LT4DH 2016

Language Technology Resources and Tools for Digital Humanities (LT4DH)

Proceedings of the Workshop

December 11-16, 2016
Osaka, Japan

ISBN978-4-87974-708-2

Preface

Language resources are increasingly used not only in Language Technology (LT), but also in other subject fields, such as the digital humanities (DH) and in the field of education. Applying LT tools and data for such fields implies new perspectives on these resources regarding domain adaptation, interoperability, technical requirements, documentation, and usability of user interfaces. The workshop Language Technology Resources and Tools for the Digital Humanities focusses on the use of LT tools and data in DH, the discussion focussing on example applications and the type and range of research questions where LT tools can be beneficial.

LT applications are often trained and adjusted to individual text types or corpora published in specific formats. Using the tools in other contexts results in a difference in the data that is to be processed, e.g. historical data or different 'genres'. Though it may seem obvious that the quality of the results may not be as high, the results may still be valuable, for example because of the sheer size of data that can be investigated rather than by manual analysis. Hence tools and resources need to be adaptable to different text types. Applying tools for data from non-LT areas such as the humanities also increases the demands on acceptable data formats, as the data to be processed may contain additional annotations or a variety of annotations. Additionally, in some cases new data conversion needs appear and the tools need to be robust enough to handle also erroneous data, giving meaningful status messages to a non-LT user. It is often also required that tools are adapted to the text types that they are intended to be used for. For example, data mining tools trained for one type of texts need to be adapted for another type.

LT tools often need to be combined in processing chains and workflows whose exact order and configuration depends on the particular LT application. The same is true for DH workflows. However, since the DH applications often significantly differ from those in LT, new configurations of tools need to be entertained and additional requirements for the interoperability of tools may arise. This is particularly the case for interfacing annotation and querying tools as well as the incorporation of data exploration and data visualization techniques.

The technical requirements of some LT tools and the considerable learning curve for its use poses another obstacle for non-expert users in the DH. This means, inter alia, that downloads of tools and complex local installations should be avoided and tools should be made available as web-applications whenever possible. Moreover, usability studies of LT tools for DH applications may give important feedback for the adaptation of user interaction, adaptation of algorithms, and the need for additional functionality.

Organisers

- Erhard Hinrichs (University of Tübingen, Germany)
- Marie Hinrichs (University of Tübingen, Germany)
- Thorsten Trippel (University of Tübingen, Germany)

Programme Committee

- Andre Blessing (University of Stuttgart, Germany)
- Mirjam Bluemm (Universtity of Göttingen, Germany)
- António Branco (University of Lisbon, Portugal)
- Thierry Declerck (DFKI, Germany)
- Stefanie Dipper (Ruhr-University Bochum, Germany)
- Thomas Gloning (University of Gießen, Germany)
- Elena Gonzalez-Blanco (National Distance Education University, Spain)
- Hanna Hedeland (University of Hamburg, Germany)
- Erhard Hinrichs (University of Tübingen, Germany)
- Marie Hinrichs (University of Tübingen, Germany)
- Nancy Ide (Vassar College, USA)
- Wiltrud Kessler (University of Stuttgart, Germany)
- Sandra Kübler (Indiana University Bloomington, USA)
- Gunn Lyse (University of Bergen, Norway)
- Monica Monachini (Institute for Computational Linguistics «A. Zampolli», Italian National Research Council, Italy)
- Stefan Schmunk (University of Göttingen, Germany)
- Stephanie Strassel (LDC, Philadelphia, USA)
- Thorsten Trippel (University of Tübingen, Germany)
- Arjan van Hessen (University of Utrecht, Netherlands)

Table of Contents

Conference Program

Sunday December 11,2016

09:00–09:15 Welcome and Introduction

09:15–10:00 Invited Keynote

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

Oral Presentations of Submitted Papers, Session 1

10:00–10:25

Finding Rising and Falling Words
Erik Tjong Kim Sang

10:25–10:50

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

Sunday December 11,2016 (continued)

10:50–11:10 Coffee Break

Oral Presentations of Submitted Papers, Session 2

11:10–11:35

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

11:35–12:00

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

12:00–14:00 Lunch Break

Oral Presentations of Submitted Papers, Session 3

14:00–14:25

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

Sunday December 11,2016 (continued)

14:25–14:50

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

14:50–15:15

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

15:15–15:40 Posterslam

15:40–16:00 Coffee Break

Poster Presentations of Submitted Papers

16:00–17:00

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

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

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

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

Geographical Visualization of Search Results in Historical Corpora
Florian Petran

Sunday December 11,2016 (continued)

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

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

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

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

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

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

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

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

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

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

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

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