COLING 2018

The 27th International Conference on Computational Linguistics

Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL-2018)

August 25, 2018
Santa Fe, New Mexico, USA

Copyright of each paper stays with the respective authors (or their employers).

ISBN 978-1-948087-61-2

Introduction

Welcome to the 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature. After a great joint event last year, we are happy to present another lineup of high-quality papers at this year's workshop. Everything you want to know about it has been placed on-line (https://sighum.wordpress.com/events/latech-clfl-2018/). We accepted 64.3% of submitted papers. You can appreciate the thematic variety in what follows.

We thank all submitters, congratulate those whose work appears here, and express our debt of gratitude to the members of the program committee. We wish to single out one participant: Ted Underwood, who has graciously accepted our invitation to be the workshop's guest speaker.

Enjoy!

Beatrice, Stefania, Anna, Anna, Nils, Stan

Invited Talk

Ted Underwood is a Professor of Information Science and English at the University of Illinois, Urbana-Champaign. He has published widely on quantitative approaches to literary history, with articles in New Literary History, Cultural Analytics, Representations, and Modern Language Quarterly. His research currently focuses on questions about literary character, and ways of using machine learning to compare situated historical perspectives. In 2019, the University of Chicago Press will publish his next book, Distant Horizons: Digital Evidence and Literary Change.

Measurement and Human Perspective

Researchers who use quantitative methods to study culture may find themselves torn between two approaches to knowledge. Quantitative disciplines often value generalizable methods that make different frames of reference comparable. But humanists have long insisted that knowledge about culture is irreducibly perspectival, relative to a particular vantage point. Of course, the tension between these two approaches is not exactly a contradiction. Contemporary quantitative methods are sometimes flexible enough to incorporate varied human perspectives as elements of a model. This talk will describe a few recent projects that have attempted perspectival modeling, and reflect on opportunities for further research.

Invited Speaker

Ted Underwood, University of Illinois, United States

Organizers

Beatrice Alex, School of Informatics, University of Edinburgh
Stefania Degaetano-Ortlieb, Dept. of Language Science & Technology, Universität des Saarlandes
Anna Feldman, Dept. of Linguistics & Dept. of Computer Science, Montclair State University
Anna Kazantseva, National Research Council of Canada
Nils Reiter, Institute for Natural Language Processing (IMS), Stuttgart University
Stan Szpakowicz, School of Electrical Engineering and Computer Science, University of Ottawa

Program Committee

Cecilia Ovesdotter Alm, Rochester Institute of Technology, USA
JinYeong Bak, KAIST, Republic of Korea
Gosse Bouma, University of Groningen, The Netherlands
Paul Buitelaar, National University of Ireland, Galway, Ireland
Thierry Declerck, Deutsche Forschungszentrum für Künstliche Intelligenz GmbH, Germany
Stefanie Dipper, Ruhr-University, Bochum, Germany
Jacob Eisenstein, Georgia Institute of Technology, United States
Micha Elsner, Ohio State University, United States
Mark Finlayson, Florida International University, United States
Serge Heiden, École normale supérieure de Lyon, France
Iris Hendrickx, Radboud University, Nijmegen, The Netherlands
Gerhard Heyer, University of Leipzig, Germany
Graeme Hirst, University of Toronto, Canada
Amy Isard, University of Edinburgh, United Kingdom
Adam Jatowt, Kyoto University, Japan
Mike Kestemont, University of Antwerp, Belgium
Dimitrios Kokkinakis, University of Gothenburg, Sweden
Stasinos Konstantopoulos, National Centre of Scientific Research “Demokritos”, Greece
John Lee, City University of Hong Kong, Hong Kong
Chaya Liebeskind, Jerusalem College of Technology, Israel
Barbara McGillivray, Alan Turing Institute/University of Cambridge, United Kingdom
Gerard de Melo, Tsinghua University, China
Rada Mihalcea, University of Michigan, United States
Borja Navarro Colorado, University of Alicante, Spain
John Nerbonne, University of Freiburg, Germany
Pierre Nugues, Lund University, Sweden
Mick O’Donnel, Universidad Autónoma de Madrid, Spain
Petya Osenova, Sofia University and IICT-BAS, Bulgaria
Michael Piotrowski, University of Lausanne, Switzerland
Georg Rehm, DFKI, Germany
Martin Reynaert, Tilburg University, Radboud University Nijmegen, The Netherlands
Marijn Schraagen, Utrecht University, The Netherlands
Sarah Schulz, University of Stuttgart, Germany

Eszter Simon, Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary
Caroline Sporleder, Göttingen University, Germany
Elke Teich, Saarland University, Germany
Sara Tonelli, FBK, Trento, Italy
Thorsten Trippel, University of Tübingen, Germany
Menno van Zaanen, Tilburg University, The Netherlands
Kalliopi Zervanou, Utrecht University, The Netherlands
Heike Zinsmeister, University of Hamburg, Germany

Table of Contents

Learning Diachronic Analogies to Analyze Concept Change
Matthias Orlikowski, Matthias Hartung and Philipp Cimiano 1

A Linked Coptic Dictionary Online
Frank Feder, Maxim Kupreyev, Emma Manning, Caroline T. Schroeder and Amir Zeldes 12

Using relative entropy for detection and analysis of periods of diachronic linguistic change
Stefania Degaetano-Ortlieb and Elke Teich 22

Cliche Expressions in Literary and Genre Novels
Andreas van Cranenburgh 34

Analysis of Rhythmic Phrasing: Feature Engineering vs. Representation Learning for Classifying Readout Poetry
Timo Baumann, Hussein Hussein and Burkhard Meyer-Sickendiek 44

Cross-Discourse and Multilingual Exploration of Textual Corpora with the DualNeighbors Algorithm
Taylor Arnold and Lauren Tilton 50

The Historical Significance of Textual Distances
Ted Underwood 60

One Size Fits All? A simple LSTM for non-literal token and construction-level classification
Erik-Lân Do Dinh, Steffen Eger and Iryna Gurevych 70

Supervised Rhyme Detection with Siamese Recurrent Networks
Thomas Haider and Jonas Kuhn 81

Normalizing Early English Letters to Present-day English Spelling
Mika Hämäläinen, Tanja Säily, Jack Rueter, Jörg Tiedemann and Eetu Mäkelä 87

Power Networks: A Novel Neural Architecture to Predict Power Relations
Michelle Lam, Catherina Xu and Vinodkumar Prabhakaran 97

Automated Acquisition of Patterns for Coding Political Event Data: Two Case Studies
Peter Makarov 103

A Method for Human-Interpretable Paraphrasticality Prediction
Maria Moritz, Johannes Hellrich and Sven Büchel 113

Exploring word embeddings and phonological similarity for the unsupervised correction of language learner errors
Ildikó Pilán and Elena Volodina 119

Towards Coreference for Literary Text: Analyzing Domain-Specific Phenomena
Ina Roesiger, Sarah Schulz and Nils Reiter 129

An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing
Thomas Schmidt and Manuel Burghardt 139

Automatic identification of unknown names with specific roles
Samia Touileb, Truls Pedersen and Helle Sjøvaag 150

Induction of a Large-Scale Knowledge Graph from the Regesta Imperii
Juri Opitz, Leo Born and Vivi Nastase 159

Conference Program

Saturday, August 25, 2018

09:00–10:30 Session 1

Learning Diachronic Analogies to Analyze Concept Change
Matthias Orlikowski, Matthias Hartung and Philipp Cimiano

A Linked Coptic Dictionary Online
Frank Feder, Maxim Kupreyev, Emma Manning, Caroline T. Schroeder and Amir Zeldes

Using relative entropy for detection and analysis of periods of diachronic linguistic change
Stefania Degaetano-Ortlieb and Elke Teich

11:00–12:30 Session 2

Cliche Expressions in Literary and Genre Novels
Andreas van Cranenburgh

Analysis of Rhythmic Phrasing: Feature Engineering vs. Representation Learning for Classifying Readout Poetry
Timo Baumann, Hussein Hussein and Burkhard Meyer-Sickendiek

Cross-Discourse and Multilingual Exploration of Textual Corpora with the DualNeighbors Algorithm
Taylor Arnold and Lauren Tilton

Saturday, August 25, 2018 (continued)

13:50–14:50 Invited Talk

Measurement and Human Perspective
Ted Underwood

14:50–15:50 Poster Teasers and Poster Session

One Size Fits All? A simple LSTM for non-literal token and construction-level classification
Erik-Lân Do Dinh, Steffen Eger and Iryna Gurevych

Supervised Rhyme Detection with Siamese Recurrent Networks
Thomas Haider and Jonas Kuhn

Normalizing Early English Letters to Present-day English Spelling
Mika Hämäläinen, Tanja Säily, Jack Rueter, Jörg Tiedemann and Eetu Mäkelä

Power Networks: A Novel Neural Architecture to Predict Power Relations
Michelle Lam, Catherina Xu and Vinodkumar Prabhakaran

Automated Acquisition of Patterns for Coding Political Event Data: Two Case Studies
Peter Makarov

A Method for Human-Interpretable Paraphrasticality Prediction
Maria Moritz, Johannes Hellrich and Sven Büchel

Exploring word embeddings and phonological similarity for the unsupervised correction of language learner errors
Ildikó Pilán and Elena Volodina

Towards Coreference for Literary Text: Analyzing Domain-Specific Phenomena
Ina Roesiger, Sarah Schulz and Nils Reiter

Saturday, August 25, 2018 (continued)

An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing
Thomas Schmidt and Manuel Burghardt

16:20–17:20 Session 3

Automatic identification of unknown names with specific roles
Samia Touileb, Truls Pedersen and Helle Sjøvaag

Induction of a Large-Scale Knowledge Graph from the Regesta Imperii
Juri Opitz, Leo Born and Vivi Nastase