

## Message from the Program Chair

The *Fourth NTCIR Workshop* (NTCIR-4) is the fourth in a series of evaluation workshops designed to enhance research in information access technologies, including text retrieval, cross-language information retrieval, automatic text summarization, information extraction, question-answering, text mining, etc., using Japanese and East Asian-languages text. It is co-sponsored by the Research Center for Information Resources at the National Institute of Informatics (RCIR/NII, Director: Jun Adachi) and the Japanese MEXT Grant-in-Aid for Scientific Research on Informatics (#13224087, Primary Investigator: Jun Adachi). And the Question Answering Challenge's third subtask was supported by NII Collaborative Research Grant Type-B.

The purposes of the NTCIR Workshop are:

- to encourage research in information access technologies by developing and providing the infrastructure for large-scale evaluation of information access technologies,
- to provide a forum for research groups interested in cross-system comparison and in exchanging research ideas, and
- to investigate methodologies and metrics for evaluation of information access technologies and methods for constructing large-scale reusable test collections.

With the Internet, the information access technologies are part of the fundamental social infrastructures of our life. The term *information access* refers the whole process that users obtained relevant information, from document collections, to solve their problems. Traditionally document retrieval has been the core technology to support such process, and then the scope of information access technologies is being evolved and diversified to include various technologies supporting users to utilize information from documents, such as information retrieval, text summarization, question answering, text mining, etc. Research and development in information access technologies always requires solid evidences based on experiments to show the superiority of the proposed method over the previous ones. A test collection is a dataset usable for such testing and experiments. The importance of large-scale test collections is widely recognized. We hope the NTCIR Workshops serve as one of the venues to produce re-usable test collections and fruitful forum of researchers, and that wide-ranging insights have been produced as results of the Workshop.

The Forth NTCIR Workshop started with document distribution in April 2003 and the final meeting was held on 2-4<sup>th</sup> June 2004, at NII, Tokyo, Japan. The NTCIR-4 consists of five tasks:

- (1) Cross-Lingual Information Retrieval Task (CLIR),
- (2) Patent Retrieval Task (PATENT),
- (3) Question-Answering Challenge (QAC2),
- (4) Text Summarization Challenge (TSC3), and
- (5) WEB Task (WEB).

This was the second NTCIR workshop hosted the same set of five tasks, and in each the size of the test collections are increased. We have taken "twice rule", in which a task continue at least two NTCIR so that the participating research groups can continuously work based on the test collection constructed and feedback of the first workshop, and such works after the first year would lead improvement and enhancement of the technologies.

Seventy-four research groups from ten different countries and areas conducted the tasks set by the organizers and submitted results. We appreciate the efforts and energy of each group in their research for this workshop.

Each task was coordinated and organized by the *task organizers*, who are listed on the "Workshop

Organization” of this volume. CLIR is organized in cooperation with National Taiwan University and the Korea Institute of Science and Technology Information (KISTI). PATENT is organized in cooperation with the Japan Intellectual Property Association (JIPA) as a joint research project with the NII and topic creation and relevance judgments on patent could not be realized without the collaboration and significant contribution from them. We deeply thank for their cooperation and collaboration.

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., Chosunilbo, Hankooki.com, Industrial Property Cooperation Center, Japan Patent Office, Japan Patent Information Organization, Korea Economic Daily, Linguistic Data Consortium, Mainichi Newspaper, Nippon Database Kaihatsu, Co. Ltd., NRI Cyber Patent, PATOLIS, the Sing Tao Group, Taiwan News, UDN.COM, Wisers Information Ltd., Yomiuri Shinbun and those who elsewhere must be acknowledged.

NTCIR-4 has two new ventures: A special issue of the ACM-Transaction on Asian Language Processing (ACM-TALIP) for NTCIR-4 and an Open Submission Session, I would like to thank the recommendation committee members and reviewers for their detailed reviews. I am grateful to the advisors and members of the WEB Task Advisory Committee. Special thanks go to Drs Ellen Voorhees and Donna Harman, National Institute of Standards and Technology (NIST), USA, for their continuous support and substantial advice based on their experience from the beginning of the project, and for permitting the use of TREC's evaluation program. We thank CLEF, the Cross-Language Evaluation Forum, for support and collaboration on various occasions.

We hope that NTCIR-4 serves to enhance the research in information access technologies, and are grateful any feedbacks from you to improve the NTCIR activities.



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