	H
IFLAB5	.2107	.5039	.1507	.2269	.0000	.2070	.1568	.2596	.5296	.2485	.3201	.1473	.1505
RDNDC517	.1999	.4956	.1395	.2128	.0362	.1859	.1665	.2175	.3044	.2528	.3160	.1341	.1527
HTC12	.1944	.4386	.1445	.2033	.0816	.1928	.1615	.2375	.4118	.2483	.3342	.1123	.1323
fj002-07	.1912	.4395	.1405	.2036	.0333	.1992	.1645	.2097	.4382	.1966	.2482	.1348	.1361
ricoh3	.1766	.4077	.1290	.1898	.0000	.1818	.1512	.1844	.3361	.2351	.2251	.1087	.1246
BOLA2	.1675	.4245	.1150	.1728	.0997	.1478	.1489	.1782	.4590	.1922	.2874	.1031	.1205
TRL1	.0849	.2087	.0596	.0888	.0355	.0579	.0806	.0698	.1264	.1659	.1448	.0580	.0607
kle-patent1	.0786	.1329	.0675	.0785	.0800	.1124	.0589	.1060	.2090	.0622	.0732	.0572	.0445
JSPAT1	.0683	.2020	.0410	.0729	.0093	.0781	.0644	.0382	.1013	.1040	.0906	.0398	.0413
TUT-K2	.0505	.1728	.0255	.0534	.0129	.0701	.0423	.0593	.2212	.0952	.0593	.0189	.0294
# of Topics	619	105	514	574	45	109	126	60	9	39	46	249	185

in "Y=B" partially relevant passages were also used to calculate MAP.

"BASE" is a control in which all passages in a target document is sorted according to the passage ID. In other words, a retrieved document is provided to a user as it is. Any passage retrieval method whose result is below that of BASE has little utility.

The best result in each column is underlined. While greater values of MAP are obtained with better methods, smaller values of CRS are obtained with better methods.

Looking at Table 7, one can see that the relative superiority between runs in CRS and MAP significantly differs. While the best CRS was obtained with IFLAB4 in both X=A and X=B, the best MAP in X=A and X=B were obtained with JSPAT1 and HTC5, respectively. However, the ranks of JSPAT1 and HTC5 were low in terms of CRS.

In summary, MAP is not desired to evaluate retrieval methods when the relevance judgement was performed based on the combinational relevance.

6 Conclusion

In the Fifth NTCIR Workshop, we organized the Patent Retrieval Task and performed three subtasks. This paper described the Document Retrieval Subtask and Passage Retrieval Subtask. Both subtasks were intended for the patent-to-patent invalidity search.

During the evaluation in the Document Retrieval Subtask, we identified that the retrieval task was difficult when a topic patent application and the citation used to reject the application do not share the same applicant or IPC codes. For the next step, the number of these types of topics should be increased to evaluate retrieval methods for hard topics.

References

[1] A. Fujii, M. Iwayama, and N. Kando. Overview of patent retrieval task at NTCIR-4. In *Proceedings of the Fourth NTCIR Workshop on Research in Information Access Technologies Information Retrieval, Question Answering and Summarization*, 2004.

Table 5. MAP for topic categories: NTC-5-B mandatory runs.

		Applicant		IPC		IPC section							
RunID	ALL	Same	Diff	Same	Diff	A	B	C	D	E	F	G	H
RDNDC505	.1619	.3891	.1075	.1685	.0814	.1601	.1535	.1984	.2154	.1765	.2000	.1127	.1164
HTC12	.1573	.3545	.1096	.1633	.0664	.1681	.1409	.2068	.2898	.1562	.2151	.1030	.1029
IFLAB1	.1539	.3751	.0997	.1613	.0536	.1634	.1308	.2082	.2978	.1588	.1976	.1098	.1051
ricoh3	.1447	.3190	.1000	.1550	.0000	.1564	.1367	.1624	.2573	.1545	.1508	.1008	.0974
kle-patent1	.0757	.1541	.0579	.0763	.0555	.0918	.0626	.1033	.1464	.0498	.0789	.0628	.0477
JSPAT1	.0548	.1533	.0308	.0587	.0040	.0621	.0639	.0268	.0707	.0631	.0709	.0354	.0357
TUT-K1	.0283	.0735	.0178	.0300	.0102	.0431	.0242	.0207	.0545	.0290	.0363	.0130	.0184
# of Topics	1189	261	1053	1131	127	201	243	110	13	81	97	488	372

Table 6. MAP for topic categories: NTC-5-B mandatory and optional runs.

		Applicant		IPC		IPC section							
RunID	ALL	Same	Diff	Same	Diff	A	B	C	D	E	F	G	H
IFLAB5	.1684	.3878	.1131	.1821	.0000	.1694	.1420	.2240	.3687	.1658	.2048	.1219	.1161
RDNDC517	.1654	.3863	.1123	.1768	.0276	.1588	.1565	.2085	.2207	.1765	.2032	.1162	.1161
HTC12	.1573	.3545	.1096	.1633	.0664	.1681	.1409	.2068	.2898	.1562	.2151	.1030	.1029
fj002-07	.1566	.3484	.1110	.1680	.0150	.1672	.1498	.1935	.3166	.1502	.1719	.1117	.1096
ricoh3	.1447	.3190	.1000	.1550	.0000	.1564	.1367	.1624	.2573	.1545	.1508	.1008	.0974
BOLA2	.1396	.3563	.0926	.1447	.0567	.1340	.1321	.1435	.3300	.1294	.1850	.0983	.0963
kle-patent1	.0757	.1541	.0579	.0763	.0555	.0918	.0626	.1033	.1464	.0498	.0789	.0628	.0477
TRL1	.0675	.1599	.0448	.0697	.0245	.0552	.0645	.0521	.0892	.0979	.0894	.0505	.0446
JSPAT1	.0548	.1533	.0308	.0587	.0040	.0621	.0639	.0268	.0707	.0631	.0709	.0354	.0357
TUT-K2	.0413	.1186	.0230	.0430	.0138	.0556	.0321	.0445	.1558	.0503	.0545	.0228	.0229
# of Topics	1189	261	1053	1131	127	201	243	110	13	81	97	488	372

[2] A. Fujii, M. Iwayama, and N. Kando. The patent retrieval task in the fourth NTCIR workshop. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 560–561, 2004.

[3] A. Fujii, M. Iwayama, and N. Kando. Test collections for patent-to-patent retrieval and patent map generation in NTCIR-4 workshop. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1643–1646, 2004.

[4] M. Iwayama, A. Fujii, and N. Kando. Overview of classification subtask at NTCIR-5 patent retrieval task. In *Proceedings of the Fifth NTCIR Workshop*, 2005.

[5] M. Iwayama, A. Fujii, N. Kando, and Y. Marukawa. Evaluating patent retrieval in the third NTCIR workshop. *Information Processing & Management*, 42(1):207–221, 2006. (To appear).

[6] M. Iwayama, A. Fujii, N. Kando, and A. Takano. Overview of patent retrieval task at NTCIR-3. In *Proceedings of the Third NTCIR Workshop on Research in Information Retrieval, Automatic Text Summarization and Question Answering*, 2003.

[7] S. Robertson and S. Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 232–241, 1994.

[8] E. M. Voorhees and D. M. Tice. Building a question answering test collection. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 200–207, 2000.

A Participating Groups

Table 8 shows the ID, name, and participating subtask(s) of each group, in which “D”, “P”, and “CL” denote Document Retrieval Subtask, Passage Retrieval Subtask, and Classification Subtask, respectively.

	IPC section	Frequency Type	Token
A	Human necessities	380 M	446 M
B	Performing operations; Transporting	915 M	1170 M
C	Chemistry; Metallurgy	480 M	749 M
D	Textiles; Paper	66 M	93 M
E	Fixed constructions	206 M	261 M
F	Mechanical engineering; Lighting; Heating; Weapons; Blasting	394 M	529 M
G	Physics	1 115 M	1480 M
H	Electricity	1009 M	1374 M

Figure 2. IPC sections and their frequencies in NTCIR-5 patent collection.

Table 7. CRS and MAP for mandatory runs in Passage Retrieval Subtask (X: relevance level of target document, Y: relevance level of passage).

X=A				X=B			
		MAP				MAP	
Run ID	CRS	Y=A	Y=B	Run ID	CRS	Y=A	Y=B
IFLAB4	12.34	.4747	.4520	IFLAB4	10.91	.4850	.4614
IFLAB5	13.06	.5072	.4713	IFLAB5	11.23	.4891	.4636
RDNDCP503	13.07	.4713	.4547	JSPAT1	11.67	.4875	.4610
RDNDCP507	13.07	.4684	.4568	HTC1	11.70	.4969	.4744
HTC1	13.24	.5043	.4735	RDNDCP503	12.10	.4320	.4231
JSPAT1	13.25	.5223	.4781	RDNDCP505	12.13	.4396	.4352
HTC2	14.41	.4848	.4562	HTC5	12.14	.5135	.4797
BASE	16.22	.3361	.3451	BASE	16.23	.3700	.3717
# of Target Docs 174				# of Target Docs 356			

Table 8. Groups and their participating subtasks (D: Document Retrieval Subtask, P: Passage Retrieval Subtask, C: Classification Subtask).

Group ID	Group Name	Subtask
BOLA	KAIST	D, C
fj002	IT Media Laboratories, Fujitsu Laboratories	D
FXDM	Fuji Xerox DMS Development	C
HTC	Hitachi	D, P
IFLAB	University of Tsukuba, Ishikawa-Fujii Laboratory	D, P
JSPAT	Research and Development Strategy Department, Justsystem Corporation	D, P, C
KLE	Knowledge and Language Engineering Lab	D
NICT	National Institute of Information and Communications Technology	C
RDNDC	Research and Development Headquarters, NTT DATA Corporation	D, P
RICOH	RICOH	D
TRL	Tokyo Research Laboratory, IBM Research	D
TUT-K	Knowledge Data Engineering Laboratory, Toyohashi University of Technology	D
WGLAB	WebGenie Information Ltd.	C