Evaluation of Information Access Technologies at NTCIR Workshop

Noriko Kando
National Institute of Informatics (NII), Tokyo
kando@nii.ac.jp

Abstract: This paper introduces the NTCIR Workshops, a series of evaluation workshops that are designed to enhance research in information access technologies, such as information retrieval, text summarization, question answering, text mining, etc., by providing infrastructure of large-scale evaluation. A brief history, test collections, and recent progress after the previous CLEF Workshop are described with highlighting the difference from CLEF in this paper. To conclude, some thoughts on future directions are suggested.

1 Introduction

The *NTCIR Workshops* [1][1] are 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, etc.

The aims of the NTCIR project are:

1. to encourage research in information access technologies by providing large-scale test collections 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.

That is to say, the main goal of the NTCIR project is to provide infrastructure of large-scale evaluation. The importance of large-scale evaluation infrastructure in IA research has been widely recognized. Fundamental text processing procedures for IA such as stemming and indexing include 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.

1.1 Information Access

The term "information access" (IA) includes a whole process to make information in the documents usable for the user who has problems or information needs. A traditional IR system returns a ranked list of retrieved documents that are likely to contain information relevant to the user's needs. This is one of the most fundamental and core processes of IA. It is however not the end of the story for the users. After obtaining a ranked list of retrieved documents, the user skims the documents, performs relevance judgments, locates the relevant information, reads, analyses, compares the contents with other documents, integrates, summarizes

[1] NTCIR-3 and 4 are sponsored by the National Institute of Informatics (NII) and *Japanese MEXT Grant-in-Aid for Scientific Research on Informatics (#13224087)* in and after FY2001. Patent task is organized by collaboration with Japan Intellectual Property Right Association and NII, and CLIR Task is organized by collaboration with National Taiwan University, Korean Institute for Scientific and Technological Information (KISTI).

and performs information-based work such as decision making, problem solving, writing, etc., based on the information obtained from the retrieved documents. We have looked at IA technologies to help users utilize the information in large-scale document collections. IR, summarization, question answering, etc are a "family", in which the same target is aimed while each of the technologies has been investigated by different communities with least interaction[2].

1.2 Focus of the *NTCIR*

As shown in Figure 1, 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 IR and CLIR with Japanese or other Asian languages and testing on various document genres. For the challenging issues, the targets are the 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 task in the experiment design. These two directions have been supported by a forum of researchers and discussion among them.

From the beginning, CLIR has been one of the central interests of the NTCIR, because CLIR between English and own-languages is critical for international information transfer in Asian countries, and it was challenging to perform CLIR between languages with completely different structures and origins such as English and Chinese or English and Japanese.

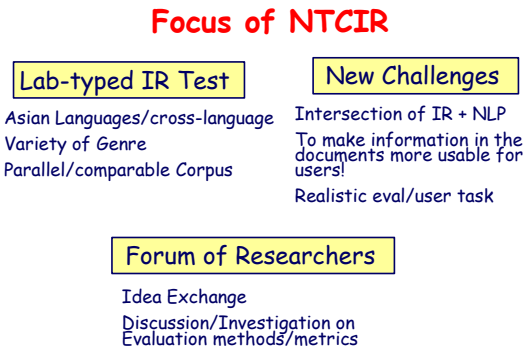

Figure 1. Focus of NTCIR

In the following, the next section provides a brief history of *NTCIR*. Section 3 describes *NTCIR Test Collections,* Section 4 reports recent progress after our reports at the previous CLFEs [2-4], and Section 5 outlines the features of the coming NTCIR Workshop, *NTCIR-4*. Section 6 is summary.

2 NTCIR

2.1 History of NTCIR

In the NTCIR, a workshop is held once per about one and half years. Since we respect the interaction between participants, we call a whole the process from document release to the final meeting as "workshop". Each workshop selects several research areas called "Task", or "Challenge" for more challenging task. Each task has been organized by the researchers of the domain and a task may consist of more than one subtasks. Figure 2 shows the evolution of the tasks in the NTCIR Workshops and Table 1 is a list of subtasks and test collections used in the tasks [5-7].

As shown in Table 1, the 4th NTCIR Workshop hosts 5 tasks, *CLIR,* Patent Retrieval Task (*PATENT),* Question Answering Challenge (*QAC),* Text Summarization Challenge (*TSC)*, and WEB Task (*WEB)* and their sub-tasks.

---

[2] In addition to the above, how to define the question of the user before the retrieval is also included in the scope of the IA although it has not been explicitly investigated in NTCIR.

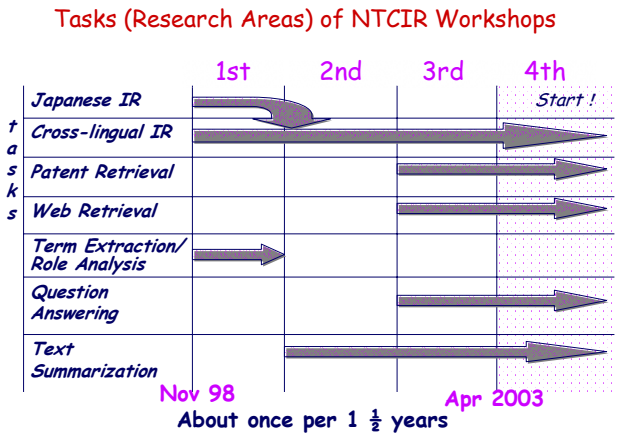

Figure 2. Tasks of NTCIR Workshops

Table 1. History of NTCIR Workshops

	Period	Tasks	Subtasks	Test collections
1	Nov.1998-Sept.1999	Ad Hoc IR	J-JE	NTCIR-1
		CLIR	J-E	
		Term Extraction	Term Extraction/ Role Analysis	
2	June 2000-March 2001	Chinese Text Retrieval	Chiniese IR: C-C	CIRB010
			CLIR: E-C	
		Japanese&English IR	Monolingual IR: J-J, E-E	NTCIR-1, -2
			CLIR: J-E, E-J, J-JE, E-JE	
		Text Summarization	Intrinsic - Extraction/Free generated	NTCIR-2Summ
			Extrinsic - IR task-based	
3	Oct. 2001-Oct. 2002	CLIR	Single Language IR:C-C,K-K,J-J	NTCIR-3CLIR
			Bilingual CLIR:x-J,x-C, x-K	
			Mulilingual CLIR:x-CJE	
		Patent	Cross Genre w/ or wo CLIR CCKE-J	NTCIR-3 PATENT
			[Optional] Alianment, RST Analysis of Claims	
		Question Answering	Subtask-1: Five Possible Answers	NTCIR-3QA
			Subtask-2: One Set of All the Answers	
			Subtask-3: Series of Questions	
		Text Summarization	Single Document Summarization	NTCIR-3 SUMM
			Multi-document Summarization	
		Web Retrieval	Survey Retrieval	NTCIR-3 WEB
			Target Retrieval	
			[Optional] Speech-Driven	
4	Apr. 2003 - June 2004	CLIR	Single Language IR:C-C,K-K,J-J	NTCIR-4CLIR
			Bilingual CLIR:x-J,x-C, x-K	
			Pivoted Bilingual CLIR	
			Mulilingual CLIR:x-CKJE	
		Patent	"Invalidity Search"= Search Patents by a Patent	NTCIR-4 PATENT
			[Feasibility] Automatic Patent Map Creation	
		Question Answering	Subtask-1: Five Possible Answers	NTCIR-4 QA
			Subtask-2: One Set of All the Answers	
			Subtask-3: Series of Questions	
		Text Summarization	Multi-document Summarization	NTCIR-4 SUMM
		Web Retrieval	Informational Retrieval	NTCIR-4 WEB
			Navigational Retrieval	
			[Pilot] Geographical Information	
			[Pilot] (Search Results) Topical Classification	

n-m: n=query language, m=document language(s), J:Japanese, E:English, C:Chinese, K:Korean, x
*: number of active participating groups that submitted task results

2.2 Participants

As shown in Figures 3 and 4, the number of participants has been gradually increasing. Different tasks attracted different research groups although many are overlapped, or changed the participating tasks over workshops. Many international participants were enrolled to CLIR. Patent Retrieval task attracted many participants from company research laboratories and "veteran" NTCIR participants. WEB task has participants from various research communities like machine learning, DBMS, and so on. The number of collaborating teams across different organizations is increasing in recent NTCIRs.

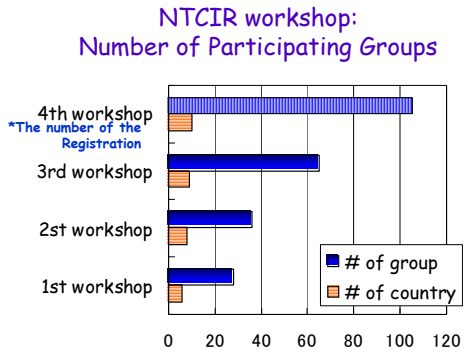

Fig 3 Number of Participating Groups

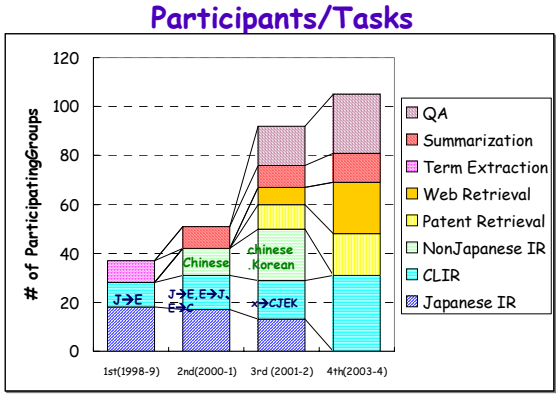

Fig 4 Participating Groups per Task

3 Test Collections

The test collections constructed for the NTCIR Workshops are listed in Table 2. In the NTCIR project the term "test collection" is used for any kind of data set usable for system testing and experiments although it often means IR test collections used in search experiments. One of our interests is to prepare realistic evaluation infrastructure, and those efforts include scaling up the document collection, document genres, languages, topic structure and relevance judgments.

3.1 Documents

Documents were collected from various domains or genres. Format of the documents are basically the same as TREC or CLEF and are plain text with SGML-like tags. Each of the specialized document genre collections contained characteristic fields for the genre – Web collection contains html tags, hyperlinks, URL of the document, etc., and patent collection has tags indicating document structure of patent, and 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 user community who use the type of documents in their everyday tasks.

3.2 Topics.

A sample topic record is shown in Figure. 5. Topics are defined as statements of "user's requests" rather than "queries", which are the strings actually submitted to the system, because we wish to allow both manual and automatic query construction from the topics. Emphasis has been shifted towards the topic structure capable more realistic experiments as well as to see the effect of background information of the topic. The characteristics are summarized as followings;

Topic Structure: Topic Structure has slightly changed in each NTCIR. A topic basically consists of a <TITLE>, a description <DESC>, and a detailed narrative <NARR> of the search request as similar to those used in CLEF and TREC. It may contain additional fields as shown in Table 3. Most of NTCIR collections contain a list of concepts <CONC>, but they are not heavily used by participants.

Table 2. Test collections constructed through NTCIR

NTCIR Test Collections; IR and QA

collection	task	documents						Task data: topic/ question		relevnce judge
		genre	filename	lang	year	# of doc	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	
	term extraction/ role analysis		ntc1-tmrc	J		2,000	-		-	-
CIRB010	IR	news	CIRB010	C_t	1998-1999	132220	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
		news	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/tex	NW100G-01	multiple *4	crawled in 2001	11,038,720	100GB	J*	47	4grades + relative
			NW10G-01			-	10GB			
NTCIR-4 CLIR	IR	news	CIRB011	Ct	1998-1999	132,173	ca.2.7GB	CtKJE	60	4 grades
			CIRB020			249,508				
			Hankookilbo +	K		220,078				
			Chosenilbo +			149,498				
			Mainichi	J		105,517				
			EIRB010	E		10,204				
			Mainichi Daily			12,723				
			Korea Times +			21,377				
			Hong Kong Standard +			96,856				
			Xinhua (AQUAINT) +			208,168				
NTCIR-4 PATENT	IR	patent full	Publication of unexamined patent	J	1993-2002	ca. 3,500,000	ca.45GB	CtCsKJE		
		abstract	Patnet 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 +			ca. 340,000				
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 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)

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						

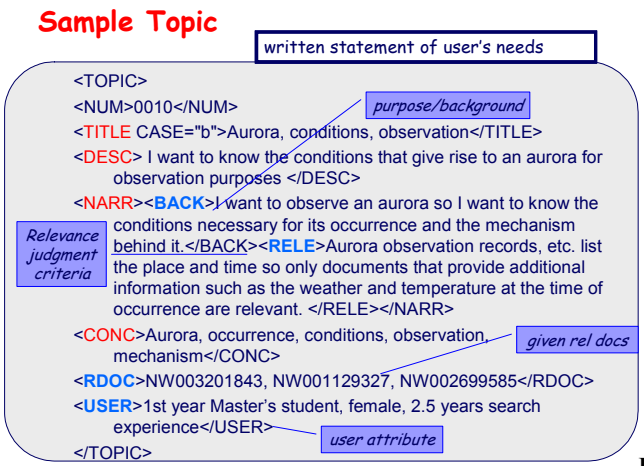

Figure 5. Sample topic (NTCIR-3 WEB)

Table 3 Topic fields

Topic Structure of NTCIR IR Test Collections

	NTCIR-1	NTCIR-2	CIRB010	NTCIR-3 CLIR	NTCIR-3 PATENT	NTCIR-3 WEB	NTCIR-4 CLIR	NTCIR-4 PATENT	NTCIR-4 WEB
Task	ad hoc, CLIR	ad hoc, CLIR	ad hoc, CLIR	CLIR	Cross-genre, CLIR	ad hoc	CLIR	invalidity	ad hoc, other
Mandatory Run *	D-only	D-only	N/A	D-only	S+A	T-only, D-only	T-only, D-only	CLAIM-only	T-only, D-only
Topic Field									
TITLE **	very short	very short	very short	very short	very short	query	query	very short	query
DESC	yes	yes	yes	yes	yes	yes	yes	yes	yes
NARR (unstructured	yes	yes	yes	yes	yes			yes	
NARR (structured)						yes	yes		yes
NARR. BACK *10						yes	yes		yes
NARR. RELE *10						yes	yes		yes
NARR. TERM *10						yes	yes		yes
PURPOSE *7								yes	
CONC	yes	yes	yes	yes	yes	yes	yes		
FIELDS	yes	yes							
TLANG / LANG *3				yes			yes		
SLANG *3				yes			yes		
RDOC *4					yes	yes			
PI *4									
USER *5						yes			yes
ARTICLE *6					yes				
DOC *9								yes	
SUPPLEMENT *6					yes				
CLAIM *8								yes	
COMP *8								yes	
COMP. CNUM *8								yes	

*: D-only=DESC only, T-only=TITLE only, A+S= run using ARTICLE and SUPPLEMENT only

**: "very short"=very short description of search request; "qeury"=comma separated term list

*3: TLANG/LANG=target language, the language of the topic; SLANG=source language, the language the topic originally constructed.

*4: RDOC=known relevant documents; PI=the patent for the invention mentioned in the news articles.

*5: USER=users' attribute

*6: ARTICLE=a news article reporting an invention; SUPPLEMENT=memorandam to focus the issues in the article relevant to the user's needs; if a human knowledgeable searcher reads ARTICLE and SUPPLEMENT, he/she understand the user's search request as specif

*7: Purpose of search (only "invalidity search" for NTCIR-4 PATENT)

*8: CLAIM=Target claim in the query patent. It was used as qeury of the search and may consists of multiple components; COMP=Component of a claim; CNUM=Claim component ID

*9: Query patent fulltext (fulltext of a patent that is used as a query of the search)

*10: BACK=Background knowledge/purpose of search; RELE=relevance judgment criteria; TERM=term definitions

<TITLE> as Query: A title is originally defined as a very short description, or "nickname" of the topic, and, since NTCIR-3 WEB[3], changed to be a "query", a string put into a search engine by users and defined as a comma-separated term lists up to three terms.

Structured <NARR>: Originally a narrative <NARR> was defined and instructed to the topic authors that it may contain background knowledge, purpose of the search, detailed explanation of the topic, criteria for relevance judgment, term definitions, etc. Since NTCIR-3 WEB, such information categories in <NARR> explicitly marked by tags like <BACK>, <RELE>, etc. as Figure 5. The purpose of this change is to examine the effect of additional information on the search effectiveness explicitly.

Mandatory runs: Any combination of topic fields is allowed to use in experiments for research purpose. In the Workshop, the *Mandatory Runs* are defined in each task, and every participant must submit at least one mandatory run using the specified topic field only. The purpose of this is to enhance the cross-system comparison based on the common condition and see the effectiveness of the additional information over it. Mandatory runs are originally "<DESC> only", then gradually shift to "<TITLE> only as well as <DESC> only".

3.3 Relevance Judgments

Relevance judgments are done by pooling, and the format and methods are basically the same as other evaluation projects including CLEF and TREC. The differences shall be summarized as follows;

1. Pooling strategies are slightly different according to each of the task
   - Additional interactive recall-oriented searches are done to improve the exhaustivity (NTCIR-1,-2) [8]
   - Additional interactive recall-oriented search are done by professional patent intermediaries (PATENT) [9]
   - "One-click distance model", in which hyperlinked documents are allowed to see in WEB [10]
   - Cross-lingual pooling for parallel or quasi-parallel documents (NTCIR-1,-2)[8]
   - Graded-depth pooling: pool creating top10, 11-20, 21-30, 31-41, (PATENT) [9]
2. Multi-grade and relative relevance judgments
   - Highly Relevant, Relevant, Partially Relevant [5-7], Irrelevant; Best Relevant, 2nd Best, 3rd Best, etc. [10]
3. Judgments includes the extracted passages to show the reason why the assessors assessed the documents as "relevant"
4. Pooled document lists to be judged are sorted in descending order of likelihood to be relevant (not the order of the document IDs)
5. Relevance judgment files may be prepared to each of the target language document sub-collections in CLIR

For 4, it helps assessors to judge consistently over a long list of pooled documents to be judged (typically 2000 - 3000 documents). Relevance judgments may change over assessors and over time. If relevant documents are appeared intensively in the first part of the list, it is easier for the non-professional assessors to set and confirm their criteria for relevance judgments, and then they can always refer those documents to re-confirm their own relevance judgment criteria when they go down to the lower ranked document. We understand they may be suffered by "order effect" of the ranked list of pooled documents in judgments, but we intentionally have used this strategy as practical and most effective one in *our* environment based on the comparative tests and interviews with assessors.

For 5, in multilingual CLIR, a topic can not always obtain sufficient number of relevant documents on every language document sub-collection, and this is the natural situation in multi-lingual CLIR. As a result, some topics can not be usable experiments on specific language documents. We can not find the way to manage this issue and only strategy we could take in NTCIR-4 CLIR is to increase the number of topics, so that larger number of topics can be used common across the document sub-collections and then improve the stableness of the evaluation.

Assessors are users of the document genre, judgments are done by the topic author except CLIR in NTCIR-3 and -4 since topics are created in cooperation of multiple countries, and then translated into each language and tested usability on each

---

[3] Topic authors are instructed to sort the terms in <TITLE> in descending order of importance to express the search request resembling the way

language document sub-collection. Judging other users' topics is sometimes hard for users and take longer time.

First two NTCIRs used two assessors per topic then tested inter-assessors consistency and found that the inconsistency among multiple assessors on a topic does not affect the stableness of the evaluation when tested on sufficient number of topics. Based on this, single assessor per topic is used in and after NTCIR-3.

3.4 Evaluation

For the evaluation, trec-eval program [11] is used by setting two threshold of the levels of relevance judgments, i.e. "*Rigid Relevance*" for "Relevant" or higher, "*Relaxed Relevant*" for "Partial Relevant" or higher ranked relevance for IR experiments. As additional metrics, several metrics for multi-grade relevance judgments are proposed including *weighted mean average precision* (wMAP), *weighted mean reciprocal rank* (wMRR, for WEB task), and used *decline cumulated gain* (DCG) [12-13].

For Question Answering, MRR is used for subtask-1, return 5 possible answers and no penalty for wrong answers, and F-measure for subtask-2, return one set of all the answers and penalty will be given for wrong answers, and subtask-3, series of question. For Text Summarization, content based and readability based intrinsic evaluation was done in NTCIR-3 for both single document and multi-document summarization, and proposed new evaluation methodology based on revision (edit distance) on system summaries by professional analysts who created the model summaries.

4. Further Analysis of NTCIR-3

After our previous reports at CLEFs [2-4] and the overview papers in the Proceedings of the NTCIR-3 [7], several additional analyses were done on the NTCIR-3 results and collection.

For *PATENT* retrieval task, though a new strategy for cross-genre retrieval called "term distillation" was proposed by Ricoh group and worked well on the collection, many research questions regarding patent retrieval were remained unsolved in NTCIR-3. The questions are, for example;

1. Is there any particular IR model (or weighting scheme) specifically effective on Patent?
2. Influence of the wide variation of document length (from 100 words to 30,000 word tokens in a document!)
3. Indexing (Character bi-gram *vs*. Word-based)
4. Target document collections: Fulltext *vs.* abstract (many commercialized systems used abstracts only)

For 1., it has been reported that ***tf*** is not effective on Patent at the *SIGIR 2000Workshop on Patent Retrieval,* but we could not find the concrete answers to the question through the NTCIR-3.

To answer these question, the NTCIR-3 Patent Task organizers conducted additional experiments on the patent collection and newspaper collection, and tested 8 different weighting schemes including both vector space as well as probabilistic models, on 6 different document collections, using 4 different indexing strategies, character bi-gram, word, compound terms, hybrid of character bi-gram and word; and 3 different topic length on a system. The results will be reported in [14].

For *WEB*, one participating group was consisted as a collaboration of research groups with strong background of content-based text retrieval and of web-link analysis, worked well at NTCIR-3 WEB. Further analysis on the effect of link on WEB collection, link-based approaches are generally worked well especially on the short queries like using TITLE only, or more specifically the first term of the TITLE, i.e. the most important terms for the users (topic authors) [15].

---

in which the users input the terms as queries. The relation between the terms is specified as an attribute of the <TITLE> in WEB Task.

5. Challenges at NTCIR-4

As shown in Table 1, the 4th NTCIR Workshop hosts 5 tasks, *CLIR, PATENT, QAC, TSC*, and *WEB* and their sub-tasks. Evaluation *schedule* varies according to each task.

April 2003: Document Release
June – September 2003: Dry Run
October – December 2003: Formal Run
20 February 2004: Evaluation Results Release
2-5 June 2004: Workshop Meeting at NII, Tokyo Japan

For the further information including late registration of the task participation, please consult NTCIR web sites at; http://research.nii.ac.jp/ntcir and http://research.nii.ac.jp/ntcir/ntc-ws4, or contact the author.

5.1 NTCIR-4CLIR

Since this is the second multilingual CLIR at NTCIR, the same task design will be continued from the previous one. Minor revision was made only to solve the major problems raised in the assessment on the NTCIR-3 as follows;

- Enlarge the English and Korea document collections comparable to Chinese and Japanese. 2.7GB in total.
- New sub-task of Pivot Language Bilingual CLIR
- Restrict the pair of topic and document languages, so that comparison will be done in fruitfully
- Set T-only run as mandatory as well as D-only run
- Question type – topics were categorized according to the nature and types of the answers in order to take a good balance of the topic set.

The new sub-task, pivot CLIR uses English as a Pivot language, then test the effectiveness of the transitive CLIR. It is one of the practical approaches of Multilingual CLIR in the environment with less availability of the direct translation resources but rich in those between each of the languages and English.

5.2 NTCIR-4 Question Answering (QAC) and Text Summarization (TSC)

QAC plans three subtasks as previous one at NTCIR-3. Among the three, subtask-1 and -2 will be done without major change. Only exceptions are; use different question sets for each of subtask-1 and -2, and increase the number of topics containing multiple answers. It was decided to avoid overestimate of the groups ignoring the possibility of multiple answers and returning the first priority answer only to the every question in subtask 2.

QAC subtask-3, answering to the series of question, is one of the major focus of the NTCIR-4 QAC. We plan to increase the number of sequence as well as task design aiming to tackle the problems resembling the real-world "Report Writing" task based on a set of relevant documents. The task design also related to the *TSC*, content-based evaluation of multi-document summarization will be done by set of questions. This is, more fundamentally, what kind of aspects of an event or topic that users want to know. Some of the questions may be more appropriate for the current factoid –oriented QA and others may covered by summarization. IR covered both and those focus of QAC and TSC has many intersection of the focus of the CLIR to see the categorization of question types.

5.3 Specialized Genre Related Tasks at NTCIR-4: Patent and WEB

Both PATENT and WEB plan (1) Main task(s) and (2) Feasibility or Pilot studies for more challenging tasks as follows;

PATENT- Main: Invalidity Task:

To search patents to invalidate the query patents. Claims of the query patents are used as query and they are segmented into components of the invention or technologies consisting of the investigation, then search related patents. A patent may be invalidated by one patent or by combination of multiple patents. Return document IDs as well as relevant passages.

PATENT - Feasibility: Long term research plan over NTCIR4-5. Automatic Patent Map Creation
A kind of Text mining -- Detect sets of technologies used in a set of patents, extract them, and make a table to show the relationship between technologies and patents, and evolution or trends among them.

WEB - Main: Informational Search and Navigation Oriented Search, in which find most informative and reliable page
WEB – Pilot: Geographical oriented and Topical Classification of the Search results

For the details, please visit the website of each task, which are linked from the NTCIR's main web site.

6. Summary

Brief history of NTCIR and recent progress after NTCIR-3 are reported in this paper. One of the characteristic features of the NTCIR is targeting "Information Access" technologies, in which a whole process for users to obtain and utilize the information in the documents are interested in and see the intersection between all the related technologies including IR, Summarization, QA, Text mining, etc., and treat them as like a "family". Other aspects are, for see the users' information task behind the laboratory-typed testing. We are in the process of the fourth-iteration in a series. Evaluation must be changed according to the technologies evolution and change of the social needs. We have been and are struggling for this. Collaboration and any leads and advices are always more than welcome.

References

1. NTCIR Project: http://research.nii.ac.jp/ntcir/.
2. Kando, N. "NTCIR Workshop: Japanese- and Chinese-English cross-lingual information retrieval and multi-grade relevance judgments ".In *Proceedings of the first Cross-Language Evaluation Forum (CLEF2000)*, Lisbon, Portugal, Sept.17-22, 2000 Springer, 2001, pp.24-33 (Lecture Notes in Computer Science; 2069)
3. Kando, N. " CLIR system evaluation at the second NTCIR workshop", In *Proceedings of the second Cross-Language Evaluation Forum (CLEF2001)* , Darmstadt, German, Sept 3-4, 2001, Springer, 2002, pp.371-388 (Lecture Notes in Computer Science; 2406)
4. Kando, N. "CLIR at NTCIR Workshop 3; Cross-Language and Cross-Genre Retrieval" In *Proceedings of 3rd Cross-Language Evaluation Forum (CLEF2002)*, Rome, Italy, Sept. 19-20, 2002 , Springer (Lecture Note in Computer Science) (to appear)
5. *NTCIR Workshop 1: Proceedings of the First NTCIR Workshop on Research in Japanese Text Retrieval and Term Recognition*. Tokyo Japan, 30 Aug.–1 Sept., 1999. ISBN 4-924600-77-6. (http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings/).
6. *NTCIR Workshop 2: Proceedings of the Second NTCIR Workshop on Research in Chinese and Japanese Text Retrieval and Text Summarization*. Tokyo Japan, June 2000 – March 2001. ISBN 4-924600-96-2.
7. *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 .
8. Kuriyama, K., Yoshioka, M., Kando, N. "Construction of a Large Scale Test Collection NTCIR-2: The Effect of Additional Interactive Search and Cross-Lingual Pooling", *IPSJ Transactions on Databases*, Vol.43, No. SIG2 (TOD13), pp.48-59, March 2002.(in Japanese)
9. Iwayama, M., Fujii, A., Kando, N., Takano, A. "Overview of Patent Retrieval Task at NTCIR-3". In *NTCIR Workshop 3: Proceedings of the Third NTCIR Workshop on Research in Information Retrieval, Question Answering and Summarization*, Tokyo Japan, Oct. 2001 – Oct. 2002 . (to appear).
10. Eguchi, K., Oyama, K., Ishida, E., Kando, N., Kuriyama, K. "Overview of Web Retrieval Task at the Third NTCIR Workshop", In *NTCIR Workshop 3: Proceedings of the Third NTCIR Workshop on Research in Information Retrieval, Question Answering and Summarization*, Tokyo Japan, Oct. 2001 – Oct. 2002 . (to appear).
11. Buckley, C. trec-eval IR evaluation package. Available from ftp://ftp.cs.cornell.edu/pub/smart.
12. Jarvelin, K., Kekalainen, J. "IR evaluation methods for retrieving highly relevant documents". In: *Proceedings of the 23rd Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*. Athens Greece, July 2000, 41–48.
13. Voorhees, E.M. "Evaluation by Highly Relevant Documents", in *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001)*, New Orleans, Sept. 2001, pp. 74-82.
14. Iwayama, M., Fujii, A., Kando, N., Marukawa, K. "Empirical study on retrieval models for different document genres: Patents and newspaper articles. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003)*, Toronto, Canada, July 2003, (to appear)
15. Eguchi, K., Oyama, K., Ishida, E., Kando, N., and Kuriyama K., "Evaluation Methods for Web Retrieval Tasks Considering Hyperlink Structure", *IEICE Transaction on Information and Systems*. (Sep. 2003, in Japanese, to appear).