

24th International Conference on Computational Linguistics

(COLING 2012)

Demonstrations

Mumbai, India
8-15 December 2012

Editors:

Martin Kay
Christian Boitet

ISBN: 978-1-62748-391-9

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2012) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2013)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Complex Predicates in Telugu: A Computational Perspective</i> Rahul Balusu	1
<i>Heloise — An Ariane-G5 Compatible Environment for Developing Expert MT Systems Online</i> Vincent Berment and Christian Boitet	9
<i>Keyphrase Extraction in Scientific Articles: A Supervised Approach</i> Pinaki Bhaskar, Kishorjit Nongmeikapam and Sivaji Bandyopadhyay	17
<i>IKAR: An Improved Kit for Anaphora Resolution for Polish</i> Bartosz Broda, Łukasz Burdka and Marek Maziarz	25
<i>Intention Analysis for Sales, Marketing and Customer Service</i> Cohan Sujay Carlos and Madhulika Yalamanchi	33
<i>Authorship Identification in Bengali Literature: a Comparative Analysis</i> Tanmoy Chakraborty	41
<i>Word Root Finder: a Morphological Segmentor Based on CRF</i> Joseph Z. Chang and Jason S. Chang	51
<i>An Efficient Technique for De-Noising Sentences using Monolingual Corpus and Synonym Dictionary</i> Sanjay Chatterji, Diptesh Chatterjee and Sudeshna Sarkar	59
<i>An Example-Based Japanese Proofreading System for Offshore Development</i> Yuchang Cheng and Tomoki Nagase	67
<i>DomEx: Extraction of Sentiment Lexicons for Domains and Meta-Domains</i> Ilia Chetviorkin and Natalia Loukachevitch	77
<i>On the Romanian Rhyme Detection</i> Alina Ciobanu and Liviu P. Dinu	87
<i>Hierarchical Dialogue Policy Learning using Flexible State Transitions and Linear Function Approximation</i> Heriberto Cuayáhuitl, Ivana Kruijff-Korabayová and Nina Dethlefs	95
<i>Automated Paradigm Selection for FSA based Konkani Verb Morphological Analyzer</i> Shilpa Desai, Jyoti Pawar and Pushpak Bhattacharyya	103
<i>Hindi and Marathi to English NE Transliteration Tool using Phonology and Stress Analysis</i> Manikrao Dhore, Shantanu Dixit and Ruchi Dhore	111
<i>Dealing with the Grey Sheep of the Romanian Gender System, the Neuter</i> Liviu P. Dinu, Vlad Niculae and Maria Sulea	119
<i>Authorial Studies using Ranked Lexical Features</i> Liviu P. Dinu and Sergiu Nisioi	125

<i>ScienQuest: a Treebank Exploitation Tool for non NLP-Specialists</i> Achille Falaise, Olivier Kraif, Agnès Tutin and David Rouquet	131
<i>An In-Context and Collaborative Software Localisation Model</i> Amel Fraise, Christian Boitet and Valérie Bellynck	141
<i>Efficient Feedback-based Feature Learning for Blog Distillation as a Terabyte Challenge</i> Dehong Gao, Wenjie Li and Renxian Zhang	147
<i>Beyond Twitter Text: A Preliminary Study on Twitter Hyperlink and its Application</i> Dehong Gao, Wenjie Li and Renxian Zhang	155
<i>Rule Based Hindi Part of Speech Tagger</i> Navneet Garg, Vishal Goyal and Suman Preet	163
<i>Fangorn: A System for Querying very large Treebanks</i> Sumukh Ghodke and Steven Bird	175
<i>CRAB Reader: A Tool for Analysis and Visualization of Argumentative Zones in Scientific Literature</i> Yufan Guo, Ilona Silins, Roi Reichart and Anna Korhonen	183
<i>Automatic Punjabi Text Extractive Summarization System</i> Vishal Gupta and Gurpreet Lehal	191
<i>Complete Pre Processing Phase of Punjabi Text Extractive Summarization System</i> Vishal Gupta and Gurpreet Lehal	199
<i>Revisiting Arabic Semantic Role Labeling using SVM Kernel Methods</i> Laurel Hart, Hassan Alam and Aman Kumar	207
<i>fokas: Formerly Known As – A Search Engine Incorporating Named Entity Evolution</i> Helge Holzmann, Gerhard Gossen and Nina Tahmasebi	215
<i>An Annotation System for Development of Chinese Discourse Corpus</i> Hen-Hsen Huang and Hsin-Hsi Chen	223
<i>Modeling Pollyanna Phenomena in Chinese Sentiment Analysis</i> Ting-Hao Huang, Ho-Cheng Yu and Hsin-Hsi Chen	231
<i>Eating Your Own Cooking: Automatically Linking Wordnet Synsets of Two Languages</i> Salil Joshi, Arindam Chatterjee, Arun Karthikeyan Karra and Pushpak Bhattacharyya	239
<i>I Can Sense It: a Comprehensive Online System for WSD</i> Salil Joshi, Mitesh M. Khapra and Pushpak Bhattacharyya	247
<i>Collaborative Computer-Assisted Translation Applied to Pedagogical Documents and Literary Works</i> Ruslan Kalitvianski, Christian Boitet and Valérie Bellynck	255
<i>Discrimination-Net for Hindi</i> Diptesh Kanojia, Arindam Chatterjee, Salil Joshi and Pushpak Bhattacharyya	261

<i>Rule Based Urdu Stemmer</i>	
Rohit Kansal, Vishal Goyal and Gurpreet Singh Lehal.....	267
<i>JMaxAlign: A Maximum Entropy Parallel Sentence Alignment Tool</i>	
Max Kaufmann.....	277
<i>MIKE: An Interactive Microblogging Keyword Extractor using Contextual Semantic Smoothing</i>	
Osama Khan and Asim Karim.....	289
<i>Domain Based Classification of Punjabi Text Documents</i>	
Nidhi Krail and Vishal Gupta.....	297
<i>Open Information Extraction for SOV Language Based on Entity-Predicate Pair Detection</i>	
Woong-Ki Lee, Yeon-Su Lee, Hyoung-Gyu Lee, Won-Ho Ryu and Hae-Chang Rim ...	305
<i>An Omni-Font Gurmukhi to Shahmukhi Transliteration System</i>	
Gurpreet Singh Lehal, Tejinder Singh Saini and Savleen Kaur Chowdhary	313
<i>THUTR: A Translation Retrieval System</i>	
Chunyang Liu, Qi Liu, Yang Liu and Maosong Sun	321
<i>Recognition of Named-Event Passages in News Articles</i>	
Luis Marujo, Wang Ling, Anatole Gershman, Jaime Carbonell, João P. Neto and David Matos	329
<i>Nonparametric Model for Inupiaq Word Segmentation</i>	
ThuyLinh Nguyen and Stephan Vogel.....	337
<i>Stemming Tigrinya Words for Information Retrieval</i>	
Omer Osman and Yoshiki Mikami.....	345
<i>OpenWordNet-PT: An Open Brazilian Wordnet for Reasoning</i>	
Valeria de Paiva, Alexandre Rademaker and Gerard de Melo	353
<i>WordNet Website Development And Deployment using Content Management Approach</i>	
Neha Prabhugaonkar, Apurva Nagvenkar, Venkatesh Prabhu and Ramdas Karmali ..	361
<i>A Demo for Constructing Domain Ontology from Academic Papers</i>	
Feiliang Ren.....	369
<i>A Practical Chinese-English ON Translation Method Based on ON's Distribution Characteristics on the Web</i>	
Feiliang Ren.....	377
<i>Elissa: A Dialectal to Standard Arabic Machine Translation System</i>	
Wael Salloum and Nizar Habash.....	385
<i>Domain Based Punjabi Text Document Clustering</i>	
Saurabh Sharma and Vishal Gupta.....	393
<i>Open source multi-platform NooJ for NLP</i>	
Max Silberstein, Tamás Váradi and Marko Tadić.....	401

<i>Punjabi Text-To-Speech Synthesis System</i>	
Parminder Singh and Gurpreet Singh Lehal	409
<i>EXCOTATE: An Add-on to MMAX2 for Inspection and Exchange of Annotated Data</i>	
Tobias Stadtfeld and Tibor Kiss	417
<i>Bulgarian Inflectional Morphology in Universal Networking Language</i>	
Velislava Stoykova.....	423
<i>Central and South-East European Resources in META-SHARE</i>	
Marko Tadić and Tamás Váradi	431
<i>Markov Chains for Robust Graph-Based Commonsense Information Extraction</i>	
Niket Tandon, Dheeraj Rajagopal and Gerard de Melo	439
<i>Visualization on Financial Terms via Risk Ranking from Financial Reports</i>	
Ming-Feng Tsai and Chuan-Ju Wang	447
<i>UNL Explorer</i>	
Hiroshi Uchida, Meiyang Zhu and Md. Anwarus Salam Khan	453
<i>An SMT-driven Authoring Tool</i>	
Sriram Venkatapathy and Shachar Mirkin	459
<i>Generating Questions from Web Community Contents</i>	
Baoxun Wang, Bingquan Liu, Chengjie Sun, Xiaolong Wang and Deyuan Zhang.....	467
<i>Demo of iMAG Possibilities: MT-postediting, Translation Quality Evaluation, Parallel Corpus Production</i>	
Ling Xiao Wang, Ying Zhang, Christian Boitet and Valérie Bellynck	475
<i>Jane 2: Open Source Phrase-based and Hierarchical Statistical Machine Translation</i>	
Joern Wuebker, Matthias Huck, Stephan Peitz, Malte Nuhn, Markus Freitag, Jan-Thorsten Peter, Saab Mansour and Hermann Ney	483
<i>Automatic Extraction of Turkish Hypernym-Hyponym Pairs From Large Corpus</i>	
Savas Yildirim and Tugba Yildiz.....	493
<i>Chinese Web Scale Linguistic Datasets and Toolkit</i>	
Chi-Hsin Yu and Hsin-Hsi Chen.....	501
<i>Developing and Evaluating a Computer-Assisted Near-Synonym Learning System</i>	
Liang-Chih Yu and Kai-Hsiang Hsu.....	509
<i>Arabic Morphological Analyzer with Agglutinative Affix Morphemes and Fusional Concatenation Rules</i>	
Fadi Zaraket and Jad Makhoul.....	517
<i>SMR-Cmp: Square-Mean-Root Approach to Comparison of Monolingual Contrastive Corpora</i>	
HuaRui Zhang, Chu-Ren Huang and Francesca Quattri.....	527
<i>A Machine Learning Approach to Convert CCGbank to Penn Treebank</i>	
Xiaotian Zhang, Hai Zhao and Cong Hui.....	535