# Preface

The *Seventh NTCIR Workshop* (NTCIR-7) is the latest in the evaluation workshop series designed to enhance research in information access technologies, including text retrieval, cross-language information access, question-answering, information extraction, text mining, etc., using texts in East Asian-languages such as Japanese, Chinese, and Korean and English. This is a community-based effort building a research infrastructure for evaluation and testing information access technologies.

The goals of the NTCIR Workshops are:

1.  Encourage research in information access technologies by providing research infrastructures for evaluation and testing,
2.  Provide a forum for research groups for exchanging research ideas, and
3.  Investigate methodologies and metrics for evaluation of information access technologies.

The term "information access" includes the whole process of making information in documents usable by users. It includes document retrieval from vast collections of documents, and the technologies to support users in utilizing information in the retrieved documents. Research and development of information access technologies require solid evidences based on experiments to show the superiority of newly proposed mechanisms over previous ones. A test collection is a data set used for such testing and a major component of the infrastructure. NTCIR Workshops aim to serve for those areas and wide-ranging insights have been produced as results.

NTCIR-7 Program Committee reviewed task proposals and suggested to merge related ones. The following research areas were selected as "tasks", and organized into the clusters.

> Cluster 1: Advanced Cross-Lingual Information Access (ACLIA)
> > Complex Cross-Lingual Question Answering (CCLQA) task
> > Information Retrieval for Question Answering (IR4QA) task
>
> Cluster 2: User Generated Contents
> > Multilingual Opinion Analysis Task (MOAT)
>
> Cluster 3: Focused Domain (Patent)
> > Patent Mining (PAT MN) task
> > Patent Translation (PAT MT) task
>
> Cluster 4: Multimodal Summarization of Trend (MuST)

Each task was coordinated and organized by the *task organizers*, a group of researchers of the topic. Each cluster originally planned to include mono- and cross-lingual information retrieval (CLIR) components[1], and designed module-based experiments in a cluster. Participants are welcome to participate in end-to-end evaluation in a cluster as well as evaluation in particular task/module in the cluster.

NTCIR-7 Research Planning and Coordinating Committee consisted of the coordinators of each cluster and a liaison from the program committee, and has led the activities for NTCIR-7. The call-for-task-participation in NTCIR-7 started in September 2007, and the final meeting is held on 16-19 December 2008, at NII, Tokyo. We consider the *workshop* to be the whole process from the call for participation in the experiments and evaluation of the tasks set by NTCIR to the final meeting.

A hundred and eighty-seven research groups registered and conducted the tasks set by the organizers, but unfortunately many groups could not continue their participation for various reasons. Eighty-two research groups from fifteen different countries and areas completed the experiments in time and remained as the "active participants". We plan to survey the reasons why so many gave up, and hope to learn from this experience to

---

[1] CLIR on Blog task in the Cluster 2 was cancelled, so there was no retrieval component for Cluster 2 outside of the optional sentence-level topic relevance subtask.

improve future NTCIR Evaluation Workshops. One piece of good news for NTCIR was that we have a large number of new participants.

The Second International Workshop on Evaluating Information Access (EVIA2008) is organized by two program co-chairs, Mark Sanderson and Tetsuya Sakai, as a refereed workshop that welcomes submissions from both the NTCIR task participants and individuals not involved in the NTCIR tasks. It contributed to enlarge the NTCIR community and we expect it to provide a wider forum to discuss related topics and research leading to future NTCIR tasks.

This volume contains the reports of the experiments and investigations performed within NTCIR-7, and the papers presented at EVIA are in a separate volume.
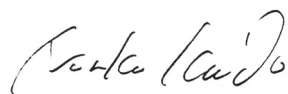
Technologies and social networking environments are rapidly changing and gaining enhanced functionality. It is obvious that the Web has become a valuable information resource, and the importance of information retrieval and access technologies has greatly increased to improve the quality of life for individuals; how to retrieve, how to create intelligence from the retrieved information, how to "weave out" the knowledge from such a vast collection of information, and how to utilize them – the technologies to support these activities are keys for a better society. Although NTCIR has worked on the aspects without interaction with users, which has provided a more reliable, sensible and stable evaluation and testing environment that has sped up technological transfer from research ideas into real-world applications, we have to tackle more complicated tasks to evaluate interactive and exploratory search which is critical to the process of information retrieval and access.

NTCIR-7 and EVIA 2008 are sponsored by the National Institute of Informatics (NII). CCLQA is partially supported by NII Joint Research Grant. PAT MT is partially supported by National Institute of Information and Communications Technology and NII Joint Research Grant. MuST is supported by joint Research Grant of Nippon Telegraph and Telephone, Corp., the University of Tokyo, and a NII Joint Research Grant.

We are extremely grateful to the following organizations for their understanding of the project and for providing substantial support in various forms: Central Daily News, China Daily News, China Times Inc., Industrial Property Cooperation Center, Japan Intellectual Property Association, Japan Parent Office, Japan Patent Information Organization, Korea Economic Daily, Linguistic Data Consortium, Mainichi Newspapers, Nippon Database Kaihatsu, Co. Ltd., the Sing Tao Group, Singapore Press Holding, Taiwan News, UDN.COM, and Yahoo! Japan. We are thankful to the Program Committee members, Advisory Committee members, the NTCIR-7 Research Planning and Coordinating Committee members, and the EVIA Program Committee members. We are thankful to Ellen Voorhees and Donna Harman, the National Institute of Standards and Technology, USA, and to Carol Peters and CLEF, the Cross-Language Evaluation Forum, for their support, advises and collaboration on various occasions. I thank for the NTCIR-7 program committee members, cluster coordinators and the task organizers for their hard work. Thanks for the task participants for scientific input and cooperation for the task organization. Especially I would like to draw attention that many of the task participants volunteered to join the dataset creation. I appreciated their great contribution and found that would be promising direction to the things put forwards in this community-based activity. I would like to thank Miho Sugimoto, Tetsuya Sakai, and Molin Du for information logistics for both printed and online proceedings and for the meeting, and Ruiko Homma, Miho Sugimoto, and Fumiko Koiwahara for their supports behind for entire process of NTCIR-7.

I do hope each of the participants will enjoy presenting their research results and discussing with each other at the Meeting, and such an occasion will lead to fruitful collaboration and research enhancement.

December 2008

Noriko Kando
Program Chair of the Seventh NTCIR Workshop
National Institute of Informatics