

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.032.26:004.934

Гаруля
Дмитрий Владимирович

Алгоритмы формирования документооборота на основе
рекуррентных нейронных сетей

АВТОРЕФЕРАТ

Диссертация на соискание академической степени магистра
технических наук
по специальности 1-40 80 02 Системный анализ, управление и
обработка информации

Магистрант Д. В. Гаруля

Научный руководитель
А. А. Навроцкий кандидат физико-
математических наук, доцент

Минск 2022

КРАТКОЕ ВВЕДЕНИЕ

В годы зарождения документооборота, он представлял собой движение бумажных документов, как между сотрудниками компаний, так и между внешними контрагентами.

Переход от традиционного бумажного документооборота к безбумажным технологиям, который наблюдается в настоящее время, не только радикально увеличивает скорость бизнес-процессов компаний и обеспечивает экономию средств, при отказе от бумажных носителей, но и скрывает в себе потенциальный эффект – возможность применения технологий искусственного интеллекта для минимизации рутинных операций в обработке электронных документов.

При переходе на электронный документооборот в информационной системе накапливается информация об уже реализованных сценариях обработки информации: данные о типовых операциях, выполняемых пользователями системы при обработке документов, решениях принимаемых участниками бизнес-процессов на основании содержимого документов и другой информации, сопровождающей бизнес-процессы. Важной особенностью этой информации, в отличие от аналогичной при бумажном документообороте является то, что она доступна для машинной обработки без дополнительных модификаций и может служить исходным материалом для применения технологий искусственных нейронных сетей.

Бурное развитие технологий искусственного интеллекта, и машинного обучения в частности, сделало доступным и относительно недорогостоящим применение этих технологий в создании специализированных решений. Именно они обеспечат максимальную эффективность работы с документами и с лёгкостью избавят сотрудников от рутинных операций.

Технологии распознавания и поиска документов уже не являются чем-то новым и весьма распространены, но в последнее время они приобрели новое звучание, связанное с развитием технологий искусственного интеллекта. Если традиционные задачи распознавания сводились к переводу в цифровой вид отдельных букв и символов – полнотекстовому распознаванию, и пределом был разбор их семантики, исходя из привязки наборов символов к определенным позициям в бумажной форме документа (распознавание форм), то в настоящее время системы искусственных нейронных сетей позволяют выделять отдельные смысловые данные из документа не в соответствии с привязкой к позиции в тексте документа, а в соответствии с их смысловой нагрузкой.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Объект исследования – система формирования документооборота.
Предмет исследования – алгоритмы поиска информации в системах формирования документооборота.

Цель данной работы заключается в создании системы поиска информации в юридических документах.

Спецификой юридических документов является то, что их длина превышает длину обработки контекста большинства существующих алгоритмов обработки естественного языка. Кроме того, актуальной задачей является обеспечение быстрого действия без использования графических процессоров.

Задачи исследования:

- поиск и изучение существующих алгоритмов обработки естественного языка;
- изучение алгоритмов обработки длинного контекста документов;
- оптимизация алгоритмов поиска информации для обеспечения достаточного быстрого действия без использования графических процессоров;

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Общий объем магистерской диссертации составляет 49 страниц, включая две таблицы, 30 рисунков, библиографический список из 25 наименований, одно приложение.

Магистерская диссертация посвящена разработке алгоритма формирования документооборота на основе рекуррентной нейронной сети.

Введение содержит исходные данные для разработки темы, дается обоснование актуальности темы магистерской диссертации.

В первой главе проведен обзор архитектур нейронных сетей для обработки естественного языка. Выявлены наиболее эффективные архитектуры нейронных сетей для поиска информации.

Во второй главе рассмотрены методы обработки длинного контекста в сетях с архитектурой *Transformer*, которые являются наиболее эффективными в настоящий момент.

В третьей главе представлена разработанная модель поиска информации на основе архитектуры *ERNIE-DOC*. Исследования подкрепляются представленными данными обучения нейронных сетей различных архитектур наиболее подходящих для решения данных задач.

В заключении кратко изложены полученные результаты.

ЗАКЛЮЧЕНИЕ

В диссертации представлена модель поиска информации на основе нейросети с архитектурой *ERNIE*, с модифицированным по методу *DeFormer* алгоритмом ввода информации.

В ходе работы над диссертацией были разработан алгоритм формирования документооборота на основе рекуррентной нейронной сети. На его основе была создана система поиска информации в юридических документах, учитывающая специфические особенности таких документов. Было обеспечено высокое быстродействие работы алгоритмов без использования графических процессоров.

Для разработки алгоритма выбрана архитектура *Transformer*, как наиболее актуальная и активно развивающаяся в настоящее время. В качестве модели, подходящей для обработки длинного контекста документов была выбрана модель *ERNIE*, одна из разновидностей архитектуры *Transformer*, позволяющая работать с контекстом неограниченной длины.

Для оптимизации работы алгоритма поиска информации сеть *ERNIE* была модифицирована по методу *DeFormer*, заключающемся в отдельной обработке запроса и контекста с кэшированием выходов n первых слоев нейросети для контекста.

Предложенный алгоритм позволил увеличить быстродействие системы поиска примерно в 4 раза. К достоинствам алгоритма следует отнести то, что исходный текст юридических документов может быть предоставлен только в виде кэшированных выходов n -го слоя нейросети и таким образом будет скрыт от пользователей и самой поисковой системы, что препятствует утечкам служебной информации.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

[1] Алгоритмы стохастического градиентного спуска обучения и тренировки нейронных сетей / Белошедов Е. С. [и др.] // Информационные технологии и системы 2020 (ИТС 2020) = *Information Technologies and Systems 2020 (ITS 2020)* : материалы международной научной конференции, Минск, 18 ноября 2020 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : Л. Ю. Шилин [и др.]. – Минск, 2020. – С. 151–152.

[2] Сравнительный анализ разновидностей роботов для автоматизации пищевого производства / Архипенко Я. С. [и др.] // Информационные технологии и системы 2020 (ИТС 2020) = *Information Technologies and Systems 2020 (ITS 2020)* : материалы международной научной конференции, Минск, 18 ноября 2020 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : Л. Ю. Шилин [и др.]. – Минск, 2020. – С. 24–25.

[3] Гаруля, Д. В. Методы оптимизации на основе градиентного спуска с моментом в задачах обучения нейронных сетей / Гаруля Д. В., Навроцкий А. А., Белошедов Е. С. // Информационные технологии и системы 2020 (ИТС 2020) = *Information Technologies and Systems 2020 (ITS 2020)* : материалы международной научной конференции, Минск, 18 ноября 2020 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : Л. Ю. Шилин [и др.]. – Минск, 2020. – С. 163–164.

[4] Гаруля, Д. В. *Equilibrium Propagation* как алгоритм обучения аналоговых нейросетей / Д. В. Гаруля // Информационные технологии и управление : материалы 57-ой научной конференции аспирантов, магистрантов и студентов по направлению 2, Минск, 19-23 апреля 2021 года / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск, 2021. – С. 89–90.