Overview of the Patent Mining Task at the NTCIR-8 Workshop

Hidetsugu Nanba
Graduate School of Information Sciences, Hiroshima City University
3-4-1 Ozukahigashi, Hiroshima 731-3194, Japan

Atsushi Fujii
Graduate School of Information Science and Engineering, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan

Makoto Iwayama
Hitachi, Ltd. / Tokyo Institute of Technology
1-280 Higashi-Koigakubo, Kokubunji 185-8601, Japan

Taiichi Hashimoto
National Institute of Advanced Industrial Science and Technology
2-3-26 Aomi, Koto-ku, Tokyo 135-0064, Japan

ABSTRACT

This paper introduces the Patent Mining Task at the Eighth NTCIR Workshop and the test collections produced in this task. The purpose of the Patent Mining Task is to create technical trend maps from a set of research papers and patents. We performed two subtasks: (1) the subtask of research papers classification and (2) the subtask of technical trend map creation. For the subtask of research papers classification, six participant groups submitted 101 runs. For the subtask of technical trend map creation, nine participant groups submitted 40 runs. In this paper, we also report on the evaluation results of the task.

Categories and Subject Descriptors

H.3.3 **[Information Search and Retrieval]**: Search process

H.3.4 [**Systems and Software**]: Performance evaluation

H.3.5 [**Online Information Services**]: Data sharing

General Terms

Measurement, Performance, Experimentation

Keywords

Text classification, information extraction, research paper, patent

1. INTRODUCTION

The Patent Mining Task at the Eighth NTCIR Workshop (NTCIR-8) investigated the effective retrieval of necessary information from research papers and patent databases. In this paper, we introduce the task and report on the evaluation results.

For a researcher in a field of high industrial relevance, retrieving research papers and patents has become an important aspect of assessing the scope of the field. Such fields include bioscience, medical science, computer science, and materials science. In fact, the development of an information retrieval system dealing with research papers and patents for academic researchers is central to the Intellectual Property Strategic Programs for 2009 [1] of the Intellectual Property Strategy Headquarters in the Cabinet Office, Japan.

In addition, research paper searches and patent searches are required by examiners in government Patent Offices, and by the intellectual property divisions of private companies. An example is the execution of an invalidity search among existing patents and research papers, which could invalidate a rival company's patents or patents under application in a Patent Office.

However, the terms used in patents are often more abstract or creative than those used in research papers, to widen the scope of the claims. Therefore, the Patent Mining Task aims to develop fundamental techniques for retrieving, classifying, and analyzing both research papers and patents.

In previous NTCIR Workshops, Patent Classification Subtasks have been conducted [1][2]. In these subtasks, participants were asked to classify Japanese patent applications in terms of the File Forming Term (F-term) system, which is a classification system for Japanese patent documents. Here, we have been focusing on the classification of research papers in addition to patents, and we conducted the Patent Mining Task at NTCIR-7 [5]. The aim of the Patent Mining Task at NTCIR-7 was the classification of research papers, written in either Japanese or English, in terms of the International Patent Classification (IPC) system. At NTCIR-8, we continued this subtask. In addition to this subtask, we started another subtask described as "technical trend map creation".

The remainder of this paper is organized as follows. In Section 2, we explain the task description. In Section 3, we describe the participants in the task. In Section 4, we report on the evaluation results. Finally, we conclude in Section 5.

2. THE PATENT MINING TASK

2.1 Task Overview

The purpose of the Patent Mining Task is to create technical trend maps from a set of research papers and patents. Figure 1 shows an example of a technical trend map. In this map, research papers and patents are classified in terms of elemental technologies and their effects.

To create a technical trend map, the following two steps are required.

(Step 1) For a given field, research papers and patents written in various languages are collected.

[1] http://www.kantei.go.jp/jp/singi/titeki2/keikaku2009_e.pdf

	Effect 1	Effect 2	Effect 3
Technology 1	[AA 1993] [US Pat. XX/XXX]		[BB 2002]
Technology 2	[CC 2000]		
Technology 3		[US Pat. YY/YYY]	[US Pat. ZZ/ZZZ] [JP Pat. WW/WWW]

Figure 1. An example of a technical trend map created from a set of research papers and patents

(Step 2) Elemental technologies and their effects are extracted from the documents collected in Step 1, and the documents are classified in terms of the elemental technologies and their effects.

For each of these steps, the following two subtasks were conducted at NTCIR-8.

- Research Paper Classification
- Technical Trend Map Creation

In the following, we describe the details of these subtasks.

2.1.1 Subtask of Research Paper Classification

The goal of this subtask was the classification of research papers into the IPC system, which is a global standard hierarchical patent classification system. One or more IPC codes are manually assigned to each patent, aiming for effective patent retrieval.

This task was to assign one or more IPC codes at subclass, main group, and subgroup levels to a given topic, expressed in terms of the title and abstract of a research paper. An example of a topic is shown in Figure 2. Here, <TOPIC-ID> specifies the topic identification number, and <TITLE> and <ABSTRACT> specify the title and abstract of the re-search paper to be classified, respectively.

<TOPIC>

<TOPIC-ID> 100 </TOPIC-ID>

<TITLE> DTMF (Dual Tone Multi-Frequency) transmission method for a mobile communication system </TITLE>

<ABSTRACT> A highly efficient speech-encoding scheme called VSELP is adopted for Japanese digital mobile communication systems. However, DTMP (Dual Tone Multi-Frequency) signals are distorted by using this encoding scheme. This paper presents a DTMF signal transmission scheme. DTMF signals are transmitted in the form of call control messages from mobile stations (MS) to the mobile control centre (MCC). In addition, necessary control capabilities in MS and MCC are described. </ABSTRACT>

</TOPIC>

Figure 2. An example of a topic in "English"

The following tasks were conducted.

- Japanese: classification of Japanese research papers using patent data written in Japanese.
- English: classification of English research papers using patent data written in English.

In addition to these tasks, we conducted the following more challenging tasks, which required both cross-genre and cross-lingual information access techniques.

- Cross-lingual (J2E): classification of Japanese research papers using patent data written in English.
- Cross-lingual (E2J): classification of English research papers using patent data written in Japanese.

These four subtasks are summarized in Figure 3.

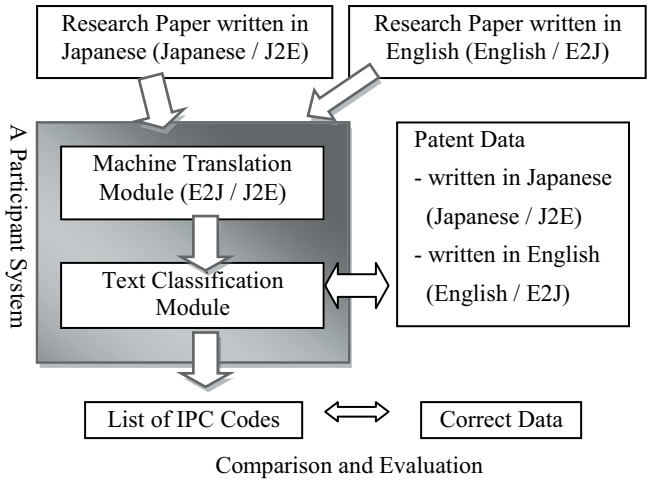

Figure 3. Summary of the subtasks for research paper classification

2.1.2 Subtask of Technical Trend Map Creation

The goal of this subtask was the extraction of expressions of elemental technologies and their effects from research papers and patents. We defined a tag set for this subtask as follows.

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

A tagged example is given in Figure 4.

[Japanese]

PM 磁束制御用コイルを設けて<TECHNOLOGY>閉ループフィードバック制御</TECHNOLOGY>を施すため、<EFFECT><ATTRIBUTION>電力損失</ATTRIBUTION>を<VALUE>最小化</VALUE></EFFECT>できる。

[English]

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

Figure 4. An example of data for the subtask "Technical Trend Map Creation"

The following tasks were conducted.

- Japanese: extraction of technologies and their effects from research papers and patents written in Japanese.
- English: extraction of technologies and their effects from research papers and patents written in English.

In the next subsection, we describe in detail the textual data used in these subtasks.

2.2 Textual Data

An overview of the textual data used in each subtask is shown in Table 1. In the following, we describe details of these data.

Table 1. Document sets

Data	Year	Size	Number	Language
(1) Unexamined Japanese patent applications	1993–2002	100 GB	3.50M	Japanese
(2) USPTO patent data	1993–2000	33 GB	0.99M	English
(3) Patent Abstracts of Japan (translated into English)	1993–2002	4.2 GB	3.50M	English
(4) NTCIR-1 and NTCIR-2 CLIR Task test collection (Abstracts of research papers)	1988–1999	1.4 GB	0.26M	Japanese/ English

2.2.1 Unexamined Japanese Patent Applications

These data were distributed to the teams participating in the subtask of patent map creation and in the "Japanese" and the "E2J" aspects of the subtask of research paper classification. To standardize the format of the documents, the organizers provided an official tool that inserts SGML-style tags into each document. Table 2 shows the tags inserted by that tool. Although passages were extracted from the specific fields, such as claims and detailed descriptions of the invention, any fields can be used for categorization purposes.

Table 2. Tags for Japanese Patent Applications

Tags	Description
<DOC>	document
<DOCNO>	document identifier
<TEXT>	text body
<PASSAGE>	passage
<PNUM>	passage identifier

2.2.2 USPTO Patent Data

These data were distributed to the groups participating in the subtask of patent map creation and in the "English" and "J2E" aspects of the subtask of research papers classification. To standardize the format of the documents, the organizers provided an official tool that inserts SGML-style tags into each document. Table 3 shows the tags inserted by that tool. Because the for-mat of the source data was more complicated than that for the Japanese patent applications, we inserted a large number of tags to enhance the readability of the USPTO patent data. The participant groups were allowed to use <DOC>, <DOCNO>, <TITLE>, <ABST>, <SPEC>, and <CLAIM> for classification purposes.

Table 3. Tags for USPTO patent data

Tags	Description
<DOC>	document
<DOCNO>	document identifier
<APP-NO>	application number
<APP-DATE>	application date
<PUB-NO>	publication number
<PUB-TYPE>	publication type
<PUB-DATE>	publication date
<PRI-IPC>	primary IPC
<IPC-VER>	IPC version
<PRI-USPC>	primary USPC
<PRIORITY>	priority information
<CITATION>	citation(s)
<INVENTOR>	inventor(s)
<ASSIGNEE>	assignee(s)
<TITLE>	title
<ABST>	abstract
<SPEC>	specification
<CLAIM>	claim(s)

2.2.3 Patent Abstracts of Japan (PAJs)

These data were distributed to the groups participating in the subtask of patent map creation and in the "English" and the "J2E" aspects of the subtask of research papers classification. The tags shown in Table 4 were assigned to each document in the PAJ. Participant groups were allowed to use all tags.

2.2.4 NTCIR-1 and NTCIR-2 Cross-lingual Information Retrieval (CLIR) Task Test Collection

This database was distributed to all participant groups, and they were allowed to use it for any purposes. The database was originally used in the CLIR tasks at the first and second NTCIR Workshops (NTCIR-1 and NTCIR-2) [3][4]. It contains 255,960 records of Japanese-English paired documents, with each record comprising a title, the author(s), an abstract, keywords, a publication year, and a conference name.

Table 4. Tags for the PAJ

Tags	Description
<B110>	number of the patent document
<B121>	plain language designation of the kind of documents
<B130>	Kind of document code according to WIPO Standard ST.16
<B190>	WIPO Standard ST3 code, or other identification, of the office or organization publishing the document
<B210>	number(s) assigned to the application(s)
<B220>	date(s) of filing the application(s)
<B310>	number(s) assigned to priority application(s)
<B320>	date(s) of filing of priority application(s)
<B511> <B512>	International Patent Classification
<B542>	title of the invention
<B711>	name(s) of applicant(s)
<B721>	name(s) of inventor(s) if known to be such

2.3 Training and Evaluation Data

2.3.1 Subtask of Research Paper Classification

Sets of topics with manually assigned IPC codes are necessary for the evaluation. However, it is very costly and time consuming to create such data sets. Therefore, we produced the data sets using the following idea.

Essentially, an invention is not patentable if it was already known before the date of filing. However, Article 30 in the Japanese patent law provides a six-month grace period for disclosures made via publication or presentation at a conference or exhibition. In this case, the applicants must mention the proceedings' title (or the conference name) and the date it was published in an "Indication of exceptions to lack of novelty" field (or *exception field*) in the patent. Figure 5 gives an example of an exception field.

We can assume that most of the content of the paper mentioned in the exception field overlaps with the patent. Therefore, if we regard the IPC codes that were assigned to the patent as the codes that should be assigned to the research paper mentioned in the exception field, it becomes possible to create a large-scale data set at low cost.

(original)

【新規性喪失の例外の表示】特許法第３０条第１項適用申請有り２０００年３月１４日 社団法人情報処理学会発行の「第６０回（平成１２年前期）全国大会講演論文集（４）」に発表

(English translation)

[Indication of exceptions to lack of novelty] The provisions set forth in Article 30, Paragraph 1 in Japanese patent law. Proceedings (Volume 4) of the 60th Annual Meeting of the Information Processing Society of Japan, published on March 14, 2000.

Figure 5. An example of the "Indication of exceptions to lack of novelty" field

The procedure used to create the data set was as follows. Firstly, we extracted publication years and proceedings titles from the exception fields in the Japanese patent applications published in the five-year period 2003-2007. Although the title and authors of a paper are not mentioned in the exception field, the authors are usually the same as the inventors of the patent. We therefore extracted and used the inventors of the patent instead of the authors' names.

Secondly, we compared these extracted data with records in a research paper database using a simple string matching method. From this automatic matching, we obtained, on average, six candidate records for each exception field.

Thirdly, we manually identified the correct match from among the candidate records, and obtained 644 pairs of matching patents and research papers.

From these pairs, we created English and Japanese topics (titles and abstracts) and their correct classifications (IPC codes extracted from patents). For each topic, averages of 1.6, 1.9, and 2.4 IPC codes were assigned at subclass, main group, and subgroup levels, respectively.

We then randomly assigned 95 topics to the "dry run" and the remaining 549 topics to the "formal run". The dry run data were provided to the participant teams as training data for the formal run. A list of pairs of a patent ID and one or more IPC codes were also provided as additional training data. These IPC codes were extracted from each patent in the data sets (1), (2), and (3).

Participant teams were asked to submit one or more ranked lists[2] of IPC codes for each topic, to be evaluated using Mean Average Precision (MAP), Recall, and Precision measurements. To calculate these measurements for each submitted run, the organizers produced a Perl program that was compatible with the trec_eval program[3]. The values for MAP, Recall, and Precision are potentially different depending on the version of trec_eval used.

2.3.2 Subtask of Technical Trend Map Creation

Sets of topics with manually assigned "TECHNOLOGY", "EFFECT", "ATTRIBUTE", and "VALUE" tags are necessary for training and evaluation. Therefore, we asked a human subject to assign these tags to the following four types of text.

[2] The maximum number of IPC codes for a single topic is 1,000.

[3] http://trec.nist.gov/trec_eval/trec_eval_latest.tar.gz

- Five hundred Japanese research papers (abstracts)
- Five hundred Japanese patents (abstracts[4])
- Five hundred English research papers (abstracts)
- Five hundred English patents (abstracts)

We then randomly selected 50 texts for the "dry run" and 200 texts for the "formal run" for each type of texts. We also provided the remaining 250 texts to the participant teams as training data.

Participant teams were asked to submit texts with automatically annotated tags, to be evaluated using Recall, Precision, and F-measure. To calculate these measurements for each submitted run, the organizers produced a Perl program.

3. PARTICIPANTS

3.1 Subtask of Research Paper Classification

There were 71 participating systems for the Japanese subtask, 24 for the English subtask, and nine for the Cross-lingual (J2E) subtask. There were six participating groups of universities and companies. Table 5 shows the breakdown of these groups.

Table 5. Breakdown of participants for the subtask of Research Paper Classification

	Japan	Other Asian Countries	Europe
University	1	3	1
Company	1	0	0

The number of runs for each subtask was as follows.

- Japanese: 71 runs from four groups
- English: 24 runs from three groups
- J2E: six runs from one group

There were no runs submitted to E2J.

3.2 Subtask of Technical Trend Map Creation

There were 27 participating systems for the Japanese subtask and 13 for the English subtask. There were nine participating groups of universities and companies. Table 6 shows the breakdown of these groups.

Table 6. Breakdown of participants for the subtask of Technical Trend Map Creation

	Japan	Other Asian Countries	Europe
University	2	3	1
Company	3	0	0

The number of runs for each subtask was as follows.

- Japanese: 27 runs from five groups
- English: 13 runs from four groups

4. RESULTS

4.1 Subtask of Research Paper Classification

We show the evaluation results for the Japanese, English, and Cross-lingual subtasks in Tables 7-9, 10-12, and 13-15, respectively.

Table 7. MAP for "Japanese" at subclass level

Run ID	MAP	Run ID	MAP
HTC04	0.7981	KECIR_JSC_OR	0.7215
HTC07	0.7941	GBCMI9	0.7047
HTC11	0.7937	GBCMI3	0.6790
HTC06	0.7932	GBCMI8	0.6384
HTC05	0.7930	GBCMI10	0.6380
HTC10	0.7918	GBCMI6	0.6355
HTC08	0.7913	GBCMI5	0.6353
HTC02	0.7894	GBCMI2	0.6347
HTC03	0.7892	GBCMI1	0.6292
HTC09	0.7848	GBCMI4	0.6241
HTC01	0.7830	GBCMI7	0.6235
*HCU	0.7289		

(HCU is the task organizer's system)

Table 8. MAP for "Japanese" subtask at main group level

Run ID	MAP	Run ID	MAP
HTC10	0.6429	GBCMI9	0.5416
HTC06	0.6418	GBCMI10	0.5140
HTC11	0.6410	GBCMI3	0.5140
HTC07	0.6409	KECIR_JMP_OR	0.5138
HTC08	0.6397	KECIR_JMP_REC	0.5111
HTC04	0.6388	GBCMI2	0.4701
HTC09	0.6387	GBCMI1	0.4700
HTC05	0.6373	GBCMI8	0.4666
HTC03	0.6290	GBCMI5	0.4598
HTC02	0.6286	GBCMI6	0.4546
HTC01	0.6263	GBCMI7	0.4459
HCU	0.5566	GBCMI4	0.4458

(HCU is the task organizer's system)

[4] Tags were assigned to the fields of "technical problem" (「発明が解決しようとする課題」), "the means for solving a technical problem" (「課題を解決するための手段」), and "effect of the invention" (「発明の効果」) in each abstract.

Table 9. MAP for "Japanese" subtask at subgroup level

Run ID	MAP	Run ID	MAP
HTC10	0.4539	GBCMI9	0.3414
HTC11	0.4525	GBCMI10	0.3131
HTC04	0.4512	GBCMI3	0.3131
HTC09	0.4506	GBCMI1	0.2994
HTC05	0.4503	GBCMI8	0.2914
HTC08	0.4492	GBCMI2	0.2884
HTC06	0.4487	GBCMI5	0.2883
HTC07	0.4487	GBCMI6	0.2788
HTC01	0.4427	GBCMI4	0.2716
HTC03	0.4425	GBCMI7	0.2705
HTC02	0.4419	KECIR_JSG_OR	0.2532
HCU	0.3639	KECIR_JSG_REC	0.2477

(HCU is the task organizer's system)

Table 10. MAP for "English" subtask at subclass level

Run ID	MAP	Run ID	MAP
KECIR_ESC_A_OR	0.7212	BiTeM_combined	0.6660
KECIR_ESC_A_REC	0.7132	BiTeM_weak	0.6612
KECIR_ESC_B_REC	0.6892	PAJ12	0.6162
BiTeM_sim	0.6833	PAJ11	0.6089

Table 11. MAP for "English" subtask at main group level

Run ID	MAP	Run ID	MAP
KECIR_EMP_A_OR	0.5474	BiTeM_combined	0.4799
KECIR_EMP_A_REC	0.5398	BiTeM_weak	0.4689
BiTeM_sim	0.4971	PAJ12	0.4338
KECIR_EMP_B_REC	0.4969	PAJ11	0.4221

Table 12. MAP for "English" subtask at subgroup level

Run ID	MAP	Run ID	MAP
KECIR_ESG_A_OR	0.3693	BiTeM_combined	0.2857
KECIR_ESG_A_REC	0.3546	BiTeM_weak	0.2819
BiTeM_sim	0.2991	PAJ12	0.2648
KECIR_ESG_B_REC	0.2925	PAJ11	0.2450

Table 13. MAP for "J2E" subtask at subclass level

Run ID	MAP
BiTeM_sim	0.7051
BiTeM_combined	0.6872
BiTeM_weak	0.6804

Table 14. MAP for "J2E" subtask at main group level

Run ID	MAP
BiTeM_sim	0.5001
BiTeM_combined	0.4837
BiTeM_weak	0.4736

Table 15. MAP for "J2E" subtask at subgroup level

Run ID	MAP
BiTeM_sim	0.3028
BiTeM_combined	0.2925
BiTeM_weak	0.2819

4.2 Subtask of Technical Trend Map Creation

We show the evaluation results for the Japanese and English subtasks in Tables 16-17 and 18-19, respectively.

5. CONCLUSION

We have given an overview of the evaluation and design of the Patent Mining Task at NTCIR-8. For the subtask of research paper classification, we focused on the "Indication of exceptions to lack of novelty" field in Japanese patent applications and thereby created 644 English and Japanese topics and their correct classifications (IPC codes). Six participant groups submitted 101 runs as formal run. For the subtask of technical trend map creation, nine participant groups submitted 40 runs as formal run.

6. REFERENCES

[1] Iwayama, M., Fujii, A., and Kando, N. 2007. Overview of Classification Subtask at NTCIR-6 Patent Retrieval Task. In *Proceedings of the 6th NTCIR Workshop Meeting*.

[2] Iwayama, M., Fujii, A., and Kando, N. 2005. Overview of Classification Subtask at NTCIR-5 Patent Retrieval Task. In *Proceedings of the 5th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access*.

[3] Kando, N., Kuriyama, K., Nozue, T., Eguchi, K., Kato, H., and Hidaka, S. 1999. Overview of IR Tasks at the First NTCIR Workshop. In *Proceedings of the 1st NTCIR Workshop on Research in Japanese Text Retrieval and Term Recognition*, 11-44.

[4] Kando, N., Kuriyama, K., and Yoshioka, M. 2001. Overview of Japanese and English Information Retrieval Tasks (JEIR) at the Second NTCIR Workshop. In *Proceedings of the 2nd NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization*, 4-37 - 4-60.

[5] Nanba, H., Fujii, A., Iwayama, M., and Hashimoto, T. 2008. Overview of the Patent Mining Task at the NTCIR-7 Workshop. In *Proceedings of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-lingual Information Access*, 325-332.

Table 16. Recall, Precision, and F-measure for "Japanese" subtask (research paper)

Run ID	Technology (Title)			Technology (Abstract)			Attribute (Abstract)		
	Recall	Precision	F-measure	Recall	Precision	F-measure	Recall	Precision	F-measure
TRL7_1	0.323	0.811	0.462	0.207	0.605	0.309	0.122	0.450	0.191
TRL6_1	0.323	0.769	0.455	0.196	0.617	0.298	0.108	0.471	0.176
TRL3_1	0.301	0.875	0.448	0.191	0.570	0.286	0.135	0.404	0.203
TRL4_1	0.301	0.903	0.452	0.191	0.590	0.288	0.115	0.425	0.181
TRL5_1	0.280	0.867	0.423	0.193	0.574	0.289	0.115	0.436	0.182
TRL8_1	0.301	0.903	0.452	0.182	0.532	0.272	0.118	0.376	0.180
TRL1_1	0.290	0.871	0.435	0.188	0.602	0.286	0.111	0.367	0.171
TRL2_1	0.290	0.871	0.435	0.188	0.607	0.287	0.108	0.376	0.168
*HCU	0.656	0.656	0.656	0.131	0.495	0.206	0.095	0.394	0.153
ONT	0.280	0.634	0.388	0.091	0.219	0.129	0.081	0.154	0.106
smlab	0.000	0.000	0.000	0.000	0.000	0.000	0.108	0.252	0.151
HTC_1	0.000	0.000	0.000	0.000	0.000	0.000	0.149	0.164	0.156
HTC_2	0.000	0.000	0.000	0.000	0.000	0.000	0.115	0.111	0.113

Run ID	Value (Abstract)			Effect (Abstract)			Average		
	Recall	Precision	F-measure	Recall	Precision	F-measure	Recall	Precision	F-measure
TRL7_1	0.163	0.539	0.251	0.051	0.500	0.093	0.181	0.573	0.275
TRL6_1	0.160	0.618	0.254	0.061	0.600	0.111	0.172	0.604	0.268
TRL3_1	0.184	0.394	0.251	0.055	0.500	0.098	0.183	0.491	0.266
TRL4_1	0.184	0.446	0.260	0.044	0.464	0.081	0.177	0.530	0.265
TRL5_1	0.170	0.467	0.249	0.041	0.429	0.075	0.172	0.534	0.260
TRL8_1	0.190	0.364	0.250	0.044	0.406	0.080	0.177	0.460	0.256
TRL1_1	0.153	0.372	0.217	0.048	0.483	0.087	0.166	0.487	0.247
TRL2_1	0.146	0.364	0.209	0.044	0.481	0.081	0.163	0.491	0.244
*HCU	0.105	0.383	0.165	0.061	0.310	0.103	0.160	0.491	0.241
ONT	0.122	0.267	0.168	0.027	0.182	0.047	0.114	0.246	0.156
smlab	0.180	0.469	0.260	0.096	0.215	0.132	0.081	0.354	0.132
HTC_1	0.207	0.210	0.209	0.055	0.112	0.073	0.100	0.188	0.131
HTC_2	0.238	0.206	0.221	0.058	0.099	0.073	0.100	0.161	0.123

(HCU is the task organizer's system)

Table 17. Recall, Precision, and F-measure for "Japanese" subtask (patent)

Run ID	Technology (Title)			Technology (Abstract)			Attribute (Abstract)		
	Recall	Precision	F-measure	Recall	Precision	F-measure	Recall	Precision	F-measure
*HCU	0.556	0.455	0.500	0.439	0.490	0.463	0.371	0.544	0.440
TRL6_2	0.000	0.000	0.000	0.399	0.407	0.403	0.403	0.523	0.455
TRL7_2	0.000	0.000	0.000	0.400	0.402	0.401	0.405	0.519	0.455
TRL8_2	0.000	0.000	0.000	0.374	0.449	0.408	0.332	0.545	0.413
TRL3_2	0.000	0.000	0.000	0.378	0.438	0.406	0.332	0.532	0.409
TRL4_2	0.000	0.000	0.000	0.378	0.445	0.409	0.330	0.535	0.408
TRL5_2	0.000	0.000	0.000	0.374	0.445	0.407	0.322	0.553	0.407
TRL2_2	0.000	0.000	0.000	0.388	0.432	0.409	0.330	0.490	0.394
TRL1_2	0.000	0.000	0.000	0.335	0.410	0.369	0.229	0.472	0.309
smlab	0.444	0.190	0.267	0.316	0.484	0.383	0.180	0.492	0.263
HTC_1_1	0.000	0.000	0.000	0.000	0.000	0.000	0.251	0.241	0.246
HTC_2_1	0.000	0.000	0.000	0.000	0.000	0.000	0.247	0.282	0.263
HTC_1_2	0.000	0.000	0.000	0.000	0.000	0.000	0.241	0.236	0.239
HTC_2_2	0.000	0.000	0.000	0.000	0.000	0.000	0.237	0.273	0.254
ONT	0.222	0.222	0.222	0.047	0.080	0.060	0.219	0.296	0.252

Run ID	Value (Abstract)			Effect (Abstract)			Average		
	Recall	Precision	F-measure	Recall	Precision	F-measure	Recall	Precision	F-measure
*HCU	0.481	0.655	0.555	0.268	0.409	0.324	0.431	0.545	0.481
TRL6_2	0.540	0.681	0.602	0.256	0.517	0.342	0.437	0.506	0.469
TRL7_2	0.542	0.676	0.602	0.264	0.531	0.352	0.438	0.501	0.468
TRL8_2	0.502	0.686	0.580	0.190	0.489	0.274	0.395	0.537	0.455
TRL3_2	0.504	0.701	0.587	0.198	0.487	0.282	0.397	0.530	0.454
TRL4_2	0.502	0.688	0.580	0.196	0.500	0.282	0.396	0.532	0.454
TRL5_2	0.494	0.705	0.581	0.202	0.553	0.296	0.390	0.539	0.453
TRL2_2	0.517	0.664	0.581	0.209	0.477	0.290	0.404	0.508	0.450
TRL1_2	0.407	0.654	0.502	0.137	0.493	0.214	0.322	0.483	0.387
smlab	0.297	0.829	0.438	0.162	0.414	0.232	0.272	0.547	0.363
HTC_1_1	0.580	0.434	0.496	0.164	0.223	0.189	0.233	0.346	0.278
HTC_2_1	0.521	0.462	0.490	0.153	0.236	0.186	0.215	0.380	0.275
HTC_1_2	0.572	0.432	0.492	0.155	0.217	0.181	0.227	0.344	0.274
HTC_2_2	0.508	0.455	0.480	0.145	0.228	0.177	0.209	0.373	0.268
ONT	0.338	0.503	0.404	0.125	0.339	0.182	0.178	0.271	0.215

Table 18. Recall, Precision, and F-measure for "English" subtask (research paper)

Run ID	Technology (Title)			Technology (Abstract)			Attribute (Abstract)		
	Recall	Precision	F-measure	Recall	Precision	F-measure	Recall	Precision	F-measure
NUSME-3	0.247	0.426	0.313	0.085	0.377	0.138	0.083	0.315	0.132
BiTeM_3	0.204	0.463	0.284	0.035	0.128	0.055	0.059	0.042	0.049
BiTeM_2	0.215	0.513	0.303	0.026	0.105	0.042	0.054	0.041	0.046
NUSME-2	0.247	0.426	0.313	0.085	0.372	0.138	0.034	0.233	0.060
BiTeM_1	0.151	0.350	0.211	0.044	0.118	0.064	0.059	0.034	0.043
ISTIC-1	0.215	0.435	0.288	0.061	0.368	0.105	0.015	0.429	0.028
ISTIC-3	0.226	0.457	0.302	0.061	0.457	0.108	0.005	0.167	0.010
ISTIC-2	0.151	0.333	0.207	0.032	0.268	0.057	0.015	0.375	0.028
BiTeM_4	0.129	0.400	0.195	0.038	0.165	0.062	0.029	0.053	0.038
NUSME-1	0.108	0.909	0.192	0.029	0.625	0.056	0.005	0.333	0.010
KAIST-IRNLP	0.140	0.812	0.239	0.000	0.000	0.000	0.000	0.000	0.000

Run ID	Value (Abstract)			Effect (Abstract)			Average		
	Recall	Precision	F-measure	Recall	Precision	F-measure	Recall	Precision	F-measure
NUSME-3	0.093	0.409	0.152	0.041	0.143	0.063	0.105	0.380	0.164
BiTeM_3	0.461	0.144	0.220	0.041	0.052	0.045	0.159	0.127	0.141
BiTeM_2	0.456	0.144	0.218	0.036	0.046	0.040	0.154	0.127	0.139
NUSME-2	0.041	0.364	0.074	0.000	0.000	0.000	0.081	0.364	0.132
BiTeM_1	0.456	0.138	0.212	0.041	0.046	0.043	0.155	0.111	0.130
ISTIC-1	0.047	0.429	0.084	0.000	0.000	0.000	0.064	0.405	0.110
ISTIC-3	0.026	0.333	0.048	0.000	0.000	0.000	0.058	0.425	0.102
ISTIC-2	0.041	0.400	0.075	0.000	0.000	0.000	0.043	0.324	0.076
BiTeM_4	0.041	0.133	0.063	0.000	0.000	0.000	0.047	0.138	0.070
NUSME-1	0.005	0.500	0.010	0.000	0.000	0.000	0.026	0.688	0.051
KAIST-IRNLP	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.812	0.031

Table 19. Recall, Precision, and F-measure for "English" subtask (patent)

Run ID	Technology (Title)			Technology (Abstract)			Attribute (Abstract)		
	Recall	Precision	F-measure	Recall	Precision	F-measure	Recall	Precision	F-measure
NUSME-3	0.385	0.231	0.288	0.390	0.288	0.331	0.174	0.366	0.236
ISTIC-1-1	0.359	0.304	0.329	0.294	0.429	0.349	0.061	0.464	0.108
NUSME-2	0.385	0.231	0.288	0.390	0.287	0.330	0.023	0.104	0.038
ISTIC-1	0.333	0.295	0.313	0.280	0.428	0.338	0.038	0.308	0.067
ISTIC-2	0.385	0.341	0.361	0.279	0.445	0.343	0.042	0.346	0.075
ISTIC-2-1	0.333	0.302	0.317	0.288	0.421	0.342	0.042	0.391	0.076
NUSME-1	0.179	0.438	0.255	0.185	0.415	0.256	0.000	0.000	0.000
ISTIC-3	0.282	0.314	0.297	0.117	0.452	0.186	0.019	0.222	0.035
BiTeM_3	0.179	0.167	0.173	0.123	0.082	0.098	0.075	0.045	0.057
BiTeM_1	0.103	0.069	0.082	0.122	0.086	0.101	0.080	0.044	0.057
BiTeM_2	0.077	0.088	0.082	0.117	0.080	0.095	0.075	0.046	0.057
BiTeM_4	0.077	0.100	0.087	0.145	0.108	0.124	0.009	0.022	0.013
KAIST_IRNLP	0.308	0.429	0.358	0.000	0.000	0.000	0.000	0.000	0.000

Run ID	Value (Abstract)			Effect (Abstract)			Average		
	Recall	Precision	F-measure	Recall	Precision	F-measure	Recall	Precision	F-measure
NUSME-3	0.364	0.585	0.449	0.110	0.190	0.139	0.350	0.316	0.332
ISTIC-1-1	0.172	0.630	0.270	0.000	0.000	0.000	0.239	0.438	0.309
NUSME-2	0.202	0.526	0.292	0.000	0.000	0.000	0.301	0.291	0.296
ISTIC-1	0.162	0.667	0.260	0.000	0.000	0.000	0.224	0.432	0.295
ISTIC-2	0.131	0.650	0.218	0.005	0.091	0.009	0.221	0.447	0.295
ISTIC-2-1	0.116	0.590	0.194	0.000	0.000	0.000	0.223	0.423	0.292
NUSME-1	0.051	0.833	0.095	0.000	0.000	0.000	0.134	0.424	0.204
ISTIC-3	0.091	0.581	0.157	0.010	0.167	0.019	0.102	0.436	0.165
BiTeM_3	0.571	0.216	0.313	0.040	0.036	0.038	0.185	0.110	0.138
BiTeM_1	0.551	0.206	0.300	0.035	0.030	0.033	0.180	0.108	0.135
BiTeM_2	0.581	0.216	0.315	0.040	0.038	0.039	0.180	0.109	0.135
BiTeM_4	0.056	0.244	0.091	0.000	0.000	0.000	0.107	0.107	0.107
KAIST_IRNLP	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.429	0.018