Lecture Notes in Computer Science 10859

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/7409

Max Silberztein · Faten Atigui
Elena Kornyshova · Elisabeth Métais
Farid Meziane (Eds.)

Natural Language Processing and Information Systems

23rd International Conference on Applications of Natural Language to Information Systems, NLDB 2018
Paris, France, June 13–15, 2018
Proceedings

Springer

Editors
Max Silberztein
Université de Franche-Comté
Besançon
France

Faten Atigui
Conservatoire National des Arts et Métiers
Paris
France

Elena Kornyshova
Conservatoire National des Arts et Métiers
Paris
France

Elisabeth Métais
Conservatoire National des Arts et Métiers
Paris
France

Farid Meziane
University of Salford
Manchester
UK

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-91946-1 ISBN 978-3-319-91947-8 (eBook)
https://doi.org/10.1007/978-3-319-91947-8

Library of Congress Control Number: 2018944303

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

© Springer International Publishing AG, part of Springer Nature 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.

Printed on acid-free paper

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

Preface

This volume of *Lecture Notes in Computer Science* (LNCS) contains the papers selected for presentation at the 23rd International Conference on Natural Language and Information Systems, held in Paris, France (CNAM), during June 13–15, 2018 (NLDB2018).

Since its foundation in 1995, the NLDB conference has brought together researchers, industry practitioners, and potential users interested in various application of natural language in the database and information systems field. The term "information systems" has to be considered in the broader sense of information and communication systems, including big data, linked data, and social networks. The field of natural language processing (NLP) has itself recently experienced several exciting developments. In research, these developments have been reflected in the emergence of neural language models (Deep Learning, Word2Vec) and a (renewed) interest in various linguistic phenomena, such as in discourse and argumentation theory (argumentation mining). Regarding applications, NLP systems have evolved to the point that they now offer real-life, tangible benefits to enterprises. Many of these NLP systems are now considered de facto offerings in business intelligence suites, such as algorithms for recommender systems and opinion mining/sentiment analysis as well as question-answering systems.

It is against this backdrop of recent innovations in NLP and its applications in information systems that the 23rd edition of the NLDB conference, NLDB 2018, took place. We have welcomed research and industrial contributions, describing novel, previously unpublished works on NLP and its applications across a plethora of topics. Indeed, this year NLDB had 99 submissions from which we selected 18 long papers (18% acceptance rate), 26 short papers, and nine posters, and each paper was reviewed by an average of three reviewers. The conference and Program Committee co-chairs had a final consultation meeting to look at all the reviews and make the final decisions.

This year, we had the opportunity to welcome two invited presentations from:

- Prof. Jean-Gabriel Ganascia (LIP6 Laboratory, Institut Universitaire de France)
- Dr. Serge Rosmorduc (Cedric Laboratory, Conservatoire National des Arts et Métiers, France)

We would like to thank all who submitted papers for reviews and for publication in the proceedings and all the reviewers for their time, effort, and for completing their assignments on time.

April 2018

Max Silberztein
Faten Atigui
Elena Kornyshova
Elisabeth Métais
Farid Meziane

Organization

General Chairs

Métais Elisabeth	Conservatoire des Arts et Métiers, France
Meziane Farid	University of Salford, UK
Silberztein Max	Université de Franche-Comté, France

Program Committee

Silberztein Max (Chair)	Université de Franche-Comté, France
Akoka Jacky	CNAM and TEM, France
Aras Hidir	FIZ Karlsruhe, Germany
Bajwa Sarwar	Imran, The Islamia University of Bahawalpur, Pakistan
Balakrishna	Mithun, Limba Corp, USA
Basile Pierpaolo	University of Bali, Italy
Béchet Nicolas	IRISA, France
Bordogna	Gloria, IREA, CNR, Italy
Bouma Gosse	Groningen University, The Netherlands
Bringay Sandra	LIRMM, France
Buitelaar Paul	Insight National University of Ireland, Galway
Caragea Cornelia	University of North Texas, USA
Christian Chiarcos	Universität Frankfurt am Main, Germany
Chiky Raja	ISEP, France
Cimiano Philipp	Universität Bielefeld, Germany
Comyn-Wattiau Isabelle	ESSEC, France
Frasincar Flavius	Erasmus University Rotterdam
Freitas André	University of Passau, Germany/Insight, Ireland
Ganguly Debasis	Dublin City University, Ireland
Gelbukh Alexander	CIC-IPN, Mexico
Guessoum Ahmed	USTHB, Algeria
Hacohen-Kerner Yaakov	Jerusalem College of Technology, Israel
Hahn Udo	University of Freiburg, Germany
Handschuh Siegfried	University of Passau, Germany
Herweg Michael	IBM, Germany
Horacek Helmut	Saarland University, Germany
Ienco Dino	IRSTEA, France
Ittoo Ashwin	HEC, University of Liege, Belgium
Johannesson Paul	Stockholm University, Sweden
Kapetanios Epaminondas	University of Westminster, UK
Kedad Zoubida	UVSQ, France
Kergosien Eric	GERiiCO, University of Lille, France
Kop Christian	University of Klagenfurt, Austria

Kordoni Valia	Humboldt University Berlin, Germany
Lafourcade Mathieu	LIRMM University College Ghent
Lefever Els	Ghent University, Belgium
Leidner Jochen	Thomson Reuters, USA
Lonsdale Deryle W.	Brigham Young University, USA
Lopez Cédric	Emvista, Montpellier, France
Mayr Heinrich C.	Alpen-Adria-Universität Klagenfurt, Austria
Mccrae John P.	National University of Ireland, Galway
Métais Elisabeth	CNAM, France
Meurs Marie-Jean	UQAM, Montreal, Canada
Meziane Farid	Salford University, UK
Mich Luisa	University of Trento, Italy
Montoyo Andres	Universidad de Alicante, Spain
Muñoz Rafael	Universidad de Alicante, Spain
Nguyen Le Minh	Japan Advanced Institute of Science and Technology, Japan
Panchenko Alexander	TU Darmstadt, Germany
Saint Dizier Patrick	IRIT-CNRS, France
Paulheim Heiko	University of Mannheim, Germany
Pei Yulong	Eindhoven University of Technology, The Netherlands
Picca Davide	Université de Lausanne, Switzerland
Prince Violaine	LIRMM, France
Qasemizadeh Behrang	Heinrich-Heine-Universität Düsseldorf, Germany
Rezgui Yacine	Cardiff University, UK
Riedl Martin	University of Hamburg, Germany
Roche Mathieu	Cirad, TETIS, France
Rosso Paolo	Technical University of Valencia, Spain
Saraee Mohamad	University of Salford, Manchester, UK
Sateli Bahar	Concordia University, Canada
Schneider Roman	IDS Mannheim, Germany
Schouten Kim	Erasmus University Rotterdam
Shaalan Khaled	The British University in Dubai, UAE
Storey Veda	Georgia State University, USA
Sugumaran Vijayan	Oakland University Rochester, USA
Teisseire Maguelonne	Irstea, TETIS, France
Thalheim Bernhard	Christian Albrechts University Kiel, Germany
Thelwall Mike	University of Wolverhampton, UK
Thirunarayan Krishnaprasad	Wright State University, USA
Tufis Dan	RACAI, Bucharest, Romania
Ureña Alfonso	Universidad de Jaén, Spain
Xu Feiyu	DFKI Saarbrücken, Germany
Zadrozny Wlodek	UNCC, USA
Zanzotto Fabio Massimo	University of Rome Tor Vergata, Italy

Organizing Committee Chairs

Atigui Faten	Conservatoire des Arts et Métiers, France
Kornyshova Elena	Conservatoire des Arts et Métiers, France

Webmaster

Lemoine Frédéric	Conservatoire des Arts et Métiers, France

Contents

Opinion Mining and Sentiment Analysis in Social Media

A Proposal for Book Oriented Aspect Based Sentiment Analysis: Comparison over Domains 3
Tamara Álvarez-López, Milagros Fernández-Gavilanes, Enrique Costa-Montenegro, and Patrice Bellot

Stance Evolution and Twitter Interactions in an Italian Political Debate 15
Mirko Lai, Viviana Patti, Giancarlo Ruffo, and Paolo Rosso

Identifying and Classifying Influencers in Twitter only with Textual Information 28
Victoria Nebot, Francisco Rangel, Rafael Berlanga, and Paolo Rosso

Twitter Sentiment Analysis Experiments Using Word Embeddings on Datasets of Various Scales 40
Yusuf Arslan, Dilek Küçük, and Aysenur Birturk

A Deep Learning Approach for Sentiment Analysis Applied to Hotel's Reviews 48
Joana Gabriela Ribeiro de Souza, Alcione de Paiva Oliveira, Guidson Coelho de Andrade, and Alexandra Moreira

Automatic Identification and Classification of Misogynistic Language on Twitter 57
Maria Anzovino, Elisabetta Fersini, and Paolo Rosso

Assessing the Effectiveness of Affective Lexicons for Depression Classification 65
Noor Fazilla Abd Yusof, Chenghua Lin, and Frank Guerin

Semantics-Based Models and Applications

Toward Human-Like Robot Learning 73
Sergei Nirenburg, Marjorie McShane, Stephen Beale, Peter Wood, Brian Scassellati, Olivier Magnin, and Alessandro Roncone

Identifying Argumentative Paragraphs: Towards Automatic Assessment of Argumentation in Theses 83
Jesús Miguel García-Gorrostieta and Aurelio López-López

Semantic Mapping of Security Events to Known Attack Patterns 91
Xiao Ma, Elnaz Davoodi, Leila Kosseim, and Nicandro Scarabeo

Neural Networks Based Approaches

Accommodating Phonetic Word Variations Through Generated Confusion Pairs for Hinglish Handwritten Text Recognition. 101
Soumyajit Mitra, Vikrant Singh, Pragya Paramita Sahu, Viswanath Veera, and Shankar M. Venkatesan

Arabic Question Classification Using Support Vector Machines and Convolutional Neural Networks 113
Asma Aouichat, Mohamed Seghir Hadj Ameur, and Ahmed Geussoum

A Supervised Learning Approach for ICU Mortality Prediction Based on Unstructured Electrocardiogram Text Reports. 126
Gokul S. Krishnan and S. Sowmya Kamath

Multimodal Language Independent App Classification Using Images and Text 135
Kushal Singla, Niloy Mukherjee, and Joy Bose

T2S: An Encoder-Decoder Model for Topic-Based Natural Language Generation. 143
Wenjie Ou, Chaotao Chen, and Jiangtao Ren

Ontology Engineering

Ontology Development Through Concept Map and Text Analytics: The Case of Automotive Safety Ontology 155
Zirun Qi and Vijayan Sugumaran

A Fuzzy-Based Approach for Representing and Reasoning on Imprecise Time Intervals in Fuzzy-OWL 2 Ontology 167
Fatma Ghorbel, Fayçal Hamdi, Elisabeth Métais, Nebrasse Ellouze, and Faiez Gargouri

Assessing the Impact of Single and Pairwise Slot Constraints in a Factor Graph Model for Template-Based Information Extraction. 179
Hendrik ter Horst, Matthias Hartung, Roman Klinger, Nicole Brazda, Hans Werner Müller, and Philipp Cimiano

NLP

Processing Medical Binary Questions in Standard Arabic Using NooJ 193
Essia Bessaies, Slim Mesfar, and Henda Ben Ghzela

Verbal Multi-Word Expressions in Yiddish 205
Chaya Liebeskind and Yaakov HaCohen-Kerner

Stochastic Approach to Aspect Tracking 217
Maoto Inoue and Takao Miura

Transducer Cascade to Parse Arabic Corpora 230
Nadia Ghezaiel Hammouda, Roua Torjmen, and Kais Haddar

Multi-Word Expressions Annotations Effect in Document Classification Task 238
Dhekra Najar, Slim Mesfar, and Henda Ben Ghezela

RDST: A Rule-Based Decision Support Tool 247
Sondes Dardour and Héla Fehri

Mention Clustering to Improve Portuguese Semantic Coreference Resolution 256
Evandro Fonseca, Aline Vanin, and Renata Vieira

Resource Creation for Training and Testing of Normalisation Systems for Konkani-English Code-Mixed Social Media Text 264
Akshata Phadte

A TF-IDF and Co-occurrence Based Approach for Events Extraction from Arabic News Corpus 272
Amina Chouigui, Oussama Ben Khiroun, and Bilel Elayeb

English Text Parsing by Means of Error Correcting Automaton 281
Oleksandr Marchenko, Anatoly Anisimov, Igor Zavadskyi, and Egor Melnikov

Annotating Relations Between Named Entities with Crowdsourcing 290
Sandra Collovini, Bolivar Pereira, Henrique D. P. dos Santos, and Renata Vieira

Automatic Detection of Negated Findings with NooJ: First Results 298
Walter Koza, Mirian Muñoz, Natalia Rivas, Ninoska Godoy, Darío Filippo, Viviana Cotik, Vanesa Stricker, and Ricardo Martínez

Part-of-Speech Tagger for Konkani-English Code-Mixed Social Media Text 303
Akshata Phadte and Radhiya Arsekar

Word2vec for Arabic Word Sense Disambiguation 308
Rim Laatar, Chafik Aloulou, and Lamia Hadrich Belghuith

Text Similarities and Plagiarism Detection

HYPLAG: Hybrid Arabic Text Plagiarism Detection System 315
Bilal Ghanem, Labib Arafeh, Paolo Rosso, and Fernando Sánchez-Vega

On the Semantic Similarity of Disease Mentions in MEDLINE® and Twitter 324
Camilo Thorne and Roman Klinger

Gemedoc: A Text Similarity Annotation Platform 333
Jacques Fize, Mathieu Roche, and Maguelonne Teisseire

Text Classification

Addressing Unseen Word Problem in Text Classification 339
Promod Yenigalla, Sibsambhu Kar, Chirag Singh, Ajay Nagar, and Gaurav Mathur

United We Stand: Using Multiple Strategies for Topic Labeling 352
Antoine Gourru, Julien Velcin, Mathieu Roche, Christophe Gravier, and Pascal Poncelet

Evaluating Thesaurus-Based Topic Models. 364
Natalia Loukachevitch and Kirill Ivanov

Classifying Companies by Industry Using Word Embeddings 377
Martin Lamby and Daniel Isemann

Towards Ontology-Based Training-Less Multi-label Text Classification 389
Wael Alkhatib, Saba Sabrin, Svenja Neitzel, and Christoph Rensing

Overview of Uni-modal and Multi-modal Representations for Classification Tasks 397
Aryeh Wiesen and Yaakov HaCohen-Kerner

Information Mining

Classification of Intangible Social Innovation Concepts 407
Nikola Milosevic, Abdullah Gok, and Goran Nenadic

An Unsupervised Approach for Cause-Effect Relation Extraction from Biomedical Text 419
Raksha Sharma, Girish Palshikar, and Sachin Pawar

A Supervised Learning to Rank Approach for Dependency Based Concept Extraction and Repository Based Boosting for Domain Text Indexing 428
U. K. Naadan, T. V. Geetha, U. Kanimozhi, D. Manjula, R. Viswapriya, and C. Karthik

[Demo] Integration of Text- and Web-Mining Results in EpidVis 437
Samiha Fadloun, Arnaud Sallaberry, Alizé Mercier, Elena Arsevska, Pascal Poncelet, and Mathieu Roche

Recommendation Systems

Silent Day Detection on Microblog Data 443
Kuang Lu and Hui Fang

Smart Entertainment - A Critiquing Based Dialog System for Eliciting User Preferences and Making Recommendations 456
Roshni R. Ramnani, Shubhashis Sengupta, Tirupal Rao Ravilla, and Sumitraj Ganapat Patil

Translation and Foreign Language Querying

Cross-Language Text Summarization Using Sentence and Multi-Sentence Compression 467
Elvys Linhares Pontes, Stéphane Huet, Juan-Manuel Torres-Moreno, and Andréa Carneiro Linhares

Arabic Readability Assessment for Foreign Language Learners 480
Naoual Nassiri, Abdelhak Lakhouaja, and Violetta Cavalli-Sforza

String Kernels for Polarity Classification: A Study Across Different Languages 489
Rosa M. Giménez-Pérez, Marc Franco-Salvador, and Paolo Rosso

Integration of Neural Machine Translation Systems for Formatting-Rich Document Translation 494
Mārcis Pinnis, Raivis Skadiņš, Valters Šics, and Toms Miks

Software Requirement and Checking

Using k-Means for Redundancy and Inconsistency Detection: Application to Industrial Requirements. 501
Manel Mezghani, Juyeon Kang, and Florence Sèdes

How to Deal with Inaccurate Service Descriptions in On-The-Fly Computing: Open Challenges 509
Frederik S. Bäumer and Michaela Geierhos

Author Index 515