LaTeCH 2013

Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2013)

August 8, 2013
Sofia, Bulgaria

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53704 USA

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 978-1-937284-62-6

Preface

We are delighted to present you with this volume containing the papers accepted for presentation at the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, endorsed by the ACL SIGHUM interest group.

Language technology has by now pervaded core processing procedures targeting the cleaning, searching, linking, enriching, and mining of digitized data from these fields. After six previous LaTeCH workshops, we are happy to carry on with aggregating and disseminating the most interesting studies – selected by a thorough peer-review process – concerning current hot topics in the area of natural language processing. Acceptance rate for LaTeCH-2013 was 62%. We would especially like to thank the members of the programme committee for willing to share their expertise by providing detailed reviews and insightful input to all the submitting authors.

On top of the regular paper presentations, the organisers are proud to integrate the SIGHUM annual business meeting into the programme.

We wish you a well-spent workshop day!

Piroska Lendvai and Kalliopi Zervanou
Chairs of LaTeCH-2013

Organizers

Workshop Chairs:

Piroska Lendvai, Research Institute for Linguistics (Hungary)
Kalliopi Zervanou, Radboud University Nijmegen (The Netherlands)

Organizing Committe:

Piroska Lendvai, Research Institute for Linguistics (Hungary)
Caroline Sporleder, Saarland University / Trier University, Germany
Antal van den Bosch, Radboud University Nijmegen (The Netherlands)
Kalliopi Zervanou, Radboud University Nijmegen (The Netherlands)

Program Committee Members:

Ion Androutsopoulos, Athens University of Economics and Business, Greece
David Bamman, Carnegie Mellon University, USA
Rens Bod, Universiteit van Amsterdam, The Netherlands
Toine Bogers, Royal School of Library and Information Science, Copenhagen, Denmark
Paul Buitelaar, DERI Galway, Ireland
Mick O'Donnell, Universidad Autonoma de Madrid, Spain
Julio Gonzalo, Universidad Nacional de Educacion a Distancia, Spain
Ben Hachey, Macquarie University, Australia
Iris Hendrickx, Radboud University Nijmegen, The Netherlands
Elias Iosif, Technical University of Crete, Greece
Jaap Kamps, Universiteit van Amsterdam, The Netherlands
Vangelis Karkaletsis, NCSR Demokritos, Greece
Mike Kestermont, University of Antwerp / Research Foundation Flanders, Belgium
Dimitrios Kokkinakis, University of Gothenburg, Sweden
Stasinos Konstantopoulos, NCSR Demokritos, Greece
Barbara McGillivray, Oxford University Press, UK
Joakim Nivre, Uppsala University, Sweden
Csaba Oravecz, Research Institute for Linguistics, Hungary
Petya Osenova, Bulgarian Academy of Sciences, Bulgaria
Katerina Pastra, Cognitive Systems Research Institute, Greece
Michael Piotrowski, University of Zurich, Switzerland
Georg Rehm, DFKI Berlin, Germany
Martin Reynaert, Tilburg University, The Netherlands
Eszter Simon, Research Institute for Linguistics, Hungary
Herman Stehouwer, Max Planck Institute for Psycholinguistics, The Netherlands
Mark Stevenson, University of Sheffield, UK
Mariët Theune, University of Twente, The Netherlands
Suzan Verberne, Radboud University Nijmegen, The Netherlands
Cristina Vertan, University of Hamburg, Germany
Menno van Zaanen, Tilburg University, The Netherlands
Svitlana Zinger, TU Eindhoven, The Netherlands

Table of Contents

Conference Program

Thursday August 8, 2013

9:00-9:15	Opening
9:15–9.35	*Generating Paths through Cultural Heritage Collections* Samuel Fernando, Paula Goodale, Paul Clough, Mark Stevenson, Mark Hall and Eneko Agirre
9:35–9.55	*Using character overlap to improve language transformation* Sander Wubben, Emiel Krahmer and Antal van den Bosch
9:55–10:15	*Comparison between historical population archives and decentralized databases* Marijn Schraagen and Dionysius Huijsmans
10:15–10:30	*Semi-automatic Construction of Cross-period Thesaurus* Chaya Liebeskind, Ido Dagan and Jonathan Schler
10:30-11:00	Coffee Break
11:00–11:15	*Language Technology for Agile Social Media Science* Simon Wibberley, David Weir and Jeremy Reffin
11:15–11:30	*Morphological annotation of Old and Middle Hungarian corpora* Attila Novák, György Orosz and Nóra Wenszky
11:30–11:45	*Argument extraction for supporting public policy formulation* Eirini Florou, Stasinos Konstantopoulos, Antonis Koukourikos and Pythagoras Karampiperis
11:45-12:30	SIGHUM annual business meeting
12:30-14:00	Lunch Break
14:00–14:20	*Towards a Tool for Interactive Concept Building for Large Scale Analysis in the Humanities* Andre Blessing, Jonathan Sonntag, Fritz Kliche, Ulrich Heid, Jonas Kuhn and Manfred Stede
14:20–14:40	*Learning to Extract Folktale Keywords* Dolf Trieschnigg, Dong Nguyen and Mariët Theune

Thursday August 8, 2013 (continued)

14:40–15:00 *Towards Creating Precision Grammars from Interlinear Glossed Text: Inferring Large-Scale Typological Properties*
Emily M. Bender, Michael Wayne Goodman, Joshua Crowgey and Fei Xia

15:00–15:15 *Using Comparable Collections of Historical Texts for Building a Diachronic Dictionary for Spelling Normalization*
Marilisa Amoia and José Manuel Martínez

15:15–15:30 *Integration of the Thesaurus for the Social Sciences (TheSoz) in an Information Extraction System*
Thierry Declerck

15:30-16:00 Coffee Break

16:00–16:15 *The (Un)faithful Machine Translator*
Ruth Jones and Ann Irvine

16:15–16:30 *Temporal classification for historical Romanian texts*
Alina Maria Ciobanu, Anca Dinu, Liviu Dinu, Vlad Niculae and Octavia-Maria Şulea

16:30–16:50 *Multilingual access to cultural heritage content on the Semantic Web*
Dana Dannells, Aarne Ranta, Ramona Enache, Mariana Damova and Maria Mateva

16:50-17:00 Closing