Report on the Patent Retrieval Task at NTCIR Workshop 3

Makoto Iwayama[†], Atsushi Fujii[††], Noriko Kando[†††], Akihiko Takano[†††]

[†] Tokyo Institute of Technology / Hitachi Ltd.
[††]Institute of Library and Information Science, University of Tsukuba
[†††] National Institute of Informatics

Motivation

While a number of commercial patent retrieval systems and services have long been operated, patent retrieval has not been paid much attention in the information retrieval community. One of the reasons is the lack of test collection targeting patent information. Although TREC test collection includes patent documents, the proportion of those documents is significantly small. Because patent documents are associated with a number of interesting characteristics from a scientific point of view, such as document length, document structures, and classifications, it is important to provide a test collection consisting of patent documents and promote research and development on patent information retrieval.

Task description

The process of patent retrieval differs significantly depending on the purposes of retrieval, such as technology survey and invalidity search. In NTCIR-3, we focused on a specific type of "technology survey," in which end-users of the retrieval task are novice users, for instance, business managers. We assumed a scenario in which a business manager reading a newspaper clips out an article about a specific technology, and the manager asks a professional searcher to find patent documents related to the clipping. The resultant patents help the manager to overview the state-of-the-art of the target technology.

In the task, each participant plays the role of the professional searcher to retrieve relevant patents by the newspaper article. This scenario can be seen as a cross-genre retrieval. The 31 search topics were created by the experts of Japan Intellectual Property Association (JIPA). We provided multi-language versions of the search topics, in Japanese, English, Korean, and traditional/simplified Chinese. The JIPA experts also performed relevant assessment.

In addition to the "technology survey" main task, we organized the optional task, in which participants were allowed to explore any research experiment as far as the official data collections are used. The results of the main and optional tasks were reported at the workshop meeting [1].

These tasks were the first serious effort to explore information retrieval targeting patent documents.

Data collections

The target documents were two years worth of unexamined Japanese patent applications (i.e., full texts) published in 1998 and 1999, containing 697,262 documents (18,139 Mbytes). Those documents were used as a collection from which relevant documents were retrieved. We also provided Japanese/English paired abstracts published in 1995-1999, containing 1,706,154 Japanese abstracts (1.833 Mbytes) and 1,701,339 English abstracts (2.711 Mbytes). Those paired abstracts were primarily intended for training cross-lingual IR systems.

Results

Eight groups submitted 36 runs to the "technology survey" task, and two groups reported their experimental results of the optional task. In summary, we could obtain the following results.

- Cross-genre retrieval was more difficult than the conventional add-hoc retrieval. RICOH investigated the effect of the term weight difference between the newspaper and the patent genres, and their framework called "term distillation" provided a promising result [4].
- Cross-lingual patent retrieval was more difficult than mono-lingual retrieval. Unfortunately only English-to-Japanese cross-lingual runs were submitted, and Korean and Chinese topics were not used. UCB created an English-Japanese dictionary from the paired abstracts distributed at the task, and applied their query translation method to cross-lingual patent retrieval [2].
- Pseudo relevance feedback was effective in specific groups. Surugadai University compared a number of methods of pseudo relevance feedback, including their method using Taylor formula [6].

However, through the task, we could find little about the scientific knowledge inherent in the patent. Thus, after the workshop the organizers conducted further experiments, through which we re-validated past experimental results (e.g., discussions associated with the effectiveness of term frequencies, inverse document frequencies, and document length) in the context of patent retrieval, and we also found that existing retrieval models (i.e., SMART and BM25) were effective in patent retrieval [5].

Challenges in NTCIR-4 Workshop

We are currently performing the "invalidity search" and "patent map generation" tasks in NTCIR-4 Workshop [3] (http://www.slis.tsukuba.ac.jp/~fujii/ntcir4/cfp-en.html).

The purpose of invalidity search is to search a patent collection for the patents that can invalidate the demand in an existing claim. This task can be seen as a patent-to-patent associative retrieval. In real world, this task is usually performed by examiners in a government patent office and searchers of the intellectual property division in private companies. In other words, unlike NTCIR-3 Workshop, we regard patents as legal documents.

In addition to the conventional document-based evaluation, we also explore the passage-based evaluation. Each group is requested to submit the relevant documents for each topic. In addition, for each document, the passage (or the passage group) that can be a ground to judge the document in question as relevant must be submitted.

We also investigate the feasibility of automatic generation of patent maps and its evaluation. The purpose of this task is to generate a patent map driven by a specific theme, such as automobiles and chemical compounds. The organizers provided participants with a set of unexamined patent applications retrieved by a specific topic, and participants are requested to organize those documents in a two-dimensional matrix. The x and y axes can vary depending on the topic, but they are usually "problems to be solved" and "solutions," respectively. This task is a pilot study and thus we do not use quantifiable evaluation measures. Instead, human experts will subjectively evaluate each submitted map.

References

[1] *Proceedings of the Third NTCIR Workshop on Research in Information Retrieval, Automatic Text Summarization and Question Answering*, 2003.
http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings3/.

[2] Aitao Chen and Fredric C. Gey. Experiments on cross-language and patent retrieval at NTCIR-3 workshop. In *Proceedings of the Third NTCIR Workshop on Research in Information Retrieval, Automatic Text Summarization and Question Answering*, 2003.

[3] Atsushi Fujii, Makoto Iwayama, and Noriko 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*, 2004.

[4] Hideo Itoh, Hiroko Mano, and Yasushi Ogawa. Term distillation for cross-DB retrieval. In *Proceedings of the Third NTCIR Workshop on Research in Information Retrieval, Automatic Text Summarization and Question Answering*, 2003.

[5] Makoto Iwayama, Atsushi Fujii, Noriko Kando, and Yuzo Marukawa. An empirical study on retrieval models for different document genres: Patents and newspaper articles. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 251–258, 2003.

[6] Kazuaki Kishida. Experiment on pseudo relevance feedback method using Taylor formula at NTCIR-3 patent retrieval task. In *Proceedings of the Third NTCIR Workshop on Research in Information Retrieval, Automatic Text Summarization and Question Answering*, 2003.