

Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

Государственное учреждение
«Администрация Парка высоких технологий»

Открытые семантические технологии проектирования интеллектуальных систем

Open Semantic Technologies for Intelligent Systems

МАТЕРИАЛЫ
МЕЖДУНАРОДНОЙ
НАУЧНО-ТЕХНИЧЕСКОЙ КОНФЕРЕНЦИИ

Выпуск 1

(Минск, 16-18 февраля 2017)

Основан в 2011 году

Минск
БГУИР 2017

УДК 004.822+004.89-027.31
ББК 32.813-02+73
О-83

ISSN 2415-7740

Редакционная коллегия:

*В. В. Голенков (отв. ред.), Л. С. Глоба, В. А. Головко, Н. А. Гулякина, О. П. Кузнецов, Б. М. Лобанов,
Г. С. Осипов, С. В. Смирнов, П. И. Соснин, Д. Ш. Сулейманов, В. Ф. Хорошевский*

Организаторы конференции:

Министерство образования Республики Беларусь

Министерство связи и информатизации

Учреждение образования «Белорусский
государственный университет информатики и
радиоэлектроники»

Российская ассоциация искусственного интеллекта

Государственное учреждение «Администрация Парка
высоких технологий» (Республика Беларусь)

Объединённый институт проблем информатики
Национальной академии наук Беларусь

Брестский государственный технологический
университет

ОАО «Савушкин продукт»

ФГБНУ «Институт управления образованием
Российской академии образования»

Кафедра молодежной политики РИВШ

ООО «Интеллектуальные процессы»

ЗАО «Qulix Systems»

ЗАО «Итранзишен»

ЧТУП «АйДжи Софт Бел»

ОДО «ВирусБлокАда»

ООО «Октонион технолоджи»

ИООО «Софтек Девелопмент»

ООО «Альторос Девелопмент»

ООО «РСЛ»

ООО «Аквео»

ООО «Датарокетс»

ООО «Соларвиндз МСП Технолоджи»

Техническая и информационная поддержка:

Научный журнал «Информатика»

Научный журнал «Онтология проектирования»

Научно-технический журнал «Автоматизация процессов
управления»

Научно-практический журнал для специалистов
«Электроника ИНФО»

Международный журнал «Программные продукты и
системы»

Научно-производственный журнал «Веснік сувязі»

*Издание осуществлено по заказу государственного учреждения
«Администрация Парка высоких технологий»*

О-83 **Открытые семантические технологии проектирования интеллектуальных
систем = Open Semantic Technologies for Intelligent Systems** : материалы
междунар. науч.-техн. конф./ редкол. : В. В. Голенков (отв. ред.) [и др.], ISSN
2415-7740; Вып.1 (Минск, 16-18 февраля 2017г.). – Минск : БГУИР, 2017. --
466 с.

ISBN 978-985-543-034-7.

Сборник включает прошедшие рецензирование статьи VII международной научно-технической
конференции «Открытые семантические технологии проектирования интеллектуальных систем».

Сборник предназначен для преподавателей высших учебных заведений, научных сотрудников,
студентов, аспирантов, магистрантов, а также для специалистов предприятий в сфере
проектирования интеллектуальных систем.

Материалы сборника одобрены Программным комитетом OSTIS и печатаются в виде,
представленном авторами.

**УДК 004.822+004.89-027.31
ББК 32.813-02+73**

ISBN 978-985-543-034-7

© УО «Белорусский государственный
университет информатики
и радиоэлектроники», 2017

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	27
ПАМЯТИ АРКАДИЯ НИКОЛАЕВИЧА БОРИСОВА	29
АРКАДИЙ НИКОЛАЕВИЧ БОРИСОВ: СТРАНИЦЫ БИОГРАФИИ И НАУЧНОЕ НАСЛЕДИЕ	32
Поспелов Д.А., Аверкин А.Н., Батыршин И.З., Тарасов В.Б., Язенин А.В.	
ОНТОЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ	37
Голенков В.В.	
ОНТОЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ БАЗ ЗНАНИЙ	57
Давыденко И.Т.	
ОНТОЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ МАШИН ОБРАБОТКИ ЗНАНИЙ	73
Шункевич Д.В.	
ОНТОЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ ПОЛЬЗОВАТЕЛЬСКИХ ИНТЕРФЕЙСОВ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ	95
Борискин А.С., Жуков И.И., Корончик Д.Н., Садовский М.Е., Хусаинов А.Ф.	
ИНТЕГРАЦИЯ ПРИКЛАДНОЙ ОНТОЛОГИИ И WIKI-РЕСУРСОВ В РАМКАХ ЕДИНОЙ БАЗЫ ЗНАНИЙ	107
Мошкин В.С., Филиппов А.А., Ярушкина Н.Г.	
ИНТЕГРАЦИЯ ОНТОЛОГИЙ И КОНЦЕПТУАЛЬНЫХ МОДЕЛЕЙ ИС	111
Гуськов Г.Ю., Наместников А.М.	
НЕЧЕТКАЯ ОНТОЛОГИЯ КАК ЯДРО БАЗЫ ЗНАНИЙ ДЛЯ ТЕХНИЧЕСКОГО ЭЛЕКТРОННОГО АРХИВА	117
Наместников А.М.	
РАЗРАБОТКА И ИССЛЕДОВАНИЕ КОМБИНИРОВАННОГО АЛГОРИТМА КЛАСТЕРИЗАЦИИ ВРЕМЕННЫХ РЯДОВ	121
Сибирев И.В., Афанасьева Т.В.	
ИСПОЛЬЗОВАНИЕ ОНТОЛОГИЧЕСКИХ ЗНАНИЙ В СЕМАНТИЧЕСКОМ ПОИСКЕ СЛОЖНЫХ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ	127
Рогушина Ю.В.	
ОНТОЛОГИЧЕСКИЙ ПОДХОД К СОПОСТАВЛЕНИЮ КОМПЕТЕНЦИЙ ПРЕДМЕТНОЙ ОБЛАСТИ	133
для специалистов в научно-исследовательских проектах	
Рогушина Ю.В. , Гладун А.Я.	
ПОСЛЕДОВАТЕЛЬНОЕ ИЗЛОЖЕНИЕ МЕТОДА ДЛЯ ИНТЕГРАЦИИ OWL DL И SWRL С ИСПОЛЬЗОВАНИЕМ PROTÉGÉ-OWL API	139
Хала Е.А.	
ВЕБ-ОРИЕНТИРОВАННАЯ ПРОГРАММНАЯ СИСТЕМА АВТОМАТИЗАЦИИ РАЗРАБОТКИ БАЗ ЗНАНИЙ НА ОСНОВЕ ТРАНСФОРМАЦИИ КОНЦЕПТУАЛЬНЫХ МОДЕЛЕЙ	145
Дородных Н.О.	

The Enlargement of Electronic Lexical Database by Computational On-line Free System

Zanouka E.

United Institute of Informatics Problems
National Academy of Sciences,
Minsk, Republic of Belarus
Email: evgeniakacan@gmail.com

Abstract—The article describes the urgency of electronic vocabularies' update. It also depicts the principle of electronic Belarusian grammar dictionaries composing by means of computational program "Word Paradigm Generator" which is the main source of Belarusian lexical database enlargement for natural language processing in computational and linguistic platform "corpus.by" [2].

Keywords—computational linguistics, lexical databases, text and speech platform.

I. INTRODUCTION

Lexicography is a branch of linguistics which studies a process of dictionary compiling and dictionaries themselves. The lexicography accomplishes different functions: it describes normalized language, teach the language, provides interlanguage communication. But the main function of the brunch is the scientific study of vocabulary. A vocabulary is in a certain way organized collection of words, usually with the addition of notes, which provides information about the features of their structure and / or their function. Today many scholars pay attention to the new scientific field – computational lexicography for many reasons: it's a new way of compiling and collecting dictionaries. Electronic dictionary is a dictionary stored as machine (computer) data instead of being printed on paper. Many researches have pointed out that these dictionaries have a number of distinctive peculiarities In contrast with printed. Electronic format:

- 1) accelerates the process of finding a required vocabulary unit;
- 2) extends the capabilities of a unit description;
- 3) electronic format makes it possible;
- 4) allows adding extra information: additional illustrations, diagrams, pictures, use recording or live-sounding speech, etc;
- 5) allows adding extra information: additional illustrations, diagrams, pictures, use recording or live-sounding speech, etc;
- 6) gives an opportunity of quick renewal of new lexicology

In this paper the author represents the program for creating new electronic grammar dictionaries for the Belarusian language and demonstrates the mechanism of its functioning on the example of her own dictionary which composes 1500 lexemes.

II. WORD PARADIGM GENERATOR

Service Word Paradigm Generator[1] is a computational program which processes unknown words. It analyses the structure of a word, defines its possible grammatical features and outputs the whole paradigm and a tag of this word (see figure 1). The tag is an annotation of a word which describes all grammatical features of a word depending of a part of speech to which it belongs to. The necessity of the tags is the next: morphological information is widely used in scientific and normative grammars, as they simplify the description of the language and reflect its systemic nature. Paradigms of one part of speech in most cases have the same features, the same structure and a set of endings for the same type of declension or conjugation, combine sets of similar words stems, and often characterized by similar stresses and / or morphonological phenomena.

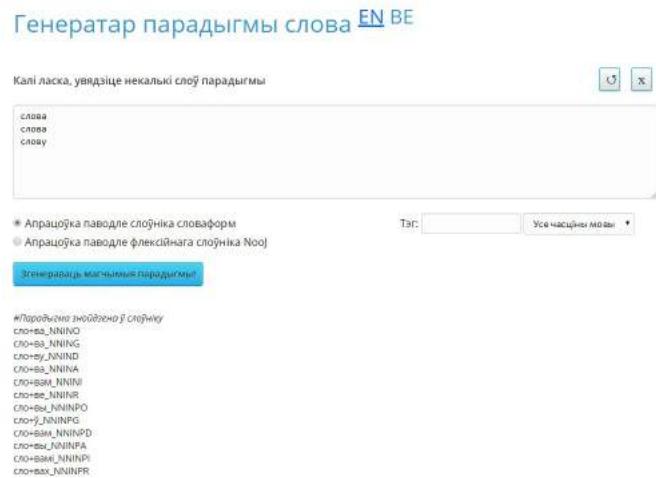


Figure 1. The lexeme which is found in the dictionary of the system

The system works on the principle of formal pattern: for a necessary word it chooses words from the database with the same grammatical characteristics and structure and outputs a list of the variants for this word. The user by himself chooses the right paradigm for this word. There are two main dictionaries which are built in the system: the dictionary of word forms and a dictionary of NooJ-format [4]. If any of these dictionaries contain a testing word, the system outputs the whole paradigm with tags (see figure 1). Otherwise it searches words corresponding to the query and outputs all

possible variants (see figure 2). In other words, the algorithm searches for the nearest paradigm(s) matching the last letters of the word, to which the user needs to get the paradigm. As it is the program for grammar dictionaries enrichment, the service processes only changeable parts of speech (noun, verb, adjective, participle, pronoun, numeral). For auxiliary parts of speech the system outputs their annotation [3].

Генератор парадигмы слова EN BE

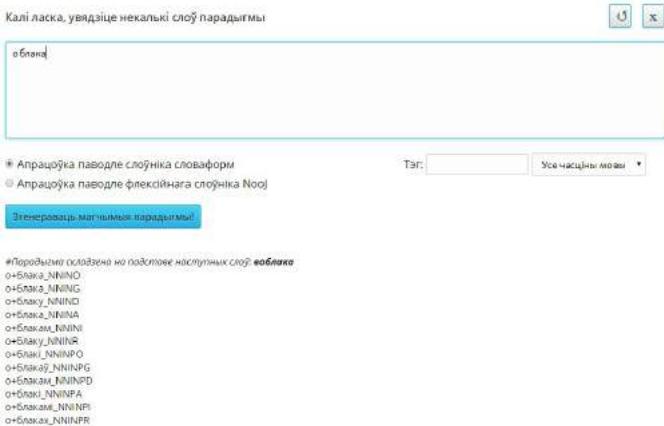


Рис. 2. The output of processing unknown word «Облака»

III. THE COMPOSITION OF LEXICAL DATABASE

Any dictionary is built according the next principles:

- 1) the nature of the dictionary;
- 2) the information they contain;
- 3) the language of the explanations;
- 4) an intended user.

While preparing our own Belarusian grammar dictionary, we have taken into account some other criteria for its compiling:

- 1) The meaning of lexical units upon which the lexicographer concentrates his attention presents very special difficulties of its own.
- 2) A lexicographer searches new words for the users whose needs are more practical.
- 3) Each entry must be self-contained, as we consult a single entry in a dictionary.
- 4) an intended user.

To correspond to all these points it was decided to analyze the vocabulary of the authors of 20th century both Belarusian and foreign (translations of foreign authors), such as Uladzimir Karatkevich (Dzikae paliavanne karalia Stakha), Vasil Bukaw (Zhurawliny kryk, Kar'er, Vawchynaia iama), Anton Adamovich (Tryvoaha), Hohal (Vechary na khutary bliz Dzikanki), Daniel Defo (Zhytstsio i dziwnyia pryhody marakhoda Rabinzona Kruza) and so on. As a result, it was examined nearly 600 pages of different books, where nearly 5 thousand of words are unknown for the system [2].

The author has chosen 1600 lexemes in a haphazard way and divided all these words into the next groups:

- 1) Non-obscene words (the dictionary contains these words).
- 2) The words that are absent in the dictionary of the system.

The main attention is paid to the last group as all these words are unknown, therefore they are the core of a new dictionary. They are also subdivided into:

- 1) Proper Nouns (Анеліна, Менск, Яніна, etc.);
- 2) Dialectisms (дамо, йшла, etc.);
- 3) Russisms (начаць, дзяйстві+цельны, палучацца etc.);
- 4) Authorial words (камчуннічайце, коця, дрыкганты, бомкнуща etc.);
- 5) Words written in Tarashkevitsa - unofficial variant of Belarusian orthography (разъбованымі, купальле, губэрнатар, госьць, дысэртацыя).

To fall under these criteria, we have created a Belarusian grammar dictionary which contains the paradigms of a lexeme with its annotation. The dictionary composes 1500 separate lexemes, including changeable and unchangeable parts of speech (see figure 3).

+		-
парсокотка_NNAMO	распляванне_NNINNO	разлегтися_VPLIFC
парсокотка_NNAMG	расплявання_NNINNG	разлагуся_VPLIF1
парсокутка_NNIND	расплявання_NNIND	разлажкіся_VPLIF2
парсокутка_NNINA	распляванне_NNINA	разлажкіся_VPLIF3
парсокутка_NNINI	распляваннем_NNINA	разлажкімся_VPLIF1P
парсокутка_NNINIR	распляваннім_NNINIR	разлажкімся_VPLIF2P
облаки_NNINNO	распляванні_НNINNO	разлажкі_ВPLIF3P
облакі_НNINNG	распляванні_НNINNG	разлажкі_ВPLIF2P
облакам_НNINRD	распляванні_НNINRD	разлажкі_ВPLIF1P
облакам_НNINPR	распляванні_НNINPR	разлажкі_ВPLIF1P
облакам_НNINPA	распляванні_НNINPA	разлажкі_ВPLIF2P
облакам_НNINPI	распляванніах_НNINPR	разлажкі_ВPLIF3P
облакам_НNINPFR	+	разлажкісь_ВPLIF2
облакам_НNINPFG	+	разлажкісь_ВPLIF1
облакам_НNINPFD	+	разлажкісь_ВPLIF2
облакам_НNINPF	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкімся_ВPLIF1P
облакам_НNINPFP	+	разлажкімся_ВPLIF2P
облакам_НNINPFP	+	разлажкімся_ВPLIF3P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF3P
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF1
облакам_НNINPFP	+	разлажкісь_ВPLIF2
облакам_НNINPFP	+	разлажкісь_ВPLIF3
облакам_НNINPFP	+	разлажкі_ВPLIF2P
облакам_НNINPFP	+	разлажкі_ВPLIF1P
облакам_НNINPFP	+	

3) A user needs a manual editing for verification.

In the process of composing new grammar electronic dictionary the author has used Word Paradigm Generator as the main system which processes unknown words and on their basis generates grammar dictionaries. It was marked some more peculiarities of the program functioning which is necessary to correct. They are presented below. The system:

- 1) gives an incomplete verb paradigm (the future or present tense form is absent, depending on the verbal aspect + not all forms of participle) (see figure 4);
- 2) gives an incomplete paradigm of the adjective (without short and comparative forms) (see figure 4);
- 3) outputs participle II paradigms (in Belarusian “дзеепрыслоўе”, which is unchangeable part of speech and has only aspect);
- 4) outputs a paradigm for unchangeable nouns (for example, са+ні, ка+ва, etc.);
- 5) outputs an incomplete paradigm of female nouns with endings -сь, -сыць);
- 6) outputs a full paradigm for complex words, where the first word is unchangeable and the second is changeable, and doesn't output a paradigm for both changeable parts of a complex word;
- 7) doesn't output a paradigm for changeable parts of speech in Tarashkevitsa.

There are also some mistakes with tags:

- 1) It outputs tags of common nouns for proper nouns.
- 2) Very often it outputs tags of participle I (in Belarusian “дзеепрыметнік”) for adjectives and vice versa.
- 3) It can confuse verb aspects, therefore output incorrect tag for verbs.

+ <u>цярпель_VPC</u> <u>цярплю_VPIF1</u> <u>церпш_VPIF2</u> <u>цярпім_VPIF2</u> <u>церпів_VPIF3</u> <u>цярпінь_VPIF3</u> <u>церпім_VPIF1P</u> <u>цярпім_VPIF1P</u> <u>церпіне_VPIF2P</u> <u>церпіне_VPIF2P</u> <u>церпіць_VPIF3P</u> <u>цярпяць_VPIF3P</u> <u>цярпі_VPM2</u> <u>цярпіце_VPM2P</u> <u>цярпей_VPIPM</u> <u>цярпела_VPIPF</u> <u>цярпела_VPIPN</u> <u>цярпелі_VPIJ</u> <u>цярпейшы_VPB</u>	+ <u>бязглуздны JJMO</u> <u>бязглузднага JJMG</u> <u>бязглузднаму JJMD</u> <u>бязглуздны JJMA</u> <u>бязглузднага JJMU</u> <u>бязглуздным JJMA</u> <u>бязглуздным JJMR</u> <u>бязглузднай JJFO</u> <u>бязглузднай JJFG</u> <u>бязглузднай JJFD</u> <u>бязглуздную JJFA</u> <u>бязглузднай JJFA</u> <u>бязглузднало JJFS</u> <u>бязглузднай JJFR</u> <u>бязглузднае JINO</u> <u>бязглузднага JJNG</u> <u>бязглузднаму JJND</u> <u>бязглузднай JJNA</u> <u>бязглуздным JJNA</u> <u>бязглуздным JJNR</u> <u>бязглуздныя JJPO</u> <u>бязглуздных JPG</u> <u>бязглуздным JPD</u> <u>бязглуздныя JPA</u> <u>бязглуздных JPU</u> <u>бязглуздным JPA</u> <u>бязглуздных JPR</u>
--	--

Рис. 4. The output of verb “Цярпець”(on the left) and the adjective “Бязглуздны” (on the right)

Among this list of program disadvantages we should distinguish two main points: the system problems, connecting with program incompleteness, and problems, connecting with unfilled database of the whole system which works only on the basis of one dictionary of word forms. All in all, the system works on 86% and this is a rather high level [3]. In future it is planned both to improve the capability of the program on the basis of this research and enlarge the lexical database of corpus.by.

V. CONCLUSION

The article covers the problem of electronic Belarusian grammar dictionaries composing by means of electronic program “Word Paradigm Generator” on the example of electronic lexical database consisting of 1500 lexemes. It is a little dictionary which contains different lexical groups including unofficial vocabulary. It provides two types of information: grammatical (the whole word paradigm) and morphological (a word annotation). Word Paradigm Generator is described as the means of electronic grammar dictionaries enrichment. The mechanism and principles of its functioning are also pointed out as well as all its deficiencies which need to be improved.

REFERENCES

- [1] Word Paradigm Generator [Электронны рэсурс]. — 2016. Рэжым доступу: <http://corpus.by/WordParadigmGenerator/>. — Дата доступу: 17/07/2016.
- [2] Камп'ютарная платформа для апрацоўкі тэксту і маўлення // <http://corpus.by/> [Электронны рэсурс]. — 2016. Рэжым доступу: <http://corpus.by/index.php?lang=be>. — Дата доступу: 16/12/2016.
- [3] Гецэвіч, Ю.С. Інтэрнэт-сістэма генерацыі парадыгмаў слова для папаўнення электронных граматычных слоўнікаў / Ю.С. Гецэвіч, В.В. Варановіч, С.І. Лысы, І.В. Рэентовіч, Я.С. Качан // Міжнародны конгрэс па інфарматыке: інформацыйныя сістэмы і тэхнологіі=International Congress on computer science: Information systems and technologies / БГУ; под ред. С.В. Абламейко. — Мінск, 2016. — С. 584-588.
- [4] Hetsevich, Yu. SEMI-AUTOMATIC PART-OF-SPEECH ANNOTATING FOR BELARUSIAN DICTIONARIES ENRICHMENT IN NOOJ / Yu. Hetsevich, V. Varanovich, E. Kachan [et al.] // NOOJ 2016 International Conference - Book of Abstracts (6-9 June, 2016, Czech Republic) / University of South Bohemia; ed. Jan Radimsky. — Ceske Budejovice, 2016. — P. 47.

ПОПОЛНЕНИЕ ЭЛЕКТРОННОЙ ЛЕКСИЧЕСКОЙ БАЗЫ ДАННЫХ ЧЕРЕЗ КОМПЬЮТЕРНУЮ СИСТЕМУ "ГЕНЕРАТОР ПАРАДИГМЫ СЛОВА"

Зеновко Е.С.

Данная статья описывает проблему создания и пополнения электронных словарей. Она также отражает принцип пополнения электронных белорусских грамматических словарей с помощью компьютерного сервиса в открытом доступе "Генератор парадигмы слова". Программа является основным источником расширения белорусской лексической базы данных для обработки естественного языка в области компьютерной и лингвистической платформы "corpus.by"[2].

Введение

Лексикография является разделом лингвистики, изучающим процесс составления словарей. Лексикография выполняет различные функции: она описывает нормированный язык, обучает языку, обеспечивает процесс коммуникации. Но основная ее функция - научное изучение лексики. Сегодня многие ученые обращают внимание на новое научное направление - компьютерную лексикографию по многим причинам: это новый способ сбора и пополнения словарей. Многие исследователи отмечают, что электронный формат словарей имеет ряд отличительных особенностей, таких как ускорение процесса поиска необходимой единицы словаря, расширение возможности описания единиц словаря, добавление дополнительной информации по словарных единицах, быстрое пополнение словарей и т.д. В данной статье автор представляет программу для создания новых электронных белорусскоязычных грамматических словарей и демонстрирует механизм его функционирования на примере своего собственного словаря, который составляет 1500 лексем.

Описание программы «Генератор парадигмы слова»

В данной главе описывается сервис "Генератор Парадигмы Слово"[1], который представляет собой компьютерную онлайн программу, которая обрабатывает неизвестные слова. Она анализирует структуру слова, определяет его возможные грамматические особенности и выводит всю парадигму и тэг этого слова. Система работает по принципу формального шаблона: для необходимого слова он выбирает из базы данных слова с одинаковыми грамматическими характеристиками и структурой и выводит список вариантов для этого слова.

Принцип и механизм создания лексической базы данных

В данном разделе описывается процесс подбора лексики для белорусскоязычного грамматического словаря, демонстрируются его особенности и механизм пополнения через Генератор Парадигмы Слова.

Особенности функционирования «Генератора парадигмы слова»

Третья падглава отражает особенности функционирования онлайн ресурса "Генератор Парадигмы Слова описывает все недостатки программы, которые должны быть исправлены, и пути их исправления.

Заключение

В статье рассматривается проблема составления электронных белорусских грамматических словарей, которые могут пополняться с помощью электронной программы "Генератор Парадигмы Слова". Данный механизм демонстрируется на примере электронной лексической базы данных, которая состоит из 1500 лексем. Это небольшой словарь, содержащий различные лексические группы слов, в том числе ненормативную лексику. Он дает два типа информации: грамматическую (вся парадигма слова) и морфологическую (аннотацию слова). Генератор Парадигмы слово описывается как средство электронного обогащения грамматических словарей. Отмечены механизм и принципы его функционирования, а также все недостатки, которые должны быть улучшены.