IJCNLP 2011

Proceedings of
the 9th Workshop on
Asian Language Resources
collocated with IJCNLP 2011

November 12-13, 2011
Shangri-La Hotel
Chiang Mai, Thailand

ALR9

Proceedings of
the 9th Workshop on Asian Language Resources
collocated with IJCNLP 2011

November 12 and 13, 2011
Chiang Mai, Thailand

We wish to thank our sponsors

Gold Sponsors

www.google.com

www.baidu.com

The Office of Naval Research (ONR)

The Asian Office of Aerospace Research and Development (AOARD)

Department of Systems Engineering and Engineering Managment, The Chinese University of Hong Kong

Silver Sponsors

Microsoft Corporation

Bronze Sponsors

Chinese and Oriental Languages Information Processing Society (COLIPS)

Supporter

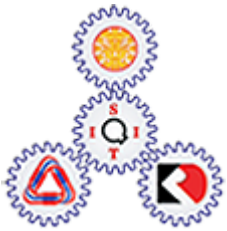
Thailand Convention and Exhibition Bureau (TCEB)

We wish to thank our sponsors

Organizers

Asian Federation of Natural Language Processing (AFNLP)

National Electronics and Computer Technology Center (NECTEC), Thailand



Sirindhorn International Institute of Technology (SIIT), Thailand

Rajamangala University of Technology Lanna (RMUTL), Thailand

Maejo University, Thailand

Chiang Mai University (CMU), Thailand

ISBN 978-974-466-565-2

Preface

We are happy to publish this volume that contains the papers presented at the 9th Workshop on Asian Language Resources hosted by the Asian Federation of Natural Language Processing, held in conjunction with the International Joint Conference on Natural Language Processing (IJCNLP 2011) in November 12-13, 2011 in Chiang Mai. The workshop and, needless to say, this volume intend to highlight the continually increasing as well as diversified efforts in Asia to build multilingual and multi-modal language resources and their applications through the use of ICTs. Corollary to these thriving endeavors, there is also a need for international standards within the region. Thus, the workshop on its second day is held in cooperation with ISO/TC37/SC4, which develops international standards for language resources management.

All papers submitted for presentation went through a double blind-review process and were evaluated by two or three members of the program committee. Twelve of the submitted papers (or 71 percent) were chosen for presentation at the conference, which are published in this volume. We want to thank all those who submitted papers for review and those who provided manuscripts for publication in these proceedings. We also want to give special thanks to the effort and hard work of our reviewers, whose commitment reflects their dedication to the growth of NLP in the region.

Rachel Edita O. Roxas (Chair)
Sarmad Hussain (Co-Chair)
Key-Sun Choi (Co-Chair)

Organizers:

Asian Language Resource Committee (ALRC)
Asian Federation of NLP (AFNLP, www.afnlp.org)

Co-Organizers:

ISO/TC37/SC4
East Asian Forum on Terminology (EAFTERM)

Program Committee:

Rachel Edita O. Roxas (Chair) - De La Salle University, Philippines
Sarmad Hussain (Co-Chair) - CLE-KICS, UET Lahore, Pakistan
Key-Sun Choi (Co-Chair) - Dept. of CS, KAIST, Korea

Mirna Adriani - University of Indonesia, Indonesia
Pushpak Bhattacharyya - IIT-Bombay, India
Miriam Butt - University of Konstanz, Germany
Thatsanee Charoenporn - NECTEC, Thailand
Rowena Cristina Guevara - University of the Philippines Diliman, Philippines
Hitoshi Isahara - NICT, Japan
Emi Izumi - NICT, Japan
Chu-Ren Huang - Hong Kong Polytechnic University, and Academia Sinica, Taiwan
Zhang Huarui - Peking University, China
Haizhou Li - I2R, Singapore
Chi Mai Luong - IOIT, Vietnamese Academy of Science and Technology, Vietnam
Ruli Marunung - University of Indonesia, Indonesia
Yoshiki Mikami - Nagaoka University of Technology, Japan
Sakrange Turance Nandasara - University of Colombo, School of Computing, Sri Lanka
Hammam Riza - IPTEKnet-BPPT, Indonesia
Kiyoaki Shirai - JAIST, Japan
Virach Sornlertlamvanich - NECTEC, Thailand
Takenobu Tokunaga - Tokyo Institute of Technology, Japan
Ruvan Weerasinghe - University of Colombo, School of Computing, Sri Lanka
Chai Wutiwiwatchai - NECTEC, Thailand
Yogendra Yadava - Tribhuvan University, Nepal

Table of Contents

Conference Program

Saturday, November 12, 2011

9:00–10:00 *Participation in Language Resource Development and Sharing (Invited Talk)*
Virach Sornlertlamvanich

Session Chair: Rachel Edita Roxas

10:00–10:30 Coffee/Tea Break

Session 1: Language Resources

Session Chair: Allan Borra

10:30–10:50 *Towards a Computational Semantic Analyzer for Urdu*
Annette Hautli and Miriam Butt

10:50–11:10 *Engineering a Deep HPSG for Mandarin Chinese (Short Paper)*
Yi Zhang, Rui Wang and Yu Chen

11:10–11:30 *Experiences in Building Urdu WordNet*
Farah Adeeba and Sarmad Hussain

11:30–11:50 *Bantay-Wika: towards a better understanding of the dynamics of Filipino culture and linguistic change (Short Paper)*
Joel Ilao, Rowena Cristina Guevara, Virgilio Llenaresas, Eilene Antoinette Narvaez and Jovy Peregrino

11:50–12:00 Panel Discussion

12:00–14:00 Lunch

Saturday, November 12, 2011 (continued)

Session 2: Corpus

Session Chair: Stephen Beale

14:00–14:20 *Error Detection for Treebank Validation*
Bharat Ram Ambati, Rahul Agarwal, Mridul Gupta, Samar Husain and Dipti Misra Sharma

14:20–14:40 *Multi-stage Annotation using Pattern-based and Statistical-based Techniques for Automatic Thai Annotated Corpus Construction*
Nattapong Tongtep and Thanaruk Theeramunkong

14:40–15:00 *Feasibility of Leveraging Crowd Sourcing for the Creation of a Large Scale Annotated Resource for Hindi English Code Switched Data: A Pilot Annotation*
Mona Diab and Ankit Kamboj

15:00–15:20 *Philippine Languages Online Corpora: Status, issues, and prospects*
Shirley Dita and Rachel Edita Roxas

15:20–15:30 Panel Discussion

15:30–16:00 Coffee/Tea Break

Session 3: NLP Tools

Session Chair: Mona Diab

16:00–16:20 *Word Disambiguation in Shahmukhi to Gurmukhi Transliteration*
Tejinder Singh Saini and Gurpreet Singh Lehal

16:20–16:40 *A Grammar Checker for Tagalog using LanguageTool*
Nathaniel Oco and Allan Borra

16:40–17:00 *Linguist's Assistant: A Resource For Linguists*
Stephen Beale and Tod Allman

17:00–17:20 *Providing Ad Links to Travel Blog Entries Based on Link Types*
Aya Ishino, Hidetsugu Nanba and Toshiyuki Takezawa

Saturday, November 12, 2011 (continued)

17:20–17:30 Panel Discussion

Sunday, November 13, 2011

10:30–12:00 *Discourse Structures (SemAF-DS)*
Koiti Hasida

12:00–14:00 Lunch

14:00–15:30 *Tool Interchange Formats*
Nancy Ide

15:30–16:00 Coffee/Tea Break

16:00–16:30 *WG Plenary*
Nancy Ide and Kiyong Lee