УДК 004.94

WORKING WITH PYTHON'S NLP LIBRARIES

Murodov Shukrilla Abdusaid o‘g‘li
Karshi International University
Senior lecturer of "Exact Sciences”

Annotation: *To provide students with practical knowledge of working with libraries related to solving NLP problems in the Python environment.*

Keywords: *NLP, SpaCy libraries, NLPub, Python, text processing, platform, software technologies, Russian language.*

Today, large-scale scientific and research work is being carried out on automatic natural language processing (NLP) on a global scale.

Creating, researching and using computer-oriented linguistic models of natural languages for practical purposes, in a word, software technologies aimed at solving linguistic problems through automated systems are called language technologies. is conducted. Today, such technologies are widely used in foreign countries. Most of the language technologies provide for the research and processing of linguistic models of foreign languages, and some of them, for example, NLTK on the Python platform, SpaCy libraries, Wolfram Alpha and similar systems have Uzbek simple studies can be conducted on some elements of the language.

To get the most information about language technologies currently operating around the world, you can refer to the NLPub catalog of electronic linguistic resources.

NLPub is a catalog of electronic resources related to natural language processing, the information in it is placed in different sections of the electronic catalog in an orderly, classified manner (Table 1).

Also, NLPub includes the following projects aimed at creating and improving linguistic resources for the Russian language:

1) RUSSE (RUSsian Semantic Evaluation) - a seminar-project on the comparison of methods of computational semantics (methods for identifying words that are semantically close to each other are compared and analyzed);

2) LRWC (Lexical Relations from the Wisdom of the Crowd) - a project related to the discussion of experts on semantic relations.

3) YARN (Yet Another RussNet) - a project to create a new open electronic thesaurus of the Russian language;

4) RTLOD (Russian Thesaurus Linked Open Data) - a project related to the creation of an open electronic thesaurus of the Russian language consisting of interrelated data;

5) RDT (Russian Distributional Thesaurus) - a project related to the creation of an open distributional thesaurus of the Russian language, etc.

Table 1

The main sections of the NLPub electronic catalog

Methods and instruments	Resources	Experts and activities	Education
Text processing, speech processing, utilities, methods, algorithms	Dictionaries, thesauruses, corpora, data bank	Organizations, consulting experts, conferences	Education, literature, diploma topics

In addition, NLPub is a permanent information partner of AINL (Artificial Intelligence and Natural Language) conferences held annually in Russia since 2012 and ISMW (Intelligence, Social Media and Web) conferences held in St. Petersburg since 2015. is considered

Below, we provide information on some of the natural language processing technologies developed by the world's largest and most famous companies, based on information from various sections of the NLPub electronic resource catalog (Table 2).

Table 2

Software technologies for text processing

Name	Function and method	Communication language	License	Platform
ABBYY Compreno	Parsing (rule-based)	Russian	Commerce	Web service
Extractor	Keyword extraction, automatic abstracting (genetic algorithm)	English, German, French, Japanese, Spanish	Commerce	Web service
MBSP	Graphematic, morphological, syntactic analysis (machine learning)	English	GPL	Python
natural	Graphematic and morphological, analysis, keyword separation (based on regular expressions, rules and TF-IFD)	English, French, German, Japanese, Spanish, Korean, Persian, Italian	MIT	Node.js
NLTK	Graphematic and morphological analysis (regular expressions, machine learning)	English	Apache License	Python
Text::Hyphen	Moving lines by syllables (based on Tex templates)	English, Russian and more than 30 languages	MIT	R and Python
Pattern	Graphematic, morphological,	English, Spanish,	BSD	Python

	syntactic analysis and spell checking (regular expressions)	German, French, Italian		
Solarix	Graphematic, morphological, syntactic analysis (based on dictionaries and rules)	Russian, English	Commerce	Windows, Linux
SpaCy	Integrated package (word processing tool)	English	Framework, MIT	Python
Twitter NLP and Part-of-Speech Tagger	Graphematic and Morphological Analysis (Machine Learning)	English	GPL	Java
AOT	Graphematic and morphological analysis (with dictionary), syntactic analysis (HPSG grammar)	Russian, English	LGPL	Linux, Windows
zamgi	Text segmentation (based on Viterbi algorithm)	all languages	MIT	Python
Tokenizer	Graphematic analysis (rule-based)	Russian, English, German	GPL	C++

TextBlob	Graphematic, morphological and tonal analysis (regular expression, machine learning	English	MIT	Python
pymorphy	Morphological analysis (based on vocabulary)	Russian, English, German	MIT	Python

In order to use existing linguistic resources in practice, it is necessary to have a strong mechanism of its management. In this case, there is a need for programming systems focused on processing more linguistic data.

LIST OF REFERENCES

1. Alexander Clark, Chris Fox, and Shalom Lappin. (Edited) The Handbook of Computational Linguistics and Natural Language Processing. 2010 Blackwell Publishing Ltd.
2. Большакова Е.И., Клышинский Э.С. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. – М.: МИЭМ, 2011. – 272 с.
3. Норов А.М. Компьютер лингвистикаси асослари. – Қарши, “Насаф”, 2017. – 128 б.
4. Пўлатов А. Компьютер лингвистикаси. –Т.: 2011.