UNITED INSTITUTE OF INFORMATICS PROBLEMS
OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

International Scientific Conference on the Automatic Processing of Natural-Language Electronic Texts “NooJ’2015”

NOOJ 2015

Abstracts

June 11–13, 2015, Minsk, Belarus

Minsk
UIIP NASB
2015

УДК 004.91

International Scientific Conference on the Automatic Processing of Natural-Language Electronic Texts "NooJ'2015" : Abstracts (11–13 June, 2015, Minsk, Belarus). – Minsk : UIIP NASB, 2015. – 80 p.
ISBN 978-985-6744-89-4.

This volume contains the abstracts of the International conference "NooJ 2015". The research presented covers different aspects of natural language processing using NooJ, including formalizing such levels of linguistic phenomena as syllabification, phonemic and prosodic transcription, multiword units and discontinuous expressions, local and structural syntax; transformational syntax and paraphrase generation, semantic analysis and machine translation, etc.

Abstracts are published in the form presented by authors.

У дадзеным зборніку прадстаўлены тэзісы дакладаў Міжнароднай канферэнцыі "NooJ 2015". Разглядаюцца розныя аспекты апрацоўкі натуральнай мовы з выкарыстаннем лінгвістычнага асяроддзя распрацоўкі NooJ, улічваючы фармалізаванне такіх напрамкаў лінгвістычнага аналізу як склададзяленне, фанетычная і прасадычная транскрыпцыі, устойлівыя выразы і дыскрэтныя слоўныя канструкцыі, лакальны і структурны сінтаксісы, трансфармацыйны сінтаксіс і перафразаванне, семантычны аналіз і машынны пераклад і г. д.

Тэзісы друкуюцца ў выглядзе, пададзеным аўтарамі.

Scientific Editors:
DSc in Engineering B.M. Lobanov,
PhD in Engineering Yu.S. Hetsevich

ISBN 978-985-6744-89-4

© United Institute of Informatics Problems of the National Academy of Sciences of Belarus, 2015

COMPARISON OF LEXICAL AND GRAMMATICAL BASE OF BELARUSIAN N-KORPUS WITH DICTIONARY PROPERTIES' DEFINITION FILE OF BELARUSIAN NOOJ MODULE

U. Koshchanka[1], Yu. Hetsevich[2], V. Varanovich[3], A. Tretyak[3]
[1] The Center for the Belarusian Culture, Language
and Literature researches of the NAS of Belarus, Minsk
e-mail: koshul@gmail.com;
[2] United Institute of Informatics Problems of the NAS of Belarus, Minsk;
[3] Belarusian State University, Minsk

In this work, we plan to compare the matching parts of the Belarusian N-korpus and dictionary properties' definition file of the Belarusian module in NooJ.

Belarusian N-korpus is the first publicly available general Belarusian language corpus [1]. The Belarusian N-korpus currently contains ~93 885 texts (~57 729 850 tokens) taken from fiction, newspapers, journals and on-line editions. The texts of the corpus are grammatically annotated and contain metatextual information. The grammar database is available under Open Database License (ODbL) v1.0. The corpus engine is available under GNU General Public License, Version 3.

The Belarusian module for NooJ has been created during the period of 2011–2012 [2]. Basic results have been presented at the 14th international NooJ conference in 2011 (Dubrovnik, Croatia). The Belarusian NooJ module consists of the following parts: Texts, Dictionary, Grammars, Samples, Projects. They were composed for a basic acquaintance with the Belarusian language. This project is highly popular, as it allows the flexibility to customize the basic functionality of NooJ by further replenishment of parts of the Belarusian module for automatic annotation of Belarusian texts.

References

1. Belarusian N-korpus [Electronic resource]. – 2012. – Mode of access : http://bnkorpus.info. – Date of access : 10.02.2015.

2. Belarusian module for NooJ / Y. Hetsevich [et al.] // NooJ web-site [Electronic resource]. – 2012. – Mode of access : http://www.nooj4nlp.net/ pages/belarusian.html. – Date of access : 10.02.2015.

CONTENTS

PREFACE 5

Ben Ali H., Rhazi A., Aouini M. Translating Arabic Active Sentences into English Passive Sentences using NooJ Platform 7

Benet V. Semantic Tags for NooJ Russian Dictionary 9

Blanco X. A Hierarchy of Semantic Labels for Spanish Dictionaries 10

Chernyshevich M., Stankevitch V. A Hybrid Approach to Extracting and Encoding Disorder Mentions from Clinical Notes 12

Collec Clerc V. Mixed Prolog and NooJ Approach in Japanese Benefactive Constructions 14

Buono di M.P. Semi-Automatic Indexing and Parsing Information on the Web with NooJ 16

Duran M. The Annotation of Compound Suffixation Structure of Quechua Verbs 18

Dzenisiuk D., Hetsevich Yu. Processing of Publication References in Belarusian and Russian Electronic Texts 20

Ghezaiel N., Haddar K. Study and Resolution of Arabic Lexical Ambiguity through the Transduction on Text Automaton 21

Hetsevich Yu., Borodina J. Using NooJ for the Processing of Satellite Data 23

Hetsevich Yu., Okrut T., Lobanov B. Grammars for the Sentence into Phrase Segmentation: Punctuation Level 25

Hiuntar A., Zahariev V. Grammars for Making Written Orthographic Words from Transcribed Spoken Language 26

Kaigorodova L., Hetsevich Yu., Nikalaenka K., Prakapovich R., Gerasuto S., Sychou U. Language Modelling for Robots-Human Interaction 28

Kirova M. Translating Spacial and Temporal Deixis in Near Languages: A Comparative Classification Approach with NooJ 30

Kocijan K., Librenjak S. Recognizing Verb-Based Croatian Idiomatic MWUs 31

Koshchanka U., Hetsevich Yu., Varanovich V., Tretyak A. Comparison of Lexical and Grammatical Base of Belarusian N-Korpus with Dictionary Properties' Definition File of Belarusian NooJ Module 33

Le Pesant D. Semantic Tagging of the Sentiment Words with NooJ 34

Loskutova A. Creation of Geographical Names Dictionary of Alaska Toponyms 35

Lysy S., Hiuntar A., Hetsevich Yu. Addition of Phonetic Transcriptions to Belarusian Module of NooJ 36

Maisto A., Guarasci R. Morpheme-Based Recognition and Translation of Medical Terms 38

Mesfar S., Najar D. How to Automatically Enrich Linguistic Resources Using NooJ: Application on Arabic Module 40

Monteleone M. Local Grammars and Formal Semantics: Past Participles Vs. Adjectives in Italian 41

Mota C., Carvalho P., Raposo F., Barreiro A. Paraphrasing Human Intransitive Adjective Constructions in Port4NooJ 43

Najar D., Mesfar S. A Large Terminological Dictionary of Arabic Compound Words..... 46

Okrut T., Lobanov B., Yakubovich Y. Context-Sensitive Homograph Disambiguation with NooJ in Belarusian and Russian Electronic Texts 48

Patsiomkin A., Hetsevich Yu. Semantic Analysis for Locating Expressive Means and Stylistic Devices in Authentic English Texts, Ranging and Classification..... 49

Pejar T., Kocijan K., Bekavac B. Normalization of Tweets in Croatian Language Using NooJ 51

Pelosi S. Morphological Relations for the Automatic Expansion of Italian Sentiment Lexicons 52