

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Alexander Gelbukh (Ed.)

Computational Linguistics and Intelligent Text Processing

13th International Conference, CICLing 2012
New Delhi, India, March 11-17, 2012
Proceedings, Part I



Springer

Volume Editor

Alexander Gelbukh
Instituto Politécnico Nacional (IPN)
Centro de Investigación en Computación (CIC)
Col. Nueva Industrial Vallejo, CP 07738, Mexico D.F., Mexico
E-mail: gelbukh@gelbukh.com

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-28603-2 e-ISBN 978-3-642-28604-9
DOI 10.1007/978-3-642-28604-9
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012932159

CR Subject Classification (1998): H.3, H.4, F.1, I.2, H.5, H.2.8, I.5

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

CICLing 2012 was the 13th Annual Conference on Intelligent Text Processing and Computational Linguistics. The CICLing conferences provide a wide-scope forum for discussion of the art and craft of natural language processing research as well as the best practices in its applications.

This set of two books contains four invited papers and a selection of regular papers accepted for presentation at the conference. Since 2001, the proceedings of the CICLing conferences have been published in Springer's *Lecture Notes in Computer Science* series as volume numbers 2004, 2276, 2588, 2945, 3406, 3878, 4394, 4919, 5449, 6008, 6608, and 6609.

The set has been structured into 13 sections:

- NLP System Architecture
- Lexical Resources
- Morphology and Syntax
- Word Sense Disambiguation and Named Entity Recognition
- Semantics and Discourse
- Sentiment Analysis, Opinion Mining, and Emotions
- Natural Language Generation
- Machine Translation and Multilingualism
- Text Categorization and Clustering
- Information Extraction and Text Mining
- Information Retrieval and Question Answering
- Document Summarization
- Applications

The 2012 event received a record high number of submissions. A total of 307 papers by 575 authors from 46 countries were submitted for evaluation by the International Program Committee, see Tables 1 and 2. This two-volume set contains revised versions of 88 papers selected for presentation; thus the acceptance rate for this set was 28.6%.

The book features invited papers by

- Srinivas Bangalore, AT&T, USA
- John Carroll, University of Sussex, UK
- Marie-Francine Moens, Katholieke Universiteit Leuven, Belgium
- Salim Roukos, IBM, USA

who presented excellent keynote lectures at the conference. Publication of extended full-text invited papers in the proceedings is a distinctive feature of the CICLing conferences. Furthermore, in addition to presentation of their invited papers, the keynote speakers organized separate vivid informal events; this is also a distinctive feature of this conference series.

Table 1. Statistics of submissions and accepted papers by country or region

Country or region	Authors		Papers ¹	Country or region	Authors		Papers ¹
	Subm.	Subm.	Accp.		Subm.	Subm.	Accp.
Argentina	1	0.5	–	Japan	25	11.5	3.5
Australia	3	1	1	Kazakhstan	10	6	–
Belgium	2	1	1	Korea, Republic of	10	5.25	2
Brazil	3	2	1	Lebanon	3	2	1
Canada	3	2.5	–	Macao	4	2	–
Chile	3	1	1	Mexico	14	7.41	1.2
China	29	12.5	5.5	Norway	1	0.5	–
Colombia	4	3	–	Poland	10	7	2
Croatia	2	1	1	Portugal	6	2	–
Cuba	1	0.33	0.33	Romania	11	10	2
Czech Republic	5	3	2	Russian Federation	9	5	–
Denmark	1	1	–	Saudi Arabia	4	2	–
Finland	7	3	2	Spain	36	11.85	8.57
France	30	12.9	7.4	Sri Lanka	4	1	1
Germany	20	8.83	4.33	Sweden	12	5	2
Greece	5	2	–	Switzerland	1	1	–
Hong Kong	1	1	1	Taiwan	2	2	–
Hungary	2	1	1	Turkey	3	1.5	1
India	196	120	18.75	United Arab Emirates	5	2	1
Indonesia	7	3	–	UK	14	4.92	2.67
Iran	11	15	2	USA	33	13.75	7.5
Ireland	2	1	1	Uruguay	5	1	1
Italy	11	4.25	2.25	Viet Nam	4	1.5	–
				<i>Total:</i>	575	307	89

¹ By the number of authors: e.g., a paper by two authors from the USA and one from UK is counted as 0.67 for the USA and 0.33 for UK.

With this event we continued with our policy of giving preference to papers with verifiable and reproducible results: we encouraged the authors to provide, in electronic form, a proof of their claims or a working description of the suggested algorithm, in addition to the verbal description given in the paper. If the paper claimed experimental results, we encouraged the authors to make available to the community all the input data necessary to verify and reproduce these results; if it claimed to advance human knowledge by introducing an algorithm, we encouraged the authors to make the algorithm itself, in some programming language, available to the public. This additional electronic material will be permanently stored on CICLing’s server, www.CICLing.org, and will be available to the readers of the corresponding paper for download under a license that permits its free use for research purposes.

In the long run we expect that computational linguistics will have verifiability and clarity standards similar to those of mathematics: in mathematics, each claim is accompanied by a complete and verifiable proof (usually much greater in size than the claim itself); each theorem – and not just its descrip-

Table 2. Statistics of submissions and accepted papers by topic²

Accepted	Submitted	% accepted	Topic
20	44	45	Text mining
18	61	30	Information extraction
18	45	40	Semantics and discourse
18	44	41	Lexical resources
16	63	25	Information retrieval
13	40	33	Practical applications
13	29	45	Opinion mining
11	35	31	Clustering and categorization
11	21	52	Acquisition of lexical resources
8	19	42	Syntax and chunking (linguistics)
8	17	47	Word sense disambiguation
8	14	57	Summarization
7	21	33	Formalisms and knowledge representation
7	16	44	Symbolic and linguistic methods
6	50	12	Other
6	23	26	Statistical methods (mathematics)
5	23	22	Morphology
5	18	28	Named entity recognition
5	15	33	POS tagging
4	30	13	Machine translation and multilingualism
4	17	24	Question answering
4	12	33	Noisy text processing and cleaning
4	5	80	Textual entailment
3	12	25	Text generation
3	10	30	Cross-language information retrieval
3	8	38	Spelling and grammar checking
2	13	15	Natural language interfaces
2	7	29	Emotions and humor
2	6	33	Parsing algorithms (mathematics)
1	9	11	Anaphora resolution
1	6	17	Computational terminology
–	4	0	Speech processing

² As indicated by the authors. A paper may belong to several topics.

tion or general idea – is completely and precisely presented to the reader. Electronic media allow computational linguists to provide material analogous to the proofs and formulas in mathematics in full length – which can amount to megabytes or gigabytes of data – separately from a 12-page description published in the book. A more detailed argumentation for this new policy can be found on www.CICLing.org/why_verify.htm.

To encourage the provision of algorithms and data along with the published papers, we selected the winner of our Verifiability, Reproducibility, and Working Description Award. The main factors in choosing the awarded submission were technical correctness and completeness, readability of the code and documenta-

tion, simplicity of installation and use, and exact correspondence to the claims of the paper. Unnecessary sophistication of the user interface was discouraged; novelty and usefulness of the results were not evaluated – those parameters were evaluated for the paper itself and not for the data.

The following papers received the Best Paper Awards, the Best Student Paper Award, as well as the Verifiability, Reproducibility, and Working Description Award, correspondingly (the best student paper was selected from papers of which the first author was a full-time student, excluding the papers that received a Best Paper Award):

- 1st Place: *Automated Detection of Local Coherence in Short Argumentative Essays Based on Centering Theory*, by Vasile Rus and Nobal Niraula, USA;
- 2nd Place: *Corpus-Driven Hyponym Acquisition for Turkish Language*, by Savaş Yıldırım and Tuğba Yıldız, Turkey;
- 3rd Place: *Towards Automatic Generation of Catchphrases for Legal Case Reports*, by Filippo Galgani, Paul Compton, and Achim Hoffmann, Australia;
- Student: *Predictive Text Entry for Agglutinative Languages Using Unsupervised Morphological Segmentation*, by Miikka Silfverberg, Krister Lindén, and Mirka Hyvärinen, Finland;
- Verifiability: *Extraction of Relevant Figures and Tables for Multi-document Summarization*, by Ashish Sadh, Amit Sahu, Devesh Srivastava, Ratna Sanyal, and Sudip Sanyal, India.

The authors of the awarded papers (except for the Verifiability Award) were given extended time for their presentations. In addition, the Best Presentation Award and the Best Poster Award winners were selected by a ballot among the attendees of the conference.

Besides their high scientific level, one of the success factors of the CICLing conferences is their excellent cultural program. The attendees of the conference had a chance to visit the main tourist attractions of the marvellous, mysterious, colorful, and infinitely diverse India: Agra with the famous Taj Mahal, Jaipur, and Delhi. They even enjoyed riding elephants!

I would like to thank all those involved in the organization of this conference. Most importantly these are the authors of the papers that constitute this book: it is the excellence of their research work that gives value to the book and sense to the work of all other people. I thank all those who served on the Program Committee, Software Reviewing Committee, Award Selection Committee, as well as additional reviewers, for their hard and very professional work. Special thanks go to Rada Mihalcea, Ted Pedersen, and Grigori Sidorov, for their invaluable support in the reviewing process.

I would like to cordially thank the Indian Institute of Technology Delhi, for hosting the conference. With deep gratitude I acknowledge the support of Prof. B.S. Panda, the Head of Department of Mathematics, IIT Delhi. My most special thanks go to Prof. Niladri Chatterjee for his great enthusiasm and hard work

on the organization of the conference, as well as to the members of the local Organizing Committee for their enthusiastic and hard work, which has led to the success of the conference.

The entire submission and reviewing process was supported for free by the EasyChair system (www.EasyChair.org). Last but not least, I deeply appreciate the Springer staff's patience and help in editing these volumes and getting them printed in record short time – it is always a great pleasure to work with Springer.

February 2012

Alexander Gelbukh

Organization

CICLing 2012 was hosted by the Indian Institute of Technology Delhi and organized by the CICLing 2012 Organizing Committee, in conjunction with the Natural Language and Text Processing Laboratory of the CIC (Center for Computing Research) of the IPN (National Polytechnic Institute), Mexico.

Organizing Chair

Niladri Chatterjee

Organizing Committee

Niladri Chatterjee (Chair)
Pushpak Bhattacharyya
Santanu Chaudhury
K.K. Biswas
Alexander Gelbukh
Arsh Sood

Ankit Prasad
Avikant Bhardwaj
Mehak Gupta
Pramod Kumar Sahoo
Renu Balyan
Susmita Chakraborty

Program Chair

Alexander Gelbukh

Program Committee

Sophia Ananiadou
Bogdan Babych
Ricardo Baeza-Yates
Sivaji Bandyopadhyay
Srinivas Bangalore
Roberto Basili
Anja Belz
Pushpak Bhattacharyya
António Branco
Nicoletta Calzolari
Sandra Carberry
Dan Cristea
Walter Daelemans
Alex Chengyu Fang
Anna Feldman
Alexander Gelbukh

Gregory Grefenstette
Eva Hajicova
Yasunari Harada
Koiti Hasida
Graeme Hirst
Aleš Horák
Nancy Ide
Diana Inkpen
Hitoshi Isahara
Aravind Joshi
Sylvain Kahane
Alma Kharraat
Philipp Koehn
Leila Kosseim
Krister Lindén
Aurelio Lopez

Cerstin Mahlow
Sun Maosong
Yuji Matsumoto
Diana McCarthy
Helen Meng
Rada Mihalcea
Ruslan Mitkov
Dunja Mladenic
Marie-Francine Moens
Masaki Murata
Vivi Nastase
Roberto Navigli
Kjetil Nørvåg
Constantin Orăsan
Patrick Saint-Dizier
Maria Teresa Pazienza
Ted Pedersen
Viktor Pekar
Anselmo Peñas
Stelios Piperidis
Irina Prodanof
Aarne Ranta
Victor Raskin
Fuji Ren

German Rigau
Fabio Rinaldi
Horacio Rodriguez
Vasile Rus
Horacio Saggion
Kepa Sarasola
Serge Sharoff
Grigori Sidorov
Thamar Solorio
John Sowa
Ralf Steinberger
Vera Lúcia Strube De Lima
Tomek Strzalkowski
Jun Suzuki
Christoph Tillmann
George Tsatsaronis
Junichi Tsujii
Dan Tufiş
Hans Uszkoreit
Felisa Verdejo
Manuel Vilares Ferro
Haifeng Wang
Bonnie Webber

Software Reviewing Committee

Ted Pedersen
Florian Holz
Miloš Jakubíček

Sergio Jiménez Vargas
Miikka Silfverberg
Ronald Winnemöller

Award Committee

Alexander Gelbukh
Eduard Hovy
Rada Mihalcea

Ted Pedersen
Yorick Wiks

Additional Referees

Adrián Blanco González
Ahmad Emami
Akinori Fujino
Alexandra Balahur
Alvaro Rodrigo
Amitava Das

Ana Garcia-Serrano
Ananthakrishnan Ramanathan
Andrej Gardon
Aniruddha Ghosh
Antoni Oliver
Anup Kumar Kolya

Arantza Casillas-Rubio
 Arkaitz Zubiaga
 Bing Xiang
 Binod Gyawali
 Blaz Novak
 Charlie Greenbacker
 Clarissa Xavier
 Colette Joubarne
 Csaba Bodor
 Daisuke Bekki
 Daniel Eisinger
 Danilo Croce
 David Vilar
 Delia Rusu
 Diana Trandabat
 Diman Ghazi
 Dipankar Das
 Egoitz Laparra
 Ekaterina Ovchinnikova
 Enrique Amigó
 Eugen Ignat
 Fazel Keshtkar
 Feiyu Xu
 Francisco Jose Ribadas Pena
 Frederik Vaassen
 Gabriela Ferraro
 Gabriela Ramirez De La Rosa
 Gerold Schneider
 Gorka Labaka
 Guenter Neumann
 Guillermo Garrido
 H M Ishrar Hussain
 Håkan Burden
 Hendra Setiawan
 Hiroya Takamura
 Hiroyuki Shindo
 Ingo Glöckner
 Ionut Cristian Pistol
 Irina Chugur
 Irina Temnikova
 Jae-Woong Choe
 Janez Brank
 Jirka Hana
 Jirka Hana
 Jordi Atserias

Julian Brooke
 K.V.S. Prasad
 Katsuhito Sudoh
 Kishiko Ueno
 Kostas Stefanidis
 Kow Kuroda
 Krasimir Angelov
 Laritza Hernández
 Le An Ha
 Liliana Barrio-Alvers
 Lorand Dali
 Luis Otávio De Colla Furquim
 Luz Rello
 Maite Oronoz Anchordoqui
 Maria Kissa
 Mario Karlovcec
 Martin Scaiano
 Masaaki Nagata
 Matthias Reimann
 Maud Ehrmann
 Maya Carrillo
 Michael Piotrowski
 Miguel Angel Rios Gaona
 Miguel Ballesteros
 Mihai Alex Moruz
 Milagros Fernández Gavilanes
 Milos Jakubicek
 Miranda Chong
 Mitja Trampus
 Monica Macoveiciuc
 Najeh Hajlaoui
 Natalia Konstantinova
 Nathan Michalov
 Nattiya Kanhabua
 Nenad Tomasev
 Niyu Ge
 Noushin Rezapour Asheghi
 Oana Frunza
 Oier Lopez De Lacalle
 Olga Kolesnikova
 Omar Alonso
 Paolo Annesi
 Peter Ljunglöf
 Pinaki Bhaskar
 Prokopis Prokopidis

Rainer Winnenburg
Ramona Enache
Raquel Martínez
Richard Forsyth
Robin Cooper
Rodrigo Agerri
Roser Morante
Ryo Otoguro
Samira Shaikh
Santanu Pal
Shamima Mithun
Sharon Small
Simon Mille
Simone Paolo Ponzetto
Siva Reddy
Somnath Banerjee
Tadej Štajner
Thierry Declerck

Ting Liu
Tom De Smedt
Tommaso Caselli
Tong Wang
Toshiyuki Kanamaru
Tsutomu Hirao
Ulf Hermjakob
Upendra Sapkota
Vanessa Murdock
Victor Darriba
V́ctor Peinado
Vít Baisa
Vojtech Kovar
Wilker Aziz
Yulia Ledeneva
Yvonne Skalban
Zuzana Neverilova

Website and Contact

The webpage of the CICLing conference series is www.CICLing.org. It contains information about past CICLing conferences and their satellite events, including published papers or their abstracts, photos, video recordings of keynote talks, as well as information about the forthcoming CICLing conferences and contact options.

Table of Contents – Part I

NLP System Architecture

Thinking Outside the Box for Natural Language Processing (Invited Paper)	1
<i>Srinivas Bangalore</i>	

Lexical Resources

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

Morphology and Syntax

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

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>	

Word Sense Disambiguation and Named Entity Recognition

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 Makhlouta, 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>Daniło 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 (Best Paper Award, First Place) . . .	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>	
Sentiment Analysis, Opinion Mining, and Emotions	
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 Heshaam Faili</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>	
Author Index	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 (Invited Paper)	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 (Invited Paper)	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 (Verifiability Award)	402
<i>Ashish Sadh, Amit Sahu, Devesh Srivastava, Ratna Sanyal, and Sudip Sanyal</i>	
Towards Automatic Generation of Catchphrases for Legal Case Reports (Best Paper Award, Third Place)	414
<i>Filippo Galgani, Paul Compton, and Achim Hoffmann</i>	

Applications

A Dataset for the Evaluation of Lexical Simplification (Invited Paper)	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 (Best Student Paper Award)	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>	
Author Index	515