

## Table of Contents

---

### Computational Linguistics Research

---

#### Computational Linguistics Formalisms

An Overview of Probabilistic Tree Transducers for Natural Language Processing.....	1
<i>Kevin Knight and Jonathan Graehl</i>	
A Modular Account of Information Structure in Extensible Dependency Grammar.....	26
<i>Ralph Debusmann, Oana Postolache, and Maarika Traat</i>	
Modelling Grammatical and Lexical Knowledge: A Declarative Approach.....	38
<i>Palmira Marrafa</i>	
Constructing a Parser for Latin.....	49
<i>Cornelis H.A. Koster</i>	
Parsing Korean Case Phenomena in a Type-Feature Structure Grammar.....	61
<i>Jong-Bok Kim and Jaehyung Yang</i>	
A Computational Model of the Spanish Clitic System.....	73
<i>Luis A. Pineda, Ivan V. Meza</i>	
A Parallel Approach to Syllabification.....	83
<i>Anca Dinu and Liviu P. Dinu</i>	

#### Semantics and Discourse

Towards Developing Probabilistic Generative Models for Reasoning with Natural Language Representations.....	87
<i>Daniel Marcu and Ana-Maria Popescu</i>	
Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing.....	99
<i>Lei Shi and Rada Mihalcea</i>	
Assigning Function Tags with a Simple Model.....	111
<i>Vasile Rus and Kirtan Desai</i>	
Finding Discourse Relations in Student Essays.....	115
<i>Rohana Mahmud and Allan Ramsay</i>	

#### Parsing and Syntactic Disambiguation

Regional vs. Global Finite-State Error Repair.....	119
<i>Manuel Vilares, Juan Otero, and Jorge Graña</i>	

Lexicalized Beam Thresholding Parsing with Prior and Boundary Estimates .....	131
<i>Deyi Xiong, Qun Liu, and Shouxun Lin</i>	
Unsupervised Evaluation of Parser Robustness.....	141
<i>Johnny Bigert, Jonas Sjöbergh, Ola Knutsson, and Magnus Sahlgren</i>	
Mutual Information Independence Model using Kernel Density Estimation for Segmenting and Labeling Sequential Data .....	153
<i>ZHOU GuoDong, YANG LingPeng, SU Jian, JI DongHong</i>	
Applying Conditional Random Fields to Chinese Shallow Parsing .....	165
<i>Yongmei Tan, Tianshun Yao, Qing Chen and Jingbo Zhu</i>	
Corpus-Based Thesaurus vs. WordNet: A Comparison of Backoff Techniques for Unsupervised PP Attachment.....	172
<i>Hiram Calvo, Alexander Gelbukh, Adam Kilgarriff</i>	
<b>Morphology</b>	
Automatic Recognition of Czech Derivational Prefixes.....	183
<i>Alfonso Medina Urrea and Jaroslava Hlaváčová</i>	
Korma 2003: Newly Improved Korean Morpheme Analysis Module for Reducing Terminological and Spacing Errors in Document Analysis .....	192
<i>Ho-cheol Choi and Sang-yong Han</i>	
Word Extraction Based on Semantic Constraints in Chinese Word-formation .....	196
<i>Maosong Sun, Shengfen Luo, and Benjamin K T'sou</i>	
Using Directed Graph based BDMM Algorithm for Chinese Word Segmentation .....	209
<i>Yaodong Chen, Ting Wang, and Huowang Chen</i>	
<b>Anaphora and Coreference</b>	
Entity-Based Noun Phrase Coreference Resolution .....	212
<i>Xiaofeng Yang, Jian Su, and Lingpeng Yang</i>	
The Right Frontier Constraint as Conditional.....	216
<i>Claudia Sassen and Peter Kühnlein</i>	
<b>Word Sense Disambiguation</b>	
Name Discrimination by Clustering Similar Contexts .....	220
<i>Ted Pedersen, Amruta Purandare, and Anagha Kulkarni</i>	
Word Sense Disambiguation by Semi-Supervised Learning .....	232
<i>Zheng-Yu Niu, Dong-Hong Ji, Chew-Lim Tan, and Ling-Peng Yang</i>	
Crossing Parallel Corpora and Multilingual Lexical Databases for WSD.....	236
<i>Alfio Massimiliano Gliozzo, Marcello Ranieri, and Carlo Strapparava</i>	
A Mapping between Classifiers and Training Conditions for WSD.....	240
<i>Aarón Pancardo-Rodríguez, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda, Paolo Rosso</i>	

Multiwords and Word Sense Disambiguation .....	244
<i>Victoria Arranz, Jordi Atserias and Mauro Castillo</i>	
Context Expansion with Global Keywords for a Conceptual Density-Based WSD .....	256
<i>Davide Buscaldi, Paolo Rosso, and Manuel Montes y Gómez</i>	
Two Web-based Approaches for Noun Sense Disambiguation.....	260
<i>Paolo Rosso, Manuel Montes-y-Gómez, Davide Buscaldi, Aarón Pancardo-Rodríguez, and Luis Villaseñor Pineda</i>	
<b>Lexical Resources</b>	
Finding Instance Names and Alternative Glosses on the Web: WordNet Reloaded .....	272
<i>Marius Paşca</i>	
Automatic Synonym Acquisition Based on Matching of Definition Sentences in Multiple Dictionaries .....	284
<i>Masaki Murata, Toshiyuki Kanamaru, and Hitoshi Isahara</i>	
Enriching WordNet with Derivational Subnets .....	296
<i>Karel Pala and Radek Sedláček</i>	
Customisable Semantic Analysis of Texts.....	303
<i>Vivi Nastase and Stan Szpakowicz</i>	
ITOLDU, a Web Service to Pool Technical Lexical Terms in a Learning Environment and Contribute to Multilingual Lexical Databases.....	315
<i>Valérie Bellynck, Christian Boitet, John Kenwright</i>	
Building a Situation-based Language Knowledge Base .....	324
<i>Qiang ZHOU, Zushun CHEN</i>	
Unsupervised Learning of P NP P Word Combinations.....	328
<i>Sofía N. Galicia-Haro, Alexander Gelbukh</i>	
<b>Natural Language Generation</b>	
Evaluating Evaluation Methods for Generation in the Presence of Variation .....	332
<i>Amanda Stent, Matthew Marge, and Mohit Singhai</i>	
Reconciling Parameterization, Configurability and Optimality in Natural Language Generation via Multiparadigm Programming .....	343
<i>Jorge Marques Pelizzoni, Maria das Graças Volpe Nunes</i>	
<b>Machine Translation</b>	
Message Automata for Messages with Variants, and Methods for their Translation.....	348
<i>Christian Boitet</i>	
The UNL Initiative: An Overview.....	368
<i>Igor Boguslavsky; Jesús Cardeñosa; Carolina Gallardo; Luis Iraola</i>	

Interactive Resolution of Intrinsic and Translational Ambiguity in a Machine Translation System .....	379
<i>Igor M. Boguslavsky, Leonid L. Iomdin, Alexander V. Lazursky, Leonid G. Mityushin, Victor G. Sizov, Leonid G. Kreydlin, Alexander S. Berdichevsky</i>	
Chinese-Japanese Clause Alignment .....	391
<i>Xiaojie Wang, Fuji Ren</i>	
Direct Combination of Spelling and Pronunciation Information for Robust Back-Transliteration .....	402
<i>Slaven Bilac and Hozumi Tanaka</i>	
<b>Speech and Natural Language Interfaces</b>	
A Prosodic Diphone Database for Korean Text-to-Speech Synthesis System .....	414
<i>Kyuchul Yoon</i>	
On a Pitch Detection Method Using Noise Reduction .....	418
<i>Jong Kuk Kim, Ki Young Lee, Myung Jin Bae</i>	
Toward Acoustic Models for Languages with Limited Linguistic Resources .....	422
<i>Luis Villaseñor-Pineda, Viet Bac Le, Manuel Montes-y-Gómez, and Manuel Pérez-Coutiño</i>	
A Study On Pitch Detection in Time-Frequency Hybrid Domain .....	426
<i>Wangrae Jo, Jongkuk Kim, Myungjin Bae</i>	
VoiceUNL: a Semantic Representation of Emotions within Universal Networking Language Formalism Based on a Dialogue Corpus Analysis .....	430
<i>Mutsuko Tomokiyo, Gérard Chollet</i>	
Combining Multiple Statistical Classifiers to Improve the Accuracy of Task Classification .....	441
<i>Wei-Lin Wu, Ru-Zhan Lu, Feng Gao, Yan Yuan</i>	
<b>Language Documentation</b>	
A Finite State Network for Phonetic Text Processing .....	452
<i>Edward John Garrett</i>	
Language Documentation: the Nahuatl Grammar .....	463
<i>Mike Maxwell, Jonathan D. Amith</i>	

---

## Intelligent Text Processing Applications

---

### Information Extraction

Creating Subjective and Objective Sentence Classifiers from Unannotated Texts .....	475
<i>Janyce Wiebe and Ellen Riloff</i>	

Instance Pruning by Filtering Uninformative Words: an Information Extraction Case Study .....	487
<i>Alfio Massimiliano Gliozzo, Claudio Giuliano, and Raffaella Rinaldi</i>	
Incremental Information Extraction Using Tree-based Context Representations.....	499
<i>Christian Siefkes</i>	
Learning Information Extraction Rules for Protein Annotation from Unannotated Corpora.....	511
<i>Jee-Hyub Kim and Melanie Hilario</i>	
Transformation-Based Information Extraction Using Learned Meta-Rules.....	523
<i>Un Yong Nahm</i>	
A Machine Learning Approach to Information Extraction.....	527
<i>Alberto Téllez-Valero, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda</i>	
Automatic Time Expression Labeling for English and Chinese Text.....	536
<i>Kadri Hacioglu, Ying Chen, Benjamin Douglas</i>	
Integrating Natural Language Techniques in OO-Method .....	548
<i>Isabel Díaz, Lidia Moreno, Inmaculada Fuentes, and Oscar Pastor</i>	
<b>Information Retrieval</b>	
Document Re-ordering Based on Key Terms in Top Retrieved Documents .....	560
<i>Yang Lingpeng, Ji Donghong, Nie Yu, Zhou Guodong</i>	
Merging Case Relations into VSM to Improve Information Retrieval Precision .....	572
<i>Wang Hongtao, Sun Maosong, Liu Shaoming</i>	
Evaluating Document-to-document Relevance based on Document Language Model: Modeling, Implementation and Performance Evaluation.....	581
<i>Ge Yu, Xiaoguang Li, Yubin Bao, Daling Wang</i>	
Retrieval Efficiency of Normalized Query Expansion .....	592
<i>Sofia Stamou and Dimitris Christodoulakis</i>	
Selecting Interesting Articles Using Their Similarity Based Only on Positive Examples .....	596
<i>Jiří Hroza and Jan Žižka</i>	
<b>Question Answering</b>	
Question Classification in Spanish and Portuguese.....	600
<i>Thamar Solorio, Manuel Pérez-Coutiño, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda, and Aurelio López-López</i>	
Learning the Query Generation Patterns.....	608
<i>Marcin Skowron and Kenji Araki</i>	
Exploiting Question Concepts for Query Expansion.....	612
<i>Hae-Jung Kim, Ki-Dong Bu, Junghyun Kim and Sang-Jo Lee</i>	
Experiment on Combining Sources of Evidence for Passage Retrieval .....	616
<i>Alexander Gelbukh, NamO Kang, SangYong Han</i>	

## **Summarization**

Summarisation through Discourse Structure .....	620
<i>Dan Cristea, Oana Postolache, and Ionuț Pistol</i>	
LexTrim: A Lexical Cohesion based Approach to Parse-and-Trim Style Headline Generation .....	632
<i>Ruichao Wang, Nicola Stokes, William Doran, Eamonn Newman, John Dunnion, Joe Carthy</i>	
Generating Headline Summary from a Document Set.....	636
<i>Kamal Sarkar, Sivaji Bandyopadhyay</i>	
Extractive Summarization Based on Word Information and Sentence Position.....	640
<i>Carlos MÉNDEZ Cruz and Alfonso MEDINA Urrea</i>	
Automatic Extraction and Learning of Keyphrases from Scientific Articles .....	644
<i>Yaakov HaCohen-Kerner, Zuriel Gross, Asaf Masa</i>	
Automatic Annotation of Corpora for Text Summarisation: A Comparative Study.....	657
<i>Constantin Orăsan</i>	

## **Text Classification, Categorization, and Clustering**

Techniques for Improving the Performance of Naive Bayes for Text Classification .....	669
<i>Karl-Michael Schneider</i>	
Efficient Modeling of Analogy.....	681
<i>Lars G. Johnsen and Christer Johansson</i>	
A Supervised Clustering Method for Text Classification .....	691
<i>Umarani Pappuswamy, Dumisizwe Bhembe, Pamela W. Jordan and Kurt VanLehn</i>	
Unsupervised Text Classification using Kohonen's Self Organizing Network.....	702
<i>Nirmalya Chowdhury and Diganta Saha</i>	
Enhancement of DTP Feature Selection Method for Text Categorization .....	706
<i>Edgar Moyotl-Hernández, Héctor Jiménez-Salazar</i>	
FASiL Adaptive Email Categorization System .....	710
<i>Yunqing Xia, Angelo Dalli, Yorick Wilks, Louise Guthrie</i>	
ESPClust: An Effective Skew Prevention Method for Model-based Document Clustering .....	722
<i>Xiaoguang Li, Ge Yu, Daling Wang, Yubin Bao</i>	
A Method of Rapid Prototyping of Evolving Ontologies .....	734
<i>Pavel Makagonov, Alejandro Ruiz Figueroa</i>	

## **Named Entity Recognition**

Resolution of Data Sparseness in Named Entity Recognition using Hierarchical Features and Feature Relaxation Principle.....	738
<i>ZHOU GuoDong, SU Jian, YANG LingPeng</i>	

Learning Named Entity Recognition in Portuguese from Spanish .....	750
<i>Thamar Solorio and Aurelio López López</i>	
A Simple Rule-based Approach to Organization Name Recognition in Chinese Text.....	757
<i>Wang Houfeng and Shi Wuguang</i>	
<b>Language Identification</b>	
Disentangling from Babylonian Confusion – Unsupervised Language Identification.....	761
<i>Chris Biemann, Sven Teresniak</i>	
On the Syllabic Similarities of Romance Languages.....	773
<i>Anca Dinu and Liviu P. Dinu</i>	
Automatic Language Identification Using Multivariate Analysis .....	777
<i>Vinosh Babu James J. and Baskaran Sankaran</i>	
<b>Spelling and Style Checking</b>	
Design and Development of a System for the Detection of Agreement Errors in Basque .....	781
<i>Arantza Díaz de Ilarraza, Koldo Gojenola and Maite Oronoz</i>	
An Experiment in Detection and Correction of Malapropisms through the Web.....	791
<i>Igor A. Bolshakov</i>	
A Paragraph Boundary Detection System .....	804
<i>Dmitriy Genzel</i>	
<b>Author Index</b> .....	815