Report from the NTCIR Workshop 3

Noriko Kando, Jun Adachi
National Institute of Informatics
{Noriko.Kando, adachi}@nii.ac.jp

1. Introduction

The *NTCIR* Workshop [1] is a series of evaluation workshops designed to enhance research in information access (IA) technologies including information retrieval (IR), cross-lingual information retrieval (CLIR), automatic text summarization, question answering, text mining and so on.

The aims of the *NTCIR* project are:

1. to encourage research in information access technologies by providing large-scale test collections that are reusable for experiments;
2. to provide a forum for research groups interested in cross-system comparisons and exchanging research ideas in an informal atmosphere; and
3. to investigate methodologies and metrics for evaluation of information access technologies and methods for constructing large-scale reusable test collections.

The main goal of the *NTCIR* project is to provide infrastructure for large-scale evaluations of IA technologies. The importance of such infrastructure in IA research has been widely recognized. Fundamental text processing procedures for IA, such as indexing includes language-dependent procedures. In particular, processing texts written in Japanese or other East Asian languages such as Chinese is quite different from processing English, French or other European languages, because there are no explicit boundaries (i.e., no spaces) between words in a sentence. The *NTCIR* project therefore started in late 1997 with emphasis on, but not limited to, Japanese or other East Asian languages, and its series of workshops has attracted international participation.
In *NTCIR*, a workshop is held about once every one and a half years, i.e., with intervals of about 18 months. Because we respect the interaction between participants, we consider the whole process from initial document release to the final meeting to be the "workshop". Each workshop selects several research areas called "*tasks*", or a "*challenges*" for the more challenging tasks. Each task has been organized by the researchers of the domain and a task may consist of more than one subtask.

1.1 Information Access

The term "information access" (IA) refers the whole process from when a user realizes his/her information needs, through the activity of searching for and finding relevant documents, and then utilizing information extracted from them. We have looked at IA technologies to help users utilize the information in large-scale document collections. IR, summarization and question answering are part of a "family", aiming at the same target, although each of them has been investigated by rather different communities.

[1] In addition, the problem of "how to define a user's search question" is also included in the scope of IA although this has not been explicitly investigated in NTCIR so far.

1.2 Focus of *NTCIR*

From the beginning of the project, we have looked at both traditional laboratory-type IR system testing and the evaluation of challenging technologies. For the laboratory-type testing, we placed emphasis on text retrieval and CLIR with Japanese or other Asian languages and testing on various document genres. For the challenging issues, the target is to shift from document retrieval to technologies that utilize "information" in documents, and investigation of methodologies and metrics for more realistic and reliable evaluation. For the latter, we have paid attention to users' information-seeking tasks in the experiment design. These two directions have been supported by a forum of researchers who are interested in cross-system comparison and by their discussions.

2. The Third NTCIR Workshop

For the *third NTCIR Workshop* [2], the process started from October 2001 and the final meeting was held on 8-10 October 2002, at National Institute of Informatics (NII) in Tokyo. It was sponsored by the NII and Japan's MEXT Grant-in-Aid for Scientific Research on Informatics (#13224087). The Patent Retrieval task was organized in cooperation with the Japan Intellectual Property Association and NII, and the *CLIR* task was organized in cooperation with the National Taiwan University and the Korean Institute for Scientific and Technological Information (KISTI).

The third *NTCIR Workshop* selected five areas of research as "tasks":

1. Cross-Lingual Information Retrieval Task (*CLIR*),
2. Patent Retrieval Task (*PATENT*),
3. Question Answering Challenge (*QAC*),
4. Text Summarization Challenge (*TSC*), and
5. Web Retrieval Task (*WEB*).

As shown in **Figure 1**, this was the first workshop to include *PATENT*, *QAC* and *WEB* tasks. *TSC* has a new subtask of multi-document summarization, and *CLIR* task includes a new subtask of multilingual *CLIR*. Not only *QAC* and *TSC*, we found that all five tasks contained new and challenging issues regarding technologies as well as their evaluation, and therefore each task was a "challenge" for both the participants and the task organizers.

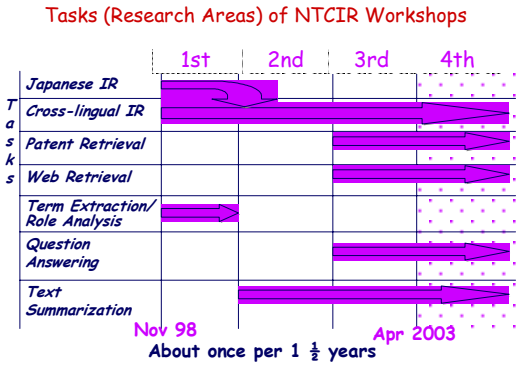

Fig. 1. Tasks at *NTCIR Workshops*

At the meeting on 8-10 October 2002 [3], the first day was an "Open Forum", i.e. no restriction on attendance. The keynote and invited speeches were given by Amit Singhal and Donna Harman. These were followed by an overview of the third *NTCIR Workshop* and discussion. The second and third days were "Semi-Open Sessions" for the task presentations; the participants were restricted to the active participants of the third NTCIR Workshop and the special participants whose applications were approved by the program committee. For each task, the overview report was followed by oral presentations from several participants and a discussion. Each participant prepared a poster presentation and/or a demo, and significant time was devoted to the digital poster sessions and the interactive discussions between participants. The achievements from each task and those of each participant are presented in the reports from each task in this issue and in the Proceedings [4].

2.1 Participants

Table 1 is a list of the active participating research groups in the third *NTCIR Workshop*. Sixty-five groups from nine different countries and areas submitted task results. Among these, 14 groups are from companies, seven are from national or independent research institutes, and 44 are from universities

.

Table 1: Active Participating Groups of the Third NTCIR Workshop

Chungnam National University (Korea) & ETRI+ (Korea)	NTT DATA* (Japan)
Carnegie Mellon University (USA)	New York University (USA) & CRL+ (Japan)
Communication Research Laboratory+ (3 groups) (Japan)	Oki Electric* (Japan)
CRL+ (Japan) & New York University (USA)	Osaka Kyoiku Univeristy (3 groups) (Japan)
Fu Jen Catholic University (Taiwan ROC)	POSTECH (2 groups) (Korea)
Hitachi* (Japan)	Queen College City University of New York (USA)
Hong Kong Polytechnic University (Hong Kong)	RICOH* (Japan)
Hummingbird* (Canada)	Ritsumeikan University (2 groups) (Japan)
Institute of Software, Chinese Academy of Sciences+ (China, PRC)	SICS+ (Sweden)
Johns Hopkins University (USA)	Surugadai University (Japan)
Keio University (2 groups) (Japan)	Thomson Legal and Regulatory* (USA)
Kent Ridge Digital Labs+ (Singapore)	Tianjin University (China PRC)
Kochi University of Technology (Japan)	Tokyo Institute of Technology (Japan)
Korea University (Korea)	Tokai University & Beijin Japan Center (China PRC)
Matsushita Electric Industrial* (Japan)	Toshiba* (Japan)
Microsoft Research Asia* (China PRC)	Toyohashi University of Technology (4 groups) (Japan)
Mie University (Japan)	ULIS & AIST+ (2 groups) (Japan)
Nara Advanced Institute of Science and Technology (Japan)	University Aizu (2 groups)
NAIST & CRL+ (Japan)	University of California Berkeley (2 groups) (USA)
National Taiwan University (Taiwan ROC)	University of Tokyo (2 groups) (Japan)
NEC Kansai* (Japan)	University of Lib and Information Science (2 groups) (Japan)
NEC MRL* (Japan)	University of Tokyo (Japan) & RICOH* (Japan)
NTT Data Technology* (Japan)	Waterloo University (Canada)
NTT-CS* (Japan)	Yokohama National University (2 groups) (Japan)
NTT-CS* (Japan) & NAIST (Japan)	65 groups from 9 countries, *: company, +: national or independent research institute, without-symbol: university

As shown in **Figure 2**, the number of participants has gradually increased. Different tasks attracted different research groups. Many international participants enrolled in *CLIR*. The *PATNET* task attracted participants from company research laboratories and “veteran” *NTCIR* participants. The *WEB* task had participants from various research communities such as machine learning and DBMS.

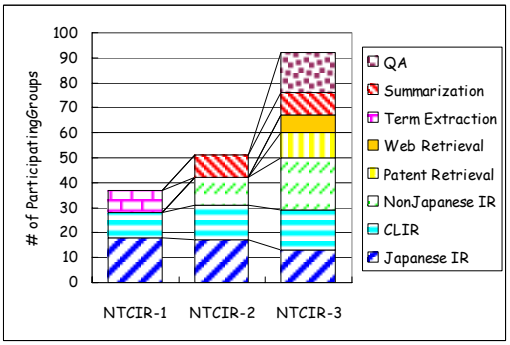

Fig. 2. Number of Participating Groups, by Task

2.2 Test Collections

Table 2 shows the test collections constructed through the series of *NTCIR workshops*. In the *NTCIR* the term "*test collection*" is used for any kind of data set usable for system testing and experiments. One of our interests is to prepare realistic evaluation infrastructures and efforts include scaling up the document collection and increasing variety of document genres and languages. Both patent and scientific document collections have *parallel corpora* of English and Japanese abstracts. The task (experiment) design and relevance judgment criteria were set according to the nature of the document collection and of the user community who use this type of document in their everyday life.

The structure of the topic record in the IR test collections is similar to that used in TREC [5] and CLEF [6]. Topics are defined as statements of "user's requests" rather than "queries" (the strings actually submitted to the system), because we wished to allow both manual and automatic query construction from the topics. Emphasis has been shifted towards the topic structure to allow more realistic experiments and to gauge the effect of background information on the topic. For example, the narrative <NARR>, longer natural language explanation for each topic, may be structured using tags indicating subfields in <NARR>, such as "Background/Purpose of Search", "Relevance Judgment Criteria", or "Term Definition". Most *NTCIR* collections contain a list of concepts <CONC>, but they are not heavily used by participants. The topics in the *PATENT* and *WEB* collections contain documents or document IDs usable as queries.

In the Workshop, *Mandatory Runs* are defined for each IR-related task, and every participant must submit at least one mandatory run using the specified topic field only. The purpose of this is to enhance cross-system comparisons by basing them on common conditions, and to judge the effectiveness of the additional information. Mandatory runs were originally designated "<DESC> only", but the WEB Retrieval task was designated "<TITLE> only" and "<DESC> only". Any combination of topic fields may be used in experiments for research purposes.

For multi-document summarization in the TSC, both the documents themselves and the topics of the document sets were given to the participants.

Table 2. Test collections constructed by *NTCIR*

NTCIR Test Collections; IR and QA										
Collection	Task	Documents						Task data		
								Topic/		Relevance judge
		Genre	Filename	Lang.	Year	# of docs	Size	Lang.	#	
NTCIR-1	IR	Sci. abstract	ntc1-je	JE	1988-1997	339,483	577MB	J	83	3 grades
			ntc1-j	J		332,918	312MB			
			ntc1-e	E		187,080	218MB		60	
	TE*5		ntc1-tmrc	J		2,000	-		-	-
CIRB010	IR	News	CIRB010	C_t	1998-1999	132,173	132MB	C_tE	50	4 grades
NTCIR-2	IR	Sci. abstract	ntc2-j	J	1986-1999**	400,248	600MB	JE	49	4 grades
			ntc2-e	E		134,978	200MB			
NTCIR-3 CLIR	IR	News	KEIB010	K	1994	66,146	74MB	C_tKJE	30	4 grades
			CIRB011	C_t	1998-1999	132,173	870MB	C_tKJE	50	4 grades
			CIRB020			249,508				
			Mainichi	J		220,078				
			EIRB010	E		10,204				
			Mainichi Daily			12,723				
NTCIR-3 PATENT	IR	Patent full	kkh *3	J	1998-1999	697,262	18GB	C_tC_sKJE	31	3 grades
		Abstract	jsh *3	J	1995-1999	1,706,154	1,883MB			
		Abstract	paj *3	E	1995-1999	1,701,339	2,711MB			
NTCIR-3 QA	QA	News	Mainichi	J	1998-1999	220,078	282MB	J*	1200	exact answer
NTCIR-3 WEB	IR	Web (html/text)	NW100G-01	multiple*4	crawled in 2001	1,445,466	10GB	J*	47	4 grades + relative
						11,038,720	100GB			
NTCIR-4 CLIR	IR	News	CIRB011	Ct	1998-1999	132,173	ca.3GB	CtKJE	60	4 grades
			CIRB020			249,203				
			Hankookilbo +	K		149,921				
			Chosenilbo +			104,517				
			Mainichi	J		220,078				
			Yomiuri +			373,558				
			EIRB010	E		10,204				
			Mainichi Daily			12,723				
			Korea Times +			19,599				
			Hong Kong Standard +			96,683				
			Xinhua +			208,167				
NTCIR-4 PATENT	IR	patent full	Publication of unexamined patent application +	J	1993-2002	ca. 3,500,000	ca.45GB	CtCsKJE		
		Abstract	Patent Abstracts of Japan (PAJ) +	E	1993-2002	ca. 3,500,000	ca.10GB			
NTCIR-4 QA	QA	News	Mainichi	J	1998-1999	220,078	ca.776MB	J*		
			Yomiuri +			373,558				
NTCIR-4 WEB	IR	Web (html/text)	NW100G-01	multiple*4	crawled in 2001	11,038,720	100GB	J*		
J:Japanese, E:English, C:Chinese (C_t:Traditional Chinese, C_s: Simplified Chinese), K:Korean;										
"+" indicates the document collection was newly added for NTCIR-4										
* English translation is available										
** gakkai subfiles: 1997-1999, kaken subfiles: 1986-1997										
*3: kkh : Publication of unexamined patent application, jsh: Japanese abstract, paj: English translation of jsh										
*4: almost Japanese or English (some in other languages)										
*5: Term extraction/ role analysis										

NTCIR Text Summarization

Collection	Task	Documents					Summaries		
		Genre	Filename	Lang	Year	# of doc	Types	Analysts	total#
NTCIR-2 SUMM	Single doc	News	Mainichi	J	1994.1995 .1998	180 doc	7	3	3780
NTCIR-2 TAO	Single doc	News	Mainichi	J	1998	1000 doc	2	1	2000
NTCIR-3 SUMM	Single doc	News	Mainichi	J	1998-1999	60 docs	7	3	1260
	Multi doc		Mainichi	J		50 sets	2	3	300
NTCIR-4 SUMM	Multi doc	News	Mainichi	J	1998-1999				
			Yomiuri						

2.3 Relevance Judgments and Evaluation

In IR-related tasks, relevance judgments were graded using a scale similar to previous *NTCIR* workshops: highly relevant, relevant, partially relevant and irrelevant. For the *Patent Retrieval* task, professional patent intermediaries conducted judgments on the pooled documents consisting of the documents listed in the higher ranks in each submitted run, together with intensive interactive search and judgments using several commercial patent retrieval systems and the system provided by the task organizers. Such integration of the two different strategies was found to improve the completeness of the relevance judgments for a large-scale document collection with longer documents.

For the *QAC*, the mean reciprocal rank (MRR) is used for subtask 1, returning five possible answers with no penalty for wrong answers, and the F-measure is used for subtask 2, returning one set of all the answers with penalties given for wrong answers. For subtask 3, a series of questions are used. For *Text Summarization*, content-based and readability-based intrinsic evaluation was used for both single document and multi-document summarization, and a new evaluation methodology was proposed and conducted based on the revision (edit distance) of system summaries by the professional analysts who created the model summaries.

3. Discussion

A brief overview of the *third NTCIR Workshop* is reported here. In the *third NTCIR Workshop,* all five tasks included new components. The details of the achievements from each task and those of each participant are reported in the reports from each task in this issue, the papers in the Proceedings [4] and in several follow-up analyses published in elsewhere.

We are now in the fourth iteration of the series, with the final meeting will be held in Tokyo on 2–5 June 2004. Based on discussions during the final meeting and several *Round-Table Discussions*, we decided to continue these five tasks at the *fourth NTCIR Workshop.* Most of the tasks were continued with minor changes in task design to remedy the major problems found in the third workshop.

To enhance the research in the fourth workshop, special attentions were paid (1) to provide longer time period for experiments, and (2) to enhance the document collections. In the *third NTCIR Workshop*, lots had to be done by the participants for the new tasks and new task components. As results, participants could implement some of their research ideas, but generally such new task components had not been fully investigated and analyzed because of tight schedule of the workshop. Moreover, for some of the new components like “passage-level

relevance judgments" in the *PATENT*, *QAC* and *WEB* tasks and "Search Results Classification" at the *WEB* task, none of the participants fully accomplished. Then, in the *fourth NTCIR Workshop,* we released the document collection as early as possible and omitted the dry runs for the tasks in which the *NTCIR-3 collection* were usable for training. For much more challenging issue, we set the "*feasibility study*" subtask, in which the investigation is performed through the two consecutive workshops, i.e. for three years. In such ways, we expected that each participant could spend sufficient time for experimentation and implementation.

For the document collections, as shown in **Table 2**, we prepared the enlarged, well-balanced collections of Chinese, Korean, Japanese and English news article document collections -- the size of each language collection increased and consisted of the documents from multiple sources. The Patent document collection increased the size to 10 years.

The details of the task descriptions at the *fourth NTCIR Workshop* are found on the websites for the *NTCIR Workshop* [7], together with each task and some further information. The final meeting is also open for the researchers of the domain. Evaluation must adapt to technological evolution and the change in social needs. We are working towards this goal, and suggestions are always welcome.

4. References

1. NTCIR Project: http://research.nii.ac.jp/ntcir/
2. NTCIR Workshop 3 (2001-2002) : http://research.nii.ac.jp/ntcir-ws3/work-en.html
3. NTCIR Workshop 3 Meeting (8-10 October 2002) : http://research.nii.ac.jp/ntcir/ntcir-ws3/
4. Oyama, K., Ishida, E., Kando, N. (eds): NTCIR Workshop 3: Proceedings of the Third NTCIR Workshop on Research in Information Retrieval, Question Answering and Summarization, Tokyo Japan, Oct. 2001–Oct. 2002. ISBN-4-86049-016-9, NII, Tokyo (2003) (http://research.nii.ac.jp/ntcir/ workshop/OnlineProceedings3/)
5. TREC: http://trec.nist.gov/
6. CLEF: http://clef.iei.pi.cnr.it/
7. NTCIR Workshop 4 (2003-2004): http://research.nii.ac.jp/ntcir-ws4/work-en.html