24th International Conference on Computational Linguistics

(COLING 2012)

Demonstrations

Mumbai, India
8-15 December 2012

Editors:

Martin Kay
Christian Boitet

# Table of Contents

Complex Predicates in Telugu: A Computational Perspective  
Rahul Balusu ................................................................. 1

Heloise — An Ariane-G5 Compatible Environment for Developing Expert MT Systems Online  
Vincent Berment and Christian Boitet ......................................... 9

Keyphrase Extraction in Scientific Articles: A Supervised Approach  
Pinaki Bhaskar, Kishorjit Nongmeikapam and Sivaji Bandyopadhyay ............. 17

IKAR: An Improved Kit for Anaphora Resolution for Polish  
Bartosz Broda, Łukasz Burdka and Marek Maziarz .................................. 25

Intention Analysis for Sales, Marketing and Customer Service  
Cohan Sujay Carlos and Madhulika Yalamanchi ..................................... 33

Authorship Identification in Bengali Literature: a Comparative Analysis  
Tanmoy Chakraborty ........................................................................ 41

Word Root Finder: a Morphological Segmentor Based on CRF  
Joseph Z. Chang and Jason S. Chang................................................ 51

An Efficient Technique for De-Noising Sentences using Monolingual Corpus and Synonym Dictionary  
Sanjay Chatterji, Diptesh Chatterjee and Sudeshna Sarkar ......................... 59

An Example-Based Japanese Proofreading System for Offshore Development  
Yuchang Cheng and Tomoki Nagase................................................. 67

DomEx: Extraction of Sentiment Lexicons for Domains and Meta-Domains  
Ilia Chetviorkin and Natalia Loukachevitch ......................................... 77

On the Romanian Rhyme Detection  
Alina Ciobanu and Liviu P. Dinu ..................................................... 87

Hierarchical Dialogue Policy Learning using Flexible State Transitions and Linear Function Approximation  
Heriberto Cuayáhuitl, Ivana Kruijff-Korbayová and Nina Dethlefs ..................... 95

Automated Paradigm Selection for FSA based Konkani Verb Morphological Analyzer  
Shilpa Desai, Jyoti Pawar and Pushpak Bhattacharyya .............................. 103

Hindi and Marathi to English NE Transliteration Tool using Phonology and Stress Analysis  
Manikrao Dhore, Shantanu Dixit and Ruchi Dhore ................................. 111

Dealing with the Grey Sheep of the Romanian Gender System, the Neuter  
Liviu P. Dinu, Vlad Niculae and Maria Sulea ........................................ 119

Authorial Studies using Ranked Lexical Features  
Liviu P. Dinu and Sergiu Nisioi .......................................................... 125
ScienQuest: a Treebank Exploitation Tool for non NLP-Specialists
Achille Falaise, Olivier Kraif, Agnès Tutin and David Rouquet ....................... 131

An In-Context and Collaborative Software Localisation Model
Amel Fraisse, Christian Boitet and Valérie Bellynck................................. 141

Efficient Feedback-based Feature Learning for Blog Distillation as a Terabyte Challenge
Dehong Gao, Wenjie Li and Renxian Zhang ........................................ 147

Beyond Twitter Text: A Preliminary Study on Twitter Hyperlink and its Application
Dehong Gao, Wenjie Li and Renxian Zhang ........................................ 155

Rule Based Hindi Part of Speech Tagger
Navneet Garg, Vishal Goyal and Suman Preet...................................... 163

Fangorn: A System for Querying very large Treebanks
Sumukh Ghodke and Steven Bird .................................................. 175

CRAB Reader: A Tool for Analysis and Visualization of Argumentative Zones in Scientific Literature
Yufan Guo, Ilona Silins, Roi Reichart and Anna Korhonen .......................... 183

Automatic Punjabi Text Extractive Summarization System
Vishal Gupta and Gurpreet Lehal .................................................. 191

Complete Pre Processing Phase of Punjabi Text Extractive Summarization System
Vishal Gupta and Gurpreet Lehal .................................................. 199

Revisiting Arabic Semantic Role Labeling using SVM Kernel Methods
Laurel Hart, Hassan Alam and Aman Kumar ....................................... 207

fokas: Formerly Known As – A Search Engine Incorporating Named Entity Evolution
Helge Holzmann, Gerhard Gossen and Nina Tahmasebi............................ 215

An Annotation System for Development of Chinese Discourse Corpus
Hen-Hsen Huang and Hsin-Hsi Chen .............................................. 223

Modeling Pollyanna Phenomena in Chinese Sentiment Analysis
Ting-Hao Huang, Ho-Cheng Yu and Hsin-Hsi Chen ................................ 231

Eating Your Own Cooking: Automatically Linking Wordnet Synsets of Two Languages
Salil Joshi, Arindam Chatterjee, Arun Karthikeyan Karra and Pushpak Bhattacharyya.................................................. 239

I Can Sense It: a Comprehensive Online System for WSD
Salil Joshi, Mitesh M. Khapra and Pushpak Bhattacharyya ........................ 247

Collaborative Computer-Assisted Translation Applied to Pedagogical Documents and Literary Works
Ruslan Kalitvianski, Christian Boitet and Valérie Bellynck ......................... 255

Discrimination-Net for Hindi
Diptesh Kanojia, Arindam Chatterjee, Salil Joshi and Pushpak Bhattacharyya .... 261
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule Based Urdu Stemmer</td>
<td>Rohit Kansal, Vishal Goyal and Gurpreet Singh Lehal</td>
<td>267</td>
</tr>
<tr>
<td>JMaxAlign: A Maximum Entropy Parallel Sentence Alignment Tool</td>
<td>Max Kaufmann</td>
<td>277</td>
</tr>
<tr>
<td>MIKE: An Interactive Microblogging Keyword Extractor using Contextual Semantic Smoothing</td>
<td>Osama Khan and Asim Karim</td>
<td>289</td>
</tr>
<tr>
<td>Domain Based Classification of Punjabi Text Documents</td>
<td>Nidhi Krail and Vishal Gupta</td>
<td>297</td>
</tr>
<tr>
<td>Open Information Extraction for SOV Language Based on Entity-Predicate Pair Detection</td>
<td>Woong-Ki Lee, Yeon-Su Lee, Hyoung-Gyu Lee, Won-Ho Ryu and Hae-Chang Rim</td>
<td>305</td>
</tr>
<tr>
<td>An Omni-Font Gurmukhi to Shahnukhi Transliteration System</td>
<td>Gurpreet Singh Lehal, Tejinder Singh Saini and Savleen Kaur Chowdhary</td>
<td>313</td>
</tr>
<tr>
<td>THUTR: A Translation Retrieval System</td>
<td>Chunyang Liu, Qi Liu, Yang Liu and Maosong Sun</td>
<td>321</td>
</tr>
<tr>
<td>Recognition of Named-Event Passages in News Articles</td>
<td>Luis Marujo, Wang Ling, Anatole Gershman, Jaime Carbonell, João P Neto and David Matos</td>
<td>329</td>
</tr>
<tr>
<td>Nonparametric Model for Inupiaq Word Segmentation</td>
<td>Thuy Linh Nguyen and Stephan Vogel</td>
<td>337</td>
</tr>
<tr>
<td>Stemming Tigrinya Words for Information Retrieval</td>
<td>Omer Osman and Yoshiki Mikami</td>
<td>345</td>
</tr>
<tr>
<td>OpenWordNet-PT: An Open Brazilian Wordnet for Reasoning</td>
<td>Valeria de Paiva, Alexandre Rademaker and Gerard de Melo</td>
<td>353</td>
</tr>
<tr>
<td>WordNet Website Development And Deployment using Content Management Approach</td>
<td>Neha Prabhugaonkar, Apurva Nagvenkar, Venkatesh Prabhu and Ramdas Karmali</td>
<td>361</td>
</tr>
<tr>
<td>A Demo for Constructing Domain Ontology from Academic Papers</td>
<td>Feiliang Ren</td>
<td>369</td>
</tr>
<tr>
<td>A Practical Chinese-English ON Translation Method Based on ON's Distribution Characteristics on the Web</td>
<td>Feiliang Ren</td>
<td>377</td>
</tr>
<tr>
<td>Elissa: A Dialectal to Standard Arabic Machine Translation System</td>
<td>Wael Salloum and Nizar Habash</td>
<td>385</td>
</tr>
<tr>
<td>Domain Based Punjabi Text Document Clustering</td>
<td>Saurabh Sharma and Vishal Gupta</td>
<td>393</td>
</tr>
<tr>
<td>Open source multi-platform NooJ for NLP</td>
<td>Max Silberztein, Tamás Váradi and Marko Tadić</td>
<td>401</td>
</tr>
</tbody>
</table>
Punjabi Text-To-Speech Synthesis System
Parminder Singh and Gurpreet Singh Lehal ........................................ 409

EXCOTATE: An Add-on to MMAX2 for Inspection and Exchange of Annotated Data
Tobias Stadtfeld and Tibor Kiss .................................................... 417

Bulgarian Inflectional Morphology in Universal Networking Language
Velislava Stoykova ................................................................. 423

Central and South-East European Resources in META-SHARE
Marko Tadić and Tamás Váradi .................................................... 431

Markov Chains for Robust Graph-Based Commonsense Information Extraction
Niket Tandon, Dheeraj Rajagopal and Gerard de Melo ............................. 439

Visualization on Financial Terms via Risk Ranking from Financial Reports
Ming-Feng Tsai and Chuan-Ju Wang ............................................... 447

UNL Explorer
Hiroshi Uchida, Meiying Zhu and Md. Anwarus Salam Khan ....................... 453

An SMT-driven Authoring Tool
Sriram Venkatapathy and Shachar Mirkin .......................................... 459

Generating Questions from Web Community Contents
Baoxun Wang, Bingquan Liu, Chengjie Sun, Xiaolong Wang and Deyuan Zhang..... 467

Demo of iMAG Possibilities: MT-postediting, Translation Quality Evaluation, Parallel Corpus Production
Ling Xiao Wang, Ying Zhang, Christian Boitet and Valérie Bellynck ................. 475

Jane 2: Open Source Phrase-based and Hierarchical Statistical Machine Translation
Joern Wuebker, Matthias Huck, Stephan Peitz, Malte Nuhn, Markus Freitag, Jan-Thorsten Peter, Saab Mansour and Hermann Ney .................................................... 483

Automatic Extraction of Turkish Hypernym-Hyponym Pairs From Large Corpus
Savas Yildirim and Tugba Yildiz ..................................................... 493

Chinese Web Scale Linguistic Datasets and Toolkit
Chi-Hsin Yu and Hsin-Hsi Chen .................................................... 501

Developing and Evaluating a Computer-Assisted Near-Synonym Learning System
Liang-Chih Yu and Kai-Hsiang Hsu ................................................. 509

Arabic Morphological Analyzer with Agglutinative Affix Morphemes and Fusional Concatenation Rules
Fadi Zaraket and Jad Makhlouta .................................................... 517

SMR-Cmp: Square-Mean-Root Approach to Comparison of Monolingual Contrastive Corpora
HuaRui Zhang, Chu-Ren Huang and Francesca Quattri ............................... 527

A Machine Learning Approach to Convert CCGbank to Penn Treebank
Xiaotian Zhang, Hai Zhao and Cong Hui ........................................... 535