

**Белорусский государственный университет
Факультет социокультурных коммуникаций**

Кафедра компьютерной лингвистики и лингводидактики

**Реферат дипломной работы
«ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ МАШИННОГО ПЕРЕВОДА
НА БАЗЕ НЕЙРОННЫХ СЕТЕЙ»**

**Черетун Ксения Александровна,
Руководитель Воронович Валерий Викторович**

2019 год

РЕФЕРАТ

дипломной работы Черетун К.А.

ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ МАШИННОГО ПЕРЕВОДА НА БАЗЕ НЕЙРОННЫХ СЕТЕЙ

Дипломная работа включает 93 страницы, 14 рисунков, 6 приложений и основана на 45 источниках.

Ключевые слова: МАШИННЫЙ ПЕРЕВОД, НЕЙРОННЫЕ СЕТИ, КОРПУС ТЕКСТОВ, ПАРАЛЛЕЛЬНЫЙ КОРПУС, ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ, АВТОМАТИЗИРОВАННЫЙ ПЕРЕВОД.

Цель исследования – подготовка языкового материала для составления параллельного корпуса текстов, который будет служить предметом обучения системы русского-белорусского нейронного машинного перевода.

Объект исследования – машинный перевод на базе нейронных сетей.

Предмет исследования – параллельный корпус текстов юридического дискурса, предназначенный для обучения системы нейронного машинного перевода.

Актуальность исследования обусловлена отсутствием качественной системы русско-белорусского юридического перевода и необходимостью разработки лингвистического обеспечения для создания подобной системы.

В процессе исследования применялись такие **методы** как описание, анализ, обобщение, формализация.

Научная новизна состоит в том, что в данной работе впервые на основании практического исследования были сформулированы основные особенности работы систем русско-белорусского машинного перевода.

Практическая значимость дипломной работы заключается в реализации лингвистического обеспечения системы машинного перевода на нейронных сетях.

Результаты исследования были апробированы на 75-ой научной конференции студентов и аспирантов БГУ, Минск, 17.05.2018 (выступление с докладом «Использование нейронных сетей в программах лингвистической обработки текстов»); 76-ой научной конференции студентов и аспирантов БГУ, Минск, 16.05.2019 (выступление с докладом «Лингвистическое обеспечение машинного перевода на базе нейронных сетей»).

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

дыпломнай працы Чаратун К.А.

ЛІНГВІСТЫЧНАЕ ЗАБЕСПЯЧЭННЕ МАШЫННАГА ПЕРАКЛАДУ НА БАЗЕ НЕЙРОНАВЫХ СЕТАК

Дыпломная праца ўключае 93 старонкі, 14 малюнкаў, 6 дадаткаў і заснавана на 45 крыніцах.

Ключавыя словы: МАШЫННЫ ПЕРАКЛАД, АЎТАМАТЫЧНЫ ПЕРАКЛАД, НЕЙРОНАВЫЯ СЕТКІ, КОРПУС ТЭКСТАЎ, ПАРАЛЕЛЬНЫ КОРПУС, ЛІНГВІСТЫЧНАЕ ЗАБЕСПЯЧЭННЕ.

Мэта даследавання – падрыхтоўка моўнага матэрыялу для складання паралельнага корпуса тэкстаў, які будзе служыць прадметам навучання сістэмы руска-беларускага нейронавага машыннага перакладу.

Аб'ект даследавання – машынны пераклад на базе нейронавых сетак.

Прадмет даследавання – паралельны корпус тэкстаў юрыдычнага дыскурсу, прызначаны для навучання сістэмы нейронавага машыннага перакладу.

Аktуальнасць даследавання абумоўлена адсутнасцю якаснай сістэмы руска-беларускага юрыдычнага перакладу і неабходнасцю распрацоўкі лінгвістычнага забеспячэння для стварэння такой сістэмы.

У працэсе даследавання выкарыстоўваліся такія **метады** як апісанне, аналіз, абагульненне, фармалізацыя.

Навуковая навізна заключаецца ў тым, што ў дадзенай працы ўпершыню на падставе практычнага даследавання былі сфармуляваны асноўныя асаблівасці працы сістэм руска-беларускага машыннага перакладу.

Практычная значнасць дыпломнай працы заключаецца ў рэалізацыі лінгвістычнага забеспячэння сістэмы машыннага перакладу на нейронавых сетках.

Вынікі даследавання былі апрабаваны на 75-ай навуковай канферэнцыі студэнтаў і аспірантаў БДУ, Мінск, 17.05.2017 (выступ з дакладам «Выкарыстанне нейронавых сетак у праграмах лінгвістычнай апрацоўкі тэкстаў»); 76-ай навуковай канферэнцыі студэнтаў і аспірантаў БДУ, Мінск, 16.05.2018 (выступ з дакладам «Лінгвістычнае забеспячэнне машыннага перакладу на базе нейронавых сетак»).

Аўтар працы пацвярджае, што прыведзены ў працы матэрыял правільна і аб'ектыўна адлюстроўвае стан даследнага працэсу, а ўсе запазычаныя з літаратурных і іншых крыніц суправаджаюцца спасылкамі на іх аўтараў.

_____ (подпіс студэнта)

SUMMARY

of diploma paper by Kseniya Cheretun

LINGUISTIC SOFTWARE OF MACHINE TRANSLATION ON THE BASIS OF NEURAL NETWORKS

The diploma paper includes 93 pages, 14 illustrations, 6 appendixes; it is based on 45 literary sources.

Keywords: MACHINE TRANSLATION, NEURAL NETWORKS, TEXT CORPUS, PARALLEL CORPUS, LINGUISTIC SOFTWARE, AUTOMATED TRANSLATION.

The objective is to prepare language material for the compilation of a parallel corpus of text that will serve as the subject of training for the system of Russian-Belarusian neural machine translation.

The research object is machine translation based on neural networks.

The research subject is the parallel corpus of the text of legal discourse, designed for the training of the neural machine translation system.

The relevance of the study is due to the lack of a high-quality system of Russian-Belarusian legal translation and the need to provide linguistic support for the creation of such a system.

In the process of research, such **methods** as description, analysis, generalization, formalization were used.

The scientific novelty consists in the fact that for the first time in this work, on the basis of practical research, the main features of the work of the Russian-Belarusian machine translation systems were formulated.

The practical significance consists in the implementation of linguistic software for the machine translation system on neural networks.

Research results were approved at the 75th Scientific Conference of Students and Postgraduate of BSU, Minsk, 17.05.2017 (presentation of the report «The use of neural networks in the programs of linguistic processing of texts»); 76th Scientific Conference of Students and Postgraduate of BSU, Minsk, 16.05.2018 (presentation of the report «Linguistic software of machine translation based on neural networks»).

The author of the paper confirms that the material correctly and objectively reflects the state of the investigated process and all borrowed theoretical, methodological and methodical concepts from literature and other sources are accompanied by links to their authors.

(signature of the student)

