

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

U. Koshchanka, Yu. Hetsevich, V. Varanovich, A. Tretyak

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.