

Lecture Notes in Artificial Intelligence

10318

Subseries of Lecture Notes in Computer Science

LNAI Series Editors

Randy Goebel

University of Alberta, Edmonton, Canada

Yuzuru Tanaka

Hokkaido University, Sapporo, Japan

Wolfgang Wahlster

DFKI and Saarland University, Saarbrücken, Germany

LNAI Founding Series Editor

Joerg Siekmann

DFKI and Saarland University, Saarbrücken, Germany

More information about this series at <http://www.springer.com/series/1244>


Jorge Gracia · Francis Bond
John P. McCrae · Paul Buitelaar
Christian Chiarcos · Sebastian Hellmann (Eds.)


Language, Data, and Knowledge

First International Conference, LDK 2017
Galway, Ireland, June 19–20, 2017
Proceedings

Editors


Jorge Gracia 
Universidad Politécnica de Madrid
Madrid
Spain

Francis Bond 
Nanyang Technological University
Singapore
Singapore

John P. McCrae 
Insight Centre for Data Analytics
National University of Ireland, Galway
Galway
Ireland

Paul Buitelaar
Insight Centre for Data Analytics
National University of Ireland
Galway
Ireland

Christian Chiarcos 
Goethe-University Frankfurt
Frankfurt
Germany

Sebastian Hellmann 
University of Leipzig
Leipzig
Germany

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-59887-1 ISBN 978-3-319-59888-8 (eBook)
DOI 10.1007/978-3-319-59888-8

Library of Congress Control Number: 2017942992

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer International Publishing AG 2017

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 Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Welcome to Galway for the first international conference on Language, Data and Knowledge (LDK 2017). It is hosted by the Insight Centre for Data Analytics at the National University of Ireland, Galway.

This is the first of a new biennial conference series on Language, Data and Knowledge (LDK) which aims to bring together researchers from across disciplines concerned with the acquisition, curation and use of language data in the context of data science and knowledge-based applications. With the advent of the Web and digital technologies, an ever increasing amount of language data is now available across application areas and industry sectors, including social media, digital archives, company records, etc. The efficient and meaningful exploitation of this data in scientific and commercial innovation is at the core of data science research, employing NLP and machine learning methods as well as semantic technologies based on knowledge graphs.

Language data is of increasing importance to machine learning-based approaches in NLP, Linked Data and Semantic Web research and applications that depend on linguistic and semantic annotation with lexical, terminological and ontological resources, manual alignment across language or other human-assigned labels. The acquisition, provenance, representation, maintenance, usability, quality as well as legal, organizational and infrastructure aspects of language data are therefore rapidly becoming major areas of research that are at the focus of the conference.

Knowledge graphs are an active field of research concerned with the extraction, integration, maintenance and use of semantic representations of language data in combination with semantically or otherwise structured data, numerical data and multimodal data among others. Knowledge graph research builds on the exploitation and extension of lexical, terminological and ontological resources, information and knowledge extraction, entity linking, ontology learning, ontology alignment, semantic text similarity, Linked Data and other Semantic Web technologies. The construction and use of knowledge graphs from language data, possibly and ideally in the context of other types of data, is a further specific focus of the conference.

A further focus of the conference is the combined use and exploitation of language data and knowledge graphs in data science-based approaches to use cases in industry, including biomedical applications, as well as use cases in humanities and social sciences.

The LDK conferences have been initiated by a consortium of researchers from the Insight Centre for Data Analytics, InfAI (University Leipzig) and Wolfgang Goethe University and a Scientific Committee of leading researchers in Natural Language Processing, Linked Data and Semantic Web, Language Resources and Digital Humanities. LDK is endorsed by several international organisations: DBpedia, ACL SIGANN, Global Wordnet Association, CLARIN and Big Data Value Association (BDVA).

There were a total of 68 submissions reviewed, of which 15 were accepted as full papers (one was subsequently withdrawn) and 19 as short papers (an acceptance rate of 50%).

Reviewing was single blind (due to the difficulty in making language resources anonymous). There were at least 3 reviews per paper, with an average of 3.7. Papers came from all over the world.

A successful conference is the result of many people's efforts and contributions. The papers included in the program are the results of our participants' contributions as well as the insightful reviews from the program committee members. Besides the oral and poster paper presentations, the program is enriched by the keynote and invited speakers: Dr. Chris Welty, Senior Research Scientist at Google in New York, and an Endowed Professor of Cognitive Systems at the VU University, Amsterdam; Kathleen R. McKeown, Henry and Gertrude Rothschild Professor of Computer Science and Director of the Institute for Data Sciences and Engineering at Columbia University; Antal van den Bosch, director of the Meertens Institute in Amsterdam and Professor of Language and Speech Technology at the Centre for Language Studies at Radboud University, Nijmegen, the Netherlands; and Graham Isaac of the Irish Department at NUI Galway. On behalf of the program committee, we express our thanks to them all.

We would like to thank the scientific advisory committee for their guidance, and the local organizing committee led by Dr John P. McCrae and Dr Paul Buitelaar, for their great work in ensuring the success of this conference.

Finally, we wish that you will all enjoy the conference presentations, discussions, and exchanges between old and new friends at the beautiful harbor city of Galway.

April 2017

Jorge Gracia
Francis Bond

Organisation

Paul Buitelaar	Insight Centre for Data Analytics, NUI Galway, Ireland
Christian Chiarcos	Goethe Universität Frankfurt, Germany
Sebastian Hellmann	InfAI, Universität Leipzig, Germany
John P. McCrae	Insight Centre for Data Analytics, NUI Galway, Ireland

Scientific Advisory Committee

Pushpak Bhattacharyya	IITP, India
Francis Bond	Nanyang Technological University, Singapore
Key-Sun Choi	KAIST, South Korea
Philipp Cimiano	Bielefeld University, Germany
Edward Curry	Insight Centre for Data Analytics, NUI Galway, Ireland
Franciska de Jong	Utrecht University, The Netherlands
Thierry Declerck	DFKI GmbH, Saarland University, Germany
Tatjana Gornostaja	Tilde, Latvia
Jorge Gracia	Universidad Politécnica de Madrid, Spain
Nancy Ide	Vassar College, USA
Eric Nyberg	Carnegie Mellon University, USA
Felix Sasaki	DFKI GmbH, W3C Fellow, Germany
Karin Verspoor	University of Melbourne, Australia

Program Chairs

Francis Bond	Nanyang Technological University, Singapore
Jorge Gracia	Universidad Politécnica de Madrid, Spain

Program Committee

Nitish Aggarwal	IBM Watson, USA
Guadalupe Aguado	Universidad Politécnica de Madrid, Spain
Sören Auer	University of Bonn, Germany
Caroline Barrière	Computer Research Institute of Montreal, Canada
Núria Bel	Universitat Pompeu Fabra, Spain
Pushpak Bhattacharyya	IITP, India
Francis Bond	Nanyang Technological University, Singapore
Claire Bonial	U.S. Army Research Lab, USA
Carmen Brando	Institut National de L'Information Géographique et Forestière, France
Nicoletta Calzolari	ILC-CNR, Italy
Steve Cassidy	Macquarie University, Australia

Damir Cavar	Indiana University, USA
Key-Sun Choi	KAIST, South Korea
Philipp Cimiano	Bielefeld University, Germany
Kevin B. Cohen	University of Colorado School of Medicine, USA
Edward Curry	Insight Centre for Data Analytics, NUI Galway, Ireland
Brian Davis	Insight Centre for Data Analytics, NUI Galway, Ireland
Thierry Declerck	DFKI GmbH, Saarland University, Germany
Franciska de Jong	Utrecht University, The Netherlands
Gerard de Melo	Rutgers University, USA
Dongpo Deng	Institute of Information Science, Academia Sinica, Taiwan
Alexis Dimitriadis	Universiteit Utrecht, The Netherlands
Richard Eckart de Castilho	Technische Universität Darmstadt, Germany
Agata Filipowska	Poznan University of Economics, Poland
André Freitas	University of Passau, Germany
Francesca Frontini	Université Paul Valéry Montpellier, France
Elena González-Blanco García	Universidad Nacional de Educación a Distancia, Madrid, Spain
Jeff Good	University at Buffalo, USA
Tatjana Gornostaja	Tilde, Latvia
Dagmar Gromann	IIIA-CSIC, Barcelona, Spain
Yoshihiko Hayashi	Waseda University, Japan
Graeme Hirst	University of Toronto, Canada
Nancy Ide	Vassar College, USA
Hitoshi Isahara	Toyohashi University of Technology, Japan
Clement Jonquet	University of Montpellier, France
Fahad Khan	ILC-CNR, Italy
Bettina Klimek	Leipzig University AKSW, Germany
Roman Klingner	University of Stuttgart, Germany
Seiji Koide	Ontology, LLC, Japan
Dimitris Kontokostas	Leipzig University AKSW, Germany
Vanessa Lopez	IBM Europe, Ireland
Monica Monachini	Consiglio Nazionale delle Ricerche, Italy
Elena Montiel	Universidad Politécnica de Madrid, Spain
Steven Moran	University of Zürich, Switzerland
Luis Morgado Da Costa	Nanyang Technical University, Singapore
Andrea Moro	Microsoft, UK
Hatem Mousselly Sergieh	Darmstadt University, Germany
Yohei Murakami	Kyoto University, Japan
Roberto Navigli	Sapienza University of Rome, Italy
Petya Osenova	Bulgarian Academy of Sciences, Sofia, Bulgaria
Maciej Piasecki	Wroclaw University of Technology, Poland
Simone Ponzetto	University of Mannheim, Germany
Laurette Pretorius	UNISA, South Africa
Nils Reiter	University of Stuttgart, Germany
Mariano Rico	Universidad Politécnica de Madrid, Spain

Martin Riedl	University of Hamburg, Germany
Víctor Rodríguez-Doncel	Universidad Politécnica de Madrid, Spain
Harald Sack	FIZ Karlsruhe, Leibniz Center for Information Infrastructure, Germany
Felix Sasaki	DFKI GmbH, W3C Fellow, Germany
Agata Savary	University of Tours, France
Andrea Schalley	Griffith University, Australia
Sabine Schulte im Walde	University of Stuttgart, Germany
Gilles Sérasset	Grenobles Alpes University, France
Kiril Simov	Bulgarian Academy of Sciences, Sofia, Bulgaria
Milena Slavcheva	JRC-Brussels, Belgium
Armando Stellato	University of Rome, Tor Vergata, Italy
Marko Tadić	University of Zagreb, Croatia
Hideaki Takeda	National Institute of Informatics, Japan
Ricardo Usbeck	Leipzig University AKSW, Germany
Marieke van Erp	Vrije Universiteit Amsterdam, The Netherlands
Marc Verhagen	Brandeis University, USA
Karin Verspoor	University of Melbourne, Australia
Cristina Vertan	University of Hamburg, Germany
Marta Villegas	Universitat Autònoma de Barcelona, Spain
Piek Vossen	Vrije Universiteit Amsterdam, The Netherlands
Sebastian Walter	Bielefeld University, Germany
Ulli Waltinger	Siemens AG, Germany
Eveline Wandl-Vogt	Austrian Academy of Science, Austria
Haofen Wang	East China University of Science and Technology, China
Krzysztof Weceł	Poznan University of Economics, Poland
Masaharu Yoshioka	Hokkaido University, Japan

Contents

An Evaluation Dataset for Linked Data Profiling.	1
<i>Andrejs Abele, John P. McCrae, and Paul Buitelaar</i>	
Personal Research Agents on the Web of Linked Open Data	10
<i>Bahar Sateli and René Witte</i>	
Hybrid Method for Stress Prediction Applied to GLAFF-IT, a Large-Scale Italian Lexicon	26
<i>Basilio Calderone, Matteo Pascoli, Franck Sajous, and Nabil Hathout</i>	
OnLiT: An Ontology for Linguistic Terminology	42
<i>Bettina Klimek, John P. McCrae, Christian Lehmann, Christian Chiarcos, and Sebastian Hellmann</i>	
Representing and Aligning Similar Relations: Parts and Wholes in isiZulu vs. English.	58
<i>C. Maria Keet</i>	
CoNLL-RDF: Linked Corpora Done in an NLP-Friendly Way	74
<i>Christian Chiarcos and Christian Fäth</i>	
LLODifying Linguistic Glosses.	89
<i>Christian Chiarcos, Maxim Ionov, Monika Rind-Pawlowski, Christian Fäth, Jesse Wichers Schreur, and Irina Nevskaya</i>	
A Semantic Frame-Based Similarity Metric for Characterizing Technological Capabilities	104
<i>Scott Appling and Erica Briscoe</i>	
Reassembling and Enriching the Life Stories in Printed Biographical Registers: Norssi High School Alumni on the Semantic Web	113
<i>Eero Hyvönen, Petri Leskinen, Erkki Heino, Jouni Tuominen, and Laura Sirola</i>	
Named Entity Linking in a Complex Domain: Case Second World War History	120
<i>Erkki Heino, Minna Tamper, Eetu Mäkelä, Petri Leskinen, Esko Ikkala, Jouni Tuominen, Mikko Koho, and Eero Hyvönen</i>	
Using SWRL Rules to Model Noun Behaviour in Italian	134
<i>Fahad Khan, Andrea Bellandi, Francesca Frontini, and Monica Monachini</i>	

Hunger for Contextual Knowledge and a Road Map to Intelligent Entity Linking	143
<i>Filip Ilievski, Piek Vossen, and Marieke van Erp</i>	
Weak Supervision for Semi-supervised Topic Modeling via Word Embeddings	150
<i>Gerald Conheady and Derek Greene</i>	
Towards Interoperability in the European Poetry Community: The Standardization of Philological Concepts	156
<i>Helena Bermúdez-Sabel, Mariana Curado Malta, and Elena Gonzalez-Blanco</i>	
Joint Entity Recognition and Linking in Technical Domains Using Undirected Probabilistic Graphical Models	166
<i>Hendrik ter Horst, Matthias Hartung, and Philipp Cimiano</i>	
Neural Induction of a Lexicon for Fast and Interpretable Stance Classification	181
<i>Jérémie Clos and Nirmalie Wiratunga</i>	
The Colloquial WordNet: Extending Princeton WordNet with Neologisms . . .	194
<i>John P. McCrae, Ian Wood, and Amanda Hicks</i>	
Shifting Complexity from Text to Data Model: Adding Machine-Oriented Features to a Human-Oriented Terminology Resource	203
<i>Karolina Suchowolec, Christian Lang, Roman Schneider, and Horst Schwinn</i>	
Comparison of Word Embeddings from Different Knowledge Graphs	213
<i>Kiril Simov, Petya Osenova, and Alexander Popov</i>	
Multi-pass Sieve Coreference Resolution System for Polish	222
<i>Bartłomiej Nitoń and Maciej Ogrodniczuk</i>	
Assessing VocBench Custom Forms in Supporting Editing of Lemon Datasets	237
<i>Manuel Fiorelli, Tiziano Lorenzetti, Maria Teresa Paziienza, and Armando Stellato</i>	
Answering the Hard Questions	253
<i>Maria Khvalchik, Chanin Pithyaachariyakul, and Anagha Kulkarni</i>	
Multilingual Fine-Grained Entity Typing	262
<i>Marieke van Erp and Piek Vossen</i>	

AATOS – A Configurable Tool for Automatic Annotation 276
*Minna Tamper, Petri Leskinen, Esko Ikkala, Arttu Oksanen,
Eetu Mäkelä, Erkki Heino, Jouni Tuominen, Mikko Koho,
and Eero Hyvönen*

Deep Text Generation – Using Hierarchical Decomposition to Mitigate
the Effect of Rare Data Points 290
Nina Dethlefs and Alexander Turner

Identifying Equivalent Relation Paths in Knowledge Graphs. 299
*Sameh K. Mohamed, Emir Muñoz, Vít Nováček,
and Pierre-Yves Vandenbussche*

SRDF: A Novel Lexical Knowledge Graph for Whole Sentence
Knowledge Extraction 315
Sangha Nam, GyuHyeon Choi, and Key-Sun Choi

Letting the Genie Out of the Lamp: Using Natural Language
Processing Tools to Predict Math Performance 330
Scott Crossley and Victor Kostyuk

Measuring Accuracy of Triples in Knowledge Graphs 343
Shuangyan Liu, Mathieu d’Aquin, and Enrico Motta

Exploring the Role of Gender in 19th Century Fiction Through the Lens
of Word Embeddings. 358
*Siobhán Grayson, Maria Mulvany, Karen Wade, Gerardine Meaney,
and Derek Greene*

Scalable Disambiguation System Capturing Individualities of Mentions 365
*Tiep Mai, Bichen Shi, Patrick K. Nicholson, Deepak Ajwani,
and Alessandra Sala*

Multi-label Text Classification Using Semantic Features and
Dimensionality Reduction with Autoencoders 380
Wael Alkhatib, Christoph Rensing, and Johannes Silberbauer

Author Index 395