Overview of NTCIR-9

Tetsuya Sakai
Microsoft Research Asia
No.5 Danling Street, Haidian District, Beijing, 10080 P.R.China
tetsuyasakai@acm.org

Hideo Joho
Faculty of Library, Information and Media Science, University of Tsukuba
1-2 Kasuga, Tsukuba, Ibaraki 305-8550 JAPAN
hideo@slis.tsukuba.ac.jp

ABSTRACT

This is an overview of NTCIR-9, the Ninth NTCIR Workshop. It touches upon a brief history of NTCIR, introduces the tasks run at NTCIR-9, and reports on some statistics across these tasks. For details on the individual tasks, we refer the reader to the task overview papers and the participants' papers.

Keywords: evaluation, information access, information retrieval, natural language processing, test collection

1. INTRODUCTION

This is an overview of NTCIR-9, the Ninth NTCIR[1] Workshop. It touches upon a brief history of NTCIR, introduces the tasks run at NTCIR-9, and reports on some statistics across these tasks. "NTCIR" initially stood for *NACSIS/NII Test Collections for Information Retrieval systems*[2], but recently it has been given new interpretations, including *NII Testbeds and Community for Information Access research*. Thus, NTCIR aims to advance the state-of-the-art in information access research and its evaluation methodology by providing shared experimental environments and thereby forming an active, interdisciplinary research community.

NTCIR is a sesquiannual evaluation workshop, and each round of NTCIR has hosted several information access *tasks*. At NTCIR-9, in addition to the second round of GeoTime (GeoTemporal Information Retrieval)[3], we had six new tasks: INTENT[4] with 1CLICK (One Click Access)[5], CrossLink (Crosslingual Link Discovery)[6], SpokenDoc (IR for Spoken Documents)[7], RITE (Recognizing Inference in Text)[8], PatentMT (Patent Machine Translation)[9] and *VixEx* (Interactive Visual Exploration)[10].

Tables 1 and 2 show a complete list of active NTCIR-9 participants (i.e. those who actually participated in at least one of the tasks). Table 3 shows the number of teams by country: as it shows, there are 90 different teams in total, from 14 countries[11].

The remainder of this paper is organised as follows. Section 2 provides a brief summary of previous NTCIRs and some statistics on the NTCIR-9 tasks. Sections 3-9 very briefly discuss the NTCIR-9 tasks. For more details on the individual tasks, we refer the reader to the task overview papers and the participants' papers. All NTCIR papers are available online[12].

Table 3: Number of participating teams by country/region.

country/region	#teams
Japan	34
China	23
Taiwan	9
USA	6
Korea	4
Germany	3
Australia	2
Spain	2
UK	2
Canada	1
France	1
India	1
Ireland	1
Portugal	1
TOTAL	90

2. TASKS: PAST AND PRESENT

Figure 1 provides a summary of the tasks and the number of participating teams for all NTCIRs. The NTCIR-9 tasks are shown in red. Although all tasks except GeoTime are new at NTCIR-9, some of them are highly related to previous NTCIR tasks. For example, INTENT uses a Web corpus just like the previous WEB tasks, but focusses on recent challenges including search result diversification; PatentMT is similar to PATMT but handles one additional language (Chinese), puts an emphasis on manual evaluation and is run by a new group of organisers; VisEx inherits legacies from MuST and ACLIA. Moreover, GeoTime is actually a CLIR

[1] http://research.nii.ac.jp/ntcir/index-en.html
[2] NACSIS (National Center for Science Information System) was renamed to NII (National Institute of Informatics) in 2000.
[3] http://metadata.berkeley.edu/NTCIR-GeoTime/
[4] http://www.thuir.org/intent/ntcir9/
[5] http://research.microsoft.com/en-us/people/tesakai/1click.aspx
[6] http://ntcir.nii.ac.jp/CrossLink/
[7] http://www.cl.ics.tut.ac.jp/~sdpwg/index.php?ntcir9
[8] http://artigas.lti.cs.cmu.edu/rite/Main_Page
[9] http://ntcir.nii.ac.jp/PatentMT/. This is not a continuation of the PATMT task run at NTCIR-7 and -8.
[10] http://must.c.u-tokyo.ac.jp/visex/hiki.cgi?FrontE
[11] If a team comprises research teams from multiple countries, the country of the team's principal investigator is considered here.
[12] http://research.nii.ac.jp/ntcir/publication1-en.html

NTCIR-9 Workshop Meeting, 2011, Tokyo, Japan.

Table 1: NTCIR-9 participating teams (Part I).

team ID	organisation	country/region
AKBL	Akiba Laboratory, Toyohashi University of Technology	Japan
ALPS	ALPS Lab. at University of Yamanashi	Japan
ASR	Team ASR, Gifu University	Japan
BBN	Raytheon BBN Technologies	USA
BJTUX	Beijing Jiaotong University	China
BRKLY	University of California, Berkeley	USA
BUAA	Institute of Intelligent Information Processing, Beihang University	China
DBIIR	Information School, Renmin University of China	China
DCU	Dublin City University	Ireland
DUIIS	Daegu University	Korea
EIWA	Yamanashi Eiwa College	Japan
FudanNLP	Fudan University	China
FRDC	Fujitsu R&D Center CO., LTD	China
FX	Fuji Xerox	Japan
GETUA	University of Alicante	Spain
HIT2jointNLPLab	Heilongjiang Institute of Technology / Harbin Institute of Technology	China
HITIR	Research Center for Information Retrieval, Harbin Institute of Technology	China
HITS	Heidelberg Institute for Theoretical Studies	Germany
HU-KB	Hokkaido University	Japan
IASLD,nthuisa	Academia Sinica	Taiwan
IBM	IBM Research - Tokyo / Preferred Infrastructure	Japan
IBM	IBM Research	USA
ICL	Key Laboratory of Computational linguistics, Peking University / Ministry of Education	China
ICRC_HITSZ	Intelligence Computing Research Center, Harbin Institute of Technology Shenzhen Graduate School	China
ICT	Institute of Computing Technology, Chinese Academy of Sciences	China
ICTIR	Institute of Computing Technology (Information Retrieval Group), Chinese Academy of Sciences	China
IDEAS, III_CYUT_NTHU	Institute for Information Industry / Chaoyang University of Technology / National Tsing Hua University	Taiwan
IISR	Yuan Ze University	Taiwan
IMTKU	Information Management, Tamkang University	Taiwan
INESC	National Institute of Electroniques & Computer Systems	Portugal
IRNLP	Korea Advanced Inst. for Science & Technology	Korea
ISCAS	Institute of Software, Chinese Academy of Sciences	China
ISTIC	Institute of Scientific and Technical Information of China	China
IWAPU	Iwate Prefectural University	Japan
JAIST	Japan Advanced Institute of Science and Technology	Japan
JAPIO	Japan Patent Information Organization	Japan
JLTKB	University of Tsukuba	Japan
JUCS	Jadavpur University, Computer Sc. & Engineering	India
KECIR	Shenyang Aerospace University	China
KLE	Pohang University of Science and Technology (POSTECH)	Korea
KMI	The Open University	UK
KOLIS	Keio University, Library Science	Japan
kslab_nut	Nagaoka University of Technology	Japan

Table 2: NTCIR-9 participating teams (Part II).

team ID	organisation	country/region
KSLP	Kyungsung University	Korea
KUIDL	Kyoto University (Tanaka Laboratory)	Japan
KUTC	Kansai University	Japan
KYOTO,KyotoUniversity	Kyoto University (Kurohashi Laboratory)	Japan
LIUM	University of Le Mans	France
LTI	Language Technologies Institute, Carnegie Mellon University	USA
MCU	Ming-Chuan University	Taiwan
MSINT	Microsoft Research Asia	China
MSRA1click	Microsoft Research Asia (Virtual Sakai Lab)	China
NAIST	Nara Institute of Science and Technology	Japan
NAK	Keio University, Science and Technology	Japan
NCW	NTNU / NCCU / WebGenie Information Ltd.	Taiwan
NEU	Northeastern University, China	China
NKGW	NAKAGAWA LAB, Toyohashi University of Technology	Japan
NKI11	NKI-Lab Toyohashi University of Technology	Japan
NSNG	Northeastern University, USA / Wuhan University	USA
NTOU	National Taiwan Ocean University	Taiwan
NTTCS	Nippon Telegraph and Telephone Corporation	Japan
NTTUT	NTT Communication Science Labs. / the University of Tokyo	Japan
NTU	National Taiwan University	Taiwan
OKSAT	Osaka Kyoiku University	Japan
ORG/WSE	INTENT Organisers (Microsoft Research Asia)	China
QUT	Queensland University of Technology	Australia
RMIT	RMIT University	Australia
RWTH	RWTH Aachen University	Germany
RYSDT	Ryukoku NL-SLP Lab, Ryukoku University	Japan
SINAI	University of Jaén	Spain
SITLP	Shibaura Institute of Technology LP lab	Japan
SJTUB,SJTUBCMI	Center for Brain-like Computing and Machine Intelligence, Shanghai Jiao Tong University	China
TBFD	Team Big Four Dragons: Daido University / The University of Tokushima / Nagoya University	Japan
THU,THUIR	Tsinghua University	China
TORI	Tottori University	Japan
TOTLA	Tokyo Metropolitan University	Japan
TTOKU	Tokyo Institute of Technology	Japan
TU	Tohoku University	Japan
TUTA1	University of Tokushima	Japan
UIOWA	University of Iowa	USA
UKP	TU Darmstadt	Germany
uogTr	University of Glasgow	UK
UOTTS	The University of Tokyo (Tsujii Laboratory)	Japan
UTLIS	Language Information Sciences, The University of Tokyo	Japan
UWat	University of Waterloo	Canada
WHUTE	Wuhan University	China
WUST	Wuhan University of Science and Technolog	China
YLAB	Yamashita Laboratory, Ritsumeikan University	Japan
Yuntech	National Yunlin University of Science and Technology	Taiwan
ZSWSL	Beijing University of Posts and Telecommunications	China

Table 4: Number of runs submitted to each task.

task	#runs
RITE	212
PatentMT	130
INTENT/1CLICK	106
GeoTime	68
CrossLink	57
SpokenDoc	39
VisEx	8
TOTAL	620

task that specialises in WHEN and WHERE queries. The number of unique participating teams is shown for each task in the figure.

Figure 2 shows the number of participating groups for each task by country. Here, note that the TOTAL column counts the same team multiple times if the teams participated in multiple tasks, to represent how "active" each country is at NTCIR-9. Hence the grand total shown is 102 rather than 90, the number of unique teams at NTCIR-9.

Figure 3 shows the languages covered by each task. Crosslingual tasks are indicated in the cells: for example, "EJ" means English to Japanese crosslingual retrieval for machine translation.

Table 4 shows the number of runs submitted to each task.

The following sections briefly describe each task in turn.

3. GEOTIME

As was shown in Figure 1, GeoTime is the only second-round task at NTCIR-9. This is an ad hoc and crosslingual IR task that concerns queries with WHEN and/or WHERE components, e.g. "When and where were the last three Winter Olympics held?". The task covers English and Japanese[13], and the target documents are newspaper articles. The target document collections are larger than those of NTCIR-8 GeoTime. For more details on GeoTime, please refer to the overview paper [2].

4. INTENT AND 1CLICK

The INTENT task and the 1CLICK (One Click Access) task were proposed as separate tasks, but the NTCIR-9 Programme Committee accepted 1CLICK as a pilot subtask of INTENT. The overview of 1CLICK is given separately from that of INTENT.

The INTENT task has two subtasks:

Subtopic Mining Given a query, return a ranked list of possible subtopics strings (e.g. given "apple", return "apple ipad", "apple ceo", "apple granny smith" etc.).

Document Ranking Given a query, return a selectively diversified Web search result.

The Document Ranking task is similar to the diversity task of the TREC Web track[14], but differs in the following aspects:

- The probability of each intent given the query is considered;
- Per-intent graded relevance is considered; and
- New evaluation metrics [7] are used to handle the above two extensions.

[13] A Korean language subtask had also been planned but was eventually canceled.

[14] http://plg.uwaterloo.ca/~trecweb/2010.html

Both of the above tasks covered Chinese and Japanese (queries and Web documents).

The 1CLICK task requires the system to return a single textual output that presents important pieces of information first and to try to minimise the amount of text the user has to read. In contrast to traditional Web search where the user has to scan the ranked list and read several long Web pages to gather information, 1CLICK aims to satisfy the user immediately after the search button click. A recently-proposed, nugget-based evaluation methodology is used [5]. This year, only Japanese queries were considered.

For more details on INTENT and 1CLICK, please refer to the overview papers [6, 9].

5. CROSSLINK

CrossLink (Crosslingual Link Discovery) is a new task that aims to provide bridges across hyperlinked documents in multiple languages. Wikipedia comprises multilingual hypertext documents, but many of the links are confined within one language. Thus, CrossLink requires systems to automatically identify anchor texts within an input text together with some destination documents written in a different language. Using such a technology, a Chinese person reading an English Wikipedia page, for example, could click on an English phrase to jump to a Chinese explanation page for that phrase. This task can be regarded as an extension of the INEX 2010 Link-The-Wiki track[15] which did not have a crosslingual component.

At NTCIR-9, CrossLink has three subtasks. The input to a CrossLink system is a set of English Wikipedia articles, from which it is expected to extract prospective anchor texts together with outgoing links to Wikipedia pages in a foreign language. The three subtasks are:

E2C Provide links to Chinese Wikipedia pages;

E2J Provide links to Japanese Wikipedia pages;

E2K Provide links to Korean Wikipedia pages.

The evaluation is done at two levels: *file-to-file* which is based on whether an English page should be linked to a Chinese/Japanese/Korean page or not; and *anchor-to-file* which is based on whether the anchor text is relevant to the English topic *and* the link is appropriate. Average precision and other information retrieval metrics are used for computing the effectiveness.

For more details on CrossLink, please refer to the overview paper [10].

6. SPOKENDOC

SpokenDoc (IR for Spoken Documents) is a new task that tackles the problem of information retrieval from spontaneous speech data for which the word error rate and the out-of-vocabulary rate of automatic speech recognition may be very high. At NTCIR-9, SpokenDoc has two tasks targetting Japanese spoken lectures:

Spoken Term Detection Given a query, locate the occurrences of the query terms within the spoken documents.

Spoken Document Retrieval Given a query, retrieve a ranked list of documents (lectures) or passages from the spoken document corpus.

[15] http://www.inex.otago.ac.nz/tracks/wiki-link/wiki-link.asp

		1999	2001	2002	2004	2005	2007	2008	2010	2011
		NTCIR-1	NTCIR-2	NTCIR-3	NTCIR-4	NTCIR-5	NTCIR-6	NTCIR-7	NTCIR-8	NTCIR-9
TMREC	Automatic Term Recognition and Role Analysis	9								
Ad hoc/Crosslingual IR(1)> Chinese/English/Japanese IR(2)->CLIR(3-6)	Crosslingual IR	28	30	20	26	25	22			
TSC	Text Summarization Challenge		9	8	9					
WEB	Web Retrieval			7	11	7				
QAC	Question Answering Challenge			16	18	7	8			
PATENT	Patent Retrieval (and Classification)			10	10	13	12			
MuST	Multimodal Summarization for Trend Information					13	15	13		
Opinion(6)->MOAT(7,8)	(Multilingual) Opinon Analysis						12	21	16	
CLQA(5,6)->CCLQA@ACLIA(7,8)	(Complex) Crosslingual Question Answering					14	12	9	6	
IR4QA@ACLIA	IR for Question Answering							12	12	
CQA	Community Question Answering								4	
PAT-MN	Patent Mining							12	11	
PAT-MT(7,8)->PatentMT(9)	Patent Translation							15	8	21
GeoTime	Geotemporal IR								13	12
INTENT/1CLICK	Intent/One Click Access									20
VisEx	Interactive Visual Exploration									4
RITE	Recognizing Inference in Text									24
CrossLink	Crosslingual Link Discovery									11
SpokenDoc	IR for Spoken Documents									10

Figure 1: Tasks from NTCIR1-9 and the number of participating teams. For the previous NTCIRs, the statistics were obtained by reading the overview papers, but not all of them may be completely accurate as not all overviews explictly give the number of unique teams across all subtasks.

Participants can choose whether to use the reference automatic transcriptions provided by the organisers or to use their own speech recognition module.

Some variants of information retrieval metrics such as F1-measure and Average Precision are computed to evaluate these tasks, for evaluating based on *inter-posal units* (atomic segments within spoken documents) as well as passages that contain these units.

For more details on SpokenDoc, please refer to the overview paper [1].

7. RITE

RITE (Recognising Inference in Text) a new natural language processing task at NTCIR-9 which expects systems to recognise the semantic relationship between a given pair of texts containing statements. Such a deep understanding of texts will be useful for various information access applications that go beyond exact matches of bag of words. RITE is similar to the Recognising Textual Entailment Challenge at TAC (Text Analysis Conference)[16] but covers multiple languages (Japanese, Simplified and Traditional Chinese) as well as multiple types of relation between text pairs.

RITE has four subtasks:

BC (binary-class) Given a text pair (t_1, t_2), determine whether t_1 *entails* t_2. For example, t_1: "Yasunari Kawabata won the Nobel Prize in Literature for his novel Snow Country" entails t_2: "Yasunari Kawabata is the author of Snow Country" (English examples are given here for convenience).

MC (multi-class) Given (t_1, t_2), classify the relation between the two texts into one of the following: **F** (forward entailment), **R** (reverse entailment), **B** (bidirectional entailment), **C** (contradiction) or **I** (independence).

Entrance Exam Similar to the BC subtask, but the text pairs are from actual Japanese college-level entrance exams.

RITE4QA Similar to the BC subtask, but the text pairs are from Question Answering data from past NTCIR, where t_1 is an answer-candidate-bearing sentence (e.g. See the above Nobel Prize example) and t_2 is a sentence reconstructed using the original question and its answer (e.g. "Who wrote Snow Country?" would be transformed into "Yasunari Kawabata wrote Snow Country").

For all subtasks, the accuracy of the relation labels are evaluated. For more details on RITE, please refer to the overview paper [8].

8. PATENTMT

The task specifications of PatentMT (Patent Machine Translation) are similar to those of PATMT at NTCIR-8, but the language scope has been extended to handle Chinese and the main focus is

[16] http://www.nist.gov/tac/2010/RTE/index.html

	RITE	PatentMT	INTENT/1CLICK	GeoTime	CrossLink	SpokenDoc	VisEx	TOTAL
Japan	7	7	4	4	1	9	4	36
China	6	7	10	1	2			26
Taiwan	7	2	2		2			13
USA	3	2		2				7
Korea		1	1	1	2			5
Germany		1			2			3
Australia			1	1	1			3
Spain				2				2
UK			1		1			2
Canada			1					1
France		1						1
India	1							1
Ireland						1		1
Portugal				1				1
TOTAL	24	21	20	12	11	10	4	102

Figure 2: NTCIR-9 country/region by task matrix

	RITE	Patent MT	INTENT	1CLICK	GeoTime	CrossLink	Spoken Doc	VisEx
Chinese (simplified)		CE						
Chinese (traditional)								
English		EJ			EJ	EC, EJ, EK		Event collection only
Japanese		JE			JE			
Korean								

Figure 3: Languages covered by each NTCIR-9 task. Grey cells mean "not covered."

manual evaluation from the viewpoint of adequacy and acceptability. In this task, adequacy is manually determined based on relative comparison between systems with consideration of clause-level meanings and acceptability is manually determined based on several if-then rules that test properties of the machine translation output such as completeness, readability and grammatical correctness. Standard automatic evaluation methods for machine translation are also utilised.

The subtasks of PatentMT are:

C-E Translate each Chinese patent sentence into English;

J-E Translate each Japanese patent sentence into English;

E-J Translate each English patent sentence into Japanese.

Parallel patent sentence corpora are provided for training. Also, participating systems can optionally utilise the contexts of the test sentences rather than take each input sentence as an isolated text.

For more details on PatentMT, please refer to the overview paper [3].

9. VISEX

The Interactive Visual Exploration (VisEx) task is a pilot task that aimed to establish an evaluation framework of interactive and explorative information access environments. The task organizer provided a common framework for explorative information access environment systems (IAESs) which provided a backend retrieval engine and its API, logging system, note-taking function, etc.

Task participants were required to develop an interactive component to support either (or both) of the two subtasks: Event Collection and Trend Summarization. The former subtask asked subjects to make a report of events specified as a topic by collecting their characteristics such as times and places of their occurrences, while the latter subtask asked to make a report on summarization of the trend related to time-series statistical information given as a topic by collecting not only values and changes of the statistics but also their reasons of changes and influences.

Four teams participated in the task and a total of six runs was submitted for the two subtasks. With two additional runs of the baseline system (provided by the task organizer), 40 subjects were recruited in a user study where each system was tested by five subjects. Each subject carried out a set of four topics based on either of the subtasks. The performance measures of participated systems looked at different aspects such as task performance, search behavior, and subjective assessments.

For more details on VisEx, please refer to the overview paper [4].

10. SUMMARY

This paper serves as an overview of the NTCIR-9 tasks which involved 90 different research teams from 14 different countries/regions. Based on the experiences and the analyses of the NTCIR-9 tasks as well as new task proposals for NTCIR-10, we hope that the NTCIR community will identify open and important problems for information access and continue to tackle them by balancing between competition and collaboration.

11. ACKNOWLEDGMENTS

As the Program Chairs of NTCIR-9, we thank the NTCIR-9 General Chairs for their advice and support. Noriko Kando, in particular, has always been our greatest supporter. We also thank the NTCIR secretaries, above all Miho Sugimoto, who is always there to support the chairs.

We thank the NTCIR-9 task organisers for their tremendous efforts and their leadership. We also thank the participants for their interesting work and their willingness to collaborate.

Last but not least: we thank the following people who served in the NTCIR-9 Programme Committee to help us choose such exciting tasks from the submitted proposals: Charlie Clarke, Fred Gey, Gareth Jones, Noriko Kando, Sung-Hyon Myaeng, Stephen Robertson and Eiichiro Sumita.

12. REFERENCES

[1] T. Akiba, H. Nishizaki, K. Aikawa, T. Kawahara and T. Matsui. Overview of the IR for Spoken Documents Task in NTCIR-9 Workshop. In *Proceedings. of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologiesss*, to appear, 2011.

[2] F. Gey, R. Larson, J. Machado and M. Yoshioka. NTCIR9-GeoTime Overview – Evaluating Geographic and Temporal Search: Round 2. In *Proceedings. of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologiess*, to appear, 2011.

[3] I. Goto, B. Lu, K. P. Chow, E. Sumita and B. K. Tsou. Overview of the Patent Machine Translation Task at the NTCIR-9 Workshop. In *Proceedings. of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologiess*, to appear, 2011.

[4] T. Kato, M. Matsushita and H. Joho. Overview of the VisEx task at NTCIR-9. In *Proceedings. of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologiess*, to appear, 2011.

[5] T. Sakai, M. P. Kato, Y. -I. Song. Click the Search Button and Be Happy: Evaluating Direct and Immediate Information Access. In *Proceedings of ACM CIKM 2011*, pages 621–630, 2011.

[6] T. Sakai, M. P. Kato, Y. -I. Song. Overview of NTCIR-9 1CLICK. In *Proceedings. of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, to appear, 2011.

[7] T. Sakai and R. Song. Evaluating Diversified Search Results Using Per-Intent Graded Relevance. In *Proceedings of ACM SIGIR 2011*, pp.1043-1052, 2011.

[8] H. Shima, H. Kanayama, C.-W. Lee, C.-J. Lin, T. Mitamura, Y. Miyao, S. Shi, and K. Takeda. Overview of NTCIR-9 RITE: Recognizing Inference in TExt. In *Proceedings. of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, to appear, 2011.

[9] R. Song, M. Zhang, T. Sakai, M. P. Kato, Y. Liu, M. Sugimoto, Q. Wang, N. Orii. Overview of the NTCIR-9 INTENT Task. In *Proceedings. of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, to appear, 2011.

[10] L. X. Tang, S. Geva, A. Trotman, Y. Xu and K. Y. Itakura. Overview of the NTCIR-9 Crosslink Task: Cross-lingual Link Discovery. In *Proceedings. of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologiess*, to appear, 2011.