Patent-related Tasks at NTCIR

Mihai Lupu, Atsushi Fujii, Douglas W. Oard, Makoto Iwayama and Noriko Kando

Abstract The NII Testbeds and Community for Information access Research (NTCIR) has been the first benchmarking campaign that created a test collection specifically for patent retrieval, in 2001/2002. Over the course of just over a decade, organisers and participants at NTCIR patents-related challenges have addressed the problem of mono- and multi-lingual patent search and automated translation. In doing so, the only available East-Asian language patent test collections have been created and made publicly available for research purposes. This chapter provides a reference summary of the efforts undertaken in NTCIR, helping the reader understand the challenges addressed, the datasets created, and the solutions observed.

1 Introduction

The current NTCIR Conference, an event every 18 months attracting researchers interested in the evaluation of information access technologies from Japan, Asia and the world, started as the NTCIR Workshop in 1999, co-sponsored by the National Center for Science Information Systems (NACSIS), the former organization of the

Mihai Lupu
Vienna University of Technology (TUW), Vienna, Austria, e-mail: lupu@ifs.tuwien.ac.at

Atsushi Fujii
Tokyo Institute of Technology, Tokyo 152-8550 Japan, e-mail: fujii@cs.titech.ac.jp

Douglas W. Oard
University of Maryland, College Park, MD, USA, e-mail: oard@umd.edu

Makoto Iwayama
Hitachi, Ltd., Tokyo 185-8601 Japan, e-mail: makoto.iwayama.nw@hitachi.com

Noriko Kando
National Institute of Informatics, Tokyo 101-8430 Japan, e-mail: kando@nii.ac.jp

National Institute of Informatics (NII) and the Japan Society for the Promotion of Science. Its goals, as stated in the first edition, are:

- to encourage research in information retrieval, cross-lingual information retrieval and related areas by providing a large-scale Japanese test collection and a common evaluation setting that allows cross-system comparison;
- to provide a forum for research groups interested in comparing results and exchanging research ideas and opinions in an informal atmosphere;
- to improve the quality of the test collections based on feedback from participants;
- to investigate methods for constructing a large-scale test collection and corpus including Japanese text and evaluation methods.

Hereafter, the notation NTCIR-X is used to refer to the X-th running of NTCIR workshop.

In NTCIR-1 and NTCIR-2, academic research abstracts and newspaper articles were used to produce test collections. In NTCIR-3, the use of Web pages and patents was introduced. The use of patents in information retrieval research dates back to at least the 1970s, with 76 US patents [3] being used to evaluate the effectiveness of local feedback techniques.

Since then, a number of research papers on the processing of patents have been published, but they have been relatively infrequent in the field of information processing. In spite of its importance to science, engineering, and industry, it was not until the NTCIR initiated a patent retrieval task that patent processing became a focus of interest for the information retrieval (IR) and natural language processing (NLP) research communities. The problem was partially that, unlike Web searching, for which researchers are also users, researchers had difficulty formulating problems and requirements related to the business of real-world patents (see Chapter **??** for details of the patent business).

In pursuing their research interests, researchers are often tempted to propose a fully automated system that does not allow for user involvement. Conversely, in practical situations, a user might wish to adapt the system to a particular working environment. To maintain a reasonable balance between these objectives, the organizers of the patent-related tasks have had occasional round-table conferences with patent attorneys, examiners, and searchers, as well as researchers and engineers. This chapter is devoted to these pioneering efforts of the NTCIR to define new models for the use of patents in academic research fields.

The remainder of this chapter is organized as follows. Section 2 is a brief history of the patent-related tasks at the NTCIR. Section 3 describes the available test collections, while Section 4 provides details of experiments and subsequent observations. Each of these latter two sections is subdivided into four subsections: retrieval, classification, text mining, and machine translation. Finally, Section 5 provides a short summary.

2 History of the Patent-related Tasks at the NTCIR

2.1 Preliminaries

In 2000, the Workshop on Patent Retrieval was colocated with the ACM SIGIR conference on Research and Development in Information Retrieval [30]. The purpose of this workshop was to provide a forum for researchers and practitioners associated with patent retrieval to exchange their knowledge and experiences from different perspectives, which included operational systems, research issues, and evaluation methodologies. The outcome of this workshop motivated researchers involved in the NTCIR to foster research and development in patent retrieval by means of a large, practical test collection.

2.2 NTCIR-3 (2001–2002)

As the first trial for patent retrieval, a technology survey task was performed, in which patents related to a specific technology, such as "blue light-emitting diode", could be searched for [29]. Because patent retrieval was a new research area for the NTCIR community at that time, developing a completely new task was too ambitious. Instead, the target collection was changed to a manageable number of patent documents while maintaining the retrieval task itself. Each search topic was a newspaper clipping related to a specific technology, and the document collection comprised unexamined Japanese patent applications over a two-year period.

2.3 NTCIR-4 (2003–2004)

NTCIR-3 demonstrated the feasibility of using existing IR techniques via its technology survey task. In NTCIR-4, therefore, a patent-specific task was performed, namely invalidity search, in which prior art related to a patent application was searched for [11]. Apart from academic research, invalidity searches are performed by examiners in government patent offices and investigators in the intellectual property divisions of private companies. Each search topic was a claim in a patent application that had been rejected by the Japan Patent Office (JPO). The document collection was extended to unexamined patent applications over five years because, compared with the technology survey task, an invalidity search usually requires a larger number of documents for investigative purposes. For each topic, the citations provided by JPO examiners and prior art patents found by human experts were used as relevant documents. In preparation for NTCIR-4, the organizers arranged a tutorial on patent retrieval by an ex-patent examiner and searcher to guide participants in the new task.

In addition, a patent-map generation task was performed. This called for inter-patent analysis to organize patent documents in specific technology fields. However, because systematic evaluation is inherently difficult, and although human experts subjectively assessed the patent maps generated by automatic methods, a reusable test collection and a systematic evaluation method for patent-map generation have yet to be established.

2.4 NTCIR-5 (2004–2005)

In NTCIR-5, three patent-related tasks were performed [13]. First, as in NTCIR-4, an invalidity search was performed, but using only citations provided by JPO examiners as the relevant documents. The size of the document collection was increased to comprise unexamined Japanese patent applications over 10 years. By this stage, the size of the document collection was no longer problematic for most of the active participants in the NTCIR.

Second, because patent documents are lengthy, it is useful to point out significant fragments ("passages") in a relevant patent. Therefore, passage retrieval was also performed. Each search topic involved a relevant patent for the accompanying invalidity search and the target for each topic was the set of all passages in the topic patent. The relevant passages were those that provided grounds for judging whether the patent was relevant.

Finally, patent classification was also performed [26]. The target documents were patent applications submitted to the JPO over two years, and the correct classification codes were determined according to a multidimensional classification system called "F-term" [45]. The patents, already classified into technological fields, were further classified in terms of one or more viewpoints, such as "purpose", "function", and "effect".

2.5 NTCIR-6 (2006–2007)

In NTCIR-6, both invalidity search [15] and F-term classification [27] were again performed. In the invalidity search, patent documents published over a 10-year period by the JPO and the US Patent and Trademark Office (USPTO) were independently used as target document collections. Having explored patent retrieval issues for seven years, the organizers determined that the patent retrieval task could be concluded. Figure 1 shows a summary of the patent retrieval tasks from NTCIR-3 to NTCIR-6. In Figure 1, only the major datasets are shown. In addition to the 10 years worth of data in the JPO and USPTO collections being a representative dataset for the patent retrieval task, it appeared to have great potential for the exploration of other patent-related research fields.

<table>
<tr><td></td><td>NTCIR-3</td><td>NTCIR-4</td><td>NTCIR-5</td><td>NTCIR-6</td></tr>
<tr><td>Task</td><td>Technology survey</td><td colspan="3">Invalidity search</td></tr>
<tr><td rowspan="2">Documents</td><td colspan="3" rowspan="2">JPO unexamined application</td><td>USPTO grant</td></tr>
<tr><td>2 years</td><td>5 years</td><td colspan="2">10 years</td></tr>
<tr><td rowspan="2">Relevance judgment</td><td colspan="2">By expert searcher</td><td colspan="2" rowspan="2">Cited patent</td></tr>
<tr><td colspan="2">Citation</td></tr>
<tr><td rowspan="3">Related task</td><td colspan="4">Cross-lingual patent retrieval</td></tr>
<tr><td rowspan="2">Patent map generation</td><td>Passage retrieval</td><td rowspan="2"></td></tr>
<tr><td colspan="2">F-term classification</td></tr>
</table>

Fig. 1 Overview of the patent retrieval tasks at NTCIR.

2.6 NTCIR-7 and NTCIR-8 (2007–2010)

In NTCIR-7, the organizers for the patent retrieval task determined to address other issues in patent processing, namely machine translation (MT) and text mining. For each of these tasks, a number of researchers related to the topic were invited to join the organizing team.

For the patent MT task in NTCIR-7 [16], the 10 years of data in the JPO and USPTO patent collections, which had progressively been enhanced from NTCIR-3 to NTCIR-6, were used to produce a Japanese–English (J–E) parallel corpus for training purposes. After extracting patent families, each of which is a set of patent documents for the same inventions usually in more than one language, pairs of sentences in J–E were identified automatically. Whereas NTCIR-7 involved approximately 1.8 million J–E sentence pairs, the number expanded to approximately 3.2 million in NTCIR-8 [17], with an additional five years of JPO and USPTO patent documents. This is one of the largest bilingual sentence-aligned corpora available to the public. In preparation for NTCIR-7, the organizers invited prospective participants to a hands-on MT tutorial aimed at guiding them in the new task.

The patent mining tasks in NTCIR-7 [38] and NTCIR-8 [39] were aimed at summarizing and visualizing patents and research papers in multidimensional technical-trend maps, which resembled the patent-map generation task in NTCIR-4. However, here, the evaluation and analysis were more systematic and thorough. The tasks involved categorizing research abstracts based on the International Patent Classification (IPC) so that they could be associated with patents. The 10 years of data in the JPO and USPTO patent collections were used to train a classifier that could assign one or more IPC codes to a given document.

2.7 NTCIR-9 and NTCIR-10 (2010–2013)

In NTCIR-9 [21] and NTCIR-10 [20], the patent MT task was the only patent-related task. Although the J–E bilingual corpus was the same as in NTCIR-8, patent documents in Chinese were also used for training and testing purposes. These comprised approximately one million English–Chinese sentence pairs. The participation of world-leading research groups made possible exhaustive comparisons of different systems under different conditions. One remarkable finding was that the evaluation results for statistical MT were comparable with or even better than that for commercial rule-based MT systems, under particular conditions.

2.8 Summary

Although the patent-related activity at the NTCIR has ended after 13 years, the large collection of patent documents in Japanese and English has been made available to the public. Currently, the Chinese–English sentence pairs are also available, but for a fee. Figure 2 shows a summary of the MT and text mining tasks from NTCIR-7 to NTCIR-10. As in Figure 1, Figure 2 shows only the major datasets. Details of all the datasets and evaluations are presented in subsequent sections.

The organizers of the patent retrieval tasks also organized the ACL 2003 Workshop on Patent Corpus Processing and edited a special issue on patent processing in Information Processing & Management [14]. All of these activities have contributed to establishing research trends in the IR and NLP communities and increasing the number of publications related to patent processing (including this chapter). A list of publications related to patent information processing is maintained, with occasional updating, at the following URL: http://www.cl.cs.titech.ac.jp/~fujii/pat_proc_pub.html.

<table>
<tr><td></td><td>NTCIR-7</td><td>NTCIR-8</td><td>NTCIR-9</td><td>NTCIR-10</td></tr>
<tr><td>Task</td><td colspan="4">Machine translation</td></tr>
<tr><td rowspan="2">Document</td><td colspan="3" rowspan="2">JPO application & USPTO grant</td><td>Chinese</td></tr>
<tr><td rowspan="2">1M E-C</td></tr>
<tr><td>Sentence pair</td><td>1.8M J-E</td><td colspan="2">3.2M J-E</td></tr>
</table>

<table>
<tr><td></td><td colspan="2">NTCIR-7</td><td>NTCIR-8</td></tr>
<tr><td>Task</td><td colspan="3">Text mining</td></tr>
<tr><td>Purpose</td><td colspan="3">IPC-based classification</td></tr>
<tr><td rowspan="3">Document</td><td colspan="2" rowspan="2">Scientific abstract</td><td>Technical trend map creation</td></tr>
<tr><td></td></tr>
<tr><td colspan="3">JPO application & USPTO grant</td></tr>
</table>

Fig. 2 Overview of the machine translation and text mining tasks at NTCIR.

3 Data Collections and Tools

3.1 Retrieval

The patent retrieval task was executed between NTCIR-3 (2001) and NTCIR-6 (2007), thus spanning four editions of NTCIR.

The first test collection consists of approximately 700 000 full text unexamined (at that time) patent applications from the Japan Patent Office (JPO) as well as approximately 1.7 mil. Japanese patent abstracts and their translations in English. 31 topics have been created for this first evaluation exercise, based on newspaper articles. Each topic was available in Traditional and Simplified Chinese, Korean, Japanese, and English.

The following year, a new collection was released, complementing the existing dataset with more full-text Japanese patent applications as well as more English abstracts, but removing the Japanese version of the abstracts subset [12]. The set of topics was also increased to 101, and selected in a different way. The 2001 topics were taken from newspaper articles, but for 2003 each topic is a claim extracted from Japanese patent applications. For 34 of the 101 (the so-called *main topics*), manual assessments are available. They were created by a total of 12 experts, members of the Japan Intellectual Property Association (JIPA). For the remaining 67 topics (the so-called *additional topics*), only the citations recorded in the search reports from the Japan Patent Office are used.

For NTCIR-5 the collection was further increased to 3.5 mil English abstracts of Japanese patents and the corresponding 3.5 million Japanese full text patent applications, covering 10 years of patent data from the JPO (Japan Patent Office). The full text Japanese subset includes the entire text provided by the JPO except diagrams. In addition to the 34 main topics from NTCIR-4, that year's test collection includes 1 189 new topics, of a similar nature (i.e., a claim from an existing application) and evaluated based on the existing search reports. (i.e., the same as the 67 *additional topics* from the previous year). All topics were originally in Japanese, but manual translations to English are provided by the organisers.

In addition to document retrieval, NTCIR-5 introduced a passage retrieval task, where the participants are asked to identify relevant paragraphs in 356 of the 378 documents relevant to the 34 main topics mentioned above. The 22 documents excluded passages consisting of images or diagrams—not retrievable by text search engines. The task benefits from a clear determination of passages in documents, given by the nature of their format, as well as from the exhaustive relevance evaluation at passage level done in the previous year by the 12 human assessors. The design of the task models an invalidity search [13]. The organizers provided 41 topics, of which 7 were used for training (dry-run) and the remaining 34 for testing.

Finally, NTCIR-6, the last year when a patent retrieval collection was made available, maintained the set from the previous iteration, and added 1.3 million English full text granted patents from the USPTO (United States Patent and Trademark Office). NTCIR-6 adds 1 685 new Japanese topics to the already existing set of 1 243

topics, all consisting of the first claim of a Japanese patent application. Additionally, it adds a completely new set of 3 221 English topics, each consisting of one claim from a USPTO patent.

For all topics in these test collections, NTCIR provides graded relevance judgments, with 3 or 4 levels of relevance.

3.2 Classification

In two of its editions (2005, 2007), NTCIR also organised a patent classification task in parallel to the retrieval task mentioned above. The data collection is of course the same, and only the task definition changes, as well as the topics.

As a Japanese led evaluation effort, NTCIR created patent classification tasks against the classifications used at the Japan Patent Office (JPO), namely the F-terms (File forming terms). The reader may of course be aware that JPO also uses a File Index (FI) classification, which is an extension of the IPC (International Patent Classification). Efforts on classifying against the IPC have been reported as early as 2003 [6] and have been evaluated extensively in the CLEF-IP track discussed in Chapter **??**. Unlike the FI, F-terms are less dependent on the IPC. As indicated on the JPO website, *"F-terms re-classify or further segment each specific technical field of IPC from a variety of viewpoints (i.e., objective, application, structure, material, manufacturing process, processing and operation method, control method, etc.). Combining F-terms with IPC effectively narrows down relevant documents in prior art search."*. The essential difference compared to the IPC is that F-terms, in addition to a 5-digit theme code which could essentially be compared to the IPC subclasses or groups (there are about 1800 theme codes), adds a 4-character term code which is composed of a 2-character *viewpoint symbol* and a 2-digit numerical code. Optionally, an additional 1-character extension code can be added to the F-term. Table 1 shows a small subset of the F-term information for Theme 2H050 (Optical fiber cores).

In NTCIR-5 there are 2 008 patent applications to be classified according to theme and 500 to be classified according to the F-term. The task follows closely the behaviour of patent experts in performing their classification according to the JPO practice: first a theme classification, assigning the application in one of the technology themes defined by the JPO (e.g., 2H050 in Table 1), and then a more refined classification within each theme, indicated by a term code (e.g., AB02, BB22 in Table 1).

NTCIR-6 continued the experiments on classification by asking participants to classify 21 606 patent applications, but only against the F-terms (i.e., provided the themes for each application).

Table 1 F-terms (extract)

Theme		Theme title			
2H050		OPTICAL FIBER CORES			
Viewpoint	Description	Term codes			Additional codes
AB	OPTICAL FIBER STRAND MATERIALS	AB01 .Optical fiber core or cladding materials	AB02 ..Glass core or cladding materials	AB03 ...Quartz core or cladding materials	X: Those for core only Y: Those for clad only Z: Those for core and clad
				AB23 ...Containing Ge	
				AB33 ...Containing fluorides	
BB	COATING MATERIALS FOR THE OPTICAL FIBER CORE	BB01 .Materials for coating optical fiber core	BB02 ..Resin materials	BB03 ...Polyamide (i.e., nylon) materials	Q: Those used as the innermost layer R: Those used for intermediate layers S: Those used as the outermost layer W: Those not limited to specific layer
				BB13 ...Polyester materials	
			BB22 ..Glass materials		

3.3 Text Mining

After the initial classification tasks of NTCIR-5 and -6, NTCIR-7 and -8 expanded the scope from classification alone to classification plus (in NTCIR-8) extraction, and the combination was referred to collectively as the "Patent Mining" task. In contrast to the original classification task in which the objective was to assign theme and F-term codes to patents, the classification task in Patent Mining is to classify a set of abstracts from research papers according to the IPC. The abstracts to be classified are English and Japanese abstracts of papers presented at conferences in Japan between 1988 and 1999. These abstracts had been originally used for retrieval experiments in the first two editions of NTCIR: NTCIR-1 provided about 300 000 abstracts published between 1988 and 1997, of which over 150 000 are in both languages, while NTCIR-2 provided about 530 000 abstracts from 1997-1999 and extended summaries of grant reports for the entire period (1988-1999). In this second collection, about 400 000 of the abstracts and extended summaries are in Japanese, and the rest in English.

NTCIR-7 provides 1 956 topics chosen from among the research articles (divided evenly between English and Japanese). For each of them a set of IPCs is used as ground truth. For 1 050 topics the IPCs are highly relevant, while for the rest they are relevant but not highly relevant. Expertise in patent laws was used to identify potentially correct IPC codes for each topic in an efficient manner. In Japanese patent law, an applicant is not granted a patent for an invention available to the public. However, Article 30 permits a six-month grace period during which an invention will not lose its novelty if the disclosure was made by an inventor through publication for a designated association. Patent applications filed by means of this

exception must indicate the name of the publication and when the invention claimed was disclosed.

From the 10 years' worth of JPO documents, more than 9 000 applications via Article 30 were collected automatically and then associated manually with the corresponding research abstract. The manual verification step was necessary because details of the disclosure contain only the name of the journal or proceedings and the date of publication, but the authors and titles of the paper are not available. In summary, for each topic, the IPC codes assigned with the corresponding research abstract were used as the correct answers, with the average number of correct IPC codes per topic being 2.3.

The same Patent Mining classification task as NTCIR-7 was repeated at NTCIR-8 using the same test collection. In addition, the NTCIR-8 Patent Mining task also added an extraction subtask called Technical Map Creation. This is a very ambitious task, as it requires the participants to extract "Technology", "Effect", "Attribute", and "Value" entities from the plain text of the patents and research articles in Japanese and English. A total of 2000 documents were manually analysed and annotated (500 for each of the four types of documents). Half are provided as training data, 10% for the "dry run", and the remaining 40% for the test itself.

3.4 Machine Translation

Starting from 2008, NTCIR introduced a machine translation benchmark, based on the existing sets of patent data. For NTCIR-7 the training data consists of 1.8 million Japanese-English sentence pairs, while the test set consists of 1 381. The gold standard for intrinsic measurement of translation accuracy is inherent in the paired sentences, and participants could use either language as the source language and the other as the target language. Additionally, a set of 124 search topics from the previous year (i.e., claims from Japanese patent applications) were provided to participants, together with manual translations of those topics into English. Cross-lingual retrieval results obtained with English queries are then be used as a basis for extrinsic evaluation by assessing the effect of different automated translation techniques on the Mean Average Precision (MAP) for retrieval of Japanese patents.

For NTCIR-8 the collection was expanded with patent applications from JPO and granted patents from USPTO up to and including 2007, reaching over 5 million Japanese and over 2 million English documents. Consequently, the training set increased to 3.2 million Japanese-English sentence pairs extracted from the 1993-2005 sub-collections of JPO and USPTO patent documents. The test data is different compared to the previous year: 1 251 Japanese-English and 1 119 English-Japanese aligned sentence pairs are provided to participants, to make sure that the analysis can identify any difference in performance due to the original language of the sentences. As in the previous year, evaluation is done both intrinsically based on the BLUE score, as well as extrinsically using 91 of the NTCIR-6 topics.

Chinese was added to the Machine Translation collection of NTCIR-9, though the training data was initially only available to registered participants. Currently, the Chinese–English sentence pairs are also available, but for a fee. About 1 million Chinese-English sentence pairs were made available for training, and participants had to provide translations for 2 000 Chinese sentences. For the English-Japanese and Japanese-English subtasks, the training data for NTCIR-9 remained the same as in the previous edition, and for each of direction the test data consisted of 2 000 sentences to be translated.

4 Experiments and Observations

Now that we have an overview of all the available patent-related data generated and made available in the context of NTCIR events between 2000 and 2014, we can look at some of the results obtained over the years. In each of the following sections we first review the ground truth creation and then summarize the results. This is intended to put into context those results and to allow the reader a critical perspective on the observations.

4.1 Retrieval

For retrieval experiments, we further need to subdivide the analysis in three categories: monolingual document retrieval, monolingual passage retrieval, and cross-lingual document retrieval.

4.1.1 Monolingual

When discussing the results of the retrieval tasks, it is worth making a distinction between the experiments of NTCIR-3 and those that came afterwards. Such a distinction is necessary because of the different nature of the retrieval tasks. NTCIR-3 considered patent retrieval from the perspective of a technically-savvy user who is not necessarily a patent examiner. To model that, the trigger for the request for information is a newspaper article, and the expected results are related patents that may provide additional information about the item described in the article. For NTCIR-4, -5, and -6, the user modelled is a patent examiner tasked with identifying other patents that may invalidate a specific claim of a given patent application.

News article based task

This particular task and its corresponding topics are quite different from everything else that happened afterwards in evaluation campaigns that focused on patent documents, with the exception perhaps of the TREC Chemical Retrieval track [35] which also had a "Technology Survey" task focusing on the type of information requests specific to a technical user who is not necessarily a patent examiner.

An example topic can be found in the final report of the NTCIR-3 patent retrieval task [29], but it essentially contains a title, the headline and the text of the article that triggered the request for information, a description and narrative in their traditional TREC meaning, as well as a set of concepts pertinent to the topic, and a "supplement" with more information about what should be considered relevant.

The top performing system focused on re-weighting the terms based on their statistics in the different collections (patents vs newspaper articles). The insight is based on the observation that the nature of the texts are significantly different and therefore the weighting of the terms should take into account the frequency of terms in the two collections. The authors called this *"term distillation"* [25], but essentially it is the explicit combination of a weight based on the domain of the query (i.e., articles) and that of the target document (i.e., patents). The issue of terminology has been later revisited and confirmed by Nanba et al.[40], Mahdabi et al. [36], and Andersson et al. [1], so it appears to be a reasonable conclusion to draw from the first benchmark on patent retrieval.

At the other end of the spectrum, the lowest results were obtained by a method based on Random Indexing [44]. This low performance of statistical semantics on patent retrieval had been also reproduced for Latent Semantic Indexing by Moldovan et al. in 2005 [37], by Aono in NTCIR-6 in 2007 [2], and by a more recent revisiting of random indexing [34], albeit each using different collections. This is not to say that statistical semantics do not have a word to say in the problem of patent retrieval, but rather that perhaps their direct application to the problem needs to be more nuanced.

For the rest of the spectrum, the relatively small set of topics and the large variance in the intermediary steps taken by each participant (tokenisers, stemmers, filters of various kinds), makes it risky to draw any conclusions. The organisers of NTCIR-3 observed this and made their own study, keeping everything fixed except the retrieval model [28]. The conclusion they draw is that the methods which are known to perform best on other tasks, also perform best on this particular test case (e.g., BM25, among all the probabilistic models tried in the study). In particular, the methods that perform best are those that control for document length. This is of course reminiscent of the discussions in the early years of TREC [22], when it was observed that different methods would perform significantly differently on sub-collections that differed in their average document length or in their document length distribution.

For the patent data, the length aspect was revisited in more detail by one of the participants, Fujita [18], who reapplied the analysis performed in the early TREC test collections. The study complemented the one done two years before by the task

organisers [28] by also considering language modelling in addition to the different variants of the TF*IDF. After having observed no correlation between relevance and document length in terms of words, Fujita also considered document length in terms of claims—a very patent specific approach—under the assumption that it is actually the number of claims that models the multiple topicality present in longer newspaper articles, but this showed similar results: no correlation with relevance, a tendency of TF*IDF methods to retrieve longer documents, and a tendency of LM methods to retrieve shorter documents. In the end the author concludes that simply using a higher document length penalty in the TF*IDF model (i.e., a higher *b* parameter in BM25) is enough to obtain good performance, but reasonably stops short of claiming that language modelling will not perform better if more efforts are directed towards it.

Patent application claim based tasks

Starting with NTCIR-4, the patent retrieval tasks moved away from the general, technology survey model of information need, towards the specific model of a patent examiner [11]. This is referred to here as *invalidity* search, and corresponds to the *prior art* task organised later in CLEF-IP [42] or TREC-CHEM [35].

Table 2 Ranking and MAP scores of systems across years and test collections according to the "relaxed" relevance criteria. Consecutive runs of the same group with difference less than 10% are omitted

evaluation reported	May 2004		Dec 2005				May 2007			
document set	NTCIR-4		NTCIR-5				NTCIR-6			
topics	NTCIR-4		NTCIR-4		NTCIR-5		NTCIR-4		NTCIR-5		NTCIR-6	
# topics	34		34		1189		34		1189		1685	
1	RDNDC9	.27	HTC10[#]	.25	AFLAB5[+]	.17	HTC10	.26	HTC10	.20	HTC10	.12
2	RDNDC2	.25	RDNDC501	.24	RDNDC517	.17	HTC06	.23	HTC05	.17	HTC06	.10
3	HTC20[*]	.25	fj002-02	.22	HTC12	.16	AFLAB1	.16	AFLAB1	.15	HTC04	.08
4	ricoh[$]3	.22	ricoh2	.20	fj002-07	.16	hcu1	.16	JSPAT3	.09	AFLAB1	.08
5	AFLAB11[+]	.20	AFLAB3[+]	.18	ricoh3	.15	JSPAT3	.12	hcu1	.08	hcu1	.05
6	fj002-10	.19	kle-patent1	.16	BOLA2	.14	JSPAT1	.11	BETA6-1	.06	JSPAT0	.04
7	PLLS6	.17	BOLA3	.15	kle-patent1	.08	BETA6-1	.11			BETA6-1	.04
8	TRL8	.13	TRL12	.11	TRL1	.07						
9	NUT1	.08	TUT-K2	.09	JSPAT1	.05						
10			JSPAT2	.08	TUT-K2	.04						

[*]In the NTCIR-4 proceedings, the system was referred to as 'JAPIO'. We rename it here to make it consistent with NTCIR-5 and -6

[#]HTC10 in NTCIR-5 is different from HTC10 in NTCIR-6

[+]In the NTCIR-4, and -5 proceedings, the system was referred to as 'IFLAB'. We rename it here to make it consistent with its latest version in NTCIR-6

[$]In the NTCIR-4 the system was referred to as 'LAPIN'. We rename it here to make it consistent with its latest version in NTCIR-5

It is particularly instructional to look at participants' systems over the three evaluation campaigns. Among them, the teams from Hitachi (HTC) and the one from the Graduate School of Library at the University of Tsukuba (AFLAB) submitted runs in all three years.

HTC observed that the number of stop words did not have a significant effect in 2004, and therefore the reduced that number significantly (from approximately 3000 to only 30) in later years. Their experiments also show that using only the claim as input to the search system is not recommended because it does not contain sufficient information. Not only are the claims in general rather information sparse, but the use of only the first claim may, in hindsight, also be problematic.

The best method from the HTC group was the one that used all of their filters: stop words, special weights for measurement terms, TF calculated based on the entire query document, addition of terms from abstract and the entire document to the query, co-occurrence based term weighting, and, finally, filtering or score adjustment using theme codes. These observations are consistent across the different query sets.

The group at NTT DATA (RDNDC runs in Table 2) also obtained consistently good results in the two years it participated in the track. The characteristic feature of their system was query expansion with keywords from the *"detailed description of the invention"*. They show that this provides more useful keywords compared to a standard query term expansion based on Local Context Analysis (LCA) [49]. In both NTCIR-4 and NTCIR-5, the team put an impressive amount of effort into manual morpho-syntactic rules to both extract the components of the invention from the claim (241 patterns) and to identify the sentences in the detailed description of the invention that correspond to the previously identified components of the invention (104 patterns). While the first appear to be reasonably feasible due to the nature of the genre in the patent claims, the second are, as the authors point out, a "challenging problem". They therefore provide an alternative which removes the general rules and replaces them with a greedy approach: find those sentences which contain the most terms of the component of the invention, in the same order.

RDNDC runs also use IPC information. In NTCIR-4 this was used to re-weight terms based on their frequency in different IPC classes, while in NTCIR-5 the RDNDC team used the IPC information as a basis for reranking: after their ranking systems provided initial results, the retrieved patents having at least one IPC class in common with the query patent application received a multiplicative boost. According to their experiments, this added 2.5–5% to Mean Average Precision for all their runs.

The team at the University of Tsukuba (runs denoted by IFLAB or AFLAB in NTCIR reports) also used a module to split the claim into its constituent components as a first step. In 2004 they compared a simple punctuation-based method with a more complex set of morpho-syntactic patterns based on rhetorical structure theory initially introduced by Shinmori and colleagues [46, 47] the year before in NTCIR-3. They found that the simpler method worked just as well, thanks to the regularity of the rules of proper patent claim editing, and in subsequent years they continued only with this simpler method.

Perhaps the most interesting thing to observe in the experiments performed at the University of Tsukuba is that the effect of the IPC differed across the years. In NTCIR-4 the use of the IPC (as a hard filter) was apparently detrimental to the precision of the results (5.9% and 3.5% reduction in the rigid "highly relevant" and the relaxed "relevant" evaluation, respectively), in NTCIR-5 the same approach to the use of IPC codes showed an apparent improvement (8% and 7.5% on the rigid and relaxed evaluation, respectively). This difference might be explained by the nature of the ground truth in the two years: in NTCIR-4 it was (partially) manually created, while in NTCIR-5 it was completely automatic (based on citations). If we further imagine that search patterns at a patent office often rely on metadata (IPC or related classifications), we could reasonably hypothesise that there is a bias towards patents in the same class in the ground truth. This is not a problem of NTCIR (nor is it certain to be a problem at all), but rather an issue that has to be considered in all evaluation campaigns using the citations.

Finally another system that was consistently among the top performers in terms of MAP was the one created at RICOH Ltd [24]. In their first year of participation they considered whether it is sufficient to index only the abstract and claims of the patent collection. Their experiments showed that in fact the information present in the entire patent is needed for better relevance estimation. This confirms the findings of the other systems presented above, and complements them because if the others had considered this additional information on the query side, RICOH experiments consider it on the target document side.

In the following year, RICOH experiments also confirmed that the use of IPC codes as filters, either on the query side (as usual filters on the retrieved documents), or on the target documents side (as a form of pseudo-relevance feedback), improves the precision of the results. Echoing the University of Tsukuba results, they observe that this improvement is clearly visible for NTCIR-5, but arguable for NTCIR-4. Additionally, RICOH conducted experiments with the use of synonyms for query expansion. Synonyms were generated based on an English-Japanese dictionary (by collecting all terms which appeared in the definition of English terms containing one of the query terms to collect "term siblings" from the dictionary). This yielded only marginal improvements, in marked contrast to all of the methods discussed above, which had also considered query expansion and had observed more marked improvements. The difference here is probably in the fact that the others had selected query expansion terms on a query-by-query basis (or just re-weighted them), while in this case the synonym set was created a priori and used consistently for all queries. It is easy to imagine why this might be less effective: if we were to take an example in English, consider the synonyms on the term "bank": "depository", "exchequer", or "beach", "shore", or "chair", "seat".

All the experiments mentioned so far were Japanese monolingual. While topics were always available in Japanese and English, and some documents existed in English as well (i.e., for the PAJ subcollection), the focus of NTCIR-4 and -5 had been on Japanese monolingual retrieval. In NTCIR-6 the organisers introduced a separate task for English monolingual patent retrieval, with its own set of topics and its own target document collection (from the USPTO). Five teams participated in

this English retrieval sub-task at NTCIR-6. Table 3 shows the top results for each participant.

Table 3 Best performing results for each of the participants in the NTCIR-6 English Retrieval Sub-task

Run ID	strict	Run ID	relaxed
AFLAB2	0.04	AFLAB2	0.08
hcu1	0.03	NTNU	0.07
KLE1	0.03	KLE1	0.07
NTNU	0.02	JSPAT2	0.06
JSPAT0	0.01	hcu1	0.02

The best performing system integrated content and citation information in scoring. Fujii compares no citation information with PageRank and with a domain-specific method and observes the most improvement with the domain-specific method. This is the first system that explicitly uses patent citations in ranking, and the use of this kind of information has been proven beneficial both at CLEF-IP in the system built by Lopez and Romary [33] and in TREC-CHEM in the system built by the group at Geneva University Hospitals [19].

Overall, the results from the English Retrieval sub-task are hard to qualify. The values are certainly lower, but they are a different task, so a direct comparison cannot be made. Based on the participants report, it seems that there was some unfamiliarity with the nature of the USPTO documents. For instance, if, as we have seen, most of the systems had used IPC codes to improve their precision when searching JPO patents, this was no longer as useful because the USPTO, at the time, primarily used a very different classification scheme and only assigned one IPC code to each patent. Another example is the APP-DATE field, which does not necessarily have the same meaning as the FDATE field of the Japanese applications. The experiments done at Pohang University of Science and technology (POSTECH) [31] had shown that the use of the APP-DATE field actually reduced precision, and this, in principle, should never happen if it had the meaning that team had expected it to have.

4.1.2 Passage Retrieval

In NTCIR-3 it was observed that patents are significantly longer documents than newspaper articles, and, apart from the implications of this in the document scoring methods, it was also decided to have a subtask on retrieving passages as opposed to full documents. NTCIR-4 first defined such a sub-task, but it was not evaluated that year, so NTCIR-5 considered it again. This time, participants were given both topics and relevant documents, and the passage retrieval task consisted of retrieving relevant passages from the known relevant documents. Therefore, there were 41 topics (7 for the dry-run and 34 for the formal run of NTCIR-4) and 378 relevant documents. For each of the relevant documents, participants had to rank its paragraphs in order of their expected utility as a basis for judging the relevance of

the document. A new evaluation metric was defined—the Combinational Relevance Score (CRS)—proportional to the rank at which the list of paragraphs contains at least one relevant paragraph (or set of paragraphs, if the evaluators considered a set instead of just one). Table 4 shows the results of the runs, for each document and paragraph relevance category (strict or relaxed). Here, the Mean Average Precision (MAP) is calculated on the ranking of the paragraphs within a target document.

Most participants' runs were essentially the same as for document retrieval, with the difference that instead of indexed documents, they indexed passages as documents. IPC codes were no longer used because they were irrelevant given that the re-ranking was taking place inside a target document known to be relevant to the query. Only HTC substantially changed their indexing scheme and moved from a term-based index to a character n-gram index. The motivation for this was the relatively small amount of text in a passage, and the resulting desire to have a more flexible matching scheme. Unfortunately, a direct comparison with a term-based index was not made, and it is consequently difficult to estimate the benefit of this approach.

Table 4 Passage Retrieval evaluation results in NTCIR-5

Document relevance: strict				Document relevance: relaxed			
174 documents				356 documents			
		MAP, passage relevance:				MAP, passage relevance:	
Run	CRS	Strict	Relaxed	Run	CRS	Strict	Relaxed
IFLAB4	12.34	0.47	0.45	IFLAB4	10.91	0.49	0.46
IFLAB5	13.06	0.51	0.47	IFLAB5	11.23	0.49	0.46
RDNDCP503	13.07	0.47	0.45	JSPAT1	11.67	0.49	0.46
RDNDCP507	13.07	0.47	0.46	HTC1	11.70	0.50	0.47
HTC1	13.24	0.50	0.47	RDNDCP503	12.10	0.43	0.42
JSPAT1	13.25	0.52	0.48	RDNDCP505	12.13	0.44	0.44
HTC2	14.41	0.48	0.46	HTC5	12.14	0.51	0.48
BASE	16.32	0.34	0.35	BASE	16.23	0.37	0.37

4.1.3 Cross-lingual

NTCIR had organised cross-lingual evaluation tracks before, and continued to organise one in parallel to the patent retrieval track [32]. The organisers of the early Patent Retrieval tracks had encouraged participants to use the multilingual collections that were provided in their experiments. Some did used those collections to enhance system performance on a a monolingual retrieval task [9, 10], and others did provide a few cross-lingual runs for the sole purpose of exploring and evaluating cross-lingual systems. Nevertheless, the number of cross-lingual runs was considerably smaller than that of monolingual runs: NTCIR-3 had some cross-lingual runs (3 of 8 participants submitted such runs), NTCIR-4 had only one cross-lingual run of the 111 runs submitted, and NTCIR-5 had none.

In NTCIR-3 IFLAB [8] created a query translation engine based on both a commercial dictionary and language and translation models built on the available corpora. In particular, their translation engine kept the word order of the source language because it had been previously observed [7] that between English and Japanese technical terms use the same word order about 95% of the time.

The groups at the University of California, Berkeley [4] and the Swedish Institute of Computer Science (SICS) [44], while not having Japanese-speaking members, attempted the task of cross-lingual retrieval. Berkeley used external dictionaries (Babelfish) to translate the queries for both English-to-Chinese and English-to-Japanese retrieval. The innovative part was that when the dictionary did not find a translation, the team submitted the query to a Chinese or Japanese search engine and took, from the top 200 documents, the Chinese or Japanese terms surrounding the English terms, weighting them by the distance to the English terms. This amounts to a cross-lingual pseudo-relevance feedback.

The SICS team used random indexing [43] to construct a bilingual thesaurus, which they then used to generate cross-lingual queries. The approach was purely statistical and, in the absence of a manual check on the results of the bilingual thesaurus generation process, the results were significantly poorer than those using existing dictionaries.

In NTCIR-4 RICOH [24] performed English-to-Japanese cross-lingual retrieval. They did query translations and search on a multilingual database. Their officially submitted English-to-Japanese run (LATIN5) obtained a P@10 score comparable to their English-to-English run (0.16 and 0.17, respectively), but P@10 scores on different collections cannot be directly compared and their English-to-Japanese results were significantly lower than those of their Japanese-to-Japanese run (0.20). Given that queries were available in both English and Japanese, they were able to compare the performance of the query translator with the results obtained in retrieval. They showed that these two elements do not necessarily correlate: the query translation (from English to Japanese) closest to the original Japanese query did not obtain the best result in terms of P@10.

4.2 Classification

As mentioned in Section 3.2, the initial patent classification tasks at NTCIR addressed two classification problems: first, a classification of patents against the set of themes (technology areas) present in the F-Terms, and second, a classification against the set of term codes (i.e., viewpoint + 2-digit code) known as F-terms. The first one can be seen as a coarse classification based exclusively on topicality, while the second one a refinement of the first, aiming to identify different aspects within the same technical domain. Theme classification was only evaluated in 2005 at NTCIR-5 (Table 5), in the following NTCIR the focus was exclusively on the more challenging F-term classification (Table 6).

Table 5 Results of Theme classification tasks

NTCIR-5				
Runid	model	MAP	R-Precision	F-measure
BOLA1	K-NN	0.69	0.59	0.27
JSPAT2	Naive Bayes	0.66	0.56	0.53
WGLAB9	K-NN	0.62	0.53	0.07
FXDM3	VSM	0.49	0.39	0.38

From Tables 5 and 6 we can see that simplest vector similarity methods (Vector Space Model, χ^2) are not up to par with typical machine learning methods (K-Nearest Neighbour, Naive Bayes, Support Vector Machine). We should note that what is denoted here by χ^2 (run NUT05 [23] in NTCIR-6) is actually similar to a typical Vector Space Model (VSM) but with a change in the weighting function, reminiscent of information content studies.

Table 6 Results of Term classification tasks

NTCIR-5					NTCIR-6					
			R-	F-			Exact match		Relaxed match	
RunID	model	MAP	Prec.	measure	RunID	model	MAP	F-measure	MAP	F-measure
NICT5	K-NN	0.50	0.46	0.44	NCS02	N. Bayes	0.49	0.40	0.58	0.50
JSPAT1	SVM	0.40	0.39	0.28	GATE03	SVM	0.48	0.41	0.58	0.51
FXDM10	VSM	0.21	0.20	0.16	NICT01	K-NN	0.45	0.38	0.55	0.48
	-	-			JSPAT01	SVM	0.44	0.30	0.54	0.37
	-	-			NUT05	χ^2	0.41	0.24	0.51	0.38
	-	-			RDNDC14	K-NN	0.27	0.24	0.36	0.34
	-	-			baseline		0.28		0.37	

For this study it is particularly of interest to look at K-Nearest Neighbour (K-NN) methods, since they have obtained both very good and very poor results in experiments in both years. As usual, it is not straightforward to compare two systems, even if they use the same method, because there are numerous components or steps that can change. Nevertheless, we can see that the differences between BOLA1 and WGLAB9 in NTCIR-5 are: 1. the information sources from the document (i.e., PAJ, "technological field", "purpose", or "method"); and 2. the similarity function (cosine similarity for a vector space based on BM11 versus a similarity function based on structural similarity between documents). More subtle is the difference observed between RDNDC14 and NICT01 in NTCIR-6. Both systems used K-NN on top of a vector space built on BM25, with terms extracted using the same NLP tool (ChaSen) and from the same parts of the document (abstract and claims). Yet their results are significantly different (a drop of 30-40% in scores). The difference probably lies in the fact that RDNDC14 only used the first claim, as opposed to the entire set of claims used in NICT01, and the latter weighted the score of each F-term by a constant determined using experiments.

4.3 Text Mining

The first attempt to do text mining (in the general sense) had actually been at NTCIR-4 with the Patent Map Generation Task. Its purpose was to generate a patent map driven by a specific theme (e.g., automobiles), in an automatic or semi-automatic way. The desired map is a two-dimensional plot generated by considering pairs of relevant concepts. For one topic this might mean that on one axis different "problems to be solved" are to be placed and on the second axis the "solutions" are expected. For another topic the axes might be "form of product" and "date of publication". The cells were then to indicate patent numbers connecting the two concepts, in the context of that topic. From the outset, this was a difficult task, both for the organisers as well as the participants. It requires a much deeper understanding of the content of the patent than relevance evaluation, and a sufficiently large set of topically relevant documents for each topic.

The organisers selected six topics from NTCIR-3, each having at least 100 relevant documents, and the participants had to define the axes on their own and populate the matrix correspondingly. This required experience in a large number of domains related to information access, which resulted in only two teams participating, each consisting of several institutions.

The task can be treated in two steps: identification of meaningful concepts and population of the cells with patents connecting the two concepts. One team focused on clustering, using, among other methods, Latent Semantic Analysis. The other team focused on claim analysis, using morpho-syntactic patterns. In hindsight we may argue that a combination of the two methods would potentially bring even better results.

The organisers created reference patent maps which were used to guide the assessors in their evaluation, but given the nature of the task there was only a qualitative assessment of the results, not a quantitative one. Participants received their evaluation as statements of the assessors, for five of the six topics. It was observed that in the absence of an ontology, it becomes extremely difficult to populate the axes meaningfully. Both participants received positive and negative comments on the different topics, and probably the lesson learned is that, while the task is quite challenging, current tools may assist a user who has to create such a map manually.

In NTCIR-7, the long-term goal was the automatic production of technical-trend maps. These resemble the patent maps described above but with the source of the maps not being restricted to patents. As a first step in this research, the task was to classify research papers according to their IPCs to enable technical or technological trends in academia and industry to be summarized together in a single map. One challenge was the need for cross-genre classification involving research papers and patents.

Another challenge was cross-lingual classification. To train a classifier that can assign one or more IPC codes to an input document, the JPO and USPTO patent collections were used. Each document was a research abstract in Japanese or English, resulting in four combinations of languages for the training and test documents. Table 7 shows the Mean Average Precision for different runs and combinations of

Table 7 Evaluation for the IPC-based classification of research abstracts

J-to-J		E-to-E		J-to-E	
RunID	MAP	RunID	MAP	RunID	MAP
HTC13	0.44	NEUN1_S1	0.49	xrce_j2e	0.44
HTC11	0.44	NEUN1_S2	0.47	AINLP05	0.11
HTC12	0.44	NEUN1_S3	0.45	AINLP06	0.10
HTC07	0.44	xrce_e2j2e	0.42	AINLP02	0.09
HTC01	0.43	xrce_en_lm	0.42	AINLP03	0.09
HTC06	0.43	xrce_en_filter	0.42		
HTC05	0.43	xrce_en_pp	0.41		
HTC08	0.43	nttcs2	0.35		
HTC10	0.43	nttcs1	0.34		
HTC03	0.43	KECIR	0.29		
HTC02	0.43	rali2	0.14		
HTC09	0.42	ICL07	0.14		
HTC04	0.42	rali1	0.14		
nttcs4	0.40	ICL07_2	0.13		
HCU1	0.39	BRKLYPM-EN-02	0.13		
HCU2	0.39	AINLP04	0.10		
HTC14	0.39	BRKLYPM-EN-04	0.10		
nttcs3	0.36	AINLP01	0.10		
nttcs2	0.34	BRKLYPM-EN-03	0.09		
nttcs1	0.33	PI-5b	0.04		
KECIR	0.27				
HCU3	0.14				
nut1-1	0.07				
nut2-1	0.04				

languages [38]. In Table 7, "J-to-J" and "E-to-E" indicate that both the test and training documents were in the same language, while "J-to-E" indicates a cross-lingual classification for which the training documents were in English. There were no submissions to an "E-to-J" classification.

Table 7 shows that the MAP of the top run for J-to-E closely matched those for E-to-E (and for J-to-J). All of these runs used variations of the K-Nearest Neighbor method. The MAP for each of the top systems was fairly high compared with that for ad hoc retrieval, which makes sense because the use of multiple training examples made the task more like relevance feedback than ad-hoc retrieval.

The MAP of the top J-to-E run, xrce_j2e (0.44), was higher than those of E-to-E runs by the same group, such as xrce_e2j2e (0.42). This system [5] used a language modeling information retrieval approach, calculating the similarity between an input document q_s and a particular training document d_t as the probability $P(q_s|d_t)$ that q_s would be generated from d_t. For the cross-lingual runs, the NTCIR-1 bilingual document collection was used to estimate the probability that a source-language word w_s would be translated into a target-language word w_t. The resultant probability $P(w_t|w_s)$ was summed over w_s and w_t to calculate $P(q_s|d_t)$. Therefore, using more than one w_s and w_t led to an effect similar to query expansion, which presumably accounts for xrce_j2e outperforming the corresponding monolingual runs.

To compare the MAP of the paper-to-patent cross-genre runs with that of a patent-to-patent classification, one of the organizers who submitted HCU1 in Table 7 performed a classification of JPO patent applications, obtaining a MAP of 0.37, which was comparable to that of HCU1 for J-to-J runs (0.39).

The IPC-based classification was also performed in NTCIR-8, with a variable granularity for the IPC codes (e.g., subclass, main group, and subgroup) being used for evaluation purposes. As expected, the MAP was generally higher for the coarse-grained classes. As in NTCIR-7, there were no E-to-J submissions, but three J-to-E runs were submitted by one participating group, which also submitted E-to-E runs [48]. Although it is not clear how this group matched an input document in Japanese to documents in English, their presentation slide at the NTCIR-8 meeting suggested the use of Google language tools[1]. Comparing the runs for this group, the MAPs for J-to-E runs were slightly higher than those for E-to-E runs, irrespective of the granularity of the IPC codes.

In NTCIR-8, the creation of technical-trend maps was also undertaken. The purpose was to extract fundamental technologies and their effects from the research abstracts or patent documents in question. The effect of a technology is represented by an attribute and its value. These were the definitions of the elements to be extracted [39]:

- TECHNOLOGY: algorithms, tools, materials, and data used in each study or invention.
- EFFECT: pairs of ATTRIBUTE and VALUE tags.
- ATTRIBUTE and VALUE: effects of a technology that can be expressed by a pair comprising an attribute and a value.

The following is an example sentence annotated with the above tag set:

```
Through <TECHNOLOGY>closed-loop feedback control
</TECHNOLOGY>, the system could<EFFECT><VALUE>
minimize</VALUE> the <ATTRIBUTE>power loss
</ATTRIBUTE></EFFECT>.
```

Although the input documents were not actually organized as a map, the extracted elements could be of help in determining appropriate axes for a map. The submitted runs were evaluated by recall, precision, and F-measure on an element-by-element basis for different combinations of document types (research abstract or patent) and languages (Japanese or English).

The general trends present in the evaluation results were as follows. First, the precision was higher than the recall, irrespective of the document type, language, or element type. This suggests that it was difficult to identify exhaustively the various technical terms and expressions used to describe technologies. Moreover, because recall and precision were calculated on an element-by-element basis, the recall becomes zero if even a single word in an element is mislabeled. Second, the evaluation

[1] http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings8/NTCIR/03-NTCIR8-PATMN-TeodoroD_slides.pdf

Table 8 Evaluation for the technical-trend map creation (R: recall, P: precision, and F: F-measure).

	Japanese						English					
	Research			Patent			Research			Patent		
RunID	R	P	F	R	P	F	R	P	F	R	P	F
TRL7	0.18	0.57	0.28	0.41	0.52	0.46	—	—	—	—	—	—
HCU	0.16	0.49	0.24	0.43	0.55	0.48	—	—	—	—	—	—
NUSME-3	—	—	—	—	—	—	0.11	0.38	0.16	0.17	0.37	0.24

results for the patents were higher than that for the research abstracts, which suggests that technical terms and expressions in patents are more standardized than those in research papers. Finally, the evaluation results for documents in Japanese were higher than those for documents in English. However, this tendency could be caused by differences between participating systems, because no group submitted runs involving both languages.

Table 8 shows the evaluation for the groups that achieved the best F-measure in any configuration. The complete evaluation is available in the overview paper [39]. In Table 8, all the groups formulated the extraction task in terms of "BIO" chunking, which labels each token in a sentence as being the beginning (B), inside (I), or outside (O) of the span of interest. Whereas TRL and NUSME used CRF (Conditional Random Field) models to perform sequential labeling, HCU used an SVM (Support Vector Machine) to classify individual words according to the BIO labels. The general trends described above can also be observed in Table 8. In addition, HCU [41] identified typical causes of errors, as follows:

- Specific function words, such as "by" and "of", may occur inside or outside an element.
- Technologies can be expressed by a long noun phrase, such as "a device equipped with functions A, B, ... and Z", especially in patents.
- The order of an attribute and its value can vary depending on the grammatical construction, such as in "high recognition rate" and "the recognition rate becomes high".

These individual errors are ultimately caused by the target-element structure not necessarily being a simple sequence of content words.

4.4 Machine Translation

While NTCIR started as an evaluation series for information retrieval, it quickly expanded to incorporate other tasks that were also related to the broader topic of information access. Part of this broader information access focus is machine translation (MT), which is both a research topic on its own (and as such is evaluated intrinsically) and a tool for other information access systems (e.g., cross-lingual IR, and as such is evaluated extrinsically). The following two sections cover these two approaches to evaluation.

4.4.1 Intrinsic Evaluation

In NTCIR-7, all reference translations used for intrinsic evaluation of machine translation are influenced by Rule-Based MT systems. This includes both the S600 set of 600 Japanese sentences (3 translators, each used Rule-Based MT systems) and the S300 set of 300 Japanese sentences (3 translators, one of whom used a Rule-Based MT system). Participating systems were compared using BLUE scores. Additionally, human translators evaluated 100 sentences of each participant and assessed them for *adequacy* (essentially, how much of the original information is present in the translation?) and *fluency*, each with a score between 1 (not good) to 5 (good).

In the following year, BLUE was again used as a metric for intrinsic evaluation, and an additional effort was made to invite participants to propose new evaluation metrics. However, this approach resulted in only one participant, and it was not continued in subsequent years. Instead, the following two years disposed with the BLUE metric and used only adequacy and *acceptability* (i.e., to what extent can the meaning be understood by a human user?). Table 9 shows the BLEU scores from 2008 and 2010, while Table 10 shows the adequacy scores from 2008, 2011, and 2013. Comparing the two sets of results there is one thing that stands out: while statistical machine translation systems (SMTs) are clearly outperforming the rule-based or example-based systems (RBMTs or EBMTs) in terms of the BLUE scores, the opposite is the case for manual evaluations of adequacy.

Only in NTCIR-9 and -10 did one SMT system manage to outperform RBMTs. This system (NTT) also obtained good scores in the automatic evaluation. Nevertheless, we do not go here into the details of the machine translation methods, but rather refer the reader to Chapter **??** which address this technology at length. Direct comparison between the numbers obtained in each year is not recommended, because the sets to be translated are different, but the organisers of NTCIR-10 also asked participants to translate the test set of the previous year and in their track report [20] present these results, showing that the vast majority of participants had managed to increase the performance of their systems.

4.4.2 Extrinsic Evaluation

All cross-lingual retrieval evaluations are, in essence, extrinsic evaluation of some form of translation technology, but in the NTCIR Patent MT tasks the MT technology was foregrounded and thus the role of cross-lingual IR as extrinsic evaluation of MT was foregrounded.

The NTCIR-7 PatentMT task included an extrinsic evaluation of patent translation that the organizers called Cross-Language Patent Retrieval (CLPR). The key idea was to view the purpose of Machine Translation (MT) as being to support ranked retrieval of existing patents to identify previously awarded patents that invalidate some claim in a new patent application. The specific design of the task was:

- The first claim for each of 124 rejected patent applications was obtained from the Japan Patent Office (JPO) and manually translated into English.

Table 9 BLEU scores for intrinsic evaluation of MT

NTCIR-7					NTCIR-8		
Group	Method	BLEU-SRB	BLEU-MRB300	BLEU-MRB600	Group	Method	BLEU
Japanese - English translation							
NTT	SMT	27.20	35.93	43.72	EIWA-1	RBMT^{+}	34.3
Moses*	SMT	27.14	36.02	43.40	NICT-1	SMT	30.32
(MIT)	SMT	27.14	36.02	44.69	Moses*	SMT	29.08
NAIST-NTT	SMT	25.48	34.66	41.89	KLE-1	SMT	27.75
NiCT-ATR	SMT	24.79	32.29	39.40	DCU-1	SMT	27.61
KLE	SMT	24.49	33.59	40.20	TUTA-2	SMT	26.27
(tsbmt)	RBMT	23.10	37.51	48.02	NICT-4	SMT	25.79
tori	SMT	22.29	27.92	35.02	(TORI-1)	RBMT^{+}	25.65
Kyoto-U	EBMT	21.57	29.35	35.49	NICT-3	SMT	24.96
(MIBEL)	SMT	19.93	27.84	32.99	DCU-3	SMT	24.01
HIT2	SMT	19.48	29.33	33.60	TUTA-1	SMT	22.66
JAPIO	RBMT	19.46	32.62	41.77	TORI-2	RBMT^{+}	21.56
TH	SMT	15.90	24.20	28.72	KYOTO-1	EBMT	21.23
FDU-MCand	SMT	9.55	19.94	20.27	DCU-4	SMT	20.68
(NTNU)	SMT	1.41	2.48	2.63			
English - Japanese translation							
Moses*	SMT	30.58	-	-	NICT-2	SMT	35.87
HCRL	SMT	20.97	-	-	Moses*	SMT	35.27
NiCT-ATR	SMT	29.15	-	-	DCU-1	SMT	33.03
NTT	SMT	28.07	-	-	DCU-7	SMT	30.08
NAIST-NTT	SMT	27.19	-	-	KLE-1	SMT	29.18
KLE	SMT	26.93	-	-	TUTA-2	SMT	28.5
tori	SMT	25.33	-	-	DCU-6	SMT	27.93
(MIBEL)	SMT	23.72	-	-	TUTA-1	SMT	27.82
HIT2	SMT	22.84	-	-	DCU-9	SMT	27.23
(Kyoto-U)	SMT	22.65	-	-	TORI-1	RBMT^{+}	26.02
(tsbmt)	RBMT	17.46	-	-	KYOTO-1	EBMT	24.13
FDU-MCand	SMT	10.52	-	-	DCU-14	SMT	1.27
TH	SMT	2.23	-	-			

* This Moses system was not part of the official runs

$^{+}$ These RBMT systems also contained a statistical component

- This English claim was transited by MT into Japanese by each participating team.
- A standard patent retrieval system was used to search the patent collection with the MT-generated Japanese claim as a bag-of-words query.
- Each patent (in Japanese) that was cited in the decision document rejecting the application was treated as a relevant document, and all other documents were treated as not relevant.
- Mean Average Precision was reported as an evaluation measure.

The NTCIR-8 PatentMT task included an extrinsic evaluation of patent translation using the same CLPR design, this time with 91 rather than 124 claims. In NTCIR-7, the claims were selected to be relatively easy (monolingual Average Precision (AP) between 0.3 and 0.9); in NTCIR-8 the claims were selected to be relatively hard (monolingual AP below 0.4).

Table 10 Adequacy scores for intrinsic evaluation of MT

NTCIR-7			NTCIR-9			NTCIR-10		
Group	Method	score	Group	Method	score	Group	Method	score
Japanese - English translation								
(tsbmt)	RBMT	3.81	JAPIO-1	RBMT	3.67	JAPIO-1	RBMT	3.67
JAPIO	RBMT	3.71	RBMT1-1	RBMT	3.51	RBMT1-1	RBMT	3.57
(MIT)	SMT	3.15	EIWA-1	Hybrid	3.43	EIWA-1	Hybrid	3.53
NTT	SMT	2.96	RBMT3-1	RBMT	3.13	TORI-1	Hybrid	3.48
Kyoto-U	EBMT	2.85	NTT-UT-1	SMT	2.75	NTITI-1	SMT	3.32
Moses*	SMT	2.81	TORI-1	Hybrid	2.73	RWTH-1	SMT	3.07
NAIST-NTT	SMT	2.66	RWTF-1	SMT	2.66	HDU-1	SMT	3.01
KLE	SMT	2.59	Baseline1-1	SMT	2.62	ONLINE1-1	SMT	2.94
tori	SMT	2.58	NAIST-1	SMT	2.61	FUN-NRC-1	SMT	2.89
NiCT-ATR	SMT	2.47	FRDC-1	SMT	2.52	NTITI-2	SMT	2.87
HIT2	SMT	2.44	Baseline2-1	SMT	2.43	Baseline1-1	SMT	2.81
(MIBEL)	SMT	2.38	KYOTO-2	SMT	2.41	KYOTO-1	EBMT	2.74
TH	SMT	1.87	KYOTO-1	EBMT	2.38	Baseline2-1	SMT	2.68
FDU-Mcand	SMT	1.75	UOTTS-1	SMT	2.38	OKAPU-1	SMT	2.61
(NTNU)	SMT	1.08	NEU-1	SMT	2.37	TRGTK-1	SMT	2.55
			ONLINE1-1	SMT	2.27	BJTUX-1	SMT	2.25
			ICT-1	SMT	2.27	ISTIC-1	SMT	1.08
			KLE-1	SMT	2.04			
English - Japanese translation								
(tsbmt)	RBMT	3.53	NTT-UT-1	SMT	3.67	NTITI-2	SMT	3.84
Moses*	SMT	2.90	RBMT6-1	RBMT	3.51	JAPIO-1	RBMT	3.53
NTT	SMT	2.74	JAPIO-1	RBMT	3.46	RBMT6-1	RBMT	3.47
NiCT-ATR	SMT	2.59	RBMT4-1	RBMT	3.25	EIWA-1	Hybrid	3.42
(Kyoto-U)	EBMT	2.42	RBMT5-1	RBMT	2.84	ONLINE1-1	SMT	3.38
			ONLINE1-1	SMT	2.67	BJTUX-1	SMT	2.84
			Baselins1-1	SMT	2.69	TSUKU-1	SMT	2.79
			TORI-1	Hybrid	2.60	Baseline1-1	SMT	2.69
			Baseline2-1	SMT	2.48	FUN-NRC-1	SMT	2.67
			KLE-1	SMT	2.35	Baseline2-1	SMT	2.53
			FRDC-1	SMT	2.35	KYOTO-1	EBMT	2.50
			ICT-1	SMT	2.32	TRGTK-1	SMT	2.45
			UOTTS-1	SMT	2.19	ISTIC-1	SMT	2.30
			KYOTO-2	SMT	2.18			
			KYOTO-1	EBMT	2.05			
			BJTUX-1	SMT	1.80			
Chinese - English translation[†]								
			BBN-1	SMT	4.03	BBN-1	SMT	4.15
			NEU-1	SMT	3.51	RWSYS-1	HYBRID	3.52
			RWTH-1	SMT	3.42	SRI-1	SMT	3.51
			LIUM-1	SMT	3.40	HDU-1	SMT	3.5
			IBM-1	SMT	3.39	RWTH-1	SMT	3.49
			FRDC-1	SMT	3.34	ONLINE1-1	SMT	3.45
			KLE-1	SMT	3.34	ISTIC-1	SMT	3.39
			ICT-1	SMT	3.30	SJTU-1	SMT	3.32
			BUAA-1	HYBRID	3.30	TRGTK-1	SMT	3.3
			UOTTS-1	SMT	3.29	BASELINE1-1	SMT	3.23
			BASELINE1-1	SMT	3.29	BJTUX-1	SMT	3.19
			NTT-UT-1	SMT	3.23	MIG-1	SMT	3.05
			ISTIC-1	HYBRID	3.19	BASELINE2-1	SMT	2.82
			NTHU-1	SMT	3.13	EIWA-1	HYBRID	2.8
			BJTUX-1	SMT	3.11	BUAA-1	SMT	2.3
			EIWA-1	HYBRID	3.05	BJTUX-2	EBMT	2.26

[†] Only the top 16 systems represented here

The NTCIR-10 PatentMT task included an extrinsic evaluation of patent translation that the organizers call the Patent Examination Evaluation (PEE). The key idea is to view the purpose of MT as being to support making a decision on whether to grant a new patent based on an understanding of whether some other (existing) patent invalidates the claims of the new patent application. The specific design of the task is:

- Some number of rejected patent applications to the JPO are selected.
- Bilingual volunteers from the Nippon Intellectual Property Translation Association served as the assessors.
- For each rejected patent, the assessor is given:
  - The decision document (in Japanese) that identifies specific facts found in some specific prior patent that led (perhaps in part) to the rejection of the patent application.
  - The translated patent (translated by MT from Japanese to English) in which those specific facts were found.
- The assessor is asked to determine (on a graded scale) whether the degree to which those specific facts could have been ascertained from the translated patent.
- A second version of PEE, in which the prior patent is first manually translated by hand from Japanese to Chinese and then by machine from Chinese to English was also run.

It is worth noting that the CLEF-2010 and CLEF-2011 Intellectual Property lab (CLEF-IP, see also Chapter **??**) has produced a test collection that could be (but has not yet been) used for extrinsic evaluation of Patent MT. That test collection includes a patent application as a query document, and citations from various sources as relevance judgments. The query document is available in a single language (English, French, or German) but the EPO granted patents contain two fields (title and claims) in all three languages. These could be suppressed for experimental purposes (although that is not done at CLEF).

5 Summary

NTCIR has been a pioneer in creating test collections for patent retrieval. The NTCIR-3 retrieval task, based on information needs extracted from newspaper articles was not repeated, neither in NTCIR, nor in CLEF-IP, primarily due to the cost of assessment. Another common observation with CLEF-IP was the rather reduced interest or ability of teams to provide cross-lingual systems. For monolingual retrieval task, query expansion was one of the features that appears to consistently improve results (see the RDNDC runs in both NTCIR-4, and -5, as well as the HTC runs in NTCIR-4, -5, and -6). For machine translation, intrinsic evaluation based on BLEU shows equal results between rule-based an statistical MT systems (see the NTT run in NTCIR-7 and the EIWA-1 run in NTCIR-8). Classification seems to

depend less on the algorithm itself (K-NN, Naive Bayes, SVM have obtained comparable results in NTCIR-5 and -6) but, unsurprisingly, depend more on the features used, though no clear trend can be observed. Finally, text mining is a difficult task to both address and evaluate. A qualitative evaluation performed in NTCIR-4 on 6 topics from NTCIR-3 provides a starting point, on which further efforts can be built.

References

1. Andersson, L., Lupu, M., Palotti, J.R.M., Piroi, F., Hanbury, A., Rauber, A.: Insight to hyponymy lexical relation extraction in the patent genre versus other text genres. In: Proc. of IPaMin@KONVENS (2014)
2. Aono, M.: Leveraging category-based LSI for patent retrieval. In: Proc. of NTCIR-6 (2007)
3. Attar, R., Fraenkel, A.S.: Local feedback in full-text retrieval systems. Journal of the Association for Computing Machinery pp. 397–417 (1977)
4. Chen, A., Gey, F.: Experiments on cross-language and patent retrieval at NTCIR-3 workshop. In: Proc. of NTCIR-3 (2002)
5. Clinchant, S., Renders, J.M.: XRCE's participation to patent mining task at NTCIR-7. In: Proc. of NTCIR-7 (2008)
6. Fall, C., Torcsvari, A., Benzineb, K., Karetka, G.: Automated categorization in the international patent classification. ACM SIGIR Forum **37**(1) (2003)
7. Ferber, G.: English-Japanese, Japanese-English Dictionary of Computer and Data-Processing Terms. MIT Press (1989)
8. Fujii, A., Ishikawa, T.: Patent retrieval experiments at ULIS. In: Proc. of NTCIR-3 (2002)
9. Fujii, A., Ishikawa, T.: Document structure analysis in associative patent retrieval. In: Proc. of NTCIR-4 (2004)
10. Fujii, A., Ishikawa, T.: Document structure analysis for the NTCIR-5 patent retrieval task. In: Proc. of NTCIR-5 (2005)
11. Fujii, A., Iwayama, M., Kando, N.: Overview of patent retrieval task at NTCIR-4. In: Proc. of NTCIR-4 (2004)
12. Fujii, A., Iwayama, M., Kando, N.: 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, pp. 1643–1646 (2004)
13. Fujii, A., Iwayama, M., Kando, N.: Overview of patent retrieval task at NTCIR-5. In: Proc. of NTCIR (2005)
14. Fujii, A., Iwayama, M., Kando, N.: Introduction to the special issue on patent processing. Information Processing & Management **43**(5), 1149–1153 (2007)
15. Fujii, A., Iwayama, M., Kando, N.: Overview of the patent retrieval task at the NTCIR-6 workshop. In: Proc. of NTCIR-6 (2007)
16. Fujii, A., Utiyama, M., Yamamoto, M., Utsuro, T.: Overview of the patent translation task at the NTCIR-7 workshop. In: Proc. of NTCIR-7 (2008)
17. Fujii, A., Utiyama, M., Yamamoto, M., Utsuro, T., Ehara, T., Echizen-ya, H., Shimohata, S.: Overview of the patent translation task at the NTCIR-8 workshop. In: Proc. of NTCIR-8 (2010)
18. Fujita, S.: Revisiting document length hypotheses: A comparative study of Japanese newspaper and patent retrieval. ACM Transactions on Asian Language Information Processing **4**(2), 207–235 (2005)
19. Gobeill, J., Gaudinat, A., Ruch, P., Pasche, E., Teodoro, D., Vishnyakova, D.: BiTeM site report for TREC Chemistry 2010: Impact of citations feeback for patent prior art search and chemical compounds expansion for ad hoc retrieval. In: Proc. of TREC (2010)
20. Goto, I., Chow, K.P., Lu, B., Sumita, E., Tsou, B.K.: Overview of the patent machine translation task at the NTCIR-10 workshop. In: Proc. of NTCIR-10 (2013)

21. Goto, I., Lu, B., Chow, K.P., Sumita, E., Tsou, B.K.: Overview of the patent machine translation task at the NTCIR-9 workshop. In: Proc. of NTCIR-9 (2011)
22. Harman, D.: Overview of the second text retrieval conference (TREC-2). In: Proc. of TREC (1993)
23. Hashimoto, K., Yukawa, T.: Term weighting classification system using the chi-square statistic for the classification subtask at NTCIR-6 patent retrieval task. In: Proc. of NTCIR-6 (2007)
24. Itoh, H.: NTCIR-4 patent retrieval experiments at RICOH. In: Proc. of NTCIR-4 (2004)
25. Itoh, H., Mano, H., Ogawa, Y.: Term distillation in patent retrieval. In: Proc of The ACL Workshop on Patent Corpus Processing (2003)
26. Iwayama, M., Fujii, A., Kando, N.: Overview of classification subtask at NTCIR-5 patent retrieval task. In: Proc. of NTCIR-5 (2005)
27. Iwayama, M., Fujii, A., Kando, N.: Overview of classification subtask at NTCIR-6 patent retrieval task. In: Proc. of NTCIR-6 (2007)
28. Iwayama, M., Fujii, A., Kando, N., Marukawa, Y.: Evaluating patent retrieval in the third NTCIR workshop. Information Processing & Management **42**(1), 207–221 (2006)
29. Iwayama, M., Fujii, A., Kando, N., Takano, A.: Overview of patent retrieval task at NTCIR-3. In: Proc. of ACL Workshop on Patent Process Processing (2003)
30. Kando, N., Leong, M.K.: Workshop on patent retrieval SIGIR 2000 workshop report. ACM SIGIR Forum **34**(1), 28–30 (2000)
31. Kim, J., Lee, Y., Na, S.H., Lee, J.H.: POSTECH at NTCIR-6 English patent retrieval subtask. In: Proc. of NTCIR-6 (2007)
32. Kishida, K., Chen, K.h., Lee, S., Chen, H.H., Kando, N., Kuriyama, K., Myaeng, S.H., Eguchi, K.: Cross-lingual information retrieval (CLIR) task at the NTCIR workshop 3. ACM SIGIR Forum **38**(1), 17–20 (2004)
33. Lopez, P., Romary, L.: PATATRAS: Retrieval Model Combination and Regression Models for Prior Art Search. In: C. Peters, G. Di Nunzio, M. Kurimo, T. Mandl, D. Mostefa, A. Peñas, G. Roda (eds.) Multilingual Information Access Evaluation I. Text Retrieval Experiments, *LNCS*, vol. 6241. Springer (2010)
34. Lupu, M.: On the usability of random indexing in patent retrieval. In: Proc. of ICCS (2014)
35. Lupu, M., Jiashu, Z., Huang, J., Gurulingappa, H., Filipov, I., Tait, J.: Overview of the TREC 2011 chemical IR track. In: Proc. of TREC (2011)
36. Mahdabi, P., Andersson, L., Keikha, M., Crestani, F.: Automatic refinement of patent queries using concept importance predictors. In: Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, pp. 505–514 (2012)
37. Moldovan, A., Bot, R.I., Wanka, G.: Latent semantic indexing for patent documents. Int. J. Appl. Math. Comput. Sci. **15**(4) (2005)
38. Nanba, H., Fujii, A., Iwayama, M., Hashimoto, T.: Overview of the patent mining task at the NTCIR-7 workshop. In: Proc. of NTCIR-7 (2008)
39. Nanba, H., Fujii, A., Iwayama, M., Hashimoto, T.: Overview of the patent retrieval task at the NTCIR-8 workshop. In: Proc. of NTCIR-8 (2010)
40. Nanba, H., Kamaya, H., Takezawa, T., Okumura, M., Shinmori, A., Tanigawa, H.: Automatic translation of scholarly terms into patent terms. In: Proceeding of the 2nd international workshop on patent information retrieval, pp. 21–24 (2009)
41. Nanba, H., Kondo, T., Takezawa, T.: Hiroshima City University at NTCIR-8 patent mining task. In: Proc. of NTCIR-7 (2010)
42. Piroi, F., Tait, J.: CLEF-IP 2010: Retrieval experiments in the intellectual property domain. In: Proc. of CLEF (2010)
43. Sahlgren, M.: An introduction to random indexing. Tech. rep., SICS, Swedish Institute of Computer Science (2005)
44. Sahlgren, M., Hansen, P., Karlgren, J.: English-Japanese cross-lingual query expansion using random indexing of aligned bilingual text data. In: Proc of NTCIR (2002)
45. Schellner, I.: Japanese File Index classification and F-terms. World Patent Information **24**, 197–201 (2002)
46. Shinmori, A., Okumura, M., Marukawa, Y., Iwayama, M.: Rhetorical structure analysis of Japanese patent claims using cue phrases. In: Proc. of NTCIR-3 (2002)

47. Shinmori, A., Okumura, M., Marukawa, Y., Iwayama, M.: Patent claim processing for readability - structure analysis and term explanation. In: Proc of The ACL Workshop on Patent Corpus Processing (2003)
48. Teodoro, D., Pasche, E., Vishnyakova, D., Gobeill, J., Ruch, P., Lovis, C.: Automatic ipc encoding and novelty tracking for effective patent mining. In: Proc. of NTCIR-8 (2010)
49. Xu, J., Croft, W.B.: Query expansion using local and global document analysis. In: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 4–11 (1996)