

What Shall We Evaluate? -- Preliminary Discussion for the NTCIR Patent IR Challenge (PIC) Based on the Brainstorming with the Specialized Intermediaries in Patent Searching and Patent Attorneys

Noriko Kando

kando@nii.ac.jp

National Institute of Informatics (NII), Japan

1. Introduction

This paper reports the preliminary results of the brainstorming on "what we shall evaluate on patent retrieval systems" with ten real patent users including specialized intermediaries in patent searching and patent attorneys in the course of planning construction of a patent test collection and an evaluation of patent retrieval systems using it in the future NTCIR Workshop. To set up realistic search requests appropriate for the document types and the real tasks of users is critical to validate the laboratory-typed testing using the test collection. Especially retrieving documents to do with patent or other intellectual property rights has highly specific characteristics in both the nature of the documents themselves and the way of usage. To understand the real life tasks with patent retrieval is extremely important.

2. NTCIR Workshop-- NII Test Collection for IR Systems Workshop

The NTCIR Workshop is a series of evaluation workshops designed to enhance research in retrieval from Japanese or other documents, cross-lingual information retrieval, and related text processing like information processing or text summarization by providing large-scale reusable data set for testing.

The First NTCIR Workshop was held on August 30-September 1, 1999, in Tokyo[1]. It hosted three tasks, i.e., (1) Ad Hoc Information retrieval Task, which is retrieval from mixture of Japanese-English documents, (2) Cross-Lingual Information Retrieval Task, which retrieving English documents using Japanese topics, and (3) Term Extraction Task. Thirty-three groups enrolled and 28 groups submitted results. The participation to the Workshop was limited to the active participants, *i.e.* the members of the research groups that submitted the results of the tasks, advisors and members of the organizing group. Many interesting papers with various approaches were presented at the Workshop and it ended in great enthusiasm. The third day of the Workshop was organized as the NTCIR/IREX Joint Workshop. IREX Workshop, the another evaluation workshop of IR and information extraction (named entity) using Japanese newspaper articles were held consecutively.

Regarding the Second NTCIR Workshop, the process has already started. The Call for Participation has started from June, 2000. It hosts three tasks, i.e. (1) Chinese Information Retrieval Tasks is organized by the research group in the National Taiwan University, including Chinese monolingual retrieval and English-Chinese Cross-Lingual Retrieval using news documents from Taiwan; (2) Japanese-English Information Retrieval Tasks, any combination of retrieval using Japanese and English documents and topics using scientific documents; and (3) Text Summarization Task, or called "Text Summarization Challenge (TSC)"[2], which using various types of newspaper articles.

One of the characteristic aspects of the NTCIR Workshop is to host the IR evaluation as well as NLP evaluation at the same time. We are looking for some promising issues will come out from the collaboration with IR and NLP. Another characteristic aspects is to try to incorporate the situation of users or users point of view to the IR testing using the test collection by collecting realistic topics appropriate to the type of documents and the real users' task with the document type. In this direction, even for the document retrieval, we are keen to "what kind of tasks are done with the document in real life", "what kind of topics are more natural to the documents", and "who can assess the relevance of the type of documents" and so on. These are all traditional questions for the test collection and still challenging issues.

3. Challenges

For the future NTCIR Workshops, many challenging issues have been proposed. Each challenging issue is called "Challenge" in the NTCIR, This year we host "Text Summarization Challenge" which evaluates the automatic summarization of Japanese text. The research of automatic summarization has started from 1950's but how to evaluate the technology is still very challenging issue. From last August, the interested researchers organized "Text Summarization Challenge (TSC)" and have spent lots of time to discuss in detail of "what we shall evaluate" and "how we shall evaluate" through actual meetings and the mailing list.

The examples of the other challenges that we would like to host in the next or near future are the evaluations of Web document retrieval and patent retrieval, which are called "Web IR Challenge (WIC)" and "Patent IR Challenge (PIC)", respectively. For either of them, there are many aspects which are completely different from other rather traditional document retrieval of newspaper documents or scientific documents including the nature of documents and the real users tasks using the document. To identify these differences and to decide which aspects we should evaluate are one of the most important issues to design the evaluation.

3.1 Patent IR Challenge (PIC)

PIC was proposed as the need to research of specialized techniques for the patent documents. Retrieving documents to do with patent or other intellectual property rights is critical in several fields including science and technology, marketing, intellectual property rights management, and business. The importance of the international transfer of patent information has also been widely recognized and the Trilateral Concurrent Search Pilot Project was recently begun by the US Patent and Trademark Office (USPTO), the European Patent Office (EPO), and the Japanese Patent Office (JPO), using a database of one-page English abstracts prepared by the patent office of each country and this occasion encourage the interest toward the cross-lingual access to the patent documents in various languages.

However, patent retrieval has not been so far focused in the area of information retrieval (IR), but may be in the area of database management systems (DBMS). This has come about partly because research and development in information retrieval has tended to place emphasis on generalized systems that are effective for any document or query. As well, the text genre of the patents has highly specific characteristics in both semantics and syntax, and patents are used in highly specialized ways by professional users. Exploration of the genre of patents is vital for more effective patent retrieval.

Future directions for research include systems that can provide an analysis of the retrieved documents to support the usage and strategic analysis of the information in the patent documents, automatic summarization, and comparison across related documents.

IR research has so far placed emphasis on the generalized systems. However, when a focus is placed on a specific genre, various possibilities arise for providing more sophisticated search functionality and supporting users' analysis of the retrieved documents, based on the users' perspectives and characteristic aspects of the text genre.

To survey the characteristic aspects of the patent document, especially for the ones affect to the relevance judgments by the real users and the real users' tasks with patent documents, we organized the groups of professional patent searchers and had a brainstorming with part of the group on "what we shall evaluate" from the users' point of view. The preliminary results are reported in the Section 5.

4. Target Levels of Technology

Other aspect we have to be consider to set up the evaluation than "the document types" and "the real users' tasks using the type of the document" is the target levels of technology to be evaluated. It is critical to make the evaluation fruitful in the IR research community.

It decides, for example, to evaluate (1) the technology which is readily available on the operational systems; (2) the technology which has almost achieved on research systems but there may be some points to be improved; (3) the one that is attractive for users but has not achieved even on research systems and some promising techniques or models usable for the purpose were known; or (4) the one that is attractive for users but is quite challenging for implementation. Probably something between (2) and (4) are appropriate for the purpose.

5. Preliminary Results and Discussion

To survey "what we shall evaluate", we had a brainstorming with real patent users and freely proposed the characteristic aspects of the document type of the patent and real-life usage or tasks with the document of the type. The preliminary results are as followings;

(1) The Stages in the Patent Application Process:

The users' task and purpose of search and the situation of each search are heavily depend on the stage in the whole process of research and development and the process of patent application. According to the purpose and the stage in the patent application process, the factors affect the criteria of the relevance judgments and the most important components of the patent documents to assess the relevance and usability of the retrieved documents varies widely.

(2) Industries

The tasks of retrieving documents to do with patents are completely different, for example, between chemical / pharmaceutical industries and computers / machinery / electric industries.

(3) Terminology

The terms used in the patents are quite unlike the other documents like newspaper or scientific articles, many vague or general terms are often used in order to avoid making narrower the scope of the invention. Combination of general terms may contains a special meaning. To identify these combination are important. At the same time, patent documents included many acronyms and new terminology.

(4) High Recall

High recall, or exhaustive search is strongly requested for the patent retrieval. Especially the need are extremely critical in the Chemical or pharmaceutical industries in the searches done before the research process has started.

(5) Classification Scheme

Classification is important to make the search exhaustive or to sense the related patent application from other companies. But the frequent updates and changes in the classification scheme and codes are problematic. The way overcome there problems are needed.

(6) Support Tools

Support tools like dictionaries, thesaurus of chemicals, proper nouns and so on, are also very useful to do exhaustive searches. To construct and to keep these tools are labor-intensive work

(7) High Precision

At the same time, high precision searches are needed. Especially to discriminate, for example, the following issues is critical in the assessment of the search results;

- The role of each term (agent, object, etc.) or the case of each noun in the search request
- ALL or EXIST: "all the elements in the request are needed" or "one or some of the elements specified in the request are needed."
- the order of the elements (for example, the order of the layer in the some material, or the order in the description of the procedure is critical)

(8) Document Structure

Document structures are important. According to the purpose of search or the stage in the process of patent application, the weights to be given to each component of the patent documents varies.

(9) Images

Images in the patent documents are important to relevance judgments, especially in the electric / machinery / computer industries. If the search using part of the image (for example, a component in an electric circuit) would be effective and useful. Sometimes the most important terms to identify the invention appeared only in the image. So search using the characters appeared in the images is also useful and technically feasible

(10) Task-Oriented Evaluation

Task-oriented evaluation should be interesting. For example, to "Opposition", a system accepts a patent to be oppositioned, then it retrieves every related patents usable to support the opposition.

To make a realistic search request, statement of the purpose of the search or stage in the process of the application is needed. Task-oriented evaluation would be interesting idea. Under the some restricted conditions, search using image components are feasible with current technology but it is not so likely implimented on the operational systems for patent searching.

For the further studies, the opinion or idea or the IR systems researchers are also important. Based on both examination, we would like to design the evaluation of patent retrieval. Any comments, advises, leads, or discussion on this are always welcome.

[1] NTCIR Workshop: <http://www.rd.nacsis.ac.jp/~ntcadm/index-en.html>

[2] Text Summarization Challenge: <http://galaga.jaist.ac.jp:8000/tsc/index-en.html>