

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ
БЕЛАРУСЬ**

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

**ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И
ИНФОРМАТИКИ**

Кафедра биомедицинской информатики

Аннотация к дипломной работе

**Разработка алгоритма обработки и анализа коллекции
бизнес-документов с учетом семантического сходства**

Гаргома Артём Олегович

Научный руководитель — доцент, кандидат технических наук
Свирид Ю.В.

Минск, 2022

РЕФЕРАТ

Дипломная работа, 37 страниц, 11 источников.

Ключевые слова: ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА, ВОПРОСНО-ОТВЕТНЫЕ СИСТЕМЫ, МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ, ТРАНСФОРМЕРЫ, МЕХАНИЗМ ВНИМАНИЯ, ЮРИДИЧЕСКИЕ ДОГОВОРА

Объект исследования: вопросно-ответные системы, модели трансформера

Цель работы: изучение и обучение вопросно-ответной системы на юридических договорах

Результат: Проведен предварительный анализ данных юридических договоров. Рассмотрена область вопросно-ответных систем, их основные проблемы, метрики для оценки качества моделей. Изучены архитектуры трансформера и BERT, выявлены основные недостатки BERT и описано, как другие модели разрешают их. Проведен эксперимент по оценке качества моделей с большей длиной входных представлений на юридических договорах. Разобрана, обучена и оценена модель Longformer.

Область применения: вопросно-ответные системы в программах для работы с большими объемами текстовой информации, голосовых помощниках, чатботах и т.д.

РЭФЕРАТ

Дыпломная праца, 37 старонка, 11 крыніц.

Ключавыя словы: АПРАЦОЎКА НАТУРАЛЬнай МОвы, Пытанне-адказныя сістэмы, мадэлі машыннага навучання, трансфармеры, механізм увагі, юрыдычныя дагаворы

Аб'ект даследавання: пытанне-адказныя сістэмы, мадэлі трансфармера

Мэта работы: вывучэнне і навучанне пытанне-адказнай сістэмы па юрыдычных дагаворах

Вынік: Праведзены папярэдні аналіз дадзеных юрыдычных дагаворах. Разгледжана вобласць пытанне-адказных сістэм, іх асноўныя праблемы, метрыкі для ацэнкі якасці мадэляў. Вывучаны архітэктурны трансформера і BERT, выяўлены асноўныя недахопы BERT і апісана, як іншыя мадэлі дазваляюць іх. Праведзены эксперымент па ацэнцы якасці мадэляў з большай даўжынёй уваходных уяўленняў на юрыдычных дагаворах. Разабрана, навучана і ацэнена мадэль Longformer.

Вобласць прымянення: пытанне-адказныя сістэмы ў праграмах для працы з вялікімі аб'ёмамі тэкставай інфармацыі, галасавых памочніках, чатботах і г.д.

ABSTRACT

Diploma thesis, 37 pages, 11 sources.

Keywords: NATURAL LANGUAGE PROCESSING, QUESTION-ANSWER SYSTEMS, MACHINE LEARNING MODELS, TRANSFORMERS, ATTENTION MECHANISM, LEGAL CONTRACTS

Object of research: question-answer systems, transformer models

Objective: study and training of a question-answer system in legal contracts

The result: A preliminary analysis of these legal agreements was carried out. The area of question-answer systems, their main problems, metrics for assessing the quality of models is considered. The architectures of the transformer and BERT have been studied, the main disadvantages of BERT have been identified and it has been described, how other models resolve them. An experiment was carried out to assess the quality of models with longer input representations on legal contracts. Reviewed, trained and evaluated the Longformer model

The scope: question-answer systems in programs for working with large amounts of textual information, voice assistants, chatbots, etc.