

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

Elisabeth Métais Farid Meziane
Mohamad Saraee Vijayan Sugumaran
Sunil Vadera (Eds.)

Natural Language Processing and Information Systems

18th International Conference on Applications
of Natural Language to Information Systems, NLDB 2013
Salford, UK, June 19-21, 2013
Proceedings



Springer

Volume Editors

Elisabeth Métais
Conservatoire National des Arts et Métiers
Paris, France
E-mail: metais@cnam.fr

Farid Meziane
Mohamad Saraee
Sunil Vadera
University of Salford
Salford, Lancashire, UK
Email: {f.meziane, m.saraee, s.vadera}@salford.ac.uk

Vijayan Sugumaran
Oakland University
Rochester, MI, USA
E-mail: sugumara@oakland.edu

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-38823-1 e-ISBN 978-3-642-38824-8
DOI 10.1007/978-3-642-38824-8
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2013939734

CR Subject Classification (1998): I.2.7, H.3, H.2.8, I.5, J.5, I.2, I.6, J.1

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

© Springer-Verlag Berlin Heidelberg 2013

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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

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

This volume of *Lecture Notes in Computer Science* (LNCS) contains the papers presented at the 18th International Conference on Application of Natural Language to Information Systems, held at MediacityUK, University of Salford during June 19–21, 2013 (NLDB2013). Since its foundation in 1995, the NLDB conference has attracted state-of-the-art presentations and followed closely the developments of the application of natural language to databases and information systems in the wider meaning of the term.

The current conference proceedings reflect the development in the field and encompass areas such as sentiment analysis and mining, forensic computing, the Semantic Web and information search. This is in addition to the more traditional topics such as requirements engineering, question answering systems, and named entity recognition. NLDB is now an established conference and attracts researchers and practitioners from all over the world. Indeed, this year's conference saw submission for works using a large number of natural languages that include Chinese, Japanese, Arabic, Hebrew, and Farsi.

We received 80 papers and each paper was reviewed by at least three reviewers with the majority having four or five reviews. The Conference Co-chairs and Program Committee Co-chairs had a final consultation meeting to look at all the reviews and made the final decisions on the papers to be accepted. We accepted 21 papers as long/regular papers, 15 short papers, and 17 poster presentations.

We would like to thank all the reviewers for their time, effort, and for completing their assignments on time despite tight deadlines. Many thanks to the authors for their contributions.

June 2013

Elisabeth Métais
Farid Meziane
Mohamed Sarace
Vijay Sugumaran
Sunil Vadera

Organization

Conference Chairs

Elisabeth Métais	Conservatoire National des Arts et Metiers, Paris, France
Farid Meziane	University of Salford, UK
Sunil Vadera	University of Salford, UK

Programme Committee Chairs

Mohamed Saraee	University of Salford, UK
Vijay Sugumaran	Oakland University Rochester, USA

Programme Committee

Jacky Akoka	CNAM, France
Frederic Andres	National Institute of Informatics, Japan
Apostolos Antonacopoulos	University of Salford, UK
Eric Atwell	University of Leeds, UK
Abdelmajid Ben Hamadou	Sfax University, Tunisia
Bettina Berendt	Leuven University, Belgium
Johan Bos	Groningen University, The Netherlands
Goss Bouma	Groningen University, The Netherlands
Philipp Cimiano	Universität Bielefeld, Germany
Isabelle Comyn-Wattiau	ESSEC, France
Walter Daelemans	University of Antwerp, Belgium
Zhou Erqiang	University of Electronic Science and Technology, China
Stefan Evert	University of Osnabrück, Germany
Vladimir Fomichov	National Research University Higher School of Economics Russia
Alexander Gelbukh	Mexican Academy of Science, Mexico
Jon Atle Gulla	NTNU, Norway
Karin Harbusch	Koblenz University, Germany
Dirk Heylen	University of Twente, The Netherlands
Helmut Horacek	Saarland University, Germany

Ashwin Ittoo	Groningen University, The Netherlands
Paul Johannesson	Stockholm University, Sweden
Sophia Katrenko	Utrecht University, The Netherlands
Epaminondas Kapetanios	University of Westminster, UK
John Kean	University of Manchester, UK
Zoubida Kedad	Université de Versailles, France
Christian Kop	University of Klagenfurt, Austria
Valia Kordoni	Saarland University, Germany
Leila Kosseim	Concordia University, Canada
Zornitsa Kozareva	University of Southern California, USA
Nadira Lammari	CNAM, France
Dominique Laurent	University of Cergy Pontoise, France
Jochen Leidner	Thomson Reuters, USA
Piroska Lendvai	Hungarian Academy of Sciences, Hungary
Johannes Leveling	Dublin City University, Ireland
Deryle Lonsdale	Brigham Young University, USA
Rob Malouf	San Diego State University, USA
Farhi Marir	London Metropolitan University, UK
Heinrich C. Mayr	University of Klagenfurt, Austria
Farid Meziane	Salford University, UK
Elisabeth Métais	CNAM, France
Marie-Jean Meurs	Concordia University, Canada
Luisa Mich	University of Trento, Italy
Shamima Mithun	Concordia University, Canada
Marie-Francine Moens	Katholieke Universiteit Leuven, Belgium
Andres Montoyo	Universidad de Alicante, Spain
Rafael Muñoz	Universidad de Alicante, Spain
Guenter Neumann	DFKI, Germany
Jan Odijk	Utrecht University, The Netherlands
Jim O'Shea	Manchester Metropolitan University, UK
Karim Ouazzane	London Metropolitan University, UK
Pit Pichappan	Al Imam University, Saudi Arabia
Shaolin Qu	Michigan State University, USA
Mike Rosner	University of Malta, Malta
German Rigau	University of the Basque Country, Spain
Fabio Rinaldi	University of Zurich, Switzerland
Mathieu Roch	Université Montpellier 2, France
Patrick Saint-Dizier	Université Paul Sabatier, France
Mohamed Saraee	University of Salford, UK
Hide Sasaki	Ritsumeikan University, Kyoto, Japan
Khaled Shaalan	The British University in Dubai, UAE
Max Silberztein	Université de Franche-Comté, France

Samira Si-Said Cherfi	CNAM, France
Veda Storey	Georgia State University, USA
Vijayan Sugumaran	Oakland University Rochester, USA
Bernhard Thalheim	Kiel University, Germany
Michael Thelwall	University of Wolverhampton, UK
Krishnaprasad Thirunarayan	Wright State University, USA
Juan Carlos Trujillo	Universidad de Alicante, Spain
Christina Unger	Universität Bielefeld, Germany
Alfonso Ureña	University of Jaén, Spain
Sunil Vadera	University of Salford, UK
Panos Vassiliadis	University of Ioannina, Greece
Robert Wagner	Linz University, Austria
René Witte	Concordia University, Canada
Magdalena Wolska	Saarland University, Germany
Bing Wu	Dublin Institute of Technology, Republic of Ireland
Jim Yip	University of Salford, UK

Table of Contents

Full Papers

Extraction of Statements in News for a Media Response Analysis	1
<i>Thomas Scholz and Stefan Conrad</i>	
Sentiment-Based Ranking of Blog Posts Using Rhetorical Structure Theory	13
<i>Jose M. Chenlo, Alexander Hogenboom, and David E. Losada</i>	
Automatic Detection of Ambiguous Terminology for Software Requirements	25
<i>Yue Wang, Irene L. Manotas Gutiérrez, Kristina Winbladh, and Hui Fang</i>	
An OpenCCG-Based Approach to Question Generation from Concepts	38
<i>Markus M. Berg, Amy Isard, and Johanna D. Moore</i>	
A Hybrid Approach for Arabic Diacritization	53
<i>Ahmed Said, Mohamed El-Sharqwi, Achraf Chalabi, and Eslam Kamal</i>	
EDU-Based Similarity for Paraphrase Identification	65
<i>Ngo Xuan Bach, Nguyen Le Minh, and Akira Shimazu</i>	
Exploiting Query Logs and Field-Based Models to Address Term Mismatch in an HIV/AIDS FAQ Retrieval System	77
<i>Edwin Thuma, Simon Rogers, and Iadh Ounis</i>	
Exploring Domain-Sensitive Features for Extractive Summarization in the Medical Domain	90
<i>Dat Tien Nguyen and Johannes Leveling</i>	
A Corpus-Based Approach for the Induction of Ontology Lexica	102
<i>Sebastian Walter, Christina Unger, and Philipp Cimiano</i>	
SQUALL: A Controlled Natural Language as Expressive as SPARQL 1.1	114
<i>Sébastien Ferré</i>	
Evaluating Syntactic Sentence Compression for Text Summarisation	126
<i>Prasad Perera and Leila Kosseim</i>	

An Unsupervised Aspect Detection Model for Sentiment Analysis of Reviews	140
<i>Ayoub Bagheri, Mohamad Saraee, and Franciska de Jong</i>	
Cross-Lingual Natural Language Querying over the Web of Data	152
<i>Nitish Aggarwal, Tamara Polajnar, and Paul Buitelaar</i>	
Extractive Text Summarization: Can We Use the Same Techniques for Any Text?	164
<i>Tatiana Vodolazova, Elena Lloret, Rafael Muñoz, and Manuel Palomar</i>	
Unsupervised Medical Subject Heading Assignment Using Output Label Co-occurrence Statistics and Semantic Predications	176
<i>Ramakanth Kavuluru and Zhenghao He</i>	
Bayesian Model Averaging and Model Selection for Polarity Classification	189
<i>Federico Alberto Pozzi, Elisabetta Fersini, and Enza Messina</i>	
An Approach for Extracting and Disambiguating Arabic Persons' Names Using Clustered Dictionaries and Scored Patterns	201
<i>Omnia Zayed, Samhaa El-Beltagy, and Osama Haggag</i>	
ANEAR: Automatic Named Entity Aliasing Resolution	213
<i>Ayah Zirikly and Mona Diab</i>	
Improving Candidate Generation for Entity Linking	225
<i>Yuhang Guo, Bing Qin, Yuqin Li, Ting Liu, and Sheng Li</i>	
Person Name Recognition Using the Hybrid Approach	237
<i>Mai Oudah and Khaled Shaalan</i>	
A Broadly Applicable and Flexible Conceptual Metagrammar as a Basic Tool for Developing a Multilingual Semantic Web	249
<i>Vladimir A. Fomichov</i>	

Short Papers

MOSAIC: A Cohesive Method for Orchestrating Discrete Analytics in a Distributed Model	260
<i>Ransom Winder, Joseph Jubinski, John Prange, and Nathan Giles</i>	
Ranking Search Intents Underlying a Query	266
<i>Yunqing Xia, Xiaoshi Zhong, Guoyu Tang, Junjun Wang, Qiang Zhou, Thomas Fang Zheng, Qinan Hu, Sen Na, and Yaohai Huang</i>	

Linguistic Sentiment Features for Newspaper Opinion Mining	272
<i>Thomas Scholz and Stefan Conrad</i>	
Text Classification of Technical Papers Based on Text Segmentation . . .	278
<i>Thien Hai Nguyen and Kiyooki Shirai</i>	
Product Features Categorization Using Constrained Spectral Clustering	285
<i>Sheng Huang, Zhendong Niu, and Yulong Shi</i>	
A New Approach for Improving Cross-Document Knowledge Discovery Using Wikipedia	291
<i>Peng Yan and Wei Jin</i>	
Using Grammar-Profiles to Intrinsically Expose Plagiarism in Text Documents	297
<i>Michael Tschuggnall and Günther Specht</i>	
Feature Selection Methods in Persian Sentiment Analysis	303
<i>Mohamad Sarraee and Ayoub Bagheri</i>	
Towards the Refinement of the Arabic Soundex	309
<i>Nedjma Djouhra Ousidhoum and Nacéra Bensaou</i>	
An RDF-Based Semantic Index	315
<i>F. Amato, F. Gargiulo, A. Mazzeo, V. Moscato, and A. Picariello</i>	
Experiments in Producing Playful “Explanations” for Given Names (Anthroponyms) in Hebrew and English	321
<i>Yaakov HaCohen-Kerner, Daniel Nisim Cohen, and Ephraim Nissan</i>	
Collaborative Enrichment of Electronic Dictionaries Standardized-LMF	328
<i>Aida Khemakhem, Bilel Gargouri, and Abdelmajid Ben Hamadou</i>	
Enhancing Machine Learning Results for Semantic Relation Extraction	337
<i>Ines Boujelben, Salma Jamoussi, and Abdelmajid Ben Hamadou</i>	
GENDESC: A Partial Generalization of Linguistic Features for Text Classification	343
<i>Guillaume Tisserant, Violaine Prince, and Mathieu Roche</i>	
Entangled Semantics	349
<i>Diana Tanase and Epaminondas Kapetanios</i>	

Poster Papers

Phrase Table Combination Deficiency Analyses in Pivot-Based SMT	355
<i>Yiming Cui, Conghui Zhu, Xiaoning Zhu, Tiejun Zhao, and Dequan Zheng</i>	
Analysing Customers Sentiments: An Approach to Opinion Mining and Classification of Online Hotel Reviews	359
<i>Juan Sixto, Aitor Almeida, and Diego López-de-Ipiña</i>	
An Improved Discriminative Category Matching in Relation Identification	363
<i>Yongliang Sun, Jing Yang, and Xin Lin</i>	
Extracting Fine-Grained Entities Based on Coordinate Graph	367
<i>Qing Yang, Peng Jiang, Chunxia Zhang, and Zhendong Niu</i>	
NLP-Driven Event Semantic Ontology Modeling for Story	372
<i>Chun-Ming Gao, Qiu-Mei Xie, and Xiao-Lan Wang</i>	
The Development of an Ontology for Reminiscence	376
<i>Collette Curry, James O'Shea, Keeley Crockett, and Laura Brown</i>	
Chinese Sentence Analysis Based on Linguistic Entity-Relationship Model	380
<i>Dechun Yin</i>	
A Dependency Graph Isomorphism for News Sentence Searching	384
<i>Kim Schouten and Flavius Frasinca</i>	
Unsupervised Gazette Creation Using Information Distance	388
<i>Sangameshwar Patil, Sachin Pawar, Girish K. Palshikar, Savita Bhat, and Rajiv Srivastava</i>	
A Multi-purpose Online Toolset for NLP Applications	392
<i>Maciej Ogrodniczuk and Michał Lenart</i>	
A Test-Bed for Text-to-Speech-Based Pedestrian Navigation Systems	396
<i>Michael Minock, Johan Mollevik, Mattias Åsander, and Marcus Karlsson</i>	
Automatic Detection of Arabic Causal Relations	400
<i>Jawad Sadek</i>	
A Framework for Employee Appraisals Based on Inductive Logic Programming and Data Mining Methods	404
<i>Darah Aqel and Sunil Vadera</i>	
A Method for Improving Business Intelligence Interpretation through the Use of Semantic Technology	408
<i>Shane Givens, Veda Storey, and Vijayan Sugumaran</i>	

Code Switch Point Detection in Arabic	412
<i>Heba Elfardy, Mohamed Al-Badrashiny, and Mona Diab</i>	
SurveyCoder: A System for Classification of Survey Responses	417
<i>Sangameshwar Patil and Girish K. Palshikar</i>	
Rhetorical Representation and Vector Representation in Summarizing Arabic Text	421
<i>Ahmed Ibrahim and Tarek Elghazaly</i>	
Author Index	425