IJCNLP-05

Fifth Workshop on Asian Language Resources (ALR-05)

and

First Symposium on Asian Language Resources Network (ALRN)

Proceedings of the Workshop

14 October 2005
Jeju Island, Korea

© 2005 Asian Federation of Natural Language Processing

These workshop and symposium are supported by Special Coordination Funds for Promoting Science and Technology, Ministry of Education, Culture, Sport, Science and Technology, MEXT Japan.

PREFACE

It is increasingly convinced that language resources, as well as corpus-based, stochastic, and learning approaches, play a significantly important role in Natural Language Processing (NLP) research. There have been several reports on the success of constructing and using corpora in many dimensions. How to effectively re-organizing the existing resources into a unified framework and establishing the guideline for corpus development has become more and more important, which will be highly helpful for sharing resources and coping with cross-language problems.

Motivated by this background, the 5th Workshop on Asian Language Resources (ALR) and 1st Symposium on Asian Language Resources Network (ALRN) are organized under the auspices of the Asian Language Resources Committee of Asia Federation of Natural Language Processing (AFNLP) in conjunction with IJCNLP2005.

The purposes of the workshop and symposium are as follows.
(1) To investigate the situation of Asian Language Resources, and to make a catalog of the result of this investigation;
(2) To investigate and discuss the problems related to the standards and specification on creating various kinds of language resources;
(3) To promote communications between developers and users of various language resources in order to fill the gap between language resources and practical applications;
(4) To launch a roadmap for Asian Language Resources.

ALR-05 accepts 10 regular papers. We are so sure that the selected papers for presentation are informative and can gain much potential for further research.

We hope to meet worldwide active researchers working on Asian languages to promote the research on linguistic resources and related fields. We are sure that the workshop and symposium will fruitfully contribute to construct a unifying architecture and mechanism for Asian Language Resources development, management and sharing.

Our workshop and symposium would not have been succeeded without the hard work of the program committee. Also, we would like to express our great thanks to the arrangement of the IJCNLP-05 organizing committee and the secretariat. Finally, we wish that all the participants can benefit a lot and enjoy themselves in the workshop and symposium.

Bo Xu (chair)
Chu-Ren Huang (co-chair)
Takenobu Tokunaga (co-chair)
Jun Zhao (co-chair)

PROGRAMME COMMITTEE

Bo Xu (chair)	Chinese Academy of Sciences
Chu-Ren Huang (co-chair)	Academia Sinica
Takenobu Tokunaga (co-chair)	Tokyo Institute of Technology
Jun Zhao (co-chair)	Chinese Academy of Sciences
Nicoletta Calzolari	Istituto di Linguistica Computazionale del CNR
Baobao Chang	Peking University
Shuichi Itahashi	National Institute of Advanced Industrial Science and Technology
Donghong Ji	Institute for Inforcomm Research Singapore
Kiyong Lee	Korea University
Qin Lu	Polytechnique University of Hong Kong
Nguyen Thi Minh Huyen	Hanoi University of Sciences
Hae-Chang Rim	Korea University
Kiyoaki Shirai	Japan Advanced Institute of Science and Technology
Nashunwuritu	Inner Mongolia University
Virach Sornlertlamvanich	Thai Computational Linguistics Laboratory, NICT
Maosong Sun	Tsinghua University
Jane Tsay	Chung-Cheng University
Hsiao-chuan Wang	Tsing Hua Unversity
Elizabeth Zeitoun	Academia Sinica

PROGRAMME

Friday, October 14, 2005

Time	Event	Authors
8:30	*Registration*	
9:00	*Opening*	
9:10	*Keynote Speech*	Nicoletta Calzolari
9:50	Break	
10:00	Domain Knowledge Engineering Based on Encyclopedias and the Web Text	Sui, Z., Cui, G., Ding, W., Zhang, Q.
10:20	Evaluation of a Japanese CFG Derived from a Syntactically Annotated Corpus with Respect to Dependency Measures	Noro, T., Koike, C., Hashimoto, T., Tokunaga, T., Tanaka, H.
10:40	Corpus-oriented Acquisition of Chinese	Zhang, Y., Kashioka, H.
11:00	Break	
11:20	The Standard of Chinese Corpus Metadata	He, T., Xu, X.
11:40	An Integrated Framework for Archiving, Processing and Developing Learning Materials for an Endangered Aboriginal Language in Tai-	Yang, M., Rau, D. V.
12:00	Construction of Structurally Annotated Spoken Dialogue Corpus	Kato, S., Matsubara, S., Yamaguchi, Y.,
12:20	Lunch	
13:40	Cross-lingual Conversion of Lexical Semantic Relations: Building Parallel Wordnets	Huang, C., Su, I., Hong, J., Li, X.
14:00	Taiwan Child Language Corpus: Data Collection and Annotation	Tsay, J. S.
14:20	Question Classification using Multiple	Li, X., Huang, X., Wu, L.
14:40	Harvesting the Bitexts of the Laws of Hong Kong From the Web	Kit, C., Liu, X., Sin, K., Webster, J.J.
15:00	Break	
15:10	*Symposium* : Asian language resources: Infrastructure towards a multilingual language processing environment in Asia	
16:50	Closing	

Table of Contents

PREFACE i

PROGRAMME COMMITTEE ii

PROGRAMME iii

Domain Knowledge Engineering Based on Encyclopedias and the Web Text
Zhifang Sui, Gaoying Cui, Wansong Ding and Qinlong Zhang 1

Evaluation of a Japanese CFG Derived from a Syntactically Annotated Corpus with Respect to Dependency Measures
Tomaya Noro, Chimato Koike, Taiichi Hashimoto, Takenobu Tokunaga and Hozumi Tanaka 9

Corpus-oriented Acquisition of Chinese Grammar
Yan Zhang and Hideki Kashioka 17

The Standard of Chinese Corpus Metadata
Tingting He and Xiaoqi Xu 24

An Integrated Framework for Archiving, Processing and Developing Learning Materials for an Endangered Aboriginal Language in Tai-wan
Meng-Chien Yang and D. Victoria Rau 32

Construction of Structurally Annotated Spoken Dialogue Corpus
Shingo Kato, Shigeki Matsubara, Yukiko Yamaguchi and Nobuo Kawaguchi 40

Cross-lingual Conversion of Lexical Semantic Relations: Building Parallel Wordnets
Chu-Ren Huang, I-Li Su, Jia-Fei Hong and Xiang-Bing Li 48

Taiwan Child Language Corpus: Data Collection and Annotation
Jane S. Tsay 56

Question Classification using Multiple Classifiers
Xin Li, Xuan-jing Huang and Li-de Wu 64

Harvesting the Bitexts of the Laws of Hong Kong From the Web
Chunyu Kit, Xiaoyue Liu, KingKui Sin and Jonathan J. Webster 71

AUTHOR INDEX 79