IJCNLP 2008

Workshop on Technologies and Corpora for Asia-Pacific Speech Translation (TCAST)

Proceedings of the Workshop

Organizer

Asian Speech Translation Advanced Research Consortium
(A-Star)

Local Host

International Institute of Information Technology, India

January 11 2008
Hyderabad, India

Preface

This volume contains the paper accepted for presentation at the 2008 Workshop on Technologies and Corpora for Asia-Pacific Speech Translation (TCAST), which is part of the The Third International Joint Conference on Natural Language Processing held on January 7-12, 2008, in Hyderabad, India (IJCNLP2008). This workshop took place on January 11 2008.

In an age of global communication, information exchange by means of speech-to-speech technology is playing an increasingly important role. This technology is vital in breaking down language barriers and facilitating better social interaction and exchange in business and other areas. Research programs have been launched in many different countries and efforts have been made to develop successful speech-to-speech systems for several languages around the world. In the Asia-Pacific region, extensive efforts are needed to develop the field. In the region a large number of languages and dialects are spoken, some of these languages have a very rich cultural heritage. However, many of these languages have been neglected and information resources are not available.

Given this background, the objective of the workshop was to present the research and development work currently in progress for the development of corpora, data tools and techniques for the processing of Asian languages and their standardisation for applications in speech translation between Asian languages. The main aims of this workshop were to allow participants to interact and share knowledge of available resources and ongoing research, and to discuss possible avenues for future development in the field. This workshop was a part of the activities of the expert group on "Speech and Natural Language Processing" created under the ASTAP program, APEC-TEL and the A-Star project.

We would like to acknowledge the exceptional cooperation of our organizing committee members during the organization of this workshop.

Andrew Finch
Workshop Organizer
November 2007

Organization

Workshop Chair:

Satoshi Nakamura (NiCT-ATR, Japan)

Organizing Committee:

Satoshi Nakamura (NiCT-ATR, Japan)

Andrew Finch (NiCT-ATR, Japan)

Sakriani Sakti (NiCT-ATR, Japan)

Program Committee:

Satoshi Nakamura (NiCT-ATR, Japan)

S.S. Agrawal (CDAC, India)

Hammam Riza (BPPT, Indonesia)

Jun Park (ETRI, Korea)

Chai Wutiwiwatchai (NECTEC, Thailand)

Bo Xu (CAS, China)

Linshan Lee (NTU, Taipei)

Workshop Website:

http://www.slc.atr.jp/TCAST/TCAST2008/TCAST_Home.html

Workshop Program

09:00-09:30 Workshop Registration

09:30-10:00 Opening Speech
Satoshi Nakamura (NiCT-ATR, Japan)

Session 1: Machine Translation

10:00-10:30 *Transformation-based Sentence Splitting method for Statistical Machine Translation*
Jonghoon Lee, Donghyeon Lee and Gary Geunbae Lee

10:30-11:00 Coffee Break

11:00-11:30 *Speech-to-Speech Translation Activities in Thailand*
Chai Wutiwiwatchai, Thepchai Supnithi and Krit Kosawat

11:30-12:00 *Phrase-based Machine Transliteration*
Andrew Finch and Eiichiro Sumita

12:00-13:30 Lunch

Session 2: Speech Recognition

13:30-14:00 *Development of Indonesian Large Vocabulary Continuous Speech Recognition System within A-STAR Project*
Sakriani Sakti, Eka Kelana, Hammam Riza, Shinsuke Sakai, Konstantin Markov and Satoshi Nakamura

14:00-14:30 *Using Confidence Vector in Multi-Stage Speech Recognition*
Hyungbae Jeon, Kyuwoong Hwang, Hoon Chung, Seunghi Kim, Jun Park and Yunkeun Lee

14:30-15:00 *Toward Asian Speech Translation System: Developing Speech Recognition and Machine Translation for Indonesian Language*
Hammam Riza and Oskar Riandi

15:00-15:45 Discussion and Closing

Table of Contents

Author Index