

# Lecture Notes in Artificial Intelligence

10713

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>

Georg Rehm · Thierry Declerck (Eds.)

# Language Technologies for the Challenges of the Digital Age

27th International Conference, GSCL 2017  
Berlin, Germany, September 13–14, 2017  
Proceedings

*Editors*  
Georg Rehm  
DFKI GmbH  
Berlin  
Germany

Thierry Declerck  
DFKI GmbH  
Saarbrücken  
Germany



ISSN 0302-9743                      ISSN 1611-3349 (electronic)  
Lecture Notes in Artificial Intelligence  
ISBN 978-3-319-73705-8              ISBN 978-3-319-73706-5 (eBook)  
<https://doi.org/10.1007/978-3-319-73706-5>

Library of Congress Control Number: 2017963770

LNCS Sublibrary: SL7 – Artificial Intelligence

© The Editor(s) (if applicable) and The Author(s) 2018. This book is an open access publication.

**Open Access** This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

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

GSCL 2017 was the 27th edition of the biennial conference of the German Society for Computational Linguistics and Language Technology (GSCL e.V.). It was also the first to take place in Berlin. The conference offered, once again, high-quality articles covering a wide spectrum of computational linguistics research.

The unprecedented social, political, and economic relevance of online media and digital communication technologies inspired this year's main conference topic: "Language Technologies for the Challenges of the Digital Age." This relevance has, in turn, given rise to the still ongoing debate around several phenomena relating to online media, most importantly, around the topic of false news ("fake news"), online rumors, abusive language, and hate speech as well as filter bubbles and echo chambers. We have been actively engaged in several research activities around these topics, most notably in the project *Digitale Kuratierungstechnologien* (DKT; Digital Curation Technologies, funded by BMBF and coordinated by DFKI) and the European project *PHEME*, in which the University of Saarland participated. The two invited talks of GSCL 2017 examined aspects of our main conference topic from different perspectives: Holger Schwenk (Facebook AI Research, Paris, France) talked about "Learning Deep Representations in NLP," especially with regard to the aspect of overcoming language barriers with the help of machine translation. Kalina Bontcheva (University of Sheffield, UK) focused on "Automatic Detection of Rumours in Social Media."

It would have been impossible to organize this conference without the help and commitment of many people. We would like to thank all speakers, participants, workshop organizers (details at <http://gsc12017.dfki.de>), as well as the members of the Program Committee. We would also like to thank the people who made it possible to organize GSCL 2017 at the Humboldt University of Berlin: Stefan Müller, Anke Lüdeling, Manfred Krifka, Steffen Hofmann, and Dagmar Oehler. Of crucial importance in the planning and organization of the conference was the Executive Committee of GSCL e.V., especially Heike Zinsmeister, Torsten Zesch, Andreas Witt, and Roman Schneider. Furthermore, we would like to thank Nieves Sande, who organized all the practical aspects of the event and coordinated the conference office, Julian Moreno Schneider, who not only created and maintained the conference website but who also played a central role in assembling this proceedings volume, as well as Stefanie Hegele, Charlene Röhl, and Peter Bourgonje, who helped on site at the conference desk and with the logistics. Thanks are also due to the three gold and seven silver sponsors without whom the conference would not have been possible.

This Open Access conference proceedings volume was made possible with the help of the BMBF project *Digitale Kuratierungstechnologien* (DKT). We are grateful to Alexandra Buchsteiner (Pt-J) for supporting the initial suggestion to publish this book.

# Organization

GSCL 2017 was organized by Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI) in cooperation with Humboldt-Universität zu Berlin.

## Conference Chairs

Georg Rehm	DFKI, Germany
Thierry Declerck	DFKI, Germany

## Organizing Committee

Georg Rehm	DFKI, Germany
Thierry Declerck	DFKI, Germany
Nieves Sande	DFKI, Germany
Julian Moreno-Schneider	DFKI, Germany
Stefan Müller	Humboldt-Universität zu Berlin, Germany

## Program Committee

Michael Beisswenger	Universität Duisburg-Essen, Germany
Delphine Bernhard	Université de Strasbourg, France
Chris Biemann	Universität Hamburg, Germany
Aljoscha Burchardt	DFKI, Germany
Christian Chiarcos	Goethe-Universität Frankfurt, Germany
Philipp Cimiano	Universität Bielefeld, Germany
Bertold Crysmann	CNRS, France
Michael Cysouw	University of Marburg, Germany
Thierry Declerck (Chair)	DFKI, Germany
Stefanie Dipper	Ruhr-Universität Bochum, Germany
Christian Federmann	Microsoft, USA
Bernhard Fisseni	Uni Duisburg-Essen, Germany
Alexander Fraser	Ludwig-Maximilians-Universität München, Germany
Jost Gippert	Universität Frankfurt, Germany
Dagmar Gromann	Artificial Intelligence Research Institute, Spain
Iryna Gurevych	Technische Universität Darmstadt, Germany
Gerhard Heyer	Universität Leipzig, Germany
Roman Klinger	Universität Stuttgart, Germany
Brigitte Krenn	Austrian Research Institute for Artificial Intelligence, Austria
Sandra Kübler	Indiana University, USA
Jonas Kuhn	Universität Stuttgart, Germany

Lothar Lemnitzer	BBAW, Germany
Piroska Lendvai	University of Göttingen, Germany
Nils Lenke	Nuance Communications, Aachen, Germany
Henning Lobin	Universität Gießen, Germany
Cerstin Mahlow	Universität Stuttgart, Germany
Alexander Mehler	Goethe-Universität Frankfurt, Germany
Friedrich Neubarth	Austrian Research Institute for Artificial Intelligence, Austria
Sebastian Padó	Universität Stuttgart, Germany
Karola Pitsch	Universität Duisburg-Essen, Germany
Uwe Quasthoff	Universität Leipzig, Germany
Georg Rehm (Chair)	DFKI, Germany
Frank Richter	Goethe-Universität Frankfurt, Germany
Manfred Sailer	Goethe-Universität Frankfurt, Germany
David Schlangen	Universität Bielefeld, Germany
Helmut Schmid	CIS München, Germany
Thomas Schmidt	IDS Mannheim, Germany
Ulrich Schmitz	Universität Duisburg-Essen, Germany
Roman Schneider	IDS Mannheim, Germany
Bernhard Schröder	Universität Duisburg-Essen, Germany
Sabine Schulte im Walde	Universität Stuttgart, Germany
Manfred Stede	Universität Potsdam, Germany
Benno Stein	Universität Weimar, Germany
Angelika Storrer	Universität Mannheim, Germany
Elke Teich	Universität des Saarlandes, Germany
Mihaela Vela	Universität des Saarlandes, Germany
Michael Wiegand	Universität des Saarlandes, Germany
Andreas Witt	Universität zu Köln, Germany
Marcos Zampieri	University of Cologne, Germany
Torsten Zesch	Universität Duisburg-Essen, Germany
Heike Zinsmeister	Universität Hamburg, Germany

## GSCL 2017 – Sponsors

### Gold Sponsors



facebook



### Silver Sponsors



**3pc**  
Neue Kommunikation

European Media  
Laboratory GmbH



Lin|gu|is|tik  
PORTAL FÜR  
SPRACHWISSENSCHAFT  
[www.linguistik.de](http://www.linguistik.de)

parlamind

text & form



### Proceedings Sponsors



GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

WK · POTENZIAL  
UNTERNEHMEN  
REGION  
Die BMBF-Innovationsinitiative  
Neue Länder

The project “Digitale Kuratierungstechnologien” (DKT) is supported by the German Federal Ministry of Education and Research (BMBF), “Unternehmen Region”, instrument Wachstumskern-Potenzial (no. 03WKP45). More information: <http://www.digitale-kuratierung.de>.



# Contents

## Processing German: Basic Technologies

Reconstruction of Separable Particle Verbs in a Corpus of Spoken German . . . . .	3
<i>Dolores Batinić and Thomas Schmidt</i>	
Detecting Vocal Irony . . . . .	11
<i>Felix Burkhardt, Benjamin Weiss, Florian Eyben, Jun Deng, and Björn Schuller</i>	
The Devil is in the Details: Parsing Unknown German Words . . . . .	23
<i>Daniel Dakota</i>	
Exploring Ensemble Dependency Parsing to Reduce Manual Annotation Workload . . . . .	40
<i>Jessica Sohl and Heike Zinsmeister</i>	
Different German and English Coreference Resolution Models for Multi-domain Content Curation Scenarios . . . . .	48
<i>Ankit Srivastava, Sabine Weber, Peter Bourgonje, and Georg Rehm</i>	
Word and Sentence Segmentation in German: Overcoming Idiosyncrasies in the Use of Punctuation in Private Communication. . . . .	62
<i>Kyoko Sugisaki</i>	
Fine-Grained POS Tagging of German Social Media and Web Texts. . . . .	72
<i>Stefan Thater</i>	
Developing a Stemmer for German Based on a Comparative Analysis of Publicly Available Stemmers . . . . .	81
<i>Leonie Weissweiler and Alexander Fraser</i>	
Negation Modeling for German Polarity Classification. . . . .	95
<i>Michael Wiegand, Maximilian Wolf, and Josef Ruppenhofer</i>	

## Processing German: Named Entities

NECKAr: A Named Entity Classifier for Wikidata . . . . .	115
<i>Johanna Geiß, Andreas Spitz, and Michael Gertz</i>	

Investigating the Morphological Complexity of German Named Entities: The Case of the GermEval NER Challenge. . . . .	130
<i>Bettina Klimek, Markus Ackermann, Amit Kirschenbaum, and Sebastian Hellmann</i>	
Detecting Named Entities and Relations in German Clinical Reports . . . . .	146
<i>Roland Roller, Nils Rethmeier, Philippe Thomas, Marc Hübner, Hans Uszkoreit, Oliver Staeck, Klemens Budde, Fabian Halleck, and Danilo Schmidt</i>	
In-Memory Distributed Training of Linear-Chain Conditional Random Fields with an Application to Fine-Grained Named Entity Recognition . . . . .	155
<i>Robert Schwarzenberg, Leonhard Hennig, and Holmer Hemsén</i>	
<b>Online-Media and Online-Content</b>	
What Does This Imply? Examining the Impact of Implicitness on the Perception of Hate Speech . . . . .	171
<i>Darina Benikova, Michael Wojatzki, and Torsten Zesch</i>	
Automatic Classification of Abusive Language and Personal Attacks in Various Forms of Online Communication . . . . .	180
<i>Peter Bourgonje, Julian Moreno-Schneider, Ankit Srivastava, and Georg Rehm</i>	
Token Level Code-Switching Detection Using Wikipedia as a Lexical Resource . . . . .	192
<i>Daniel Claeser, Dennis Felske, and Samantha Kent</i>	
How Social Media Text Analysis Can Inform Disaster Management . . . . .	199
<i>Sabine Gründer-Fahrer, Antje Schlaf, and Sebastian Wustmann</i>	
A Comparative Study of Uncertainty Based Active Learning Strategies for General Purpose Twitter Sentiment Analysis with Deep Neural Networks . . . . .	208
<i>Nils Haldenwang, Katrin Ihler, Julian Kniephoff, and Oliver Vornberger</i>	
An Infrastructure for Empowering Internet Users to Handle Fake News and Other Online Media Phenomena. . . . .	216
<i>Georg Rehm</i>	

Different Types of Automated and Semi-automated Semantic Storytelling:  
 Curation Technologies for Different Sectors . . . . . 232  
*Georg Rehm, Julián Moreno-Schneider, Peter Bourgonje,  
 Ankit Srivastava, Rolf Fricke, Jan Thomsen, Jing He,  
 Joachim Quantz, Armin Berger, Luca König, Sören Räuchle,  
 Jens Gerth, and David Wabnitz*

Twitter Geolocation Prediction Using Neural Networks . . . . . 248  
*Philippe Thomas and Leonhard Hennig*

**Miscellaneous**

Diachronic Variation of Temporal Expressions in Scientific Writing  
 Through the Lens of Relative Entropy . . . . . 259  
*Stefania Degaetano-Ortlieb and Jannik Strötgen*

A Case Study on the Relevance of the Competence Assumption  
 for Implicature Calculation in Dialogue Systems . . . . . 276  
*Judith Victoria Fischer*

Supporting Sustainable Process Documentation . . . . . 284  
*Markus Gärtner, Uli Hahn, and Sibylle Hermann*

Optimizing Visual Representations in Semantic Multi-modal Models  
 with Dimensionality Reduction, Denoising and Contextual Information . . . . . 292  
*Maximilian Köper, Kim-Anh Nguyen, and Sabine Schulte im Walde*

Using Argumentative Structure to Grade Persuasive Essays . . . . . 301  
*Andreas Stiegelmayr and Margot Mieskes*

**Author Index** . . . . . 309