На правах рукописи

Забавин Алексей Сергеевич

**Разработка программы «Базы Знаний ТОГУ» с использованием полнотекстового поиска с помощью лексем естественного языка**

09.04.04 – Программная инженерия

Автореферат  
магистерской диссертации

Хабаровск – 2025

**Общая характеристика работы**

Диссертация посвящена вопросам обработки естественного языка и методиках поиска и индексации текстовых данных.

**Актуальность темы исследования**. Актуальность продиктована необходимостью разработки автоматического ассистента службы поддержки ТОГУ, который бы с большей релевантностью предоставлял пользователю ответы на поставленные вопросы. Существующие сторонние разработки менее ориентированы на контекст работы и информационную специфику вуза и требуют больших усилий по интеграции.

**Цель и задачи исследования**. Цель данного исследования состоит в разработке программного средства, использующего RNN искусственную нейронную сеть семейств Word2vec и GloVe для синтаксического и семантического анализа и оптимизации поисковых запросов.

**Научная новизна:**

разработанная программа и модель Word2Vec осуществляет хранение и полнотекстовый поиск ответов на вопросы на естественном языке;

предложен метод оптимизации поискового запроса улучшает качество поиска в базе знаний и может в дальнейшем может быть использован для генерации RAG контекста к современным LLM для создания виртуального ассистента для студентов ТОГУ.

**Теоретическая и практическая значимость**. Предложенные модели позволяют осуществлять синтаксический, морфологический и семантический анализ предложений

**Методология и методы исследования.** В работе использована теория искусственных нейронных сетей, методы обработки естественного языка, в том числе методы векторизации слов в пространстве языка с помощью моделей со скользящим окном.

**Основные положения, выносимые на защиту.**

1. Использование полнотекстового PostgreSQL поиска позволяет эффективнее индексировать и искать информацию в документах текстового типа;
2. Применение синтаксического и семантического анализа пользовательского запроса позволило улучшить поисковую выдачу путем оптимизации исходного запроса;
3. Разработанный метод оптимизации запроса позволяет поднять качество релевантных ответов в два раза перед полнотекстовым поиском без модификаций.

**Апробация работы и публикации.** Основные результаты докладывались на конференции «VI региональной научно-практической конференции» (г. Хабаровск, 2025). По материалам работы опубликованы статьи в сборнике СНПК 2025.

**Объем и структура диссертации**. Диссертация состоит из введения, трех глав, заключения, списка использованных источников и двух приложений. Полный объем диссертации составляет 81 страницы, включая 16 рисунков и 5 таблиц. Список использованных источников содержит 33 наименований.

**Содержание работы**

**Во введении** описывается актуальность исследований, определяются цель и задачи, обосновывается научная новизна, теоретическая и практическая значимость работы.

**Первая глава** посвящена изучению предметной области. Приведены основный понятия, методы для анализа трансформаций изображений, метрики для оценки качества/незаметности встраивания и извлечения ЦВЗ. Также рассмотрены вейвлет и дискретное косинусное преобразования, сингулярное разложение матриц, что является теоретической основой исследования и используется в разработанных алгоритмах.

**Во второй главе** приведено описание и схемы работы для разрабо–танного метода маркирования отдельных наиболее значимых объектов изображения и, в частности, для используемого им частотного алгоритм маркирования. Данные алгоритмы используются в практической части исследования (в третьей и четвертой главах).

**Третья глава** наглядно демонстрирует пример работы программы, разработанной для внедрения и извлечения ЦВЗ и использующей алгоритмы, описанные во второй главе. Рассмотрены результаты тестирования разработанных методов путем сравнения качества исходного и маркированного изображений, исходного и извлеченного ЦВЗ. Приведено подробное описание технической реализации программы.

**Четвертая глава** демонстрирует пример работы программы, разработанной для анализа трансформаций изображений с применением методов ORB и преобразования Фурье-Меллина. Проведено тестирование и сравнение найденных с помощью данных методов трансформаций в равных условиях, т.е. на одних и тех же изображениях. Приведено подробное описание технической реализации программы.

**В заключении** приведены основные результаты, полученные в дис–сертационном исследовании.

**Приложение А** включает реализованные классы и функции разрабо–танных программ.

**Приложение Б** содержит проверку на заимствования.

**Публикации автора по теме диссертации**

1. Забавин А.С. Вихтенко Э.М. «VI региональная научно-практическая конференция» // «ИССЛЕДОВАНИЕ ТЕХНОЛОГИЙ МОРФОЛОГИЧЕСКОГО И СЕМАНТИЧЕСКОГО АНАЛИЗА И ИХ ПРИМЕНЕНИЕ В ЗАДАЧЕ ИНФОРМАЦИОННОГО ПОИСКА»)