LaTeCH 2015

Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2015)

July 30, 2015
Beijing, China

©2015 The Association for Computational Linguistics
and The Asian Federation of Natural Language Processing

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 978-1-941643-63-1

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).

The current, ninth edition of the LaTeCH workshop was organised for the first time in Asia and we were delighted to present humanities research from Indonesia, Singapore, China and Korea, covering topics such as Chinese poetry, Korean History and folktales. We have also witnessed further expansion in the resource and tool development approaches, with works now covering topics ranging from Latin and medieval manuscripts to Minoan, Linear A texts. The submissions were substantial not only in terms of quantity, but also in terms of quality and variety. Acceptance rate for LaTeCH-2015 was 61%.

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, and to the ACL-IJCNLP 2015 organisers, especially the Workshop Co-chairs, Hang Li and Sebastian Riedel for their help with administrative matters. Finally, we wish to thank Nils Reiter for volunteering to oversee the workshop local organisation, moderate the annual SIGHUM meeting and chair the workshop sessions.

Kalliopi Zervanou, Marieke van Erp and Beatrice Alex

Organizers:

Kalliopi A. Zervanou (Co-Chair), Utrecht University (The Netherlands)
Marieke van Erp (Co-Chair), VU University Amsterdam (The Netherlands)
Beatrice Alex (Co-Chair), University of Edinburgh (United Kingdom)

Local Organizer:

Nils Reiter, Stuttgart University (Germany)

Program Committee:

Toine Bogers, Aalborg University, Copenhagen (Denmark)
Antal van den Bosch, Radboud University Nijmegen (The Netherlands)
Gosse Bouma, Rijksuniversiteit Groningen (The Netherlands)
Paul Buitelaar, DERI Galway (Ireland)
Mariona Coll Ardanuy, Trier University (Germany)
Thierry Declerck, DFKI (Germany)
Stefanie Dipper, Ruhr-Universität Bochum (Germany)
Milena Dobreva, University of Malta (Malta)
Mick O'Donnell, Universidad Autonoma de Madrid (Spain)
Antske Fokkens, VU University Amsterdam (The Netherlands)
Ben Hachey, University of Sydney (Australia)
Iris Hendrickx, Radboud University Nijmegen (The Netherlands)
Adam Jatowt, Kyoto University (Japan)
Jaap Kamps, University of Amsterdam (The Netherlands)
Vangelis Karkaletsis, NCSR Demokritos (Greece)
Mike Kestemont, University of Antwerp & Research Foundation Flanders (Belgium)
Dimitrios Kokkinakis, University of Gothenburg (Sweden)
Stasinos Konstantopoulos, NCSR Demokritos (Greece)
Barbara McGillivray, Oxford University Press
Gerard de Melo, Tsinghua University (China)
Saif Mohammad, National Research Council (Canada)
Joakim Nivre, Uppsala University (Sweden)
Nelleke Oostdijk, Radboud University Nijmegen (The Netherlands)
Petya Osenova, Bulgarian Academy of Sciences (Bulgaria)
Michael Piotrowski, Leibniz Institute of European History in Mainz (Germany)
Georg Rehm, DFKI (Germany)
Martin Reynaert, Tilburg University (The Netherlands)
Eric Sanders, Radboud University Nijmegen (The Netherlands)
Marijn Schraagen, Utrecht University (The Netherlands)
Eszter Simon, Research Institute for Linguistics (HASRIL) (Hungary)
Caroline Sporleder, Trier University (Germany)
Herman Stehouwer, Max Planck for Plasmaphysics (Germany)
Mariët Theune, University of Twente (The Netherlands)
Takenobu Tokunaga, Tokyo Institute of Technology (Japan)
Cristina Vertan, University of Hamburg (Germany)
Frans Wiering, Utrecht University (The Netherlands)
Menno van Zaanen, Tilburg University (The Netherlands)
Svitlana Zinger, TU Eindhoven (The Netherlands)

Table of Contents

Catching the Red Priest: Using Historical Editions of Encyclopaedia Britannica to Track the Evolution of Reputations
Yen-Fu Luo, Anna Rumshisky and Mikhail Gronas 1

Five Centuries of Monarchy in Korea: Mining the Text of the Annals of the Joseon Dynasty
JinYeong Bak and Alice Oh 10

Analyzing Sentiment in Classical Chinese Poetry
Yufang Hou and Anette Frank 15

Measuring the Structural and Conceptual Similarity of Folktales using Plot Graphs
Victoria Anugrah Lestari and Ruli Manurung 25

Towards Annotating Narrative Segments
Nils Reiter 34

Ranking Relevant Verb Phrases Extracted from Historical Text
Eva Pettersson, Beáta Megyesi and Joakim Nivre 39

Ranking election issues through the lens of social media
Stephen Wan and Cécile Paris 48

Word Embeddings Pointing the Way for Late Antiquity
Johannes Bjerva and Raf Praet 53

Enriching Interlinear Text using Automatically Constructed Annotators
Ryan Georgi, Fei Xia and William Lewis 58

Automatic interlinear glossing as two-level sequence classification
Tanja Samardzic, Robert Schikowski and Sabine Stoll 68

Enriching Digitized Medieval Manuscripts: Linking Image, Text and Lexical Knowledge
Aitor Arronte Alvarez 73

A preliminary study on similarity-preserving digital book identifiers
Klemo Vladimir, Marin Silic, Nenad Romic, Goran Delac and Sinisa Srbljic 78

When Translation Requires Interpretation: Collaborative Computer–Assisted Translation of Ancient Texts
Andrea Bellandi, Davide Albanesi, Giulia Benotto, Emiliano Giovannetti and Gianfranco Di Segni 84

Integrating Query Performance Prediction in Term Scoring for Diachronic Thesaurus
Chaya Liebeskind and Ido Dagan 89

Minoan linguistic resources: The Linear A Digital Corpus
Tommaso Petrolito, Ruggero Petrolito, Francesco Perono Cacciafoco and Gregoire Winterstein 95

Lexicon-assisted tagging and lemmatization in Latin: A comparison of six taggers and two lemmatization models
Tim vor der Brück, Steffen Eger and Alexander Mehler 105

Workshop Program

Thursday 30th July 2015

9:00–10:30 Session I

9:00–9:10 *Welcome*
Nils Reiter

9:10–9:40 *Catching the Red Priest: Using Historical Editions of Encyclopaedia Britannica to Track the Evolution of Reputations*
Yen-Fu Luo, Anna Rumshisky and Mikhail Gronas

9:40–10:00 *Five Centuries of Monarchy in Korea: Mining the Text of the Annals of the Joseon Dynasty*
JinYeong Bak and Alice Oh

10:00–10:30 *Analyzing Sentiment in Classical Chinese Poetry*
Yufang Hou and Anette Frank

10:30–11:00 Coffee Break

11:00–12:40 Session II

11:00–11:30 *Measuring the Structural and Conceptual Similarity of Folktales using Plot Graphs*
Victoria Anugrah Lestari and Ruli Manurung

11:30–11:50 *Towards Annotating Narrative Segments*
Nils Reiter

11:50–12:20 *Ranking Relevant Verb Phrases Extracted from Historical Text*
Eva Pettersson, Beáta Megyesi and Joakim Nivre

12:20–12:40 *Ranking election issues through the lens of social media*
Stephen Wan and Cécile Paris

Thursday 30th July 2015 (continued)

12:40–14:00 Lunch

14:00–14:40 Session III: SIGHUM Annual Meeting

14:40–15:10 Session IV: Poster Boosters

14:40–14:45 *Word Embeddings Pointing the Way for Late Antiquity*
Johannes Bjerva and Raf Praet

14:45–14:50 *Enriching Interlinear Text using Automatically Constructed Annotators*
Ryan Georgi, Fei Xia and William Lewis

14:50–14:55 *Automatic interlinear glossing as two-level sequence classification*
Tanja Samardzic, Robert Schikowski and Sabine Stoll

14:55–15:00 *Enriching Digitized Medieval Manuscripts: Linking Image, Text and Lexical Knowledge*
Aitor Arronte Alvarez

15:00–15:05 *A preliminary study on similarity-preserving digital book identifiers*
Klemo Vladimir, Marin Silic, Nenad Romic, Goran Delac and Sinisa Srbljic

15:05–15:10 *When Translation Requires Interpretation: Collaborative Computer–Assisted Translation of Ancient Texts*
Andrea Bellandi, Davide Albanesi, Giulia Benotto, Emiliano Giovannetti and Gianfranco Di Segni

Thursday 30th July 2015 (continued)

15:10–16:00 Poster Session and Coffee

16:00–17:30 Session V

16:00–16:20 *Integrating Query Performance Prediction in Term Scoring for Diachronic Thesaurus*
Chaya Liebeskind and Ido Dagan

16:20–16:50 *Minoan linguistic resources: The Linear A Digital Corpus*
Tommaso Petrolito, Ruggero Petrolito, Francesco Perono Cacciafoco and Gregoire Winterstein

16:50–17:20 *Lexicon-assisted tagging and lemmatization in Latin: A comparison of six taggers and two lemmatization models*
Tim vor der Brück, Steffen Eger and Alexander Mehler

17:20–17:30 *Closing*
Nils Reiter