

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, Lancaster, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Friedemann Mattern

ETH Zurich, Zurich, Switzerland

John C. Mitchell

Stanford University, Stanford, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

C. Pandu Rangan

Indian Institute of Technology Madras, Chennai, India

Bernhard Steffen

TU Dortmund University, Dortmund, Germany

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbrücken, Germany

More information about this series at <http://www.springer.com/series/7407>

Alexander Gelbukh (Ed.)

Computational Linguistics and Intelligent Text Processing

18th International Conference, CICLing 2017
Budapest, Hungary, April 17–23, 2017
Revised Selected Papers, Part I

Editor
Alexander Gelbukh
CIC, Instituto Politécnico Nacional
Mexico City, Mexico

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-77112-0 ISBN 978-3-319-77113-7 (eBook)
<https://doi.org/10.1007/978-3-319-77113-7>

Library of Congress Control Number: 2018934347

LNCS Sublibrary: SL1 – Theoretical Computer Science and General Issues

© Springer Nature Switzerland AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, 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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

CICLing 2017 was the 18th International Conference on Computational Linguistics and Intelligent Text Processing. 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 volumes 2004, 2276, 2588, 2945, 3406, 3878, 4394, 4919, 5449, 6008, 6608, 6609, 7181, 7182, 7816, 7817, 8403, 8404, 9041, 9042, 9623, and 9624.

The set has been structured into 18 sections representative of the current trends in research and applications of natural language processing:

- General
- Morphology and Text Segmentation
- Syntax and Parsing
- Word Sense Disambiguation
- Reference and Coreference Resolution
- Named Entity Recognition
- Semantics and Text Similarity
- Information Extraction
- Speech Recognition
- Applications to Linguistics and the Humanities
- Sentiment Analysis
- Opinion Mining
- Author Profiling and Authorship Attribution
- Social Network Analysis
- Machine Translation
- Text Summarization
- Information Retrieval and Text Classification
- Practical Applications

This year our invited speakers were Marco Baroni (Facebook Artificial Intelligence Research), Iryna Gurevych (Ubiquitous Knowledge Processing Lab, TU Darmstadt), Björn W. Schuller (University of Passau, Imperial College London, Harbin Institute of Technology, University of Geneva, Joanneum Research, and EERING GmbH), and Hinrich Schuetze (Center for Information and Language Processing, University of Munich). They delivered excellent extended lectures and organized lively discussions. Full contributions of these invited talks are included in this book set.

After careful reviewing, the Program Committee selected 86 papers for presentation, out of 356 submissions from 60 countries.

To encourage providing algorithms and data along with the published papers, we selected three winners 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 documentation, 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, instead, they 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 Awards, respectively:

Best Verifiability Award, First Place:

“Label-Dependencies Aware Recurrent Neural Networks”

by Yoann Dupont, Marco Dinarelle, and Isabelle Tellier

Best Paper Award, Second Place, and Best Presentation Award:

“Idioms: Humans or Machines, It’s All About Context”

by Manali Pradhan, Jing Peng, Anna Feldman, and Bianca Wright

Best Student Paper Award:

“Dialogue Act Taxonomy Interoperability Using a Meta-Model”

by Soufian Salim, Nicolas Hernandez, and Emmanuel Morin

Best Paper Award, First Place:

“Gold Standard Online Debates Summaries and First Experiments Towards Automatic Summarization of Online Debate Data”

by Nattapong Sanchan, Ahmet Aker, and Kalina Bontcheva

Best Paper Award, Third Place:

“Efficient Semantic Search over Structured Web Data: A GPU Approach” by Ha-Huyen Tran, Erik Cambria, and Hoang Giang Do.

A conference is the result of the work of many people. First of all I would like to thank the members of the Program Committee for the time and effort they devoted to the reviewing of the submitted articles and to the selection process. Obviously I thank the authors for their patience in the preparation of the papers, not to mention the very development of their scientific results that form this book. I also express my most cordial thanks to the members of the local Organizing Committee for their considerable contribution to making this conference become a reality.

Organization

CICLing 2017 was hosted by the Pázmány Péter Catholic University, Faculty of Information Technology and Bionics, Budapest, Hungary, and organized by the CICLing 2017 Organizing Committee in conjunction with the Pázmány Péter Catholic University, Faculty of Information Technology and Bionics, the Natural Language and Text Processing Laboratory of the CIC, IPN, and the Mexican Society of Artificial Intelligence (SMIA).

Organizing Committee

| | |
|----------------------|---|
| Attila Novák (Chair) | MTA-PPKE Language Technology Research Group, Pázmány Péter Catholic University |
| Gábor Prószéky | MTA-PPKE Language Technology Research Group, Pázmány Péter Catholic University |
| Borbála Siklósi | MTA-PPKE Language Technology Research Group, Pázmány Péter Catholic University |

Program Committee

| | | |
|-------------------------|----------------------|------------------------|
| Bayan Abushawar | Gregory Grefenstette | Inderjeet Mani |
| Galia Angelova | Tunga Gungor | Alexander Mehler |
| Alexandra Balahur | Eva Hajicova | Farid Meziane |
| Sivaji Bandyopadhyay | Yasunari Harada | Rada Mihalcea |
| Leslie Barrett | Karin Harbusch | Evangelos Milios |
| Roberto Basili | Koiti Hasida | Ruslan Mitkov |
| Pushpak Bhattacharyya | Ales Horak | Dunja Mladenic |
| Christian Boitet | Veronique Hoste | Marie-Francine Moens |
| Nicoletta Calzolari | Diana Inkpen | Hermann Moisl |
| Nick Campbell | Hitoshi Isahara | Masaki Murata |
| Michael Carl | Aminul Islam | Preslav Nakov |
| Violetta Cavalli-Sforza | Guillaume Jacquet | Costanza Navarretta |
| Niladri Chatterjee | Milos Jakubicek | Joakim Nivre |
| Dan Cristea | Sylvain Kahane | Kjetil Norvag |
| Walter Daelemans | Alma Kharrat | Attila Novák |
| Mike Dillinger | Philipp Koehn | Nir Ofek |
| Samhaa El-Beltagy | Valia Kordoni | Kemal Oflazer |
| Michael Elhadad | Mathieu Lafourcade | Constantin Orasan |
| Anna Feldman | Elena Lloret | Ivandre Paraboni |
| Robert Gaizauskas | Bente Maegaard | Saint-Dizier Patrick |
| Alexander Gelbukh | Cerstin Mahlow | Maria Teresa Pazzienza |
| Dafydd Gibbon | Suresh Manandhar | Ted Pedersen |

| | | |
|----------------------|-----------------------|----------------------|
| Viktor Pekar | Kepa Sarasola | Mike Thelwall |
| Anselmo Peñas | Khaled Shaalan | Juan-Manuel |
| Soujanya Poria | Serge Sharoff | Torres-Moreno |
| James Pustejovsky | Grigori Sidorov | George Tsatsaronis |
| Marta R. Costa-Jussà | Kiril Simov | Dan Tufis |
| Victor Raskin | Vaclav Snasel | Jerzy Tyszkiewicz |
| German Rigau | Efstathios Stamatatos | Manuel Vilares Ferro |
| Fabio Rinaldi | Mark Steedman | Aline Villavicencio |
| Horacio Rodriguez | Josef Steinberger | Ellen Voorhees |
| Paolo Rosso | Stan Szpakowicz | Piotr W. Fuglewicz |
| Vasile Rus | William Teahan | Annie Zaenen |

Software Reviewing Committee

Ted Pedersen
Florian Holz
Miloš Jakubíček
Sergio Jiménez Vargas
Miikka Silfverberg
Ronald Winnemöller

Best Paper Award Committee

Alexander Gelbukh
Eduard Hovy
Rada Mihalcea
Ted Pedersen
Yorrick Wiks

Contents – Part I

General

Invited Paper:

| | |
|---|----|
| Overview of Character-Based Models for Natural Language Processing | 3 |
| <i>Heike Adel, Ehsaneddin Asgari, and Hinrich Schütze</i> | |
| Pooling Word Vector Representations Across Models | 17 |
| <i>Rajendra Banjade, Nabin Maharjan, Dipesh Gautam, Frank Adrasik, Arthur C. Graesser, and Vasile Rus</i> | |
| Strategies to Select Examples for Active Learning with Conditional Random Fields | 30 |
| <i>Vincent Claveau and Ewa Kijak</i> | |

Best Verifiability Award, First Place:

| | |
|---|----|
| Label-Dependencies Aware Recurrent Neural Networks | 44 |
| <i>Yoann Dupont, Marco Dinarelli, and Isabelle Tellier</i> | |
| Universal Computational Formalisms and Developer Environment for Rule-Based NLP | 67 |
| <i>Svetlana Sheremetyeva</i> | |

Morphology and Text Segmentation

| | |
|---|-----|
| Several Ways to Use the Lingwarium.org Online MT Collaborative Platform to Develop Rich Morphological Analyzers | 81 |
| <i>Vincent Berment, Christian Boitet, Jean-Philippe Guilbaud, and Jurgita Kapočiūtė-Dzikiene</i> | |
| A Trie-structured Bayesian Model for Unsupervised Morphological Segmentation | 87 |
| <i>Murathan Kurfalı, Ahmet Üstün, and Burcu Can</i> | |
| Building Morphological Chains for Agglutinative Languages | 99 |
| <i>Serkan Ozen and Burcu Can</i> | |
| Joint PoS Tagging and Stemming for Agglutinative Languages. | 110 |
| <i>Necva Bölüçü and Burcu Can</i> | |
| Hungarian Particle Verbs in a Corpus-Driven Approach. | 123 |
| <i>Ágnes Kalivoda</i> | |

| | |
|--|-----|
| HANS: A Service-Oriented Framework for Chinese Language Processing . . . | 134 |
| <i>Lung-Hao Lee, Kuei-Ching Lee, and Yuen-Hsien Tseng</i> | |

Syntax and Parsing

| | |
|--|-----|
| Learning to Rank for Coordination Detection | 145 |
| <i>Xun Wang, Rumeng Li, Hiroyuki Shindo, Katsuhito Sudoh, and Masaaki Nagata</i> | |
| Classifier Ensemble Approach to Dependency Parsing | 158 |
| <i>Silpa Kanneganti, Vandan Mujadia, and Dipti M. Sharma</i> | |
| Evaluation and Enrichment of Stanford Parser Using an Arabic Property Grammar | 170 |
| <i>Raja Bensalem Bahloul, Nesrine Kadri, Kais Haddar, and Philippe Blache</i> | |

Word Sense Disambiguation

| | |
|--|-----|
| SenseDependency-Rank: A Word Sense Disambiguation Method Based on Random Walks and Dependency Trees | 185 |
| <i>Marco Antonio Sobrevilla-Cabezudo, Arturo Oncevay-Marcos, and Andrés Melgar</i> | |
| Domain Adaptation for Word Sense Disambiguation Using Word Embeddings | 195 |
| <i>Kanako Komiya, Shota Suzuki, Minoru Sasaki, Hiroyuki Shinnou, and Manabu Okumura</i> | |

Reference and Coreference Resolution

Invited Paper:

| | |
|---|-----|
| “Show Me the Cup”: Reference with Continuous Representations | 209 |
| <i>Marco Baroni, Gemma Boleda, and Sebastian Padó</i> | |
| Improved Best-First Clustering for Coreference Resolution in Indian Classical Music Forums | 225 |
| <i>Joe Cheri Ross and Pushpak Bhattacharyya</i> | |
| A Robust Coreference Chain Builder for Tamil. | 233 |
| <i>R. Vijay Sundar Ram and Sobha Lalitha Devi</i> | |

Named Entity Recognition

| | |
|---|-----|
| Structured Named Entity Recognition by Cascading CRFs | 249 |
| <i>Yoann Dupont, Marco Dinarelli, Isabelle Tellier, and Christian Lautier</i> | |

| | |
|---|-----|
| Arabic Named Entity Recognition: A Bidirectional GRU-CRF Approach | 264 |
| <i>Mourad Gridach and Hatem Haddad</i> | |

| | |
|---|-----|
| Named Entity Recognition for Amharic Using Stack-Based Deep Learning | 276 |
| <i>Utpal Kumar Sikdar and Björn Gambäck</i> | |

Semantics and Text Similarity

Best Paper Award, Second Place, and Best Presentation Award:

| | |
|---|-----|
| Idioms: Humans or Machines, It's All About Context | 291 |
| <i>Manali Pradhan, Jing Peng, Anna Feldman, and Bianca Wright</i> | |

Best Student Paper Award:

| | |
|---|-----|
| Dialogue Act Taxonomy Interoperability Using a Meta-model | 305 |
| <i>Soufian Salim, Nicolas Hernandez, and Emmanuel Morin</i> | |

| | |
|--|-----|
| Textual Entailment Using Machine Translation Evaluation Metrics | 317 |
| <i>Tanik Saikh, Sudip Kumar Naskar, Asif Ekbal, and Sivaji Bandyopadhyay</i> | |

| | |
|---|-----|
| Supervised Learning of Entity Disambiguation Models by Negative Sample Selection. | 329 |
| <i>Hani Daher, Romaric Besançon, Olivier Ferret, Hervé Le Borgne, Anne-Laure Daquo, and Youssef Tamaazousti</i> | |

| | |
|--|-----|
| The Enrichment of Arabic WordNet Antonym Relations | 342 |
| <i>Mohamed Ali Batita and Mounir Zrigui</i> | |

| | |
|--|-----|
| Designing an Ontology for Physical Exercise Actions | 354 |
| <i>Sandeep Kumar Dash, Partha Pakray, Robert Porzel, Jan Smeddinck, Rainer Malaka, and Alexander Gelbukh</i> | |

| | |
|---|-----|
| Visualizing Textbook Concepts: Beyond Word Co-occurrences. | 363 |
| <i>Chandramouli Shama Sastry, Darshan Siddesh Jagaluru, and Kavi Mahesh</i> | |

| | |
|--|-----|
| Matching, Re-Ranking and Scoring: Learning Textual Similarity by Incorporating Dependency Graph Alignment and Coverage Features | 377 |
| <i>Sarah Kohail and Chris Biemann</i> | |

| | |
|--|-----|
| Text Similarity Function Based on Word Embeddings for Short Text Analysis | 391 |
| <i>Adrián Jiménez Pascual and Sumio Fujita</i> | |

Information Extraction

| | |
|--|-----|
| Domain Specific Features Driven Information Extraction from Web Pages of Scientific Conferences | 405 |
| <i>Piotr Andruszkiewicz and Rafał Hazan</i> | |
| Classifier-Based Pattern Selection Approach for Relation Instance Extraction | 418 |
| <i>Angrosh Mandya, Danushka Bollegala, Frans Coenen, and Katie Atkinson</i> | |
| An Ensemble Architecture for Linked Data Lexicalization | 435 |
| <i>Rivindu Perera and Parma Nand</i> | |
| A Hybrid Approach for Biomedical Relation Extraction Using Finite State Automata and Random Forest-Weighted Fusion | 450 |
| <i>Thanassis Mavropoulos, Dimitris Liparas, Spyridon Symeonidis, Stefanos Vrochidis, and Ioannis Kompatsiaris</i> | |
| Exploring Linguistic and Graph Based Features for the Automatic Classification and Extraction of Adverse Drug Effects | 463 |
| <i>Tirthankar Dasgupta, Abir Naskar, and Lipika Dey</i> | |
| Extraction of Semantic Relation Between Arabic Named Entities Using Different Kinds of Transducer Cascades. | 475 |
| <i>Fatma Ben Mesmia, Kaouther Bouabidi, Kais Haddar, Nathalie Friburger, and Denis Maurel</i> | |
| Semi-supervised Relation Extraction from Monolingual Dictionary for Russian WordNet. | 488 |
| <i>Daniil Alexeyevsky</i> | |

Speech Recognition

| | |
|--|-----|
| ASR Hypothesis Reranking Using Prior-Informed Restricted Boltzmann Machine. | 503 |
| <i>Yukun Ma, Erik Cambria, and Benjamin Bigot</i> | |
| A Comparative Analysis of Speech Recognition Systems for the Tatar Language | 515 |
| <i>Aidar Khusainov</i> | |

Applications to Linguistics and the Humanities

Invited Papers:

| | |
|---|------------|
| Interactive Data Analytics for the Humanities | 527 |
| <i>Iryna Gurevych, Christian M. Meyer, Carsten Binnig, Johannes Fürnkranz, Kristian Kersting, Stefan Roth, and Edwin Simpson</i> | |
| Language Technology for Digital Linguistics: Turning the Linguistic Survey of India into a Rich Source of Linguistic Information | 550 |
| <i>Lars Borin, Shafqat Mumtaz Virk, and Anju Saxena</i> | |
| Classifying World Englishes from a Lexical Perspective: A Corpus-Based Approach | 564 |
| <i>Frank Z. Xing, Danyuan Ho, Diyana Hamzah, and Erik Cambria</i> | |
| Towards a Map of the Syntactic Similarity of Languages | 576 |
| <i>Alina Maria Ciobanu, Liviu P. Dinu, and Andrea Sgarro</i> | |
| Romanian Word Production: An Orthographic Approach Based on Sequence Labeling | 591 |
| <i>Liviu P. Dinu and Alina Maria Ciobanu</i> | |
| Author Index | 605 |

Contents – Part II

Sentiment Analysis

| | |
|---|-----|
| A Comparison Among Significance Tests and Other Feature Building Methods for Sentiment Analysis: A First Study. | 3 |
| <i>Raksha Sharma, Dibyendu Mondal, and Pushpak Bhattacharyya</i> | |
| BATframe: An Unsupervised Approach for Domain-Sensitive Affect Detection | 20 |
| <i>Kokil Jaidka, Niyati Chhaya, Rahul Wadbude, Sanket Kedia, and Manikanta Nallagatla</i> | |
| Leveraging Target-Oriented Information for Stance Classification | 35 |
| <i>Jiachen Du, Ruifeng Xu, Lin Gui, and Xuan Wang</i> | |
| Sentiment Polarity Classification of Figurative Language: Exploring the Role of Irony-Aware and Multifaceted Affect Features | 46 |
| <i>Delia Irazú Hernández Farías, Cristina Bosco, Viviana Patti, and Paolo Rosso</i> | |
| Sarcasm Annotation and Detection in Tweets | 58 |
| <i>Johan G. Cyrus M. Ræder and Björn Gambäck</i> | |
| Modeling the Impact of Modifiers on Emotional Statements | 71 |
| <i>Valentina Sintsova, Margarita Bolívar Jiménez, and Pearl Pu</i> | |
| CSenticNet: A Concept-Level Resource for Sentiment Analysis in Chinese Language | 90 |
| <i>Haiyun Peng and Erik Cambria</i> | |
| Emotional Tone Detection in Arabic Tweets. | 105 |
| <i>Amr Al-Khatib and Samhaa R. El-Beltagy</i> | |
| Morphology Based Arabic Sentiment Analysis of Book Reviews | 115 |
| <i>Omar El Ariss and Loai M. Alnemer</i> | |
| Adaptation of Sentiment Analysis Techniques to Persian Language | 129 |
| <i>Kia Dashtipour, Amir Hussain, and Alexander Gelbukh</i> | |
| Verb-Mediated Composition of Attitude Relations Comprising Reader and Writer Perspective. | 141 |
| <i>Manfred Klenner, Simon Clematide, and Don Tugener</i> | |

| | |
|--|-----|
| Customer Churn Prediction Using Sentiment Analysis and Text Classification of VOC | 156 |
| <i>Yiou Wang, Koji Satake, Takeshi Onishi, and Hiroshi Masuichi</i> | |
| Benchmarking Multimodal Sentiment Analysis | 166 |
| <i>Erik Cambria, Devamanyu Hazarika, Soujanya Poria, Amir Hussain, and R. B. V. Subramanyam</i> | |
| Machine Learning Approaches for Speech Emotion Recognition: Classic and Novel Advances | 180 |
| <i>Panikos Heracleous, Akio Ishikawa, Keiji Yasuda, Hiroyuki Kawashima, Fumiaki Sugaya, and Masayuki Hashimoto</i> | |
| Opinion Mining | |
| Mining Aspect-Specific Opinions from Online Reviews Using a Latent Embedding Structured Topic Model | 195 |
| <i>Mingyang Xu, Ruixin Yang, Paul Jones, and Nagiza F. Samatova</i> | |
| A Comparative Study of Target-Based and Entity-Based Opinion Extraction | 211 |
| <i>Joseph Lark, Emmanuel Morin, and Sebastián Peña Saldarriaga</i> | |
| Supervised Domain Adaptation via Label Alignment for Opinion Expression Extraction. | 224 |
| <i>Yifan Zhang, Arjun Mukherjee, and Fan Yang</i> | |
| Comment Relevance Classification in Facebook | 241 |
| <i>Chaya Liebeskind, Shmuel Liebeskind, and Yaakov HaCohen-Kerner</i> | |
| Detecting Sockpuppets in Deceptive Opinion Spam. | 255 |
| <i>Marjan Hosseinia and Arjun Mukherjee</i> | |

Author Profiling and Authorship Attribution

Invited Paper:

| | |
|--|-----|
| Reading the Author and Speaker: Towards a Holistic and Deep Approach on Automatic Assessment of What is in One's Words | 275 |
| <i>Björn W. Schuller</i> | |
| Improving Cross-Topic Authorship Attribution: The Role of Pre-Processing | 289 |
| <i>Iliia Markov, Efsthios Stamatatos, and Grigori Sidorov</i> | |
| Author Identification Using Latent Dirichlet Allocation | 303 |
| <i>Hiram Calvo, Ángel Hernández-Castañeda, and Jorge García-Flores</i> | |

| | |
|--|-----|
| Personality Recognition Using Convolutional Neural Networks. | 313 |
| <i>Maite Giménez, Roberto Paredes, and Paolo Rosso</i> | |
| Character-Level Dialect Identification in Arabic Using Long Short-Term Memory | 324 |
| <i>Karim Sayadi, Mansour Hamidi, Marc Bui, Marcus Liwicki, and Andreas Fischer</i> | |
| A Text Semantic Similarity Approach for Arabic Paraphrase Detection | 338 |
| <i>Adnen Mahmoud, Ahmed Zrigui, and Mounir Zrigui</i> | |
| Social Network Analysis | |
| Curator: Enhancing Micro-Blogs Ranking by Exploiting User's Context | 353 |
| <i>Hicham G. Elmongui and Riham Mansour</i> | |
| A Multi-view Clustering Model for Event Detection in Twitter. | 366 |
| <i>Di Shang, Xin-Yu Dai, Weiye Ge, Shujiang Huang, and Jiajun Chen</i> | |
| Monitoring Geographical Entities with Temporal Awareness in Tweets | 379 |
| <i>Koji Matsuda, Mizuki Sango, Naoaki Okazaki, and Kentaro Inui</i> | |
| Just the Facts: Winnowing Microblogs for Newsworthy Statements using Non-Lexical Features | 391 |
| <i>Nigel Dewdney and Rachel Cotterill</i> | |
| Impact of Content Features for Automatic Online Abuse Detection | 404 |
| <i>Etienne Papegnies, Vincent Labatut, Richard Dufour, and Georges Linarès</i> | |
| Detecting Aggressive Behavior in Discussion Threads Using Text Mining . . . | 420 |
| <i>Filippos Karolos Ventirozos, Iraklis Varlamis, and George Tsatsaronis</i> | |
| Machine Translation | |
| Combining Machine Translation Systems with Quality Estimation. | 435 |
| <i>László János Laki and Zijian Győző Yang</i> | |
| Evaluation of Neural Machine Translation for Highly Inflected and Small Languages. | 445 |
| <i>Mārcis Pinnis, Rihards Krišlauks, Daiga Dekšne, and Toms Miks</i> | |
| Towards Translating Mixed-Code Comments from Social Media. | 457 |
| <i>Thoudam Doren Singh and Tamar Solorio</i> | |
| Multiple System Combination for PersoArabic-Latin Transliteration | 469 |
| <i>Nima Hemmati, Hesham Faily, and Jalal Maleki</i> | |

| | |
|--|-----|
| Building a Location Dependent Dictionary for Speech Translation Systems | 482 |
| <i>Keiji Yasuda, Panikos Heracleous, Akio Ishikawa, Masayuki Hashimoto, Kazunori Matsumoto, and Fumiaki Sugaya</i> | |

Text Summarization

Best Paper Award, First Place:

| | |
|--|-----|
| Gold Standard Online Debates Summaries and First Experiments Towards Automatic Summarization of Online Debate Data | 495 |
| <i>Nattapong Sanchan, Ahmet Aker, and Kalina Bontcheva</i> | |
| Optimization in Extractive Summarization Processes Through Automatic Classification | 506 |
| <i>Angel Luis Garrido, Carlos Bobed, Oscar Cardiel, Andrea Alexendri, and Ruben Quilez</i> | |
| Summarizing Weibo with Topics Compression | 522 |
| <i>Marina Litvak, Natalia Vanetik, and Lei Li</i> | |
| Timeline Generation Based on a Two-Stage Event-Time Anchoring Model | 535 |
| <i>Tomohiro Sakaguchi and Sadao Kurohashi</i> | |

Information Retrieval and Text Classification

Best Paper Award, Third Place:

| | |
|---|-----|
| Efficient Semantic Search Over Structured Web Data: A GPU Approach | 549 |
| <i>Ha-Nguyen Tran, Erik Cambria, and Hoang Giang Do</i> | |
| Efficient Association Rules Selecting for Automatic Query Expansion | 563 |
| <i>Ahlem Bouziri, Chiraz Latiri, and Eric Gaussier</i> | |
| Text-to-Concept: A Semantic Indexing Framework for Arabic News Videos | 575 |
| <i>Sadek Mansouri, Chahira Lhioui, Mbarek Charhad, and Mounir Zrigui</i> | |
| Approximating Multi-class Text Classification Via Automatic Generation of Training Examples | 585 |
| <i>Filippo Geraci and Tiziano Papini</i> | |

Practical Applications

| | |
|--|-----|
| Generating Appealing Brand Names | 605 |
| <i>Gaurush Hiranandani, Pranav Maneriker, and Harsh Jhamtani</i> | |

| | |
|--|------------|
| Radiological Text Simplification Using a General Knowledge Base. | 617 |
| <i>Lionel Ramadier and Mathieu Lafourcade</i> | |
| Mining Supervisor Evaluation and Peer Feedback in Performance Appraisals | 628 |
| <i>Girish Keshav Palshikar, Sachin Pawar, Saheb Chourasia, and Nitin Ramrakhiyani</i> | |
| Automatic Detection of Uncertain Statements in the Financial Domain | 642 |
| <i>Christoph Kilian Theil, Sanja Štajner, Heiner Stuckenschmidt, and Simone Paolo Ponzetto</i> | |
| Automatic Question Generation From Passages. | 655 |
| <i>Karen Mazidi</i> | |
| Author Index | 667 |