

## Preface from the Program Chair

The Second NTCIR Workshop was the second in a series of evaluation workshops designed to enhance research in Japanese and Asian language text retrieval and related text processing technology such as summarization and extraction. The participating groups enrolled for the Workshop by 31 July 2000 and the results were presented at the Meeting, held in Tokyo, 7-9 March 2001. It was the first challenge to host tasks organized by separate groups outside the National Institute of Informatics (NII) and resulted in adding a variety of tasks. It was also the first experience to be organized under the support of the Research Center of the Information Resources (Director: Jun Adachi), which was launched as part of the NII in April 2000.

The goals of the NTCIR Workshops are as follows:

- to encourage research in information retrieval and related text processing technology, such as summarization by providing large-scale test collections, and a common evaluation setting that allows cross-system comparisons
- to provide a forum for research groups interested in comparing results and exchanging ideas and opinions in an informal atmosphere, and
- to investigate methods for constructing large-scale test collections and evaluation methods.

With the rapid growth of the Internet, the importance of research in IR (Information Retrieval) and related text processing technologies, to make information in the documents immediately usable, is increasing tremendously. Research and development of IR systems and text processing technology always requires solid evidence based on experiments to show the superiority of the proposed system over previous ones. A test collection is a data set used for such testing. The importance of the large-scale test collections in this research, and the forum for researchers to stimulate each other, is widely recognized. We hope the NTCIR Workshop serves as one such opportunity, and that wide-ranging insight has been produced as a result of the Workshop.

The Second NTCIR Workshop hosted three tasks: (1) Chinese Text Retrieval Tasks (CHTR), (2) Japanese and English Information Retrieval Tasks (JEIR,) and (3) Text Summarization Challenge (TSC). Three reusable data sets were created through the Workshop:

- *CIRB010*, which contains 132,173 (200MB) Chinese newspaper articles, and 50 Chinese topics and English translations,
- *NTCIR-2*, which contains 403,248 Japanese (600MB) and 134,978 English (200MB) documents, and 49 Japanese topics with English translation, and
- *NTCIR-2 Summ*, which contains about 100 hand-created summaries of various types of Japanese newspaper articles.

Forty-five groups from eight countries and areas enrolled in the Second NTCIR Workshop. Among them, 16 groups enrolled in CHTR (14 in the Chinese Monolingual Retrieval and 13 in the English-Chinese Cross-Lingual Retrieval Subtasks), 31 groups in JEIR (22 in Japanese Monolingual Retrieval, 11 in English Monolingual Retrieval, 16 in Japanese-English, 14 in English-Japanese, 11 in a Japanese-Mixture of Japanese and English, 11 in an English-Mixture of Japanese and English Cross-Lingual Retrieval Subtasks), and 15 in TSC. Of these groups, 13 are from companies, 28 from universities and four from national research institutes. Results were submitted by the following groups and we greatly appreciate the efforts and energy of each group in their research using the Collections.

ATT Labs & Duke University (US)

Communications Research Laboratory (Japan)

Fuji Xerox (Japan)

Fujitsu Laboratories (Japan)

Fujitsu R&D Center (China)

Central Research Laboratory, Hitachi Co. (Japan)

Hong Kong Polytechnic (Hong Kong, China)

Institute of Software, Chinese Academy of Sciences (China)

Johns Hopkins University (US)

JUSTSYSTEM Corp. (Japan)  
Kanagawa University (Japan)  
Korea Advanced Institute of Science and Technology (Korea)  
Matsushita Electric Industrial (Japan)  
National. TsinHua University (Taiwan, ROC)  
NEC Media Research Laboratories (Japan)  
National Institute of Informatics (Japan)  
NTT-CS & NAIST (Japan)  
OASIS, Aizu University (Japan)  
Osaka Kyoiku University (Japan)  
Queen College-City University of New York (US)  
Ricoh Co. (two groups) (Japan)  
Surugadai University (Japan)  
Trans EZ Co. (Taiwan ROC)  
Toyohashi University of Technology (two groups) (Japan)  
University of California Berkeley (US)  
University of Cambridge/Toshiba/Microsoft (UK)  
University of Electro-Communications (two groups) (Japan)  
University of Library and Information Science (Japan)  
University of Maryland (US)  
University of Tokyo (two groups) (Japan)  
Yokohama National University (Japan)  
Waseda University (Japan)

We are extremely grateful to the following organizations for their understanding of the project and for providing permission to use document data for the Test Collections:

For *CIRB010*: Central Daily News, China Daily News, Chinatimes, Chinatimes Commercial, Chinatimes Express, and United Daily News, and

For *NTCIR-2* and *NTCIR-2 Summ*: the 65 academic societies (<http://research.nii.ac.jp/ntcir/thanks1-en.html>), Heiwa Information Center Co. Ltd., Mainichi Newspaper Co., and TAO.

We would like to thank the members of the program committee of the Second NTCIR Workshop. Special thanks go to Drs Ellen Voorhees and Donna Harman, NIST, USA, for permitting the use of TREC's evaluation program and scripts to check search result files, and for substantial advice based on their experience. Analysts of the IR relevance assessments and those who undertook the evaluation of the summarization have all worked very hard. Without their contribution, the Workshop could not have become a reality. Hard work by a dedicated group is necessary for the success of any project. The hard work of task organizers, Hsin-Hsi Chen, Kuang-hua Chen, Manabu Okumura, Takahiro Fukushima and their groups, the group in the NII and those elsewhere must be acknowledged. We have been able to list some of these people on the "Workshop Organization" page.

We enjoyed the Workshop and hope that all of you had an equally fulfilling experience.



Noriko Kando, Ph.D.  
Program Chair of the Second NTCIR Workshop  
National Institute of Informatics