

ИНТЕЛЛЕКТУАЛЬНЫЙ МЕДИЦИНСКИЙ АССИСТЕНТ НА ОСНОВЕ ТЕХНОЛОГИИ RAG ДЛЯ ОБРАБОТКИ РАЗНОРОДНЫХ ДАННЫХ

Рассматривается программная реализация интеллектуального медицинского ассистента на базе технологии RAG. Предлагается использование микросервисной архитектуры и многоуровневого хранилища данных для автоматизации обработки разнородной медицинской информации

Введение

Для решения задачи автоматизации работы с медицинскими данными предлагается использовать технологию RAG, объединяющую методы извлечения информации и генерации текста. Современный этап развития здравоохранения характеризуется стремительным ростом объемов как структурированных, так и неструктурированных данных, включая рукописные записи и сканы. Для эффективной обработки таких массивов и поддержки принятия врачебных решений разработана гибридная система, способная преобразовывать разрозненные исторические документы в семантически связанную информацию.

I. Микросервисная архитектура и гибридное хранилище

Классические подходы к хранению данных не позволяют эффективно решать задачи семантического поиска и работы со слабоструктурированными медицинскими картами. Для обеспечения гибкости и семантических возможностей предлагается использовать комбинацию трех специализированных СУБД:

- SQLite выполняет роль надежного реляционного хранилища для критически важных структурированных данных;
- MongoDB используется как документо-ориентированная база для хранения разнородных отчетов в формате JSON с сохранением их оригинальной структуры и всех смысловых деталей;
- Chroma DB функционирует как высокопроизводительная векторная база данных, обеспечивающая семантический поиск по медицинским записям с использованием алгоритмов поиска ближайших соседей.

Данный подход эффективно разделяет обязанности между базами данных, позволяя системе быстро агрегировать релевантную информацию из огромного массива документов.

Рублевская Екатерина Александровна, студент кафедры интеллектуальных информационных технологий БГУИР, katerublevskya@gmail.com.

Научный руководитель: Сальников Даниил Андреевич, ассистент кафедры интеллектуальных информационных технологий БГУИР, d.salnikov@bsuir.by.

II. Интеллектуальный конвейер обработки данных

Технологический конвейер начинается с предобработки медицинских сканов алгоритмами библиотеки OpenCV для коррекции неравномерного освещения и подавления шума. Извлечение текста, включая анализ рукописных записей, осуществляется через интеграцию с Groq Vision API.

Для подготовки текста к индексации применяется алгоритм чанкирования с перекрытием, который сохраняет смысловые связи между логически завершенными фрагментами. Векторные представления генерируются моделью Sentence-BERT (all-MiniLM-L6-v2). При запросе система находит релевантные фрагменты в Chroma DB, обогащает их полными текстами из MongoDB и структурированным контекстом из SQLite, после чего передает агрегированные данные в LLM для генерации обоснованного медицинского отчета. Финальный результат форматируется в PDF с помощью библиотеки WeasyPrint.

III. Выводы

Данная система успешно автоматизирует обработку медицинских документов, формируя сквозной конвейер: от исходного изображения до структурированного PDF-отчета. Использование гибридной архитектуры хранения и микросервисного подхода обеспечивает высокую точность семантического поиска, масштабируемость и возможность независимого обновления компонентов, что значительно снижает рутинную нагрузку на медицинский персонал.

Список литературы

1. AnythingLLM. Open-source AI application for document interaction / Electronic resource – Mode of access: <https://useanything.com>. – Date of access: 15.03.2026.
2. RAGFlow. Open-source RAG engine for document understanding / Electronic resource. – Mode of access: <https://github.com/ragflow>. – Date of access: 15.03.2026.