ALR 12

The 12th Workshop on Asian Language Resources

Proceedings of the Workshop

December 12, 2016
Osaka, Japan

ISBN978-4-87974-722-8

Preface

This 12th Workshop on Asian Language Resources (ALR12) focuses on language resources in Asia, which has more than 2,200 spoken languages. There are now increasing efforts to build multi-lingual, multi-modal language resources, with varying levels of annotations, through manual, semi-automatic and automatic approaches, as the use of ICT spreads across Asia. Correspondingly, the development of practical applications of these language resources has also been rapidly advancing. The ALR workshop series aims to forge a better coordination and collaboration among researchers on these languages and in the NLP community in general, to develop common frameworks and processes for promoting these activities. ALR12 collaborates with ISO/TC 37/SC 4, which develops international standards for "Language Resources Management," and ELRA, which is campaigning LRE map, in order to integrate efforts to develop an Asian language resource map. Also, the workshop is supported by AFNLP, which has a dedicated Asian Language Resource Committee (ARLC), whose aim is to coordinate the important ALR initiatives with different NLP associations and conferences in Asia and other regions. This workshop consists of twelve oral papers and seven posters, plus a special session to introduce ISO/TC 37/SC 4 activities to the community, to stimulate further interactions between research and standardization.

ALR12 program co-chairs

Koiti Hasida
President, GSK
The University of Tokyo

Kam-Fai Wong
President, AFNLP
The Chinese University of Hong Kong

Nicoletta Calzolari
Honorary President, ELRA
ILC-CNR

Key-Sun Choi
Secretary, ISO/TC 37/SC 4
KAIST

Organisers

Koiti Hasida
Kam-Fai Wong
Nicoletta Calzolari
Key-Sun Choi

Programme Committee

Kenji Araki
Normaziah Aziz
Khalid Choukri
Kohji Dohsaka
Kentaro Inui
Hitoshi Isahara
Kai Ishikawa
Satoshi Kinoshita
Kiyoshi Kogure
Haizhou Li
Joseph Mariani
Fumihito Nishino
Win Pa Pa
Ayu Purwarianti
Lu Qin
Hammam Riza
Hiroaki Saito
Kiyoaki Shirai
Virach Sornlertlamvanich
Keh-Yih Su
Kumiko Tanaka-Ishii
Takenobu Tokunaga
Masao Utiyama

Table of Contents

Conference Program

Monday, December 12, 2016

09:00–09:05 ***Opening***

09:05–10:25 **Oral Session 1: Annotation**

An extension of ISO-Space for annotating object direction
Daiki Gotou, Hitoshi Nishikawa and Takenobu Tokunaga

Annotation and Analysis of Discourse Relations, Temporal Relations and Multi-Layered Situational Relations in Japanese Texts
Kimi Kaneko, Saku Sugawara, Koji Mineshima and Daisuke Bekki

Developing Universal Dependencies for Mandarin Chinese
Herman Leung, Rafaël Poiret, Tak-sum Wong, Xinying Chen, Kim Gerdes and John Lee

Developing Corpus of Lecture Utterances Aligned to Slide Components
Ryo Minamiguchi and Masatoshi Tsuchiya

10:25–10:35 ***Coffee Break***

10:35–11:55 **Oral Session 2: Data**

VSoLSCSum: Building a Vietnamese Sentence-Comment Dataset for Social Context Summarization
Minh-Tien Nguyen, Dac Viet Lai, Phong-Khac Do, Duc-Vu Tran and Minh-Le Nguyen

BCCWJ-DepPara: A Syntactic Annotation Treebank on the 'Balanced Corpus of Contemporary Written Japanese'
Masayuki Asahara and Yuji Matsumoto

SCTB: A Chinese Treebank in Scientific Domain
Chenhui Chu, Toshiaki Nakazawa, Daisuke Kawahara and Sadao Kurohashi

Big Community Data before World Wide Web Era
Tomoya Iwakura, Tetsuro Takahashi, Akihiro Ohtani and Kunio Matsui

Monday, December 12, 2016 (continued)

12:00–14:00 ***Lunch Break***

14:00–14:30 **Poster session**

An Overview of BPPT's Indonesian Language Resources
Gunarso Gunarso and Hammam Riza

Creating Japanese Political Corpus from Local Assembly Minutes of 47 prefectures
Yasutomo Kimura, Keiichi Takamaru, Takuma Tanaka, Akio Kobayashi, Hiroki Sakaji, Yuzu Uchida, Hokuto Ototake and Shigeru Masuyama

Selective Annotation of Sentence Parts: Identification of Relevant Sub-sentential Units
Ge Xu, Xiaoyan Yang and Chu-Ren Huang

14:35–15:55 **Oral Session 3: Analysis**

The Kyutech corpus and topic segmentation using a combined method
Takashi Yamamura, Kazutaka Shimada and Shintaro Kawahara

Automatic Evaluation of Commonsense Knowledge for Refining Japanese ConceptNet
Seiya Shudo, Rafal Rzepka and Kenji Araki

SAMER: A Semi-Automatically Created Lexical Resource for Arabic Verbal Multiword Expressions Tokens Paradigm and their Morphosyntactic Features
Mohamed Al-Badrashiny, Abdelati Hawwari, Mahmoud Ghoneim and Mona Diab

Sentiment Analysis for Low Resource Languages: A Study on Informal Indonesian Tweets
Tuan Anh Le, David Moeljadi, Yasuhide Miura and Tomoko Ohkuma

Monday, December 12, 2016 (continued)

15:55–16:55 TC37 Session

Introducing ISO/TC 37/SC 4 Language Resources Management Activities
Nicoletta Calzolari

Towards Application of ISO-TimeML and ISOspace to Korean and other Asian Languages
Kiyong Lee

Standardization of Numerical Expression Extraction and Representations in English and Other Languages
Haitao Wang

Design of ISLRN for Asian Language Resources
Khalid Choukri

16:55–17:00 ***Closing***