

ACL 2012

**50th Annual Meeting of the  
Association for Computational Linguistics**

**Proceedings of the System Demonstrations**

10 July 2012  
Jeju Island, Korea

©2012 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 978-1-937284-27-5

## **Preface**

Welcome to the proceedings of the system demonstration session. This volume contains the papers of the system demonstrations presented at the 50th Annual Meeting of the Association for Computational Linguistics, held in Jeju, Republic of Korea, on July 10, 2012.

The system demonstrations program offers the presentation of early research prototypes as well as interesting mature systems. The system demonstration chair and the members of the program committee received 49 submissions, 29 of which were selected for inclusion in the program after review by three members of the program committee.

I would like to thank the members of the program committee for their excellent job in reviewing the submissions and providing their support for the final decision.



**Chair:**

Min Zhang (Institute for Infocomm Research, Singapore)

**Program Committee:**

Ai Ti Aw (Institute for Infocomm Research, Singapore)  
Kalika Bali (Microsoft Research India, India)  
Rafael Banchs (Institute for Infocomm Research, Singapore)  
Sivaji Bandyopadhyay (Jadavpur University, India)  
Francis Bond (NTU, Singapore)  
Monojit Choudhury (Microsoft Research India, India)  
Eng Siong Chng (NTU, Singapore)  
Ido Dagan (Bar Ilan University, Israel)  
Xiangyu Duan (Institute for Infocomm Research, Singapore)  
George Foster (IIT, NRC, Canada)  
Guohong Fu (Heilongjiang University, China)  
Sarvnaz Karimi (The University of Melbourne, Australia)  
Mitesh Khapra (Indian Institute of Technology Bombay, India)  
Su Nam Kim (Monash University, Australia)  
A Kumaran (Microsoft Research India, India)  
Sadao Kurohashi (Kyoto University, Japan)  
Olivia Kwong (City University of Hong Kong, China)  
Mu Li (Microsoft Research Asia, China)  
Ziheng Lin (NUS, Singapore)  
Qun Liu (ICT, CAS, China)  
R. Costa-jussà Marta (Barcelona Media, Spain)  
Alessandro Moschitti (Trento University, Italy)  
Jong-Hoon Oh (NICT, Japan)  
Long Qiu (Institute for Infocomm Research, Singapore)  
Yan Qu (ShareThis Inc., USA)  
Xiaodong Shi (Xiamen University, China)  
Maosong Sun (Tsinghua University, China)  
Jun Sun (NUS, Singapore)  
Le Sun (IS, CAS, China)  
Chew Lim Tan (NUS, Singapore)  
Raghavendra Udupa (Microsoft Research India, India)  
yunqing xia (Tsinghua University, China)  
Deyi Xiong (Institute for Infocomm Research, Singapore)  
Myun Yang (Harbin Institute of Technology, China)  
Tiejun Zhao (Harbin Institute of Technology, China)  
Changle Zhou (Xiamen University, China)  
GuoDong Zhou (Suzhou University, China)

Chengqing Zong (IA, CAS, China)

## Table of Contents

<i>Applications of GPC Rules and Character Structures in Games for Learning Chinese Characters</i> Wei-Jie Huang, Chia-Ru Chou, Yu-Lin Tzeng, Chia-Ying Lee and Chao-Lin Liu .....	1
<i>Specifying Viewpoint and Information Need with Affective Metaphors: A System Demonstration of the Metaphor-Magnet Web App/Service</i> Tony Veale and Guofu Li .....	7
<i>QuickView: NLP-based Tweet Search</i> Xiaohua LIU, Furu WEI, Ming ZHOU and QuickView Team Microsoft .....	13
<i>NiuTrans: An Open Source Toolkit for Phrase-based and Syntax-based Machine Translation</i> Tong Xiao, Jingbo Zhu, Hao Zhang and Qiang Li .....	19
<i>langid.py: An Off-the-shelf Language Identification Tool</i> Marco Lui and Timothy Baldwin .....	25
<i>Personalized Normalization for a Multilingual Chat System</i> Ai Ti Aw and Lian Hau Lee .....	31
<i>IRIS: a Chat-oriented Dialogue System based on the Vector Space Model</i> Rafael E. Banchs and Haizhou Li .....	37
<i>LetsMT!: Cloud-Based Platform for Do-It-Yourself Machine Translation</i> Andrejs Vasiļjevs, Raivis Skadiņš and Jörg Tiedemann .....	43
<i>A Web-based Evaluation Framework for Spatial Instruction-Giving Systems</i> Srinivasan Janarthanam, Oliver Lemon and Xingkun Liu .....	49
<i>DOMCAT: A Bilingual Concordancer for Domain-Specific Computer Assisted Translation</i> Ming-Hong Bai, Yu-Ming Hsieh, Keh-Jiann Chen and Jason S. Chang .....	55
<i>The OpenGrm open-source finite-state grammar software libraries</i> Brian Roark, Richard Sproat, Cyril Allauzen, Michael Riley, Jeffrey Sorensen and Terry Tai ..	61
<i>Multilingual WSD with Just a Few Lines of Code: the BabelNet API</i> Roberto Navigli and Simone Paolo Ponzetto .....	67
<i>BIUTEE: A Modular Open-Source System for Recognizing Textual Entailment</i> Asher Stern and Ido Dagan .....	73
<i>Entailment-based Text Exploration with Application to the Health-care Domain</i> Meni Adler, Jonathan Berant and Ido Dagan .....	79
<i>CSNIPER - Annotation-by-query for Non-canonical Constructions in Large Corpora</i> Richard Eckart de Castilho, Sabine Bartsch and Iryna Gurevych .....	85

<i>ACCURAT Toolkit for Multi-Level Alignment and Information Extraction from Comparable Corpora</i> Mărcis Pinnis, Radu Ion, Dan Ștefănescu, Fangzhong Su, Inguna Skadiņa, Andrejs Vasiļjevs and Bogdan Babych . . . . .	91
<i>Demonstration of IlluMe: Creating Ambient According to Instant Message Logs</i> Lun-Wei Ku, Cheng-Wei Sun and Ya-Hsin Hsueh . . . . .	97
<i>INPRO_iSS: A Component for Just-In-Time Incremental Speech Synthesis</i> Timo Baumann and David Schlangen . . . . .	103
<i>WizIE: A Best Practices Guided Development Environment for Information Extraction</i> Yunyao Li, Laura Chiticariu, Huahai Yang, Frederick Reiss and Arnaldo Carreno-fuentes . . . . .	109
<i>A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle</i> Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar and Shrikanth Narayanan . . . . .	115
<i>Building Trainable Taggers in a Web-based, UIMA-Supported NLP Workbench</i> Rafal Rak, BalaKrishna Kolluru and Sophia Ananiadou . . . . .	121
<i>Akamon: An Open Source Toolkit for Tree/Forest-Based Statistical Machine Translation</i> Xianchao Wu, Takuya Matsuzaki and Jun'ichi Tsujii . . . . .	127
<i>Subgroup Detector: A System for Detecting Subgroups in Online Discussions</i> Amjad Abu-Jbara and Dragomir Radev . . . . .	133
<i>A Graphical Interface for MT Evaluation and Error Analysis</i> Meritxell González, Jesús Giménez and Lluís Màrquez . . . . .	139
<i>Online Plagiarized Detection Through Exploiting Lexical, Syntax, and Semantic Information</i> Wan-Yu Lin, Nanyun Peng, Chun-Chao Yen and Shou-de Lin . . . . .	145
<i>UWN: A Large Multilingual Lexical Knowledge Base</i> Gerard de Melo and Gerhard Weikum . . . . .	151
<i>FLOW: A First-Language-Oriented Writing Assistant System</i> MeiHua Chen, ShihTing Huang, HungTing Hsieh, TingHui Kao and Jason S. Chang . . . . .	157
<i>Social Event Radar: A Bilingual Context Mining and Sentiment Analysis Summarization System</i> Wen-Tai Hsieh, Chen-Ming Wu, Tsun Ku and Seng-cho T. Chou . . . . .	163
<i>Syntactic Annotations for the Google Books NGram Corpus</i> Yuri Lin, Jean-Baptiste Michel, Erez Aiden Lieberman, Jon Orwant, Will Brockman and Slav Petrov . . . . .	169





# Conference Program

July 10, 2012

## 10:30-16:00 Session 1

*Applications of GPC Rules and Character Structures in Games for Learning Chinese Characters*

Wei-Jie Huang, Chia-Ru Chou, Yu-Lin Tzeng, Chia-Ying Lee and Chao-Lin Liu

*Specifying Viewpoint and Information Need with Affective Metaphors: A System Demonstration of the Metaphor-Magnet Web App/Service*

Tony Veale and Guofu Li

*QuickView: NLP-based Tweet Search*

Xiaohua LIU, Furu WEI, Ming ZHOU and QuickView Team Microsoft

*NiuTrans: An Open Source Toolkit for Phrase-based and Syntax-based Machine Translation*

Tong Xiao, Jingbo Zhu, Hao Zhang and Qiang Li

*langid.py: An Off-the-shelf Language Identification Tool*

Marco Lui and Timothy Baldwin

*Personalized Normalization for a Multilingual Chat System*

Ai Ti Aw and Lian Hau Lee

*IRIS: a Chat-oriented Dialogue System based on the Vector Space Model*

Rafael E. Banchs and Haizhou Li

*LetsMT!: Cloud-Based Platform for Do-It-Yourself Machine Translation*

Andrejs Vasiljevs, Raivis Skadiņš and Jörg Tiedemann

*A Web-based Evaluation Framework for Spatial Instruction-Giving Systems*

Srinivasan Janarthanam, Oliver Lemon and Xingkun Liu

*DOMCAT: A Bilingual Concordancer for Domain-Specific Computer Assisted Translation*

Ming-Hong Bai, Yu-Ming Hsieh, Keh-Jiann Chen and Jason S. Chang

*The OpenGrm open-source finite-state grammar software libraries*

Brian Roark, Richard Sproat, Cyril Allauzen, Michael Riley, Jeffrey Sorensen and Terry Tai

*Multilingual WSD with Just a Few Lines of Code: the BabelNet API*

Roberto Navigli and Simone Paolo Ponzetto

**July 10, 2012 (continued)**

**10:30-16:00 Session 1**

*BIUTEE: A Modular Open-Source System for Recognizing Textual Entailment*

Asher Stern and Ido Dagan

*Entailment-based Text Exploration with Application to the Health-care Domain*

Meni Adler, Jonathan Berant and Ido Dagan

*CSNIPER - Annotation-by-query for Non-canonical Constructions in Large Corpora*

Richard Eckart de Castilho, Sabine Bartsch and Iryna Gurevych

*ACCURAT Toolkit for Multi-Level Alignment and Information Extraction from Comparable Corpora*

Mārcis Pinnis, Radu Ion, Dan Ştefănescu, Fangzhong Su, Inguna Skadiņa, Andrejs Vasiļjevs and Bogdan Babych

*Demonstration of IlluMe: Creating Ambient According to Instant Message Logs*

Lun-Wei Ku, Cheng-Wei Sun and Ya-Hsin Hsueh

*INPRO\_iSS: A Component for Just-In-Time Incremental Speech Synthesis*

Timo Baumann and David Schlangen

*WizIE: A Best Practices Guided Development Environment for Information Extraction*

Yunyao Li, Laura Chiticariu, Huahai Yang, Frederick Reiss and Arnaldo Carreno-fuentes

*A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle*

Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar and Shrikanth Narayanan

*Building Trainable Taggers in a Web-based, UIMA-Supported NLP Workbench*

Rafal Rak, BalaKrishna Kolluru and Sophia Ananiadou

*Akamon: An Open Source Toolkit for Tree/Forest-Based Statistical Machine Translation*

Xianchao Wu, Takuya Matsuzaki and Jun'ichi Tsujii

*Subgroup Detector: A System for Detecting Subgroups in Online Discussions*

Amjad Abu-Jbara and Dragomir Radev

*A Graphical Interface for MT Evaluation and Error Analysis*

Meritxell González, Jesús Giménez and Lluís Màrquez

**July 10, 2012 (continued)**

**10:30-16:00 Session 1**

*Online Plagiarized Detection Through Exploiting Lexical, Syntax, and Semantic Information*

Wan-Yu Lin, Nanyun Peng, Chun-Chao Yen and Shou-de Lin

*UWN: A Large Multilingual Lexical Knowledge Base*

Gerard de Melo and Gerhard Weikum

*FLOW: A First-Language-Oriented Writing Assistant System*

MeiHua Chen, ShihTing Huang, HungTing Hsieh, TingHui Kao and Jason S. Chang

*Social Event Radar: A Bilingual Context Mining and Sentiment Analysis Summarization System*

Wen-Tai Hsieh, Chen-Ming Wu, Tsun Ku and Seng-cho T. Chou

*Syntactic Annotations for the Google Books NGram Corpus*

Yuri Lin, Jean-Baptiste Michel, Erez Aiden Lieberman, Jon Orwant, Will Brockman and Slav Petrov