

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 фармалізаваўныя такія 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

USING NOOJ FOR THE PROCESSING OF SATELLITE DATA

Yu. Hetsevich, J. Borodina

United Institute of Informatics Problems of the NAS of Belarus, Minsk

e-mail: yury.hetsevich@gmail.com

The paper describes the processing of satellite telemetry data and its conversion from reduced form into full orthographically correct word sequences for the further use in text-to-speech synthesis or other applications.

The purpose of a telemetry system is to reliably and transparently convey measurement information from a remotely located data generating source to users located in space or on Earth. Typically, data generators are scientific sensors, science housekeeping sensors, engineering sensors and other subsystems on-board a spacecraft [1]. By satellite telemetry data in our paper we also mean results of measurements made by nanosatellites from space and received on Earth by special software.

For the research we use text corpus which was created with the data from satellite telemetry converter software. The data is provided in the form of abbreviations, numbers and measurement units along with orthographical words (e.g. *Voltage of 5V system is 4,904 [V]; Temperature of the 145 MHz TX: 16 °C*).

Main difficulty which arises during the work with quantitative expressions with measurement units (QEMU) in relation to telemetry is language. Telemetry data are mostly collected in English as an international language for science, yet Belarusian national space tradition requires this information to be available in one of the national languages of Belarus: Belarusian or Russian. Therefore our task is not only to convert QEMU from reduced form into the full one, but also to automatically translate data from English to Belarusian. The grammar made in NooJ [2] is designed to be language independent and self-sufficient, i.e. there are no dictionaries applied.

The grammar takes as an input text sequences like *0,9708 A* and transforms first numerical part, then measurement units into Belarusian phrase: *нуль цэлых дзевяць тысяч семсот восем дзесяцітысячных ампера* ‘zero point nine thousand seven hundred and eight ten-thousandths of an ampere’. Due to the fact that Belarusian and Russian languages are both synthetic, declension paradigm of this numeral will be very complex, but our grammar can handle it.

This work continues the work which was presented in the previous NooJ conference in 2014 held in Sassari, Italy.

References

1. Synchronization, T.C. Report Concerning Space Data System Standards. Channel Coding – Summary of Concept and Rationale / T.C. Synchronization // Green Book. – November, 2012.
2. Silberztein, M. Nooj Manual / M. Silberztein [Electronic resource]. – 2003. – Mode of access : <http://www.nooj4nlp.net>. – Date of access : 21.12.2014.

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., Prakupovich 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