

## Preface

The *Fifth NTCIR Workshop* (NTCIR-5) is the fifth in a series of evaluation workshops designed to enhance research in information access technologies, including text retrieval, cross-language information access, question-answering, automatic text summarization, information extraction, text mining, etc., using East Asian-languages text such as Japanese, Chinese, Korean.

It is co-sponsored by the Research Center for Information Resources at the National Institute of Informatics (RCIR/NII) and the Japanese MEXT Grant-in-Aid for Scientific Research on Informatics (#13224087). Question Answering Challenge and Patent Retrieval tasks are partially supported by NII Joint Research Grant (Type-B) as well. A pilot workshop, Multimodal Summarization for Trend Information (MuST) is supported by a Joint Research Grant of Nippon Telegraph and Telephone, Corp. and the University of Tokyo, and NII Joint Research Grant (Type-B).

The goals of the NTCIR Workshops are:

1. to encourage research in information access technologies by providing large-scale test collections reusable for experiments and common evaluation infrastructures,
2. to provide a forum for research groups interested in cross-system comparisons and exchanging research ideas, and
3. to investigate *a*) methodologies and metrics for evaluation of information access technologies and *b*) methods for constructing large-scale reusable test collections.

The term “information access” includes the whole process of making information in documents usable by users. It includes document retrieval from vast collections of various types of documents for various tasks and technologies to support users in utilizing information in the retrieved documents. Research and development of information access technologies require solid evidence based on experiments to show the superiority of the proposed system and/or strategy over previous ones. A test collection is a data set used for such testing. We hope that the NTCIR Workshops serve for those areas and that wide-ranging insights have been produced as results.

The call-for-task-participation to NTCIR-5 started in August 2004 and the final meeting is held on 6-9 December 2005, at NII, Tokyo. We consider the *workshop* to be the whole process from the call for participation to the final meeting because we have placed emphasis on the interaction and collaboration between participants and organizers through this activity.

The NTCIR-5 hosts five tasks: (1) Cross-Lingual Information Retrieval Task (CLIR), (2) Cross-Lingual Question Answering (CLQA), (3) Patent Retrieval Task (PATENT), (4) Question-Answering Challenge (QAC), and (5) WEB Task (WEB). A pilot workshop MuST is organized but the results of each participating group will be presented in a separate workshop in March 2006. Each task was coordinated and organized by the *task organizers*, a group of researchers of the topic. A hundred and two research groups registered and conducted the tasks set by the organizers, but unfortunately a number of groups could not continue the participation with various reasons. And then seventy-seven research groups from fifteen different countries and areas remained as active participants. The Open Submission Session, a refereed session that welcomes submissions from both NTCIR and non-NTCIR task participants, is a second venture in NTCIR Workshop. The session contributes to enlarge the NTCIR community and provide a wider forum to discuss related topics and researches leads to future NTCIR tasks. This *proceedings* contains reports of the experiments and investigations performed within NTCIR-5 and the papers presented at its Open Submission Session.

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 Intellectual Property Association, Japan Patent Office, Japan Patent Information Organization, Korea Economic Daily, Linguistic Data Consortium, Mainichi Newspapers, Nippon Database Kaihatsu, Co. Ltd., the Sing Tao Group, Taiwan News, UDN.COM, Wisers Information Ltd., and Yomiuri Shimbun. We are thankful to the Advisory Committee members, the NTCIR-5 Research Planning and Coordinating Committee members, and the Open Submission Session Program Committee members. We are thankful to Ellen Voorhees and Donna Harman, National Institute of Standards and Technology, USA, and to CLEF, the Cross-Language Evaluation Forum, for their support, advises and collaboration on various occasions. We also thank Hatsuko Shimanouchi, Yurika Nakayama, Ai Kuba and Shigeeko Tokuda for their supports behind.

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

December 2005

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