

# **6th Workshop on Recent Advances in Slavonic Natural Language Processing**

**(RASLAN 2012)**

**Karlova Studanka, Czech Republic  
7 - 9 December 2012**

**Editors:**

**A. Horak**

**P. Rychly**

**ISBN: 978-1-62993-832-5**

**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© (2012) by the Masaryk University Faculty of Informatics  
All rights reserved.

Printed by Curran Associates, Inc. (2014)

For permission requests, please contact the Masaryk University Faculty of Informatics  
at the address below.

Masaryk University Faculty of Informatics  
c/o Dana Hlaváčková  
Centre for Natural Language Processing  
B206 Botanická 554/68a, Ponava, Brno  
Czech Republic

Phone: +420 549 49 1864, 3329

[ydana@aurora.fi.muni.cz](mailto:ydana@aurora.fi.muni.cz)

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Table of Contents

---

## I Syntax, Morphology and Lexicon

---

Saara: Anaphora Resolution on Free Text in Czech . . . . .	3
<i>Vašek Němčík</i>	
Behaviour of the Czech Suffix <i>-ák</i> – A Case Study . . . . .	9
<i>Dana Hlaváčková, Karel Pala</i>	
Reproducing Czech Syntactic Parsing Results Published in CoNLL Tasks .	15
<i>Lucia Kocíncová</i>	
Adaptation of Czech Parsers for Slovak . . . . .	23
<i>Marek Medved', Miloš Jakubíček, Vojtěch Kovář, Václav Němčík</i>	

---

## II Logic and Language

---

Deduction System for TIL-2010 . . . . .	33
<i>Marie Duží, Marek Menšík, Lukáš Vích</i>	
Czech Knowledge-Based System with Temporal Reasoning . . . . .	43
<i>Andrej Gardoň</i>	
Linguistic Logical Analysis of Direct Speech . . . . .	51
<i>Aleš Horák, Miloš Jakubíček, Vojtěch Kovář</i>	
Building Evaluation Dataset for Textual Entailment in Czech . . . . .	61
<i>Zuzana Nevěřilová</i>	

---

## III Text Corpora and Tools

---

Detecting Spam Content in Web Corpora . . . . .	69
<i>Vít Baisa, Vít Suchomel</i>	
Recent Czech Web Corpora . . . . .	77
<i>Vít Suchomel</i>	
CzAccent – Simple Tool for Restoring Accents in Czech Texts . . . . .	85
<i>Pavel Rychlý</i>	
Towards 100M Morphologically Annotated Corpus of Tajik . . . . .	91
<i>Gulshan Dovudov, Vít Suchomel, Pavel Šmerk</i>	

---

## **IV Language Modelling**

---

Building A Thesaurus Using LDA-Frames .....	97
<i>Jiří Materna</i>	
Improving Automatic Ontology Development .....	105
<i>Marek Grác, Adam Rambousek</i>	
Authorship Verification based on Syntax Features .....	111
<i>Jan Rygl, Kristýna Zemková, Vojtěch Kovář</i>	
Segmentation from 97% to 100%: Is It Time for Some Linguistics? .....	121
<i>Petr Sojka</i>	

## **Author Index**