

**Белорусский государственный университет
Факультет социокультурных коммуникаций**

Кафедра компьютерной лингвистики и лингводидактики

**Реферат дипломной работы
«Обработка данных на белорусском языке, полученных из
социальных сетей»**

**Гизун Артур Александрович,
Руководитель Гецевич Юрий Станиславович**

2018 год

РЕФЕРАТ

Дипломная работа включает 50 страниц, 23 рисунка и основана на 30 источниках.

Ключевые слова: веб–скрейпинг, электронные словари, социальные сети, частотные словари, контекстологические словари,

Цель – формирование алгоритма по созданию белорусскоязычных частотных контекстологических словарей, на основе большого количества данных полученных из социальных сетей.

Объект исследования – частотные контекстологические словари

Предмет исследования – процесс формирования частотных контекстологических словарей.

Актуальность исследования обусловлена большим количеством необработанных лингвистических данных, которые содержат в себе социальные сети и, также отсутствием подобных исследований белорусскоязычного сегмента интернета.

В процессе исследования применялись такие **методы** как эксперимент, синтез, наблюдение, обобщение, сравнительный анализ.

Теоретическая новизна работы состоит в разработке методов и алгоритмов создания белорусскоязычных словарей на основе данных полученных из социальных сетей.

Практическая значимость дипломной работы состоит в создании белорусскоязычных частотных контекстологических словарей с помощью которых можно решать большой спектр задач от обучения до решения проблемы многозначности слов в переводе. Материалы дипломной работы были апробированы на 74-ой научной конференции студентов и аспирантов БГУ, Минск, 18.05.2017, II Международной научно-практической конференции «Лингвистика, лингводидактика, лингвокультурология: актуальные вопросы и перспективы развития», Минск, 1-2 марта 2018, 75-ой научной конференция студентов и аспирантов БГУ, Минск, 17.05.2018.

Автор работы подтверждает, что приведенный в работе материал объективно отражает состояние исследуемого процесса, а все заимствования из источников сопровождаются ссылками на их авторов.

подпись студента

РЭФЕРАТ

Дыпломная работа ўключае 50 старонак, 23 малюнка і заснавана на 30 крыніцах.

Ключавыя словы: вэб–скрэіпінг, электронныя слоўнікі, сацыяльныя сеткі, частотныя слоўнікі, кантэксталагічныя слоўнікі,

Мэта – фарміраванне алгарытму па стварэнні беларускамоўных частотных кантэксталагічных слоўнікаў, на аснове вялікай колькасці дадзеных, атрыманых з сацыяльных сетак.

Аб'ект даследавання – частотныя кантэксталагічныя слоўнікі.

Прадмет даследавання – працэс фарміравання частотных контексталагічных слоўнікаў.

Аktуальнасць даследавання абумоўлена вялікай колькасцю неапрацаваных лінгвістычных дадзеных, якія ўтрымліваюць у сабе сацыяльныя сеткі і, таксама адсутнасцю падобных даследаванняў беларускамоўнага сегмента інтэрнэту.

У працэсе даследавання ўжываліся такія метады, як эксперымент, сінтэз, назіранне, абагульненне, параўнальны аналіз.

Тэарэтычная навізна працы складаецца ў распрацоўцы метадаў і алгарытмаў стварэння беларускамоўных слоўнікаў на аснове атрыманых дадзеных з сацыяльных сетак.

Практычная значнасць дыпломнай работы складаецца ў стварэнні беларускамоўных частотных кантэксталагічных слоўнікаў з дапамогай якіх можна вырашаць вялікі спектр задач ад навучання да вырашэння праблемы шматзначнасць слоў у перакладзе. Матэрыялы дыпломнай працы былі апрабаваны на 74-й навуковай канферэнцыі студэнтаў і аспірантаў БДУ, Мінск, 18.05.2017, II Міжнароднай навукова-практычнай канферэнцыі «Лінгвістыка, лінгвадыдактыка, лінгвакультуралогія: актуальныя пытанні і перспектывы развіцця», Мінск, 1-2 сакавіка 2018, 75-й навуковая канферэнцыя студэнтаў і аспірантаў БДУ, Мінск, 17.05.2018.

Аўтар работы пацвярджае, што прыведзены у рабоце матэрыял аб'ектыўна адлюстроўвае стан доследнага працэсу, а ўсе запазычаныя з крыніц суправаджаюцца спасылкамі на іх аўтараў.

подпіс студэнта

ABSTRACT

The thesis includes 50 pages, 23 images and it is based on 30 literary source.

Keywords: web scraping, electronic dictionaries, social networks, frequency dictionaries, contextual dictionaries,

The goal is to develop an algorithm for the creation of Belarusian-language frequency contextual dictionaries based on a large amount of data obtained from social networks.

The object of research is frequency contextual dictionaries.

The subject of research is the process of formation of frequency contextual dictionaries.

The relevance of the study is due to the large number of raw linguistic data, which contain social networks and the lack of such studies of the Belarusian-language segment of the Internet.

In the course of research were used such methods as experiment, synthesis, observation, generalization, and comparative analysis.

The theoretical novelty of the work consists in the development of methods and algorithms for the creation of Belarusian-language dictionaries based on the data obtained from social networks.

The practical significance of the thesis is to create Belarusian-language frequency contextual dictionaries with which you can solve a wide range of problems from learning to solving the problem of multiple words in translation. Materials of the thesis were tested at the 74th scientific conference of students and postgraduates of the Belarusian state University, Minsk, 18.05.2017, II International scientific-practical conference "Linguistics, language pedagogy, cultural linguistics: current issues and prospects of development", Minsk, 1-2 March 2018, the 75th scientific conference of students and postgraduates of the Belarusian state University, Minsk, 17.05.2018.

The author of the work confirms that the material presented in the work objectively reflects the state of the process under study, and all borrowings from sources are accompanied by references to their authors.

student's signature