

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к магистерской диссертации

«Алгоритмы машинного перевода юридических текстов на белорусский язык на основе нейронных сетей»

Сапач Яна Петровна

Научный руководитель – кандидат технических наук,
доцент Гецевич Ю.С.

Минск, 2020

Реферат

Магистерская диссертация, 44 с., 11 рис., 12 табл., 21 источник

НЕЙРОННЫЙ МАШИННЫЙ ПЕРЕВОД, НЕЙРОННАЯ СЕТЬ, ПАРАЛЛЕЛЬНЫЙ КОРПУС ТЕКСТОВ, ПЕРЕВОД НА БЕЛОРУССКИЙ ЯЗЫК, ЮРИДИЧЕСКИЙ ДОМЕН

Объект исследования – задачи перевода текстов юридической тематики, таких как кодексы Республики Беларусь, с русского языка на белорусский и связанные с рассматриваемой задачей подготовительные этапы обработки текста и нейросетевые модели для перевода.

Цель работы – построение нейросетевого переводчика текстов юридической тематики для белорусского языка.

Методы исследования – анализ проблематики и существующих подходов, эксперимент, тестирование и сравнение.

Результат – определены и изучены основные трудности, связанные с нейросетевым переводом, переводом с русского на белорусский язык, переводом текстов юридической тематики; подготовлены данные для тренировки моделей машинного обучения; подготовлены модели для разделения слов на более мелкие частицы и для проведения обратного преобразования; подготовлена и обучена нейросетевая модель перевода; проведён сравнительный анализ экспериментов с разными подходами к обработке текстов; создано программное обеспечение для взаимодействия с результатами модели.

Область применения – прикладные задачи, связанные с переводом юридических текстов, в первую очередь кодексов Республики Беларусь.

Abstract

Master thesis, 44 p., 11 figures, 12 tables, 21 sources

NEURAL MACHINE TRANSLATION, NEURAL NETWORK, PARALLEL CORPUS, TRANSLATION TO THE BELARUSIAN LANGUAGE, LEGAL DOMAIN

Object of research – the problem of translating texts of the legal subject, such as codes of the Republic of Belarus, from Russian language into Belarusian and associated with the task text processing stages and neural network models.

Research goal – development a neural network translator of legal texts for the Belarusian language.

Research methods – the analysis of problems and existing approaches, experiment, testing and comparison.

Result – the main difficulties associated with the neural network translation, translation from Russian to Belarusian, translation of legal texts were identified and studied; data for training machine learning models was prepared; neural network translation model was prepared and trained; a comparative analysis of experiments with different approaches to word processing was carried out; developed and implemented application for interaction with the model.

Application field – applied tasks related to the translation of legal texts, primarily codes of the Republic of Belarus.