Zygmunt Vetulani · Joseph Mariani
Marek Kubis (Eds.)

Human Language Technology

Challenges for Computer Science and Linguistics

7th Language and Technology Conference, LTC 2015
Poznań, Poland, November 27–29, 2015
Revised Selected Papers

Springer

Contents

Speech Processing

Intelligent Speech Features Mining for Robust Synthesis System Evaluation . . . 3
Moses E. Ekpenyong, Udoinyang G. Inyang, and Victor E. Ekong

Neural Networks Revisited for Proper Name Retrieval
from Diachronic Documents 19
Irina Illina and Dominique Fohr

Cross-Lingual Adaptation of Broadcast Transcription System to Polish
Language Using Public Data Sources 31
Jan Nouza, Petr Cerva, and Radek Safarik

Automatic Transcription and Subtitling of Slovak Multi-genre
Audiovisual Recordings 42
Ján Staš, Peter Viszlay, Martin Lojka, Tomáš Koctúr, Daniel Hládek, and Jozef Juhár

Multiword Expressions

SEJF - A Grammatical Lexicon of Polish Multiword Expressions 59
Monika Czerepowicka and Agata Savary

Lemmatization of Multi-Word Entity Names for Polish Language
Using Rules Automatically Generated Based on the Corpus Analysis 74
Jacek Małyszko, Witold Abramowicz, Agata Filipowska, and Tomasz Wagner

Parsing

Parsing of Polish in Graph Database Environment. 87
Jan Posiadała, Hubert Czaja, Eliza Szczechla, and Paweł Susicki

Language Resources and Tools

RetroC – A Corpus for Evaluating Temporal Classifiers 101
Filip Graliński and Piotr Wierzchoń

Reinvestigating the Classification Approach to the Article
and Preposition Error Correction 112
Roman Grundkiewicz and Marcin Junczys-Dowmunt

Binary Classification Algorithms for the Detection of Sparse Word Forms in New Indo-Aryan Languages 123
Rafał Jaworski, Krzysztof Jassem, and Krzysztof Stroński

Multilingual Tokenization and Part-of-speech Tagging. Lightweight Versus Heavyweight Algorithms 137
Tiberiu Boros and Stefan Daniel Dumitrescu

Ontologies and Wordnets

A Semantic Similarity Measurement Tool for WordNet-Like Databases 155
Marek Kubis

Similarity Measure for Polish Short Texts Based on Wordnet-Enhanced Bag-of-words Representation 169
Maciej Piasecki and Anna Gut

Methods of Linking Linguistic Resources for Semantic Role Labeling 183
Balázs Indig, Márton Mihált, and András Simonyi

Machine Translation

A Quality Estimation System for Hungarian 201
Zijian Győző Yang, Andrea Dömötör, and László János Laki

Leveraging the Advantages of Associative Alignment Methods for PB-SMT Systems 214
Baosong Yang and Yves Lepage

Information and Data Extraction

Events Extractor for Polish Based on Semantics-Driven Extraction Templates 231
Jolanta Cybulka and Jakub Dutkiewicz

Understanding Questions and Extracting Answers: Interactive Quiz Game Application Design 246
Volha Petukhova, Desmond Darma Putra, Alexandr Chernov, and Dietrich Klakow

Exploiting Wikipedia-Based Information-Rich Taxonomy for Extracting Location, Creator and Membership Related Information for ConceptNet Expansion 262
Marek Krawczyk, Rafal Rzepka, and Kenji Araki

Text Engineering and Processing

Lexical Analysis of Serbian with Conditional Random Fields and Large-Coverage Finite-State Resources 277
Mathieu Constant, Cvetana Krstev, and Duško Vitas

Improving Chunker Performance Using a Web-Based Semi-automatic Training Data Analysis Tool. 290
István Endrédy

A Connectionist Model of Reading with Error Correction Properties 304
Max Raphael Sobroza Marques, Xiaoran Jiang, Olivier Dufor, Claude Berrou, and Deok-Hee Kim-Dufor

Applications in Language Learning

The Automatic Generation of Nonwords for Lexical Recognition Tests 321
Osama Hamed and Torsten Zesch

Teaching Words in Context: Code-Switching Method for English and Japanese Vocabulary Acquisition Systems 332
Michal Mazur, Rafal Rzepka, and Kenji Araki

Emotions, Decisions and Opinions

Automatic Extraction of Harmful Sentence Patterns with Application in Cyberbullying Detection. 349
Michal Ptaszynski, Fumito Masui, Yasutomo Kimura, Rafal Rzepka, and Kenji Araki

Sentiment Analysis in Polish Web-Political Discussions. 363
Antoni Sobkowicz and Marek Kozłowski

Saturation Tests in Application to Validation of Opinion Corpora: A Tool for Corpora Processing 378
Zygmunt Vetulani, Marta Witkowska, Suleyman Menken, and Umut Canbolat

Less-Resourced Languages

Issues and Challenges in Developing Statistical POS Taggers for Sambalpuri 393
Pitambar Behera, Atul Kr. Ojha, and Girish Nath Jha

Cross-Linguistic Projection for French-Vietnamese Named Entity Translation 407
Ngoc Tan Le and Fatiha Sadat

National Language Technologies Portals for LRLs: A Case Study 420
Delyth Prys and Dewi Bryn Jones

Challenges for and Perspectives on the Malagasy Language in the Digital Age 430
Joro Ny Aina Ranaivoarison

Author Index 445