Language Technology Resources and Tools for Digital Humanities (LT4DH 2016)

Osaka, Japan
11 – 16 December 2016

ISBN: 978-1-5108-3320-3

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2016) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2016)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Flexible and Reliable Text Analytics in the Digital Humanities – Some Methodological Considerations
Jonas Kuhn 1

Finding Rising and Falling Words
Erik Tjong Kim Sang 2

A Dataset for Multimodal Question Answering in the Cultural Heritage Domain
Shurong Sheng, Luc Van Gool and Marie-Francine Moens 10

Extracting Social Networks from Literary Text with Word Embedding Tools
Gerhard Wohlgenannt, Ekaterina Chernyak and Dmitry Ilvovsky 18

Exploration of register-dependent lexical semantics using word embeddings
Andrey Kutuzov, Elizaveta Kuzmenko and Anna Marakasova 26

Original-Transcribed Text Alignment for Manyosyu Written by Old Japanese Language
Teruaki Oka and Tomoaki Kono 35

Shamela: A Large-Scale Historical Arabic Corpus
Yonatan Belinkov, Alexander Magidow, Maxim Romanov, Avi Shmidman and Moshe Koppel . . 45

Feelings from the Past—Adapting Affective Lexicons for Historical Emotion Analysis
Sven Buechel, Johannes Hellrich and Udo Hahn 54

Automatic parsing as an efficient pre-annotation tool for historical texts
Hanne Martine Eckhoff and Aleksandrs Berdicevskis 62

A Visual Representation of Wittgenstein's Tractatus Logico-Philosophicus
Anca Bucur and Sergiu Nisioi 71

A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures
Richard Eckart de Castilho, Éva Mújdricza-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank and Chris Biemann 76

Challenges and Solutions for Latin Named Entity Recognition
Alexander Erdmann, Christopher Brown, Brian Joseph, Mark Janse, Petra Ajaka, Micha Elsner and Marie-Catherine de Marneffe 85

Geographical Visualization of Search Results in Historical Corpora
Florian Petran 94

Implementation of a Workflow Management System for Non-Expert Users
Bart Jongejan 101

Integrating Optical Character Recognition and Machine Translation of Historical Documents
Haithem Afli and Andy Way 109

Language technology tools and resources for the analysis of multimodal communication
László Hunyadi, Tamás Váradi and István Szekrényes 117

Large-scale Analysis of Spoken Free-verse Poetry
Timo Baumann and Burkhard Meyer-Sickendiek 125

PAT workbench: Annotation and Evaluation of Text and Pictures in Multimodal Instructions
Ielka van der Sluis, Lennart Kloppenburg and Gisela Redeker 131

Semantic Indexing of Multilingual Corpora and its Application on the History Domain
Alessandro Raganato, Jose Camacho-Collados, Antonio Raganato and Yunseo Joung 140

Tagging Ingush - Language Technology For Low-Resource Languages Using Resources From Linguistic Field Work
Jörg Tiedemann, Johanna Nichols and Ronald Sprouse 148

The MultiTal NLP tool infrastructure
Driss Sadoun, Satenik Mkhitaryan, Damien Nouvel and Mathieu Valette 156

Tools and Instruments for Building and Querying Diachronic Computational Lexica
Fahad Khan, Andrea Bellandi and Monica Monachini 164

Tracking Words in Chinese Poetry of Tang and Song Dynasties with the China Biographical Database
Chao-Lin Liu and Kuo-Feng Luo 172

Using TEI for textbook research
Lena-Luise Stahn, Steffen Hennicke and Ernesto William De Luca 181

Web services and data mining: combining linguistic tools for Polish with an analytical platform
Maciej Ogrodniczuk 187