

Alexander Gelbukh (Ed.)

Computational Linguistics and Intelligent Text Processing

6th International Conference, CICLing 2005
Mexico City, Mexico, February 13-19, 2005
Proceedings

Table of Contents

Computational Linguistics Research

Computational Linguistics Formalisms

Invited Paper

An Overview of Probabilistic Tree Transducers for Natural Language Processing

Kevin Knight, Jonathan Graehl 1

A Modular Account of Information Structure in Extensible Dependency Grammar

Ralph Debusmann, Oana Postolache, Maarika Traat 25

Modelling Grammatical and Lexical Knowledge: A Declarative Approach

Palmira Marrafa 37

Constructing a Parser for Latin

Cornelis H.A. Koster 48

Parsing Korean Case Phenomena in a Type-Feature Structure Grammar

Jong-Bok Kim, Jaehyung Yang 60

A Computational Model of the Spanish Clitic System

Luis A. Pineda, Ivan V. Meza 73

A Parallel Approach to Syllabification

Anca Dinu, Liviu P. Dinu 83

Semantics and Discourse

Invited Paper

Towards Developing Probabilistic Generative Models for Reasoning with Natural Language Representations

Daniel Marcu, Ana-Maria Popescu 88

Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing

Lei Shi, Rada Mihalcea 100

Assigning Function Tags with a Simple Model <i>Vasile Rus, Kirtan Desai</i>	112
Finding Discourse Relations in Student Essays <i>Rohana Mahmud, Allan Ramsay</i>	116
Parsing and Syntactic Disambiguation	
Regional Versus Global Finite-State Error Repair <i>Manuel Vilares, Juan Otero, Jorge Graña</i>	120
Lexicalized Beam Thresholding Parsing with Prior and Boundary Estimates <i>Deyi Xiong, Qun Liu, Shouxun Lin</i>	132
<i>Best Paper Award (2nd Place)</i>	
Unsupervised Evaluation of Parser Robustness <i>Johnny Bigert, Jonas Sjöbergh, Ola Knutsson, Magnus Sahlgren</i>	142
Mutual Information Independence Model Using Kernel Density Estimation for Segmenting and Labeling Sequential Data <i>Guodong Zhou, Lingpeng Yang, Jian Su, Donghong Ji</i>	155
Applying Conditional Random Fields to Chinese Shallow Parsing <i>Yongmei Tan, Tianshun Yao, Qing Chen, Jingbo Zhu</i>	167
Distributional Thesaurus Versus WordNet: A Comparison of Backoff Techniques for Unsupervised PP Attachment <i>Hiram Calvo, Alexander Gelbukh, Adam Kilgarriff</i>	177
Morphology	
Automatic Recognition of Czech Derivational Prefixes <i>Alfonso Medina Urrea, Jaroslava Hlaváčová</i>	189
Korma 2003: Newly Improved Korean Morpheme Analysis Module for Reducing Terminological and Spacing Errors in Document Analysis <i>Ho-cheol Choi, Sang-yong Han</i>	198
Word Extraction Based on Semantic Constraints in Chinese Word-Formation <i>Maosong Sun, Shengfen Luo, Benjamin K T'sou</i>	202
Using Directed Graph Based BDMM Algorithm for Chinese Word Segmentation <i>Yaodong Chen, Ting Wang, Huowang Chen</i>	214

Anaphora and Coreference

Entity-Based Noun Phrase Coreference Resolution <i>Xiaofeng Yang, Jian Su, Lingpeng Yang</i>	218
---	-----

The Right Frontier Constraint as Conditional <i>Claudia Sassen, Peter Kühnlein</i>	222
---	-----

Word Sense Disambiguation

Name Discrimination by Clustering Similar Contexts <i>Ted Pedersen, Amruta Purandare, Anagha Kulkarni</i>	226
--	-----

Word Sense Disambiguation by Semi-supervised Learning <i>Zheng-Yu Niu, Donghong Ji, Chew-Lim Tan, Lingpeng Yang</i>	238
--	-----

Crossing Parallel Corpora and Multilingual Lexical Databases for WSD <i>Alfio Massimiliano GlioZZo, Marcello Ranieri, Carlo Strapparava</i>	242
--	-----

A Mapping Between Classifiers and Training Conditions for WSD <i>Aarón Pancardo-Rodríguez, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda, Paolo Rosso</i>	246
--	-----

Multiwords and Word Sense Disambiguation <i>Victoria Arranz, Jordi Atserias, Mauro Castillo</i>	250
--	-----

Context Expansion with Global Keywords for a Conceptual Density-Based WSD <i>Davide Buscaldi, Paolo Rosso, Manuel Montes-y-Gómez</i>	263
--	-----

Two Web-Based Approaches for Noun Sense Disambiguation <i>Paolo Rosso, Manuel Montes-y-Gómez, Davide Buscaldi, Aarón Pancardo-Rodríguez, Luis Villaseñor Pineda</i>	267
--	-----

Lexical Resources

<i>Best Paper Award (1st Place)</i>	
Finding Instance Names and Alternative Glosses on the Web: WordNet Reloaded <i>Marius Paşca</i>	280

Automatic Synonym Acquisition Based on Matching of Definition Sentences in Multiple Dictionaries <i>Masaki Murata, Toshiyuki Kanamaru, Hitoshi Isahara</i>	293
---	-----

Enriching WordNet with Derivational Subnets <i>Karel Pala, Radek Sedláček</i>	305
Customisable Semantic Analysis of Texts <i>Vivi Nastase, Stan Szpakowicz</i>	312
ITOLDU, a Web Service to Pool Technical Lexical Terms in a Learning Environment and Contribute to Multilingual Lexical Databases <i>Valérie Bellynck, Christian Boitet, John Kenwright</i>	324
Building a Situation-Based Language Knowledge Base <i>Qiang Zhou, Zushun Chen</i>	333
Unsupervised Learning of P NP P Word Combinations <i>Sofia N. Galicia-Haro, Alexander Gelbukh</i>	337
Natural Language Generation	
Evaluating Evaluation Methods for Generation in the Presence of Variation <i>Amanda Stent, Matthew Marge, Mohit Singhai</i>	341
Reconciling Parameterization, Configurability and Optimality in Natural Language Generation via Multiparadigm Programming <i>Jorge Marques Pelizzoni, Maria das Graças Volpe Nunes</i>	352
Machine Translation	
<i>Invited Paper</i>	
Message Automata for Messages with Variants, and Methods for Their Translation <i>Christian Boitet</i>	357
The UNL Initiative: An Overview <i>Igor Boguslavsky, Jesús Cardeñosa, Carolina Gallardo, Luis Iraola</i>	377
Interactive Resolution of Intrinsic and Translational Ambiguity in a Machine Translation System <i>Igor M. Boguslavsky, Leonid L. Iomdin, Alexander V. Lazursky, Leonid G. Mityushin, Victor G. Sizov, Leonid G. Kreydlin, Alexander S. Berdichevsky</i>	388
Chinese-Japanese Clause Alignment <i>Xiaojie Wang, Fuji Ren</i>	400
Direct Combination of Spelling and Pronunciation Information for Robust Back-Transliteration <i>Slaven Bilac, Hozumi Tanaka</i>	413

Speech and Natural Language Interfaces

A Prosodic Diphone Database for Korean Text-to-Speech Synthesis System <i>Kyuchul Yoon</i>	425
On a Pitch Detection Method Using Noise Reduction <i>Jongkuk Kim, Ki Young Lee, Myung Jin Bae</i>	429
Toward Acoustic Models for Languages with Limited Linguistic Resources <i>Luis Villaseñor-Pineda, Viet Bac Le, Manuel Montes-y-Gómez, Manuel Pérez-Coutiño</i>	433
A Study on Pitch Detection in Time-Frequency Hybrid Domain <i>Wangrae Jo, Jongkuk Kim, Myung Jin Bae</i>	437
VoiceUNL: A Semantic Representation of Emotions Within Universal Networking Language Formalism Based on a Dialogue Corpus Analysis <i>Mutsuko Tomokiyo, Gérard Chollet</i>	441
Combining Multiple Statistical Classifiers to Improve the Accuracy of Task Classification <i>Wei-Lin Wu, Ru-Zhan Lu, Feng Gao, Yan Yuan</i>	452
Language Documentation	
A Finite State Network for Phonetic Text Processing <i>Edward John Garrett</i>	463
Language Documentation: The Nahuatl Grammar <i>Mike Maxwell, Jonathan D. Amith</i>	474

Intelligent Text Processing Applications

Information Extraction

<i>Invited Paper</i>	
Creating Subjective and Objective Sentence Classifiers from Unannotated Texts <i>Janyce Wiebe, Ellen Riloff</i>	486
Instance Pruning by Filtering Uninformative Words: An Information Extraction Case Study <i>Alfio Massimiliano Gliozzo, Claudio Giuliano, Raffaella Rinaldi</i>	498

Incremental Information Extraction Using Tree-Based Context Representations <i>Christian Siefkes</i>	510
<i>Best Paper Award (3rd Place)</i>	
Learning Information Extraction Rules for Protein Annotation from Unannotated Corpora <i>Jee-Hyub Kim, Melanie Hilario</i>	522
Transformation-Based Information Extraction Using Learned Meta-rules <i>Un Yong Nahm</i>	535
A Machine Learning Approach to Information Extraction <i>Alberto Téllez-Valero, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda</i>	539
Automatic Time Expression Labeling for English and Chinese Text <i>Kadri Hacioglu, Ying Chen, Benjamin Douglas</i>	548
Integrating Natural Language Techniques in OO-Method <i>Isabel Díaz, Lidia Moreno, Inmaculada Fuentes, Oscar Pastor</i>	560
Information Retrieval	
Document Re-ordering Based on Key Terms in Top Retrieved Documents <i>Yang Lingpeng, Ji Donghong, Nie Yu, Guodong Zhou</i>	572
Merging Case Relations into VSM to Improve Information Retrieval Precision <i>Hongtao Wang, Maosong Sun, Shaoming Liu</i>	584
Evaluating Document-to-Document Relevance Based on Document Language Model: Modeling, Implementation and Performance Evaluation <i>Ge Yu, Xiaoguang Li, Yubin Bao, Daling Wang</i>	593
Retrieval Efficiency of Normalized Query Expansion <i>Sofia Stamou, Dimitris Christodoulakis</i>	604
Selecting Interesting Articles Using Their Similarity Based Only on Positive Examples <i>Jiří Hroza, Jan Žižka</i>	608

Question Answering

Question Classification in Spanish and Portuguese

<i>Thamar Solorio, Manuel Pérez-Coutiño, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda, Aurelio López-López</i>	612
---	-----

Learning the Query Generation Patterns

<i>Marcin Skowron, Kenji Araki</i>	620
--	-----

Exploiting Question Concepts for Query Expansion

<i>Hae-Jung Kim, Ki-Dong Bu, Junghyun Kim, Sang-Jo Lee</i>	624
--	-----

Experiment on Combining Sources of Evidence for Passage Retrieval

<i>Alexander Gelbukh, NamO Kang, Sang-yong Han</i>	628
--	-----

Summarization

Summarisation Through Discourse Structure

<i>Dan Cristea, Oana Postolache, Ionuț Pistol</i>	632
---	-----

LexTrim: A Lexical Cohesion Based Approach to Parse-and-Trim Style

Headline Generation

<i>Ruichao Wang, Nicola Stokes, William Doran, Eamonn Newman, John Dunnion, Joe Carthy</i>	645
--	-----

Generating Headline Summary from a Document Set

<i>Kamal Sarkar, Sivaji Bandyopadhyay</i>	649
---	-----

Extractive Summarization Based on Word Information and Sentence Position

<i>Carlos Méndez Cruz, Alfonso Medina Urrea</i>	653
---	-----

Automatic Extraction and Learning of Keyphrases from Scientific Articles

<i>Yaakov HaCohen-Kerner, Zuriel Gross, Asaf Masa</i>	657
---	-----

Automatic Annotation of Corpora for Text Summarisation:

A Comparative Study

<i>Constantin Orăsan</i>	670
--------------------------------	-----

Text Classification, Categorization, and Clustering

Techniques for Improving the Performance of Naive Bayes for Text Classification

<i>Karl-Michael Schneider</i>	682
-------------------------------------	-----

Efficient Modeling of Analogy

<i>Lars G. Johnsen, Christer Johansson</i>	694
--	-----

A Supervised Clustering Method for Text Classification <i>Umarani Pappuswamy, Dumisizwe Bhembé, Pamela W. Jordan, Kurt VanLehn</i>	704
Unsupervised Text Classification Using Kohonen's Self Organizing Network <i>Nirmalya Chowdhury, Diganta Saha</i>	715
Enhancement of DTP Feature Selection Method for Text Categorization <i>Edgar Moyotl-Hernández, Héctor Jiménez-Salazar</i>	719
FASiL Adaptive Email Categorization System <i>Yunqing Xia, Angelo Dalli, Yorick Wilks, Louise Guthrie</i>	723
ESPClust: An Effective Skew Prevention Method for Model-Based Document Clustering <i>Xiaoguang Li, Ge Yu, Daling Wang, Yubin Bao</i>	735
A Method of Rapid Prototyping of Evolving Ontologies <i>Pavel Makagonov, Alejandro Ruiz Figueroa</i>	746
Named Entity Recognition	
Resolution of Data Sparseness in Named Entity Recognition Using Hierarchical Features and Feature Relaxation Principle <i>Guodong Zhou, Jian Su, Lingpeng Yang</i>	750
Learning Named Entity Recognition in Portuguese from Spanish <i>Thamar Solorio, Aurelio López-López</i>	762
A Simple Rule-Based Approach to Organization Name Recognition in Chinese Text <i>Houfeng Wang, Wuguang Shi</i>	769
Language Identification	
Disentangling from Babylonian Confusion – Unsupervised Language Identification <i>Chris Biemann, Sven Teresniak</i>	773
On the Syllabic Similarities of Romance Languages <i>Anca Dinu, Liviu P. Dinu</i>	785
Automatic Language Identification Using Multivariate Analysis <i>Vinosh Babu James, Baskaran Sankaran</i>	789

Spelling and Style Checking

Design and Development of a System for the Detection of
Agreement Errors in Basque

Arantza Díaz de Ilarraza, Koldo Gojenola, Maite Oronoz 793

An Experiment in Detection and Correction of Malapropisms
Through the Web

Igor A. Bolshakov 803

A Paragraph Boundary Detection System

Dmitriy Genzel 816

Author Index..... 827