

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

Gosse Bouma Ashwin Ittoo
Elisabeth Métais Hans Wortmann (Eds.)

Natural Language Processing and Information Systems

17th International Conference on Applications
of Natural Language to Information Systems, NLDB 2012
Groningen, The Netherlands, June 26-28, 2012
Proceedings

Volume Editors

Gosse Bouma
University of Groningen
Information Science Department
Oude Kijk in 't Jatstraat 26, 9712 EK Groningen, The Netherlands
E-mail: g.bouma@rug.nl

Ashwin Ittoo
Hans Wortmann
University of Groningen
Faculty of Economics and Business
Nettelbosje 2, 9747 AE Groningen, The Netherlands
E-mail: {r.a.ittoo, j.c.wortmann}@rug.nl

Elisabeth Métais
CNAM-Laboratoire Cédric
292 rue St. Martin, 75141 Paris Cedex 03, France
E-mail: elisabeth.metais@cnam.fr

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-31177-2 e-ISBN 978-3-642-31178-9
DOI 10.1007/978-3-642-31178-9
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 20129396643

CR Subject Classification (1998): I.2.7, H.3, H.2.8, I.5, J.5, I.2.6, J.1

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

© 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

These are the proceedings of the 17th International Conference on Applications of Natural Language to Information Systems, also known as NLDB 2012, that was organized in Groningen, The Netherlands, during June 26–28. Since the first NLDB conference in 1995, the main focus of the conference has widened from using natural language processing techniques in the area of databases and information systems to more general applications of NLP that help to make large and complex collections of data and information accessible and manageable.

The rapidly evolving state of the art in NLP and the shifting interest to applications targeting document and data collections available on the Web, including an increasing amount of user-generated content, is reflected in the contributions to this conference. Topics covered are, among others, information retrieval and text classification and clustering, summarization, normalization of user generated content, ‘forensic’ NLP (addressing plagiarism, cyberbullying, and fake reviews), ontologies and natural language, sentiment analysis, question answering and information extraction, terminology and named entity recognition, and NLP tools development.

For this edition of NLDB, we received over 90 submissions. The Program Committee, consisting of renowned researchers in the area of natural language processing and information systems, did an excellent job in providing detailed and well-motivated reviews. In all, 12 papers were selected as full papers for the conference and a number of other contributions as short papers (24) or posters (16).

We would like to thank all reviewers for their time and effort, and our invited speakers, Philipp Cimiano and Bernhard Thalheim, for accepting our invitation and making the conference attractive by their contributions. Finally, we would like to thank everyone involved in the local organization, especially the Department of Business and ICT and the Department of Information Science of the University of Groningen, the *Groningen Congres Bureau*, for helping us with all practical issues, *Het Kasteel*, for hosting the conference, and *Gezamenlijk Gastheerschap Groningen*, for sponsoring the welcome reception.

June 2012

Gosse Bouma
Ashwin Ittoo
Elisabeth Métais
Hans Wortmann

Erhard Hinrichs	Tübingen University, Germany
Helmut Horacek	Saarland University, Germany
Paul Johannesson	Stockholm University, Sweden
Epaminondas Kapetanios	University of Westminster, UK
Sophia Katrenko	Utrecht University, The Netherlands
Zoubida Kedad	Université de Versailles, France
Christian Kop	University of Klagenfurt, Austria
Valia Kordoni	Saarland University, Germany
Leila Kosseim	Concordia University, Canada
Georgia Koutrika	Stanford University, USA
Zornitsa Kozareva	University of Southern California, USA
Nadira Lammari	CNAM, France
Dominique Laurent	Synapse, 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
Heinrich C. Mayr	University of Klagenfurt, Austria
Farid Meziane	Salford University, UK
Luisa Mich	University of Trento, Italy
Andres Montoyo	Universidad de Alicante, Spain
Rafael Muñoz	Universidad de Alicante, Spain
John Nerbonne	University of Groningen, The Netherlands
Guenter Neumann	DFKI, Germany
Gertjan van Noord	University of Groningen, The Netherlands
Jan Odijk	Utrecht University, The Netherlands
Stephan Oepen	Oslo University, Norway
Manuel Palomar Sanz	Universidad de Alicante, Spain
Pit Pichappan	Al Imam University, Saudi Arabia
Lonneke van der Plas	Université de Genève, Switzerland
Gabor Proszeky	Morphologic, Hungary
Mike Rosner	University of Malta, Malta
Fabio Rinaldi	University of Zurich, Switzerland
German Rigau	University of the Basque Country, Spain
Patrick Saint-Dizier	Université Paul Sabatier, France
Max Silberztein	Université de Franche-Comté, France
Ielka van der Sluis	University of Groningen, The Netherlands
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
Jörg Tiedemann	Uppsala University, Sweden

Juan Carlos Trujillo
Christina Unger
Panos Vassiliadis
Piek Vossen
Robert Wagner
René Witte
Magdalena Wolska
Jakub Zavrel

Universidad de Alicante, Spain
Universität Bielefeld, Germany
University of Ioannina, Greece
Free University, The Netherlands
Linz University, Austria
Concordia University, Canada
Saarland University, Germany
Textkernel, The Netherlands

Table of Contents

Invited Talk

Syntax, Semantics and Pragmatics of Conceptual Modelling	1
<i>Bernhard Thalheim</i>	

Full Papers

Multi-dimensional Analysis of Political Documents	11
<i>Heiner Stuckenschmidt and Cécilia Zirn</i>	
Fake Reviews: The Malicious Perspective	23
<i>Theodoros Lappas</i>	
Polarity Preference of Verbs: What Could Verbs Reveal about the Polarity of Their Objects?	35
<i>Manfred Klenner and Stefanos Petrakis</i>	
Labeling Queries for a People Search Engine	47
<i>Antje Schlaf, Amit Kirschenbaum, Robert Remus, and Thomas Efer</i>	
Litmus: Generation of Test Cases from Functional Requirements in Natural Language	58
<i>Anurag Dwarakanath and Shubhashis Sengupta</i>	
Extracting Multi-document Summaries with a Double Clustering Approach	70
<i>Sara Botelho Silveira and António Branco</i>	
Developing Multilingual Text Mining Workflows in UIMA and U-Compare	82
<i>Georgios Kontonasios, Ioannis Korkontzelos, and Sophia Ananiadou</i>	
Geographic Expansion of Queries to Improve the Geographic Information Retrieval Task	94
<i>José M. Perea-Ortega and L. Alfonso Ureña-López</i>	
Learning Good Decompositions of Complex Questions	104
<i>Yllias Chali, Sadid A. Hasan, and Kaiser Imam</i>	
A Semi Supervised Learning Model for Mapping Sentences to Logical form with Ambiguous Supervision	116
<i>Le Minh Nguyen and Akira Shimazu</i>	

On the Effect of Stopword Removal for SMS-Based FAQ Retrieval 128
Johannes Leveling

Wikimantic: Disambiguation for Short Queries 140
Christopher Boston, Sandra Carberry, and Hui Fang

Short Papers

Polish Language Processing Chains for Multilingual Information Systems 152
Maciej Ogrodniczuk and Adam Przepiórkowski

GPU-Accelerated Non-negative Matrix Factorization for Text Mining 158
Volodymyr Kysenko, Karl Rupp, Oleksandr Marchenko, Siegfried Selberherr, and Anatoly Anisimov

Generating SQL Queries Using Natural Language Syntactic Dependencies and Metadata 164
Alessandra Giordani and Alessandro Moschitti

Two-Stage Named-Entity Recognition Using Averaged Perceptrons 171
Lars Buitinck and Maarten Marr

Using Natural Language Processing to Improve Document Categorization with Associative Networks 177
Niels Bloom

MLICC: A Multi-Label and Incremental Centroid-Based Classification of Web Pages by Genre 183
Chaker Jebari

Classifying Image Galleries into a Taxonomy Using Metadata and Wikipedia 191
Gerwin Kramer, Gosse Bouma, Dennis Hendriksen, and Mathijs Homminga

Supervised HDP Using Prior Knowledge 197
Boyi Xie and Rebecca J. Passonneau

User-Driven Automatic Resource Retrieval Based on Natural Language Request 203
Edgar Camilo Pedraza, Julián Andrés Zúñiga, Luis Javier Suarez-Meza, and Juan Carlos Corrales

Integrating Lexical-Semantic Knowledge to Build a Public Lexical Ontology for Portuguese 210
Hugo Gonçalo Oliveira, Leticia Antón Pérez, and Paulo Gomes

From Ontology to NL: Generation of Multilingual User-Oriented Environmental Reports	216
<i>Nadjet Bouayad-Agha, Gerard Casamayor, Simon Mille, Marco Rospocher, Horacio Saggion, Luciano Serafini, and Leo Wanner</i>	
Web-Based Relation Extraction for the Food Domain	222
<i>Michael Wiegand, Benjamin Roth, and Dietrich Klakow</i>	
Blog Distillation via Sentiment-Sensitive Link Analysis	228
<i>Giacomo Berardi, Andrea Esuli, Fabrizio Sebastiani, and Fabrizio Silvestri</i>	
Performing Groupization in Data Warehouses: Which Discriminating Criterion to Select?	234
<i>Eya Ben Ahmed, Ahlem Nabli, and Faïez Gargouri</i>	
The Study of Informality as a Framework for Evaluating the Normalisation of Web 2.0 Texts	241
<i>Alejandro Mosquera and Paloma Moreda</i>	
Modeling Math Word Problems with Augmented Semantic Networks . . .	247
<i>Christian Liguda and Thies Pfeiffer</i>	
Improving Document-Level Sentiment Classification Using Contextual Valence Shifters	253
<i>Sara A. Morsy and Ahmed Rafea</i>	
Comparing Different Methods for Opinion Mining in Newspaper Articles	259
<i>Thomas Scholz, Stefan Conrad, and Isabel Wolters</i>	
Extracting Social Events Based on Timeline and Sentiment Analysis in Twitter Corpus	265
<i>Bayar Tsolmon, A-Rong Kwon, and Kyung-Soon Lee</i>	
Can Text Summaries Help Predict Ratings? A Case Study of Movie Reviews	271
<i>Horacio Saggion, Elena Lloret, and Manuel Palomar</i>	
Towards User Modelling in the Combat against Cyberbullying	277
<i>Maral Dadvar, Roeland Ordelman, Franciska de Jong, and Dolf Trieschnigg</i>	
Plag-Inn: Intrinsic Plagiarism Detection Using Grammar Trees	284
<i>Michael Tschuggnall and Günther Specht</i>	
On the Application of Spell Correction to Improve Plagiarism Detection	290
<i>Daniel Micol, Óscar Ferrández, and Rafael Muñoz</i>	

PIRPO: An Algorithm to Deal with Polarity in Portuguese Online
Reviews from the Accommodation Sector 296
*Marcirio Silveira Chaves, Larissa A. de Freitas, Marlo Souza, and
Renata Vieira*

Posters

Towards Interrogative Types in Task-Oriented Dialogue Systems 302
Markus M. Berg, Antje D sterh ft, and Bernhard Thalheim

Teaching Business Systems to Agree 308
Fred van Blommestein

Improving Chinese Event Construction Extraction with Lexical
Relation Pairs 314
Liou Chen, Qiang Zhou, and Hongxian Wang

A Classifier Based Approach to Emotion Lexicon Construction 320
Dipankar Das, Soujanya Poria, and Sivaji Bandyopadhyay

Technical Term Recognition with Semi-supervised Learning Using
Hierarchical Bayesian Language Models 327
Ryo Fujii and Akito Sakurai

An Experience Developing a Semantic Annotation System in a Media
Group 333
Angel L. Garrido, Oscar G mez, Sergio Ilarri, and Eduardo Mena

From Requirements to Code: Syntax-Based Requirements Analysis for
Data-Driven Application Development 339
Goran Glavaš, Krešimir Fertalj, and Jan Šnajder

Structuring Political Documents for Importance Ranking 345
Alexander Hogenboom, Maarten Jongmans, and Flavius Frasinca

Author Disambiguation Using Wikipedia-Based Explicit Semantic
Analysis 351
In-Su Kang

Automatic Population of Korean Information in Linking Open Data 355
Shin-Jae Kang and In-Su Kang

Initial Results from a Study on Personal Semantics of Conceptual
Modeling Languages 360
Dirk van der Linden, Khaled Gaaloul, and Wolfgang Molnar

Lexical Knowledge Acquisition Using Spontaneous Descriptions
in Texts 366
Augusta Mela, Mathieu Roche, and Mohamed El Amine Bekhtaoui

Interacting with Data Warehouse by Using a Natural Language Interface	372
<i>M. Asif Naeem, Saif Ullah, and Imran Sarwar Bajwa</i>	
Processing Semantic Keyword Queries for Scientific Literature	378
<i>Ibrahim Burak Ozyurt, Christopher Condit, and Amarnath Gupta</i>	
Arabic Rhetorical Relations Extraction for Answering “Why” and “How to” Questions	385
<i>Jawad Sadek, Fairouz Chakkour, and Farid Meziane</i>	
Facets of a Discourse Analysis of Safety Requirements	391
<i>Patrick Saint-Dizier</i>	
Author Index	397