

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

УДК 004.822

Кушнеревич  
Павел Михайлович

**ФОРМАЛЬНО-ЯЗЫКОВОЙ АППАРАТ ДЛЯ ФОРМИРОВАНИЯ  
ЗАПРОСОВ К СЕМАНТИЧЕСКОЙ МОДЕЛИ**

**АВТОРЕФЕРАТ**

по специальности 1-40 80 01 – Компьютерная инженерия (Хранение и  
обработка данных)

Научный руководитель  
Герман Ю.О.  
к.т.н., доцент

Минск 2023

## ВВЕДЕНИЕ

С развитием интернет-технологий и глобализацией, электронная коммерция, или е-коммерция, стала ключевым фактором в мировой экономике. По данным статистического агентства Statista, в 2022 году объем розничных продаж через интернет составил 4,9 триллиона долларов, и ожидается, что к 2025 году этот показатель достигнет 7,4 триллиона долларов [1]. Это свидетельствует о неослабевающем росте и популярности е-коммерции среди потребителей и компаний.

В условиях жесткой конкуренции на рынке, качество содержания продукта становится основным фактором, который влияет на успех продаж и лояльность клиентов. Исследования показывают, что 87% потребителей считают описание продукта одним из самых важных аспектов при принятии решения о покупке [2]. Кроме того, наличие хорошо структурированного и информативного описания продукта способствует улучшению позиций в поисковых системах и увеличению охвата аудитории.

Создание качественного и информативного описания продукта является сложным и трудоемким процессом, который требует определенных знаний и навыков от контент-менеджеров. В то же время, на рынке существует большое количество систем управления контентом (CMS), таких как Shopify, Magento и других, которые облегчают создание и редактирование описаний продуктов.

Однако с развитием искусственного интеллекта (ИИ) и машинного обучения, в частности, таких программных средств, как ChatGPT, возникает возможность автоматизировать процесс создания описаний продуктов. Эти модели способны генерировать качественный текст, но для получения желаемого результата требуется точная формулировка запроса, что может вызвать сложности для пользователей без опыта работы с такими моделями.

Согласно исследованию, проведенному компанией Nielsen Norman Group, пользователи совершают ошибки при вводе запросов в поисковые системы в 70% случаев [3]. Это подчеркивает необходимость разработки инструментов, которые помогут пользователям формулировать правильные запросы и улучшить результаты генерации текста с помощью ИИ.

В этом контексте разработка формально-языкового аппарата для формирования запросов к программному средству, такому как ChatGPT, становится актуальной и важной задачей. Это позволит контент-менеджерам легко и эффективно создавать высококачественное описание продуктов, оптимизируя процесс генерации текста с помощью ИИ.

Кроме того, проведение исследования влияния различных параметров, таких как время продажи, тональность текста, возраст аудитории, ключевые слова, цель текста, связанные темы, конкретные вопросы, название магазина, зона доставки магазина, специализация магазина в одном предложении, название магазина, атрибуты товара (категория товара, материал, доступные цвета, бренд продукта, цена) на продажи товаров, позволит выявить наиболее значимые факторы и определить оптимальные методы описания продуктов для достижения лучших результатов в электронной коммерции.

Таким образом, актуальность данной магистерской работы обусловлена стремительным ростом е-коммерции, необходимостью создания качественных описаний продуктов для успешной конкуренции на рынке, а также возможностью автоматизации процесса создания описаний с помощью ChatGpt и формально-языкового аппарата.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Целью данного исследования является разработка формально-языкового аппарата для создания запросов к семантической модели и системе генерации текста ChatGPT. Достижение этой цели подразумевает выполнение ряда задач, включая исследование существующих подходов и методов, разработку и внедрение новых алгоритмов, а также их оценку и тестирование.

Объектом данного исследования являются запросы на естественном языке к семантической модели и ChatGPT. Предмет исследования является формальные языки, грамматики, методы принятия решений, метрики схожести слов.

Актуальность исследования обусловлена растущим использованием искусственного интеллекта для генерации текста и важностью точной формулировки запросов к программным средствам на подобии ChatGPT в различных областях, начиная от коммерции и заканчивая научными исследованиями.

Для достижения цели исследования предполагается выполнение следующих задач:

1. Исследовать особенности систем электронной коммерции.
2. Изучить и проанализировать проблему формулирования точных запросов к языковым моделям и их влияние на качество генерации описаний продуктов.
3. Выполнить проверку правильности записи слов в запросе.
4. Разработать формальную грамматику для формирования запросов к семантической модели и программному средству ChatGPT, учитывая различные параметры, влияющие на качество описаний продуктов, а также принципы обработки слов с ошибками.
5. Создать алгоритм разбора и проверки запросов к языковой модели на основе разработанной формальной грамматики.
6. Разработать программное средство для ввода информации о продукте контент-менеджером с использованием разработанной формальной грамматики и интеграции с системами электронной коммерции.
7. Провести эксперименты по генерации описаний продуктов с использованием разработанного программного средства и оценить их качество.

Тема диссертационной работы соответствует перечню приоритетных направлений, определенных Указом Президента Республики Беларусь от 07.05.2020 (№156) «О приоритетных направлениях научно-технической деятельности в Республике Беларусь на 2021-2025 годы» (направление

развитие информационного общества, электронного государства и цифровой экономики).

## **КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ**

В главах диссертационного исследования изложены теоретические основы формальных языков и грамматик, методов принятия решений и метрик схожести слов, особенности e-коммерции в контексте взаимодействия с пользователем, разработка и оценка предложенного формально-языкового аппарата для создания запросов к семантической модели и ChatGPT.

## ЗАКЛЮЧЕНИЕ

В процессе исследования были достигнуты все поставленные цели и выполнены задачи. Были изучены особенности систем электронной коммерции, такие как Shopify и Magento, и их возможности интеграции с программным средством ChatGPT. Этот этап позволил выявить возможности для оптимизации процесса генерации описаний продуктов и выявления потенциальных улучшений.

Проблема формулирования точных запросов к языковым моделям была изучена и проанализирована, что позволило выявить основные трудности и препятствия на пути к получению качественных описаний продуктов. Это в свою очередь привело к необходимости создания формальной грамматики и алгоритма разбора и проверки запросов к языковой модели.

В результате была разработана формальная грамматика для формирования запросов к семантической модели и программному средству ChatGPT. Эта грамматика учитывает различные параметры, влияющие на качество описаний продуктов.

Также в ходе выполнения данной диссертации успешно достигнута главная цель работы – создание системы для обеспечения корректности и правильности записи слов в запросах. Это позволило усовершенствовать процесс обработки запросов и увеличить точность обработки семантической модели и программному средству ChatGPT.

На основе разработанной формальной грамматики был создан алгоритм разбора и проверки запросов к семантической модели и программному средству ChatGPT. Этот алгоритм является ключевым элементом в процессе создания точных и качественных описаний продуктов.

Затем было разработано программное средство для ввода информации о продукте контент-менеджером с использованием разработанной формальной грамматики и интеграции с системами электронной коммерции. Этот шаг позволил значительно упростить процесс создания описаний продуктов и увеличить их качество.

Наконец, были проведены эксперименты по генерации описаний продуктов с использованием разработанного программного средства, результаты которых показали высокую эффективность разработанного подхода.

Таким образом, была успешно выполнена поставленная цель исследования и выполнены все поставленные задачи. Все это подтверждает актуальность и значимость проведенного исследования в области обработки текстовых данных и создания описаний продуктов с использованием языковых моделей и формальных грамматик.

## СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1-А. Метод для проверки запросов на естественном языке к семантической модели/ П.М. Кушеревич, Ю.О. Герман // СБОРНИК НАУЧНЫХ СТАТЕЙ IX МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ. В двух частях. Часть 2, 2023. – С.205-209.

2-А. Кушнеревич, П. М. Метод для проверки запросов на естественном языке к семантической модели/ П.М. Кушнеревич // 59-я конференция аспирантов, магистрантов и студентов учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», 17-21 апреля 2023 г., БГУИР, Минск, Беларусь: тез. докл. – Мн. – 2023. – в печати.