

Workshop on Language Technologies for Digital Humanities and Cultural Heritage 2011

(DigHum 2011)

**Held with the 8th International Conference on Recent Advances
in Natural Language Processing 2011 (RANLP 2011)**

**Hissar, Bulgaria
16 September 2011**

Editors:

**Cristina Vertan
Petya Osenova**

**Milena Slavcheva
Stelios Piperidis**

ISBN: 978-1-62276-461-7

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© (2011) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2012)

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

Invited Talk

<i>Endangered Uralic Languages and Language Technologies</i> Gábor Prósztéký	1
---	---

Electronic Archives

<i>A Framework for Improved Access to Museum Databases in the Semantic Web</i> Dana Dannélls, Mariana Damova, Ramona Enache and Milen Chechev	3
<i>Query classification via Topic Models for an art image archive</i> Dieu-Thu Le, Raffaella Bernardi and Ed Vald	11
<i>Unlocking Language Archives Using Search</i> Herman Stehouwer and Eric Auer	19
<i>Digital Library of Poland-related Old Ephemeral Prints: Preserving Multilingual Cultural Heritage</i> Maciej Ogrodniczuk and Włodzimierz Gruszczyński	27

Language Technology and Resources

<i>Rule-Based Normalization of Historical Texts</i> Marcel Bollmann, Florian Petran and Stefanie Dipper	34
<i>Survey on Current State of Bulgarian-Polish Online Dictionary</i> Ludmila Dimitrova, Ralitsa Dutsova and Rumiana Panova	43
<i>Language Technology Support for Semantic Annotation of Icono-graphic Descriptions</i> Kamenka Staykova, Gennady Agre, Kiril Simov and Petya Osenova	51
<i>The Tenth-Century Cyrillic Manuscript Codex Suprasliensis: the creation of an electronic corpus. UNESCO project (2010–2011)</i> Hanne Martine Eckhoff, David Birnbaum, Anissava Miltenova and Tsvetana Dimitrova	57

Computational Methods in Literary Analysis

<i>SentiProfiler: Creating Comparable Visual Profiles of Sentimental Content in Texts</i> Tuomo Kakkonen and Gordana Galic Kakkonen	62
<i>Character Profiling in 19th Century Fiction</i> Dimitrios Kokkinakis and Mats Malm	70

<i>Diachronic Stylistic Changes in British and American Varieties of 20th Century Written English Language</i> Sanja Štajner and Ruslan Mitkov	78
---	----

Multimodal Aspects in Digital Humanities

<i>AVATeCH: Audio/Video Technology for Humanities Research</i> Sebastian Tschöpel, Daniel Schneider, Rolf Bardeli, Oliver Schreer, Stefano Masneri, Peter Wittenburg, Han Sloetjes, Przemek Lenkiewicz and Eric Auer	86
<i>Handwritten Text Recognition for Historical Documents</i> Veronica Romero, Nicolas Serrano, Alejandro H. Toselli, Joan Andreu Sanchez and Enrique Vidal	90
<i>Reducing OCR Errors in Gothic-Script Documents</i> Lenz Furrer and Martin Volk	97