8th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities

(LaTeCH 2014)

Held at the 14th Conference of the European Chapter of the Association for Computational Linguistics

Gothenburg, Sweden
26 April 2014

ISBN: 978-1-63266-401-3

Printed from e-media with permission by:

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

CURRAN ASSOCIATES INC.
proceedings
.com

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

Printed by Curran Associates, Inc. (2014)

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: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Workshop Program

Saturday, April 26, 2014

8:45–8:50 Welcome

8:50–9:30 Invited Talk by Gerhard Heyer:
A New Implementation for Canonical Text Services
Jochen Tiepmar, Christoph Teichmann, Gerhard Heyer, Monica Berti and Gregory Crane

Session I: Linked data in the Humanities

9:30–9:45 *How to semantically relate dialectal Dictionaries in the Linked Data Framework*
Thierry Declerck and Eveline Wandl-Vogt

9:45–10:05 *Bootstrapping a historical commodities lexicon with SKOS and DBpedia*
Ewan Klein, Beatrice Alex and Jim Clifford

10:05–10:25 *New Technologies for Old Germanic. Resources and Research on Parallel Bibles in Older Continental Western Germanic*
Christian Chiarcos, Maria Sukhareva, Roland Mittmann, Timothy Price, Gaye Detmold and Jan Chobotsky

10:25–11:00 Coffee break

Session II: Spelling normalisation & sense disambiguation

11:00–11:20 *A Multilingual Evaluation of Three Spelling Normalisation Methods for Historical Text*
Eva Pettersson, Beáta Megyesi and Joakim Nivre

11:20–11:40 *Enhancing the possibilities of corpus-based investigations: Word sense disambiguation on query results of large text corpora*
Christian Poelitz and Thomas Bartz

11:40–12:00 *A Hybrid Disambiguation Measure for Inaccurate Cultural Heritage Data*
Julia Efremova, Bijan Ranjbar-Sahraei and Toon Calders

12:00–12:15 *Automated Error Detection in Digitized Cultural Heritage Documents*
Kata Gábor and Benoît Sagot

12:15–13:45 Lunch break

Saturday, April 26, 2014 (continued)

Session III: Social Science applications

13:45–14:05 *Mining the Twentieth Century's History from the Time Magazine Corpus*
Mike Kestemont, Folgert Karsdorp and Marten Düring

14:05–14:25 *Social and Semantic Diversity:*
Socio-semantic Representation of a Scientific Corpus
Thierry Poibeau, Elisa Omodei, Jean-Philippe Cointet and Yufan Guo

Poster Booster Session

14:25–14:35 *A Tool for a High-Carat Gold-Standard Word Alignment*
Drayton Benner

14:35–14:45 *CorA: A web-based annotation tool for historical and other non-standard language data*
Marcel Bollmann, Florian Petran, Stefanie Dipper and Julia Krasselt

14:45–14:55 *Developing a Tagalog Linguistic Inquiry and Word Count (LIWC) 'Disaster' Dictionary for Understanding Mixed Language Social Media: A Work-in-Progress Paper*
Amanda Andrei, Alison Dingwall, Theresa Dillon and Jennifer Mathieu

14:55–15:05 *Text Analysis of Aberdeen Burgh Records 1530-1531*
Adam Wyner, Jackson Armstrong, Andrew Mackillop and Philip Astley

15:05–15:15 *From Syntax to Semantics. First Steps Towards Tectogrammatical Annotation of Latin*
Marco Passarotti

15:15–15:25 *On the syllabic structures of Aromanian*
Sergiu Nisioi

15:25–16:00 Coffee break & Poster Session

Session IV: Knowledge resources acquisition

16:00–16:20 *A Gazetteer and Georeferencing for Historical English Documents*
Claire Grover and Richard Tobin

16:20–16:40 *Automatic Wayang Ontology Construction using Relation Extraction from Free Text*
Hadaiq Sanabila and Ruli Manurung

16:40–17:30 SIGHUM annual business meeting