LaTeCH 2016

Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2016)

August 11, 2016
Berlin, Germany

Production and Manufacturing by
Taberg Media Group AB
Box 94, 562 02 Taberg
Sweden

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-945626-09-8

Introduction

The LaTeCH workshop series, which started in 2007, was initially motivated by the growing interest in language technology research and applications to the cultural heritage domain. The scope quickly broadened to also include the humanities and the social sciences. LaTeCH is currently the annual venue of the ACL Special Interest Group on Language Technologies for the Socio-Economic Sciences and Humanities (SIGHUM).

LaTeCH 2016 is the tenth instalment of the LaTeCH workshop series. Fittingly, LaTeCH received the best birthday present a workshop can hope for: A record number of submissions. 48 papers have been submitted in total, 23 of them being long papers (8 pages). Overall, 21 papers have been accepted for presentation, giving this workshop an acceptance rate of about 44% (long: 47%, short: 40%, previous years: about 60%).

While we did not set a specific topic for this workshop, there is one thematic group that can be easily identified among the accepted papers: Historic languages and their processing. Apart from that, several papers deal with political/social issues and diachronic development in general.

We would like to thank all authors for the hard work that went into their submissions. We are also grateful to the members of the programme committee for their thorough reviews. Reviewing this many papers in time would not have been possible without the additional reviewers who were able to join the programme committee on a short notice and those who volunteered to review a few papers more than anticipated. We also thank the ACL 2016 organisers, in particular the Workshop Co-chairs Jun Zhao and Sabine Schulte im Walde.

Beatrice Alex and Nils Reiter

Organizers:

Nils Reiter (co-chair), Stuttgart University, Germany
Beatrice Alex (co-chair), University of Edinburgh, UK
Kalliopi A. Zervanou, Utrecht University, The Netherlands

Program Committee:

Nikolaos Aletras, University College London, UK
JinYeong Bak, KAIST Daejeon, South Korea
Chris Biemann, TU Darmstadt, Germany
André Blessing, Stuttgart University, Germany
Toine Bogers, Aalborg University Copenhagen, Denmark
Gosse Bouma, Groningen University, The Netherlands
Paul Buitelaar, Insight Centre for Data Analytics, NUI Galway, Ireland
Mariona Coll Ardanuy, Trier University, Germany
Gerard de Melo, Tsinghua University, Beijing, China
Thierry Declerck, DFKI, Germany
Stefanie Dipper, Ruhr-Universität Bochum, Germany
Jacob Eisenstein, Georgia Institute of Technology, USA
Mark Finlayson, Florida International University, USA
Antske Fokkens, VU University Amsterdam, The Netherlands
Serge Heiden, ENS de Lyon, France
Aurélie Herbelot, University of Trento, Italy
Iris Hendrickx, Radboud University Nijmegen, The Netherlands
Gerhard Heyer, Leipzig University, Germany
Yufang Hou, IBM Research, Ireland
Amy Isard, University of Edinburgh, UK
Adam Jatowt, Kyoto University, Japan
Richard Johansson, University of Gothenburg, Sweden
Jaap Kamps, Universiteit van Amsterdam, The Netherlands
Vangelis Karkaletsis, NCSR Demokritos, Athens, Greece
Mike Kestemont, Antwerp University/Research Foundation Flanders, Belgium
Dimitrios Kokkinakis, University of Gothenburg, Sweden
Stasinos Konstantopoulos, NCSR Demokritos, Athens, Greece
Jonas Kuhn, Stuttgart University, Germany
John Lee, City University of Hong Kong
Chaya Liebeskind, Bar Ilan University, Israel
Clare Llewellyn, University of Edinburgh, UK
Shervin Malmasi, Harvard Medical School, USA
Ruli Manurung, University of Indonesia, Depok, Indonesia
Barbara McGillivray, Nature Publishing Group, UK
Yusuke Miyao, National Institute of Informatics, Japan
Joakim Nivre, Uppsala University, Sweden
Pierre Nugues, Lund University, Sweden
Mick O'Donnell, Universidad Autonoma de Madrid, Spain
Petya Osenova, Bulgarian Academy of Sciences, Bulgaria

Program Committee (continued):

Michael Piotrowski, Leibniz Institute of European History, Mainz, Germany
Georg Rehm, DFKI, Germany
Martin Reynaert, Tilburg University, The Netherlands
Marijn Schraagen, Utrecht University, The Netherlands
Sarah Schulz, Stuttgart University, Germany
Eszter Simon, Hungarian Academy of Sciences, Budapest, Hungary
Caroline Sporleder, Göttingen University, Germany
Herman Stehouwer, MPI for Psycholinguistics, The Netherlands
Jannik Strötgen, MPI for Computer science, Saarbrücken, Germany
Mariët Theune, University of Twente, The Netherlands
Sara Tonelli, Fondazione Bruno Kessler, Trento, Italy
Thorsten Trippel, Tübingen University, Germany
Adam Wyner, University of Aberdeen, UK
Menno van Zaanen, Tilburg University, The Netherlands
Svitlana Zinger, TU Eindhoven, The Netherlands
Heike Zinsmeister, Hamburg University, Germany

Table of Contents

Conference Program

Thursday, August 11, 2016

9:00–10:30 **Session 1**

09:00 *Brave New World: Uncovering Topical Dynamics in the ACL Anthology Reference Corpus Using Term Life Cycle Information*
Anne-Kathrin Schumann

09:30 *Analysis of Policy Agendas: Lessons Learned from Automatic Topic Classification of Croatian Political Texts*
Mladen Karan, Jan Šnajder, Daniela Sirinic and Goran Glavaš

10:00 *Searching Four-Millenia-Old Digitized Documents: A Text Retrieval System for Egyptologists*
Estíbaliz Iglesias-Franjo and Jesús Vilares

11:00–12:30 **Session 2**

11:00 *Old Swedish Part-of-Speech Tagging between Variation and External Knowledge*
Yvonne Adesam and Gerlof Bouma

11:30 *Code-Switching Ubique Est - Language Identification and Part-of-Speech Tagging for Historical Mixed Text*
Sarah Schulz and Mareike Keller

12:00 *Dealing with word-internal modification and spelling variation in data-driven lemmatization*
Fabian Barteld, Ingrid Schröder and Heike Zinsmeister

13:30–14:00 SIGHUM Business Meeting

14:00–15:00 Session 3

14:00 *You Shall Know People by the Company They Keep: Person Name Disambiguation for Social Network Construction*
Mariona Coll Ardanuy, Maarten van den Bos and Caroline Sporleder

14:30 *Deriving Players & Themes in the Regesta Imperii using SVMs and Neural Networks*
Juri Opitz and Anette Frank

15:00–16:00 Poster Session

Semi-automated annotation of page-based documents within the Genre and Multimodality framework
Tuomo Hiippala

Nomen Omen. Enhancing the Latin Morphological Analyser Lemlat with an Onomasticon
Marco Budassi and Marco Passarotti

How Do Cultural Differences Impact the Quality of Sarcasm Annotation?: A Case Study of Indian Annotators and American Text
Aditya Joshi, Pushpak Bhattacharyya, Mark Carman, Jaya Saraswati and Rajita Shukla

Combining Phonology and Morphology for the Normalization of Historical Texts
Izaskun Etxeberria, Iñaki Alegria, Larraitz Uria and Mans Hulden

Towards Building a Political Protest Database to Explain Changes in the Welfare State
Çağıl Sönmez, Arzucan Özgür and Erdem Yörük

An Assessment of Experimental Protocols for Tracing Changes in Word Semantics Relative to Accuracy and Reliability
Johannes Hellrich and Udo Hahn

Universal Morphology for Old Hungarian
Eszter Simon and Veronika Vincze

Automatic Identification of Suicide Notes from Linguistic and Sentiment Features
Annika Marie Schoene and Nina Dethlefs

Thursday, August 11, 2016 (continued)

Towards a text analysis system for political debates
Dieu-Thu Le, Ngoc Thang Vu and Andre Blessing

Whodunit... and to Whom? Subjects, Objects, and Actions in Research Articles on American Labor Unions
Vilja Hulden

16:00–17:30 Session 4

16:00 *An NLP Pipeline for Coptic*
Amir Zeldes and Caroline T. Schroeder

16:30 *Automatic discovery of Latin syntactic changes*
Micha Elsner and Emily Lane

17:00 *Information-based Modeling of Diachronic Linguistic Change: from Typicality to Productivity*
Stefania Degaetano-Ortlieb and Elke Teich