

ИСПОЛЬЗОВАНИЕ ВЕКТОРНОЙ БАЗЫ ДАННЫХ ДЛЯ ЗАДАЧИ ПРЕДВАРИТЕЛЬНОЙ ДИАГНОСТИКИ И ОЦЕНКИ СТАДИИ ЗАБОЛЕВАНИЯ ДИСПАНСЕРНЫХ ПАЦИЕНТОВ ОФТАЛЬМОЛОГИЧЕСКОГО ПРОФИЛЯ

М.В. Петренко¹, Е.И. Чигарина¹

¹Самарский университет, г. Самара
e-mail: mariapetrenko2003@mail.ru

Аннотация

В работе рассматривается применение векторной базы данных для поддержки принятия решений в офтальмологии. Предложена модель, обеспечивающая предварительную диагностику на основе текстовых жалоб пациента и объективного статуса глаз, определение стадии заболевания по результатам обследований и формирование рекомендаций по лечению. Используется гибридный подход, сочетающий семантический поиск, лексический анализ и обработку числовых параметров.

Ключевые слова: векторные базы данных, офтальмология, семантический поиск, предварительная диагностика, стадия заболевания, поддержка принятия решений.

Введение

В настоящее время значительная часть медицинской информации представлена в неструктурированном виде, что затрудняет её автоматизированную обработку. Преимуществом использования векторных баз данных является возможность применения семантического поиска.

Целью исследования является использование векторной модели хранения данных для задачи постановки предварительного диагноза на основе текстовых жалоб пациента и объективного статуса глаз, оценки стадии заболевания на основе введённых результатов обследований, а также формирования рекомендаций по дальнейшему лечению пациента.

1. Принципы работы векторных баз данных

Векторные базы данных решают задачу поиска подходящего результата за счёт преобразования текста в числовые векторы, отражающие его смысл, что позволяет находить клинически сходные случаи независимо от различий в формулировках [1]. Поиск осуществляется на основе метрик расстояния, в основном, косинусного сходства. Для повышения производительности используются алгоритмы приближённого поиска ближайших соседей (ANN).

2. Проектирование концептуальной и логической модели базы данных

Структура базы данных основана на актуальных клинических рекомендациях [2] и охватывает не только описание заболеваний, но и данные о пациентах, врачах и приёмах. Модель включает сущности, описывающие заболевание, пациента, врача, визит, а также диагностические критерии, перечень обследований и параметры, используемые для оценки стадии.

При проектировании системы, работающей с векторной базой данных, были созданы концептуальная и логическая модели базы данных. Концептуальная схема представлена на рисунке 1.

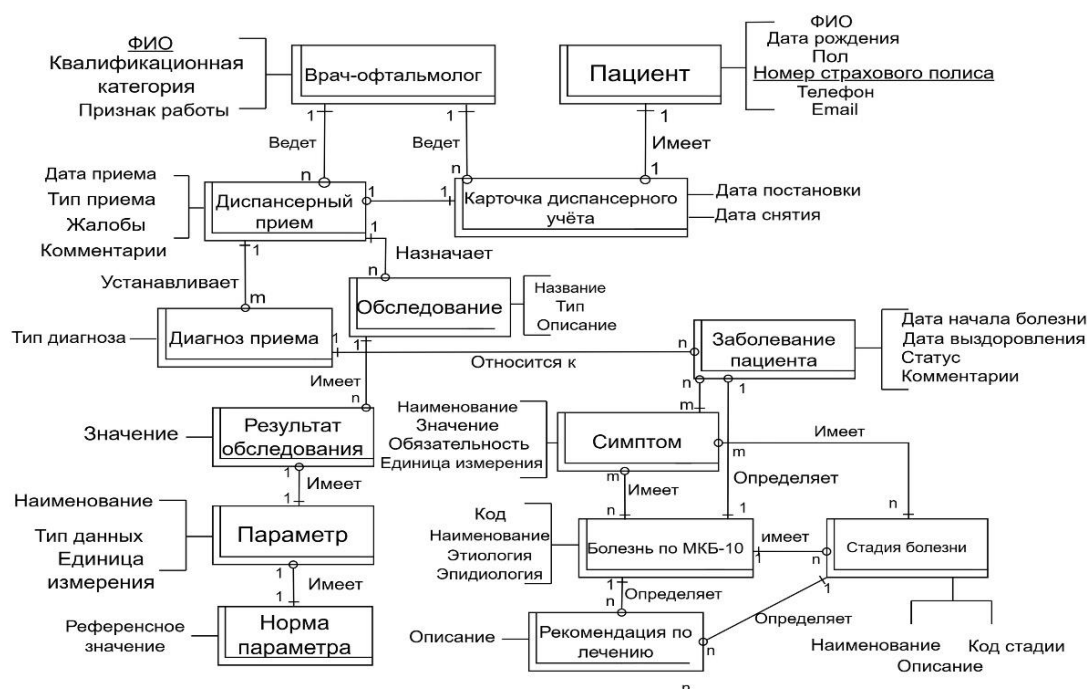


Рисунок 5 – Концептуальная схема

Для реализации векторной модели используется система Weaviate. Логическая модель построена по принципу документоориентированных баз данных, поскольку не существует унифицированных правил описания неструктурированных данных. Логическая схема векторной базы данных представлена на рисунке 2.



Рисунок 6 – Логическая схема

3. Описание математической модели задачи

Для решения задачи постановки предварительного диагноза с использованием векторной базы данных была разработана следующая модель. Пусть задано множество заболеваний $D = \{d_1, d_2, \dots, d_n\}$, где каждый элемент d_i представляет собой формализованное описание соответствующего заболевания в виде совокупности клинически значимых признаков, сформированных на основе медицинских рекомендаций и используемых при семантическом поиске. Имеется клиническое состояние пациента, которое описывается как вектор признаков: $X = (X_{compl}, X_{obj_st})$, где X_{compl} – жалобы пациента, X_{obj_st} – объективный статус глаз. Текстовые данные отображаются в общее векторное пространство фиксированной размерности с использованием предобученной трансформерной модели Sentence-BERT [3]. В частности, применяется мультиязычная модель paraphrase-multilingual-mpnet-base-v2. Формально задано отображение: $\varphi = T \rightarrow R^d$. Для каждого заболевания формируются два именованных вектора: вектор симптомов и вектор объективного статуса глаз. Для каждого заболевания формируется пара векторов $v_{d_i} = (v_{d_i}^{symp}, v_{d_i}^{obj_st})$. Аналогично для пациента формируются вектор $v_{patient} = (v_{patient}^{symp}, v_{patient}^{obj_st})$. Задача сводится к поиску заболевания с максимальным значением семантического сходства. Сходство вычисляется по косинусной мере, дополнительно учитывается лексическое совпадение ключевых слов (BM25). Дополнительно вводится проверка соответствия возрастной группе: если возраст пациента не типичен для данного заболевания, рейтинг соответствующего диагноза понижается, но не исключается из списка возможных.

После выбора предварительного диагноза определяется стадия заболевания на основе результатов обследований, перечень которых задаётся в описании болезни. Для оценки используются диагностически значимые параметры, указанные в клинических рекомендациях. Результаты обследований и критерии стадий отображаются в векторное пространство, после чего определяется степень их соответствия с использованием семантического сходства для текстовых данных и прямого сравнения для числовых параметров.

Назначение рекомендаций по лечению осуществляется на основе установленного диагноза и стадии заболевания и выбирается из данных, сформированных в соответствии с клиническими рекомендациями.

4. Описание результатов работы системы

В рамках исследования проведено экспериментальное тестирование разработанной модели на 15 клинических случаях. Корректный предварительный диагноз был определён в 12 случаях (80%), при этом в остальных наблюдениях правильный диагноз входил в число трёх наиболее вероятных. Стадия заболевания была корректно определена в 10 случаях (66,7%). В качестве иллюстрации работы системы рассмотрен пациент 45 лет с жалобами на головную боль, тошноту, радужные круги перед глазами и снижение зрения. Система выделила закрытоугольную глаукому среди наиболее вероятных диагнозов, а также предложила альтернативные варианты (миопию, макулярную дегенерацию, катаракту), поскольку симптом «снижение зрения» является неспецифичным. По результатам обследований (периметрии и офтальмоскопии) была определена развитая стадия глаукомы, что подтверждается расширением экскавации

диска зрительного нерва и сужением поля зрения (MD –8,4 дБ). После этого система формирует рекомендации по тактике лечения.

Заключение

В работе рассмотрена задача использования векторной базы данных для поддержки принятия решений в офтальмологии. Разработана модель, позволяющая осуществлять предварительную диагностику на основе текстовых жалоб пациента и объективного статуса глаз, оценку стадии заболевания по результатам обследований и формирование рекомендаций по выбору тактики лечения. Подход основан на преобразовании медицинских текстов в векторное пространство с использованием семантического поиска и гибридных методов, включающих лексический анализ и обработку числовых параметров.

В дальнейшем модель может быть расширена за счёт учёта дополнительных клинических факторов, включая сопутствующие заболевания, индивидуальные особенности пациента и динамику его состояния в процессе лечения.

Список использованных источников

1. Ковалев, А. Д. Автоматизированный подход к семантическому поиску по программной документации на основе алгоритма Doc2Vec. / А. Д. Ковалев, И. В. Никифоров, П. Д. Дробиничев // Информационно-управляющие системы. – 2021. – № 1. – С. 17–27.
2. Клинические рекомендации Минздрава Российской Федерации. – URL: <https://cr.minzdrav.gov.ru> (дата обращения: 20.01.2026).
3. Reimers, N. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks / N. Reimers, I. Gurevych // EMNLP. – 2019.

USING A VECTOR DATABASE FOR PRELIMINARY DIAGNOSIS, DISEASE STAGE ASSESSMENT, AND DECISION SUPPORT IN CHOOSING TREATMENT TACTICS FOR OPHTHALMOLOGY PATIENTS

M.V. Petrenko¹, E.I. Chigarina¹

¹*Samara University, Samara
e-mail: mariapetrenko2003@mail.ru*

Abstract

The paper presents the use of a vector database for decision support in ophthalmology. A model is proposed that enables preliminary diagnosis based on patients' textual complaints and objective ocular status, determines the stage of the disease from examination results, and generates treatment recommendations. A hybrid approach is employed, combining semantic search, lexical analysis, and numerical data processing.

Keywords: vector databases, ophthalmology, semantic search, preliminary diagnosis, disease stage, decision support.