

## Table of Contents – Part I

### NLP System Architecture

- Thinking Outside the Box for Natural Language Processing ..... 1  
*Srinivas Bangalore*

### Lexical Resources

- A Graph-Based Method to Improve WordNet Domains ..... 17  
*Aitor González, German Rigau, and Mauro Castillo*
- Corpus-Driven Hyponym Acquisition for Turkish Language ..... 29  
*Savaş Yıldırım and Tuğba Yıldız*
- Automatic Taxonomy Extraction in Different Languages Using  
Wikipedia and Minimal Language-Specific Information ..... 42  
*Renato Domínguez García, Sebastian Schmidt,  
Christoph Rensing, and Ralf Steinmetz*
- Ontology-Driven Construction of Domain Corpus with Frame Semantics  
Annotations ..... 54  
*He Tan, Rajaram Kaliyaperumal, and Nirupama Benis*
- Building a Hierarchical Annotated Corpus of Urdu: The URDU.KON-  
TB Treebank ..... 66  
*Qaiser Abbas*

### Morphology and Syntax

- A Morphological Analyzer Using Hash Tables in Main Memory  
(MAHT) and a Lexical Knowledge Base ..... 80  
*Francisco J. Carreras-Riudavets, Juan C. Rodríguez-del-Pino,  
Zenón Hernández-Figueroa, and Gustavo Rodríguez-Rodríguez*
- Optimal Stem Identification in Presence of Suffix List ..... 92  
*Vasudevan N. and Pushpak Bhattacharyya*
- On the Adequacy of Three POS Taggers and a Dependency Parser ..... 104  
*Ramadan Alfared and Denis Béchet*

Will the Identification of Reduplicated Multiword Expression (RMWE) Improve the Performance of SVM Based Manipuri POS Tagging? . . . . .	117
<i>Kishorjit Nongmeikapam, Aribam Umananda Sharma, Laishram Martina Devi, Napoleon Keisam, Khangengbam Dilip Singh, and Sivaji Bandyopadhyay</i>	
On Formalization of Word Order Properties . . . . .	130
<i>Vladislav Kuboň, Markéta Lopatková, and Martin Plátek</i>	
Core-Periphery Organization of Graphemes in Written Sequences: Decreasing Positional Rigidity with Increasing Core Order . . . . .	142
<i>Md. Izhar Ashraf and Sitabhra Sinha</i>	
Discovering Linguistic Patterns Using Sequence Mining . . . . .	154
<i>Nicolas Béchet, Peggy Cellier, Thierry Charnois, and Bruno Crémilleux</i>	
What about Sequential Data Mining Techniques to Identify Linguistic Patterns for Stylistics? . . . . .	166
<i>Solen Quiniou, Peggy Cellier, Thierry Charnois, and Dominique Legallois</i>	
Resolving Syntactic Ambiguities in Natural Language Specification of Constraints . . . . .	178
<i>Imran Sarwar Bajwa, Mark Lee, and Behzad Bordbar</i>	
A Computational Grammar of Sinhala . . . . .	188
<i>Chamila Liyanage, Randil Pushpananda, Dulip Lakmal Herath, and Ruwan Weerasinghe</i>	
Automatic Identification of Persian Light Verb Constructions . . . . .	201
<i>Bahar Salehi, Narjes Askarian, and Afsaneh Fazly</i>	
<b>Word Sense Disambiguation and Named Entity Recognition</b>	
A Cognitive Approach to Word Sense Disambiguation . . . . .	211
<i>Sudakshina Dutta and Anupam Basu</i>	
A Graph-Based Approach to WSD Using Relevant Semantic Trees and N-Cliques Model . . . . .	225
<i>Yoan Gutiérrez, Sonia Vázquez, and Andrés Montoyo</i>	
Using Wiktionary to Improve Lexical Disambiguation in Multiple Languages . . . . .	238
<i>Kiem-Hieu Nguyen and Cheol-Young Ock</i>	

Two Stages Based Organization Name Disambiguity . . . . .	249
<i>Shu Zhang, Jianwei Wu, Dequan Zheng, Yao Meng, Yingju Xia, and Hao Yu</i>	
Optimizing CRF-Based Model for Proper Name Recognition in Polish Texts . . . . .	258
<i>Michał Marcińczuk and Maciej Janicki</i>	
Methods of Estimating the Number of Clusters for Person Cross Document Coreference Task . . . . .	270
<i>Octavian Popescu and Roberto Zanolì</i>	
Coreference Resolution Using Tree CRFs . . . . .	285
<i>Vijay Sundar Ram R. and Sobha Lalitha Devi</i>	
Arabic Entity Graph Extraction Using Morphology, Finite State Machines, and Graph Transformations . . . . .	297
<i>Jad Makhoul, Fadi Zaraket, and Hamza Harkous</i>	
Integrating Rule-Based System with Classification for Arabic Named Entity Recognition . . . . .	311
<i>Sherief Abdallah, Khaled Shaalan, and Muhammad Shoaib</i>	

## Semantics and Discourse

Space Projections as Distributional Models for Semantic Composition . . . . .	323
<i>Paolo Annesi, Valerio Storch, and Roberto Basili</i>	
Distributional Models and Lexical Semantics in Convolution Kernels . . .	336
<i>Daniò Croce, Simone Filice, and Roberto Basili</i>	
Multiple Level of Referents in Information State . . . . .	349
<i>Gábor Alberti and Márton Károly</i>	
Inferring the Scope of Negation in Biomedical Documents . . . . .	363
<i>Miguel Ballesteros, Virginia Francisco, Alberto Díaz, Jesús Herrera, and Pablo Gervás</i>	
LDA-Frames: An Unsupervised Approach to Generating Semantic Frames . . . . .	376
<i>Jiří Materna</i>	
Unsupervised Acquisition of Axioms to Paraphrase Noun Compounds and Genitives . . . . .	388
<i>Anselmo Peñas and Ekaterina Ovchinnikova</i>	
Age-Related Temporal Phrases in Spanish and Italian . . . . .	402
<i>Sofía N. Galicia-Haro and Alexander Gelbukh</i>	

Can Modern Statistical Parsers Lead to Better Natural Language Understanding for Education? .....	415
<i>Umair Z. Ahmed, Arpit Kumar, Monojit Choudhury, and Kalika Bali</i>	
Exploring Classification Concept Drift on a Large News Text Corpus ...	428
<i>Artur Šilić and Bojana Dalbelo Bašić</i>	
An Empirical Study of Recognizing Textual Entailment in Japanese Text .....	438
<i>Quang Nhat Minh Pham, Le Minh Nguyen, and Akira Shimazu</i>	
Automated Detection of Local Coherence in Short Argumentative Essays Based on Centering Theory .....	450
<i>Vasile Rus and Nobal Niraula</i>	
A Symbolic Approach for Automatic Detection of Nuclearity and Rhetorical Relations among Intra-sentence Discourse Segments in Spanish .....	462
<i>Iria da Cunha, Eric SanJuan, Juan-Manuel Torres-Moreno, M. Teresa Cabré, and Gerardo Sierra</i>	
<b>Sentiment Analysis, Opinion Mining, and Emotions</b>	
Feature Specific Sentiment Analysis for Product Reviews .....	475
<i>Subhabrata Mukherjee and Pushpak Bhattacharyya</i>	
Biographies or Blenders: Which Resource Is Best for Cross-Domain Sentiment Analysis? .....	488
<i>Natalia Ponomareva and Mike Thelwall</i>	
A Generate-and-Test Method of Detecting Negative-Sentiment Sentences .....	500
<i>Yoonjung Choi, Hyo-Jung Oh, and Sung-Hyon Myaeng</i>	
Roles of Event Actors and Sentiment Holders in Identifying Event-Sentiment Association .....	513
<i>Anup Kumar Kolya, Dipankar Das, Asif Ekbal, and Sivaji Bandyopadhyay</i>	
Applying Sentiment and Social Network Analysis in User Modeling .....	526
<i>Mohammadreza Shams, Mohammadtaghi Saffar, Azadeh Shakery, and Hesham Faily</i>	
The 5W Structure for Sentiment Summarization-Visualization-Tracking .....	540
<i>Amitava Das, Sivaji Bandyopadhyay, and Björn Gambäck</i>	

The Naive Bayes Classifier in Opinion Mining: In Search of the Best Feature Set .....	556
<i>Liviu P. Dinu and Iulia Iuga</i>	
A Domain Independent Framework to Extract and Aggregate Analogous Features in Online Reviews .....	568
<i>Archana Bhattarai, Nobal Niraula, Vasile Rus, and King-Ip Lin</i>	
Learning Lexical Subjectivity Strength for Chinese Opinionated Sentence Identification .....	580
<i>Xin Wang and Guohong Fu</i>	
Building Subjectivity Lexicon(s) from Scratch for Essay Data.....	591
<i>Beata Beigman Klebanov, Jill Burstein, Nitin Madnani, Adam Faulkner, and Joel Tetreault</i>	
Emotion Ontology Construction from Chinese Knowledge .....	603
<i>Peilin Jiang, Fei Wang, Fuji Ren, and Nanning Zheng</i>	
<b>Author Index</b> .....	615



## Table of Contents – Part II

### Natural Language Generation

Exploring Extensive Linguistic Feature Sets in Near-Synonym Lexical Choice . . . . .	1
<i>Mari-Sanna Paukkeri, Jaakko Väyrynen, and Antti Arppe</i>	
Abduction in Games for a Flexible Approach to Discourse Planning . . . .	13
<i>Ralf Klabunde, Sebastian Reuße, and Björn Schlünder</i>	

### Machine Translation and Multilingualism

Document-Specific Statistical Machine Translation for Improving Human Translation Productivity . . . . .	25
<i>Salim Roukos, Abraham Ittycheriah, and Jian-Ming Xu</i>	
Minimum Bayes Risk Decoding with Enlarged Hypothesis Space in System Combination . . . . .	40
<i>Tsuyoshi Okita and Josef van Genabith</i>	
Phrasal Syntactic Category Sequence Model for Phrase-Based MT . . . . .	52
<i>Hailong Cao, Eiichiro Sumita, Tiejun Zhao, and Sheng Li</i>	
Integration of a Noun Compound Translator Tool with Moses for English-Hindi Machine Translation and Evaluation . . . . .	60
<i>Prashant Mathur and Soma Paul</i>	
Neoclassical Compound Alignments from Comparable Corpora . . . . .	72
<i>Rima Harastani, Béatrice Daille, and Emmanuel Morin</i>	
QAlign : A New Method for Bilingual Lexicon Extraction from Comparable Corpora . . . . .	83
<i>Amir Hazem and Emmanuel Morin</i>	
Aligning the Un-Alignable — A Pilot Study Using a Noisy Corpus of Nonstandardized, Semi-parallel Texts . . . . .	97
<i>Florian Petran</i>	
Parallel Corpora for WordNet Construction: Machine Translation vs. Automatic Sense Tagging . . . . .	110
<i>Antoni Oliver and Salvador Climent</i>	
Method to Build a Bilingual Lexicon for Speech-to-Speech Translation Systems . . . . .	122
<i>Keiji Yasuda, Andrew Finch, and Eiichiro Sumita</i>	

## Text Categorization and Clustering

A Fast Subspace Text Categorization Method Using Parallel Classifiers . . . . .	132
<i>Nandita Tripathi, Michael Oakes, and Stefan Wermter</i>	
Research on Text Categorization Based on a Weakly-Supervised Transfer Learning Method . . . . .	144
<i>Dequan Zheng, Chenghe Zhang, Geli Fei, and Tiejun Zhao</i>	
Fuzzy Combinations of Criteria: An Application to Web Page Representation for Clustering . . . . .	157
<i>Alberto Pérez García-Plaza, Víctor Fresno, and Raquel Martínez</i>	
Clustering Short Text and Its Evaluation . . . . .	169
<i>Prajol Shrestha, Christine Jacquin, and Béatrice Daille</i>	

## Information Extraction and Text Mining

Information Extraction from Webpages Based on DOM Distances . . . . .	181
<i>Carlos Castillo, Héctor Valero, José Guadalupe Ramos, and Josep Silva</i>	
Combining Flat and Structured Approaches for Temporal Slot Filling or: How Much to Compress? . . . . .	194
<i>Qi Li, Javier Artiles, Taylor Cassidy, and Heng Ji</i>	
Event Annotation Schemes and Event Recognition in Spanish Texts . . . . .	206
<i>Dina Wonsever, Aiala Rosá, Marisa Malcuori, Guillermo Moncecchi, and Alan Descoins</i>	
Automatically Generated Noun Lexicons for Event Extraction . . . . .	219
<i>Béatrice Arnulphy, Xavier Tannier, and Anne Vilnat</i>	
Lexical Acquisition for Clinical Text Mining Using Distributional Similarity . . . . .	232
<i>John Carroll, Rob Koeling, and Shivani Puri</i>	
Developing an Algorithm for Mining Semantics in Texts . . . . .	247
<i>Minhua Huang and Robert M. Haralick</i>	
Mining Market Trend from Blog Titles Based on Lexical Semantic Similarity . . . . .	261
<i>Fei Wang and Yunfang Wu</i>	

## Information Retrieval and Question Answering

Ensemble Approach for Cross Language Information Retrieval . . . . .	274
<i>Dinesh Mavaluru, R. Shriram, and W. Aisha Banu</i>	



Web Image Annotation Using an Effective Term Weighting . . . . .	286
<i>Vundavalli Srinivasarao and Vasudeva Varma</i>	
Metaphone-pt_BR: The Phonetic Importance on Search and Correction of Textual Information . . . . .	297
<i>Carlos C. Jordão and João Luís G. Rosa</i>	
Robust and Fast Two-Pass Search Method for Lyric Search Covering Erroneous Queries Due to Mishearing . . . . .	306
<i>Xin Xu and Tsuneo Kato</i>	
Bootstrap-Based Equivalent Pattern Learning for Collaborative Question Answering . . . . .	318
<i>Tianyong Hao and Eugene Agichtein</i>	
How to Answer Yes/No Spatial Questions Using Qualitative Reasoning? . . . . .	330
<i>Marcin Walas</i>	
Question Answering and Multi-search Engines in Geo-Temporal Information Retrieval . . . . .	342
<i>Fernando S. Peregrino, David Tomás, and Fernando Llopis Pascual</i>	

## Document Summarization

Using Graph Based Mapping of Co-occurring Words and Closeness Centrality Score for Summarization Evaluation . . . . .	353
<i>Niraj Kumar, Kannan Srinathan, and Vasudeva Varma</i>	
Combining Syntax and Semantics for Automatic Extractive Single-Document Summarization . . . . .	366
<i>Araly Barrera and Rakesh Verma</i>	
Combining Summaries Using Unsupervised Rank Aggregation . . . . .	378
<i>Girish Keshav Palshikar, Shailesh Deshpande, and G. Athiappan</i>	
Using Wikipedia Anchor Text and Weighted Clustering Coefficient to Enhance the Traditional Multi-document Summarization . . . . .	390
<i>Niraj Kumar, Kannan Srinathan, and Vasudeva Varma</i>	
Extraction of Relevant Figures and Tables for Multi-document Summarization . . . . .	402
<i>Ashish Sadh, Amit Sahu, Devesh Srivastava, Ratna Sanyal, and Sudip Sanyal</i>	
Towards Automatic Generation of Catchphrases for Legal Case Reports . . . . .	414
<i>Filippo Galgani, Paul Compton, and Achim Hoffmann</i>	

**Applications**

A Dataset for the Evaluation of Lexical Simplification . . . . .	426
<i>Jan De Belder and Marie-Francine Moens</i>	
Text Content Reliability Estimation in Web Documents: A New Proposal . . . . .	438
<i>Luis Sanz, Héctor Allende, and Marcelo Mendoza</i>	
Fine-Grained Certainty Level Annotations Used for Coarser-Grained E-Health Scenarios: Certainty Classification of Diagnostic Statements in Swedish Clinical Text . . . . .	450
<i>Sumithra Velupillai and Maria Kvist</i>	
Combining Confidence Score and Mal-rule Filters for Automatic Creation of Bangla Error Corpus: Grammar Checker Perspective . . . . .	462
<i>Bibekananda Kundu, Sutanu Chakraborti, and Sanjay Kumar Choudhury</i>	
Predictive Text Entry for Agglutinative Languages Using Unsupervised Morphological Segmentation . . . . .	478
<i>Miikka Silfverberg, Krister Lindén, and Mirka Hyvärinen</i>	
Comment Spam Classification in Blogs through Comment Analysis and Comment-Blog Post Relationships . . . . .	490
<i>Ashwin Rajadesingan and Anand Mahendran</i>	
Detecting Players Personality Behavior with Any Effort of Concealment . . . . .	502
<i>Fazel Keshtkar, Candice Burkett, Arthur Graesser, and Haiying Li</i>	
<b>Author Index</b> . . . . .	515