



**Министерство науки и высшего образования Российской Федерации**  
**Федеральное государственное бюджетное образовательное учреждение высшего образования**  
**«Московский государственный технический университет имени Н.Э. Баумана**  
**(национальный исследовательский университет)»**  
**(МГТУ им. Н.Э. Баумана)**

# **Метод построения поисковых индексов в реляционной базе данных на основе глубоких нейронных сетей**

**Студент: Маслова Марина Дмитриевна ИУ7-83Б**  
**Руководитель: Оленев Антон Александрович**

**Москва, 2023**

# Актуальность

# Цель и задачи

**Цель:** разработка метода построения поисковых индексов в реляционной базе данных на основе глубоких нейронных сетей.

## **Задачи:**

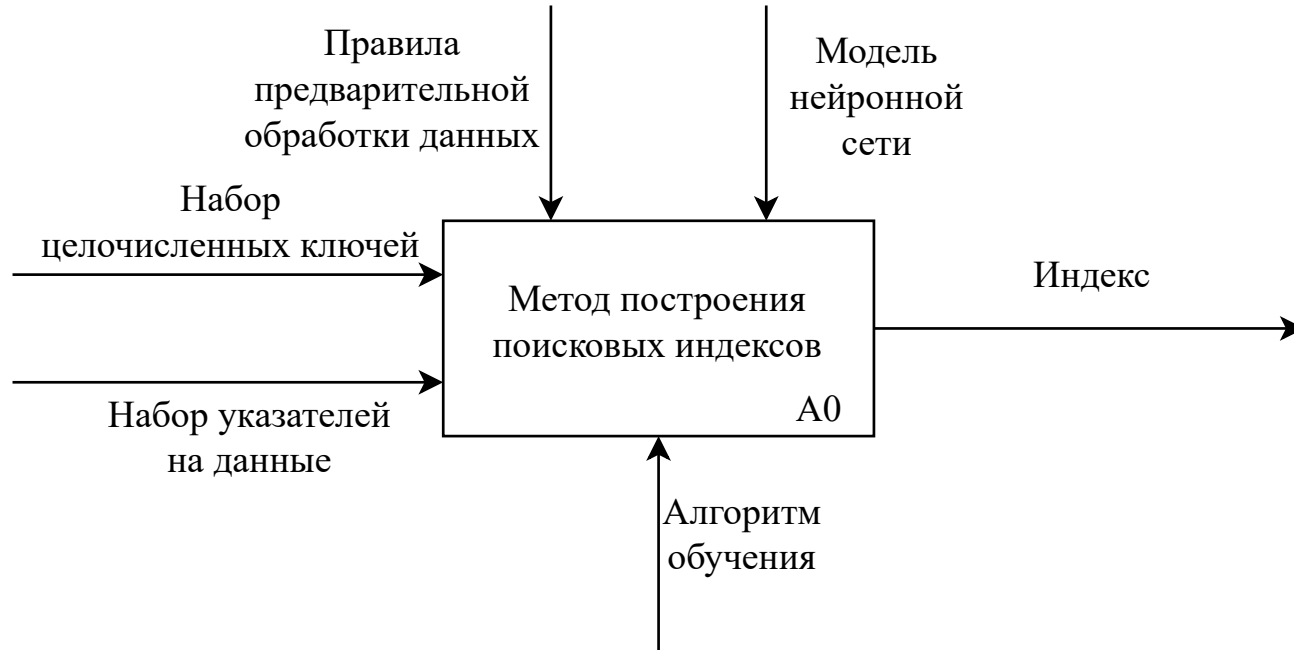
- рассмотреть и сравнить известные методы построения индексов;
- привести описание построения индексов с помощью нейронных сетей;
- разработать метод построения индексов в реляционной базе данных на основе глубоких нейронных сетей;
- разработать программное обеспечение, реализующее данный метод;
- провести исследование (по времени и памяти) операций поиска и вставки с использованием индекса, построенного разработанным методом, при различных объемах данных.

# Сравнение методов построения индексов

Метод		В-дерево	Хеш-таблица	Фильтр Блума	Обученные индексы
Временная сложность	поиска	$O(\log N)$	$O(1) / O(N)$	$O(k)$	$O(1) / O(N)$
	вставки	$O(\log N)$	$O(1) / O(N)$	$O(k)$	(*)
Память		Высокая	Средняя	Низкая	Средняя
Поиск в диапазоне		+	-	-	+
Поиск единичного ключа		+	+	-	+
Проверка существования		+	+	+	+

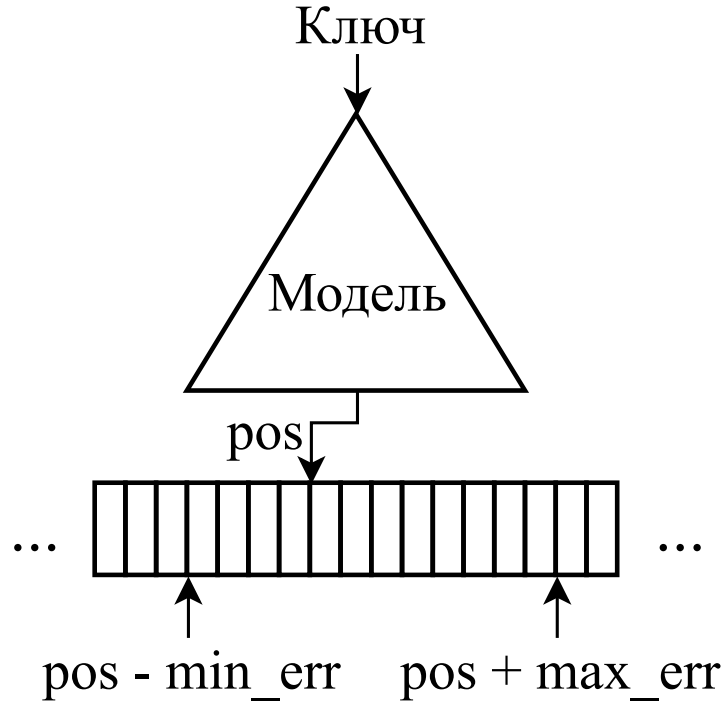
(\*) — вставка в обученный индекс требует переобучения, сложность которого зависит от архитектуры используемой модели машинного обучения.

# Постановка задачи



Ограничение: ключи — целые уникальные числа.

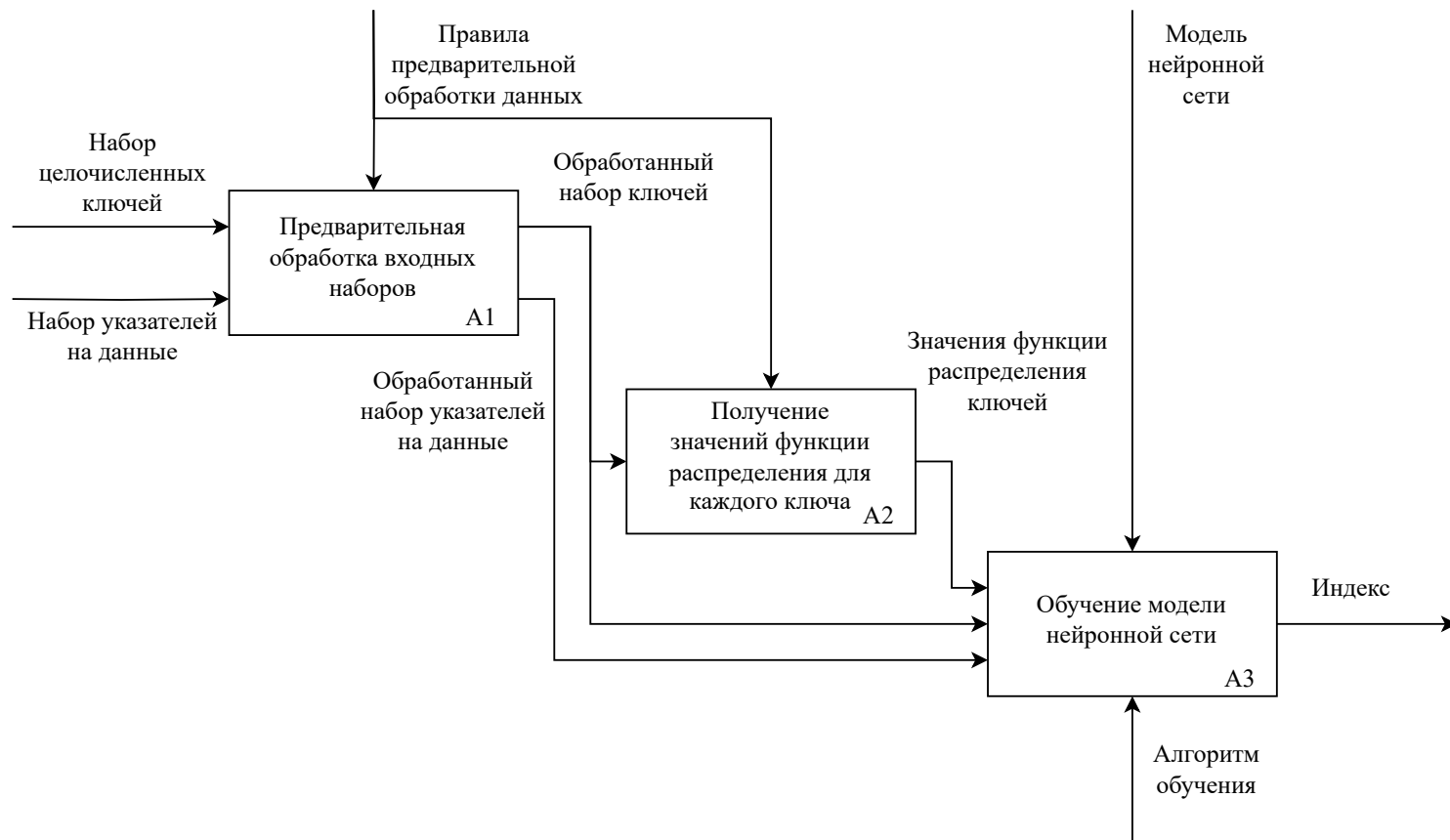
# Структура индекса



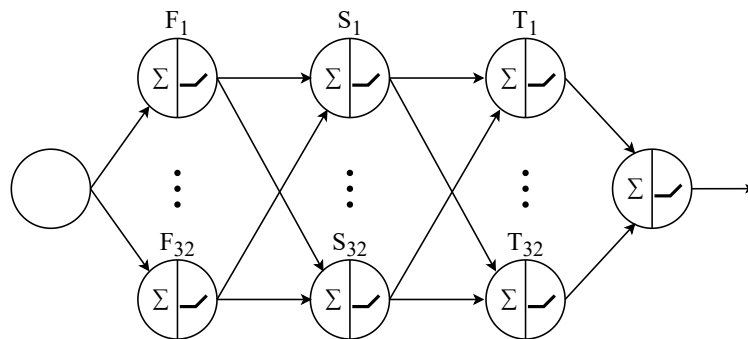
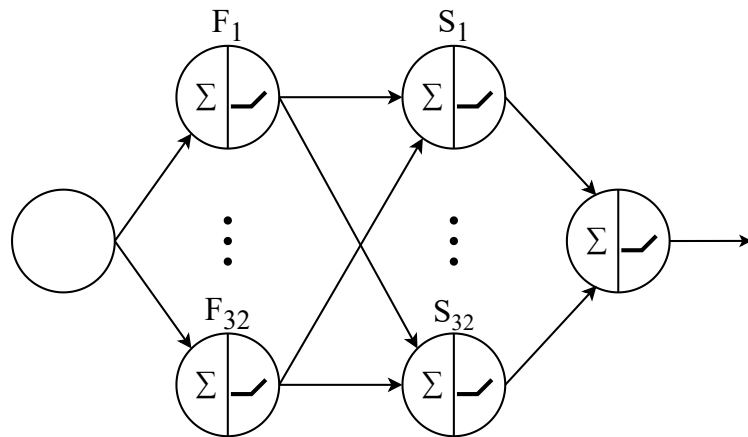
$$p = F(K) \cdot N,$$

где  $p$  — искомая позиция;  
 $K$  — ключ поиска;  
 $F(K)$  — функция распределения;  
 $N$  — количество ключей.

# Функциональная схема построения индекса

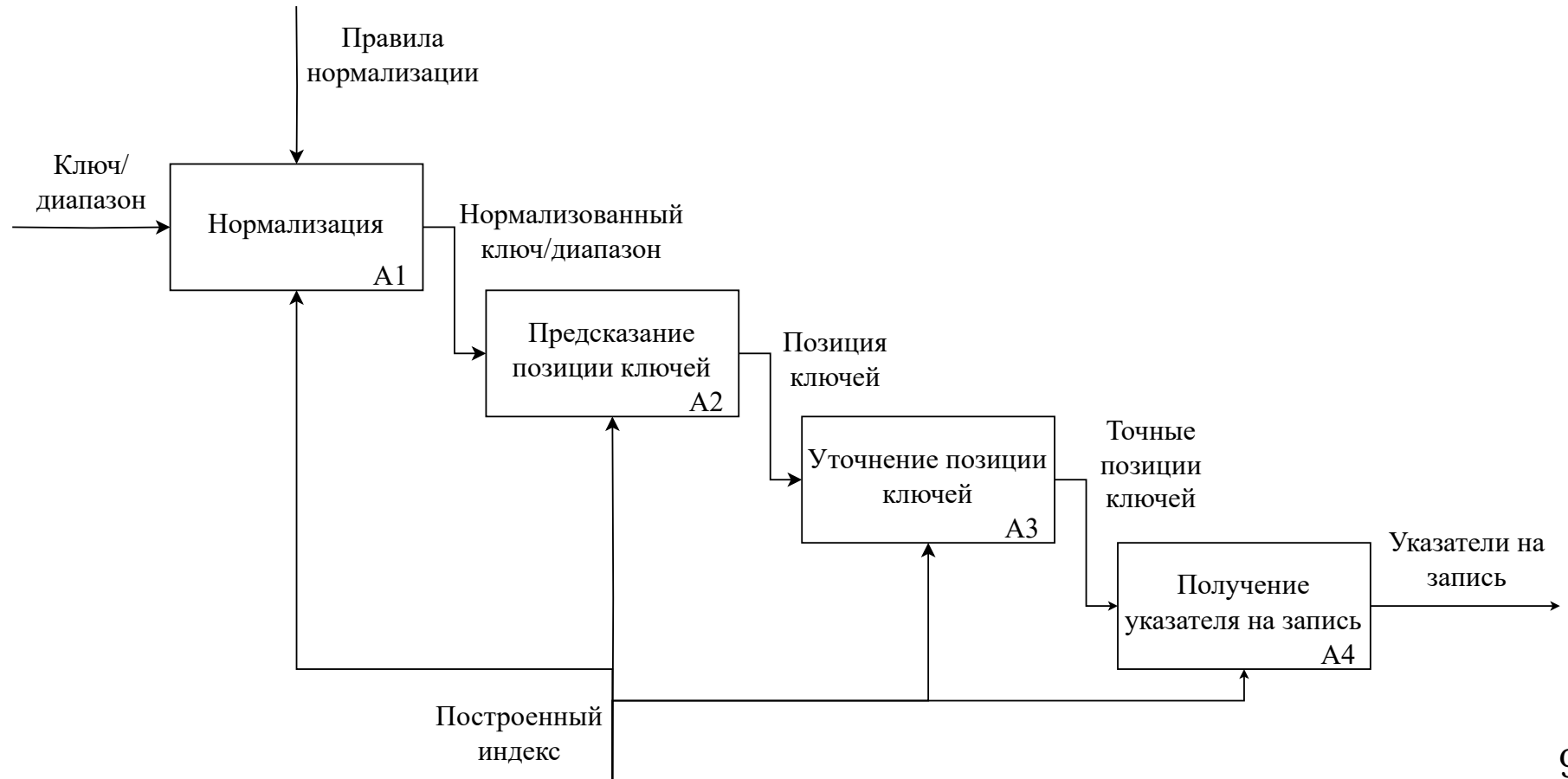


# Архитектура нейронной сети





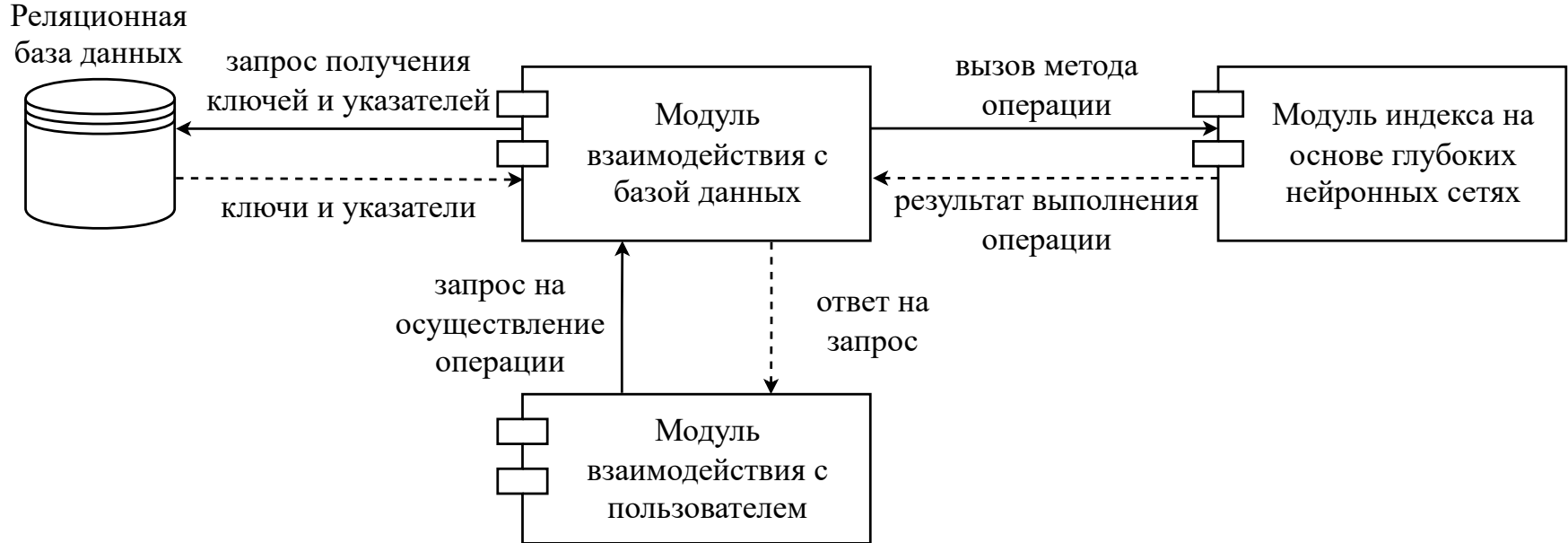
# Функциональная схема поиска



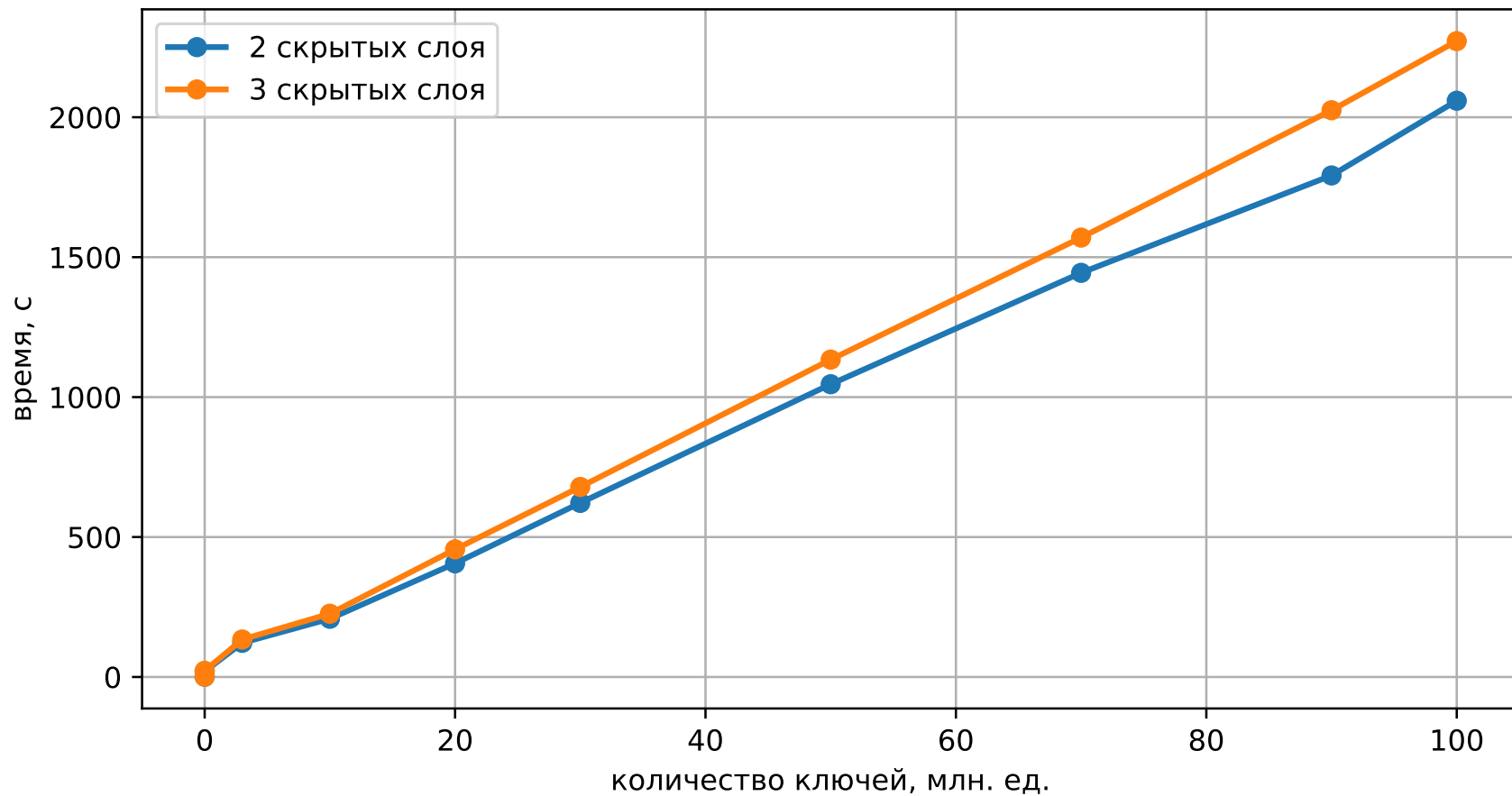
# Функциональная схема вставки



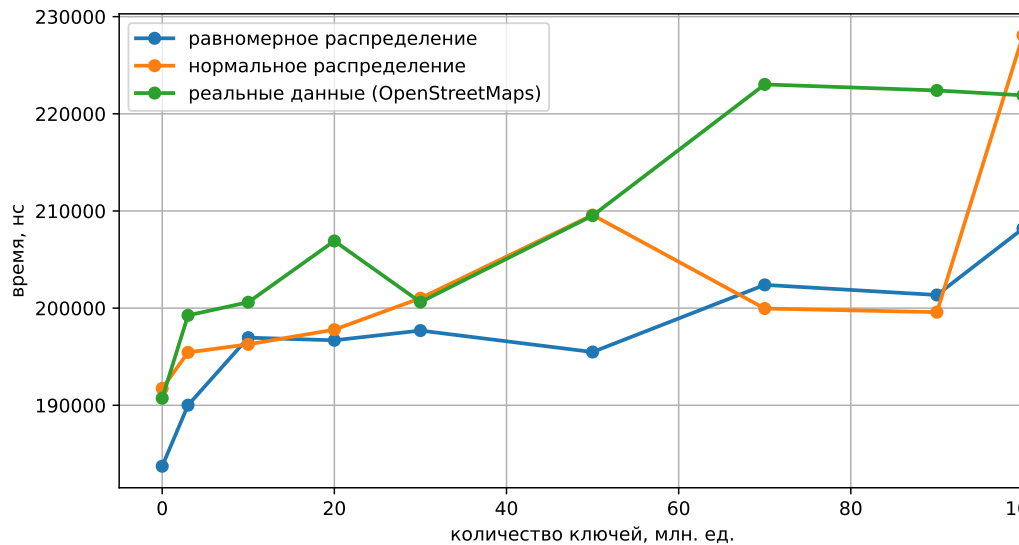
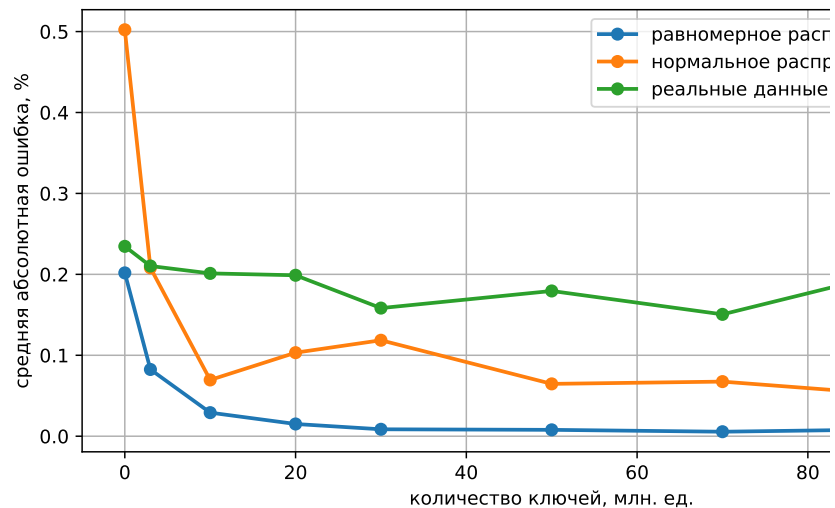
# Структура программного обеспечения



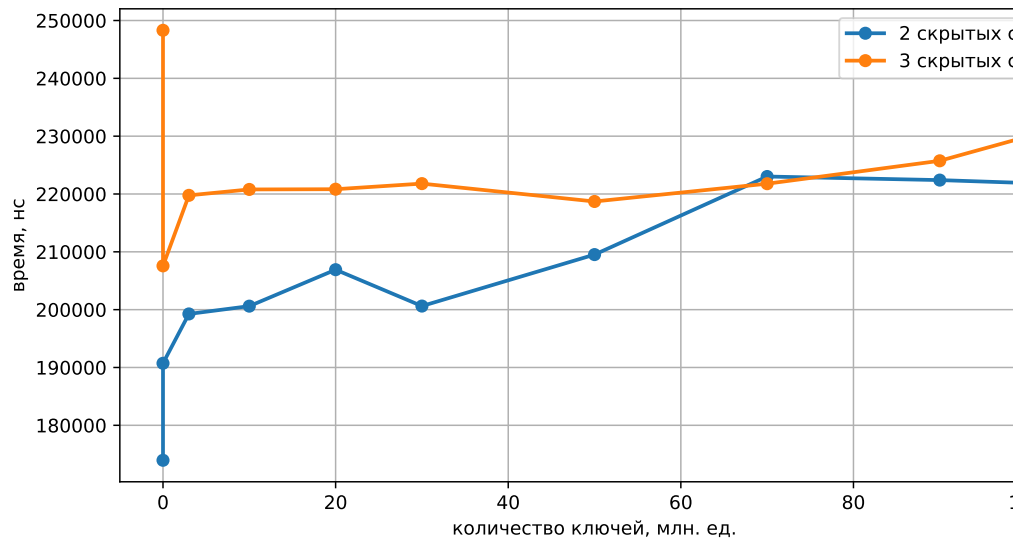
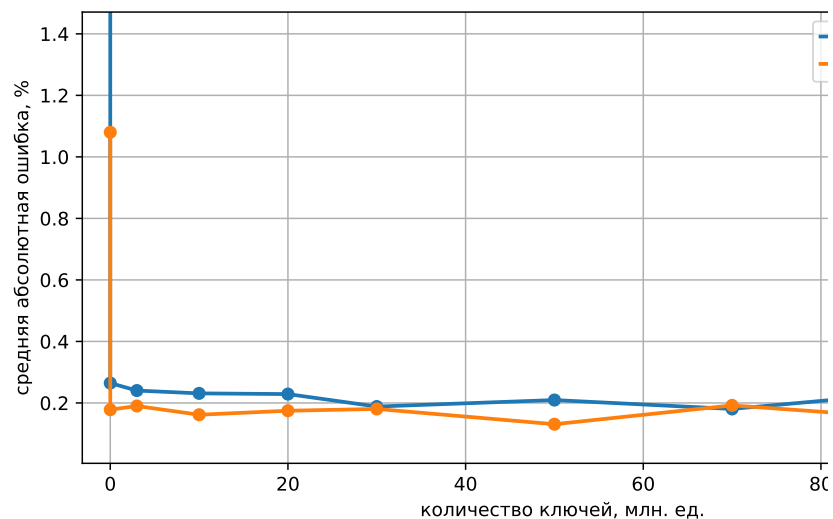
# Исследование времени построения



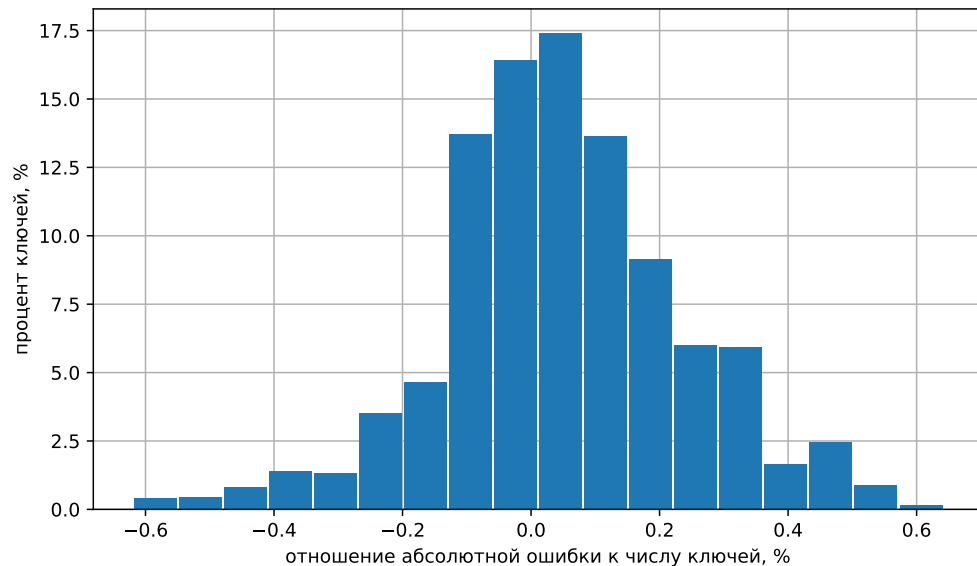
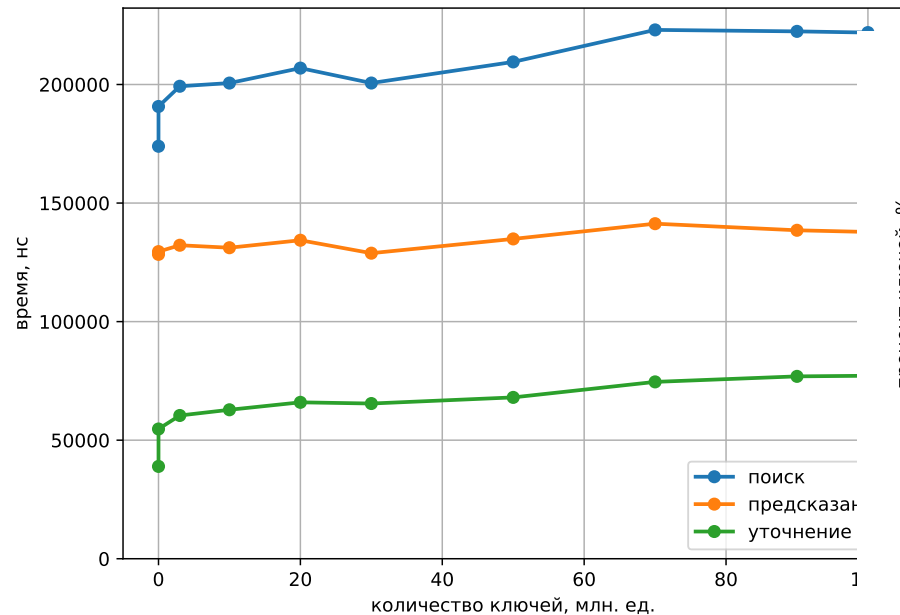
# Исследование времени поиска (распределения)



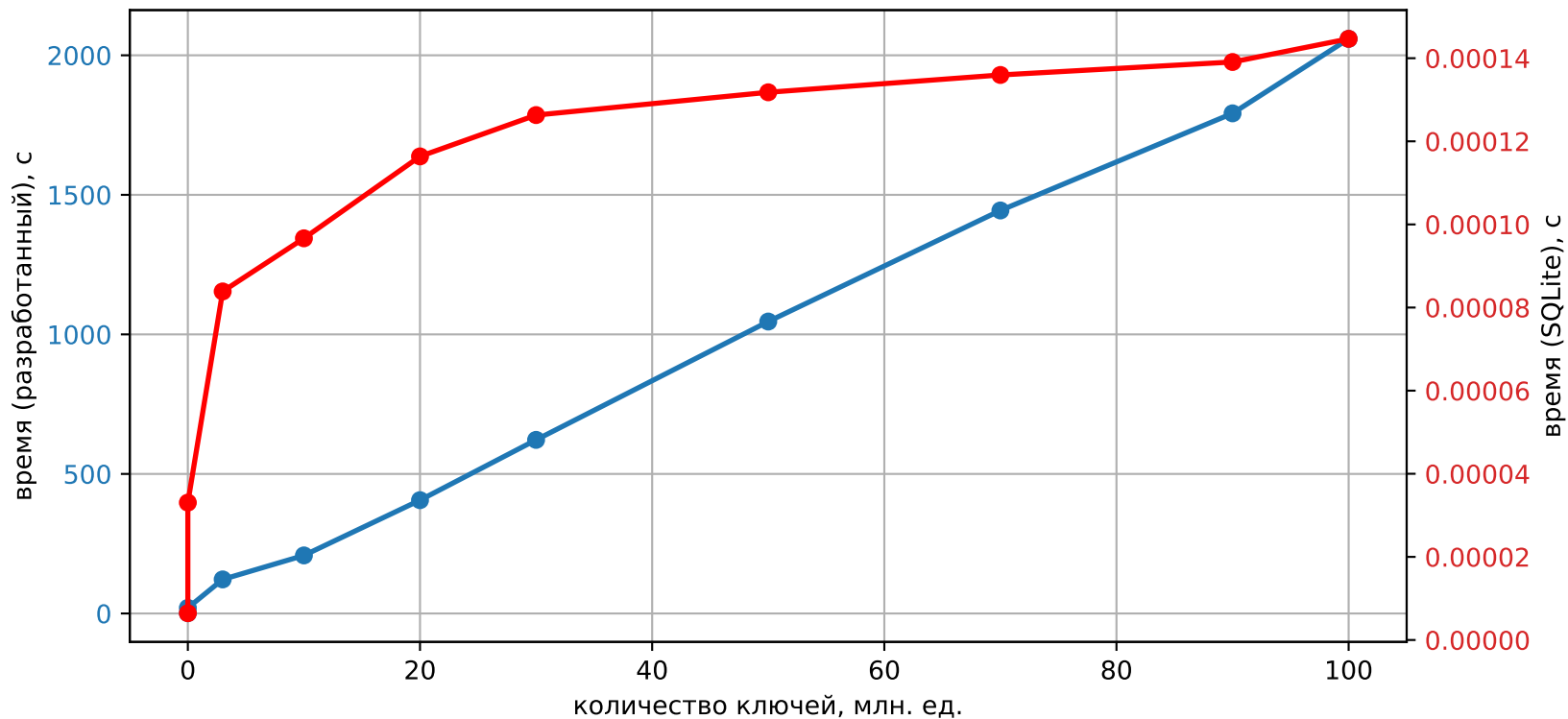
# Исследование времени поиска (модели)



# Исследование времени поиска (этапы)

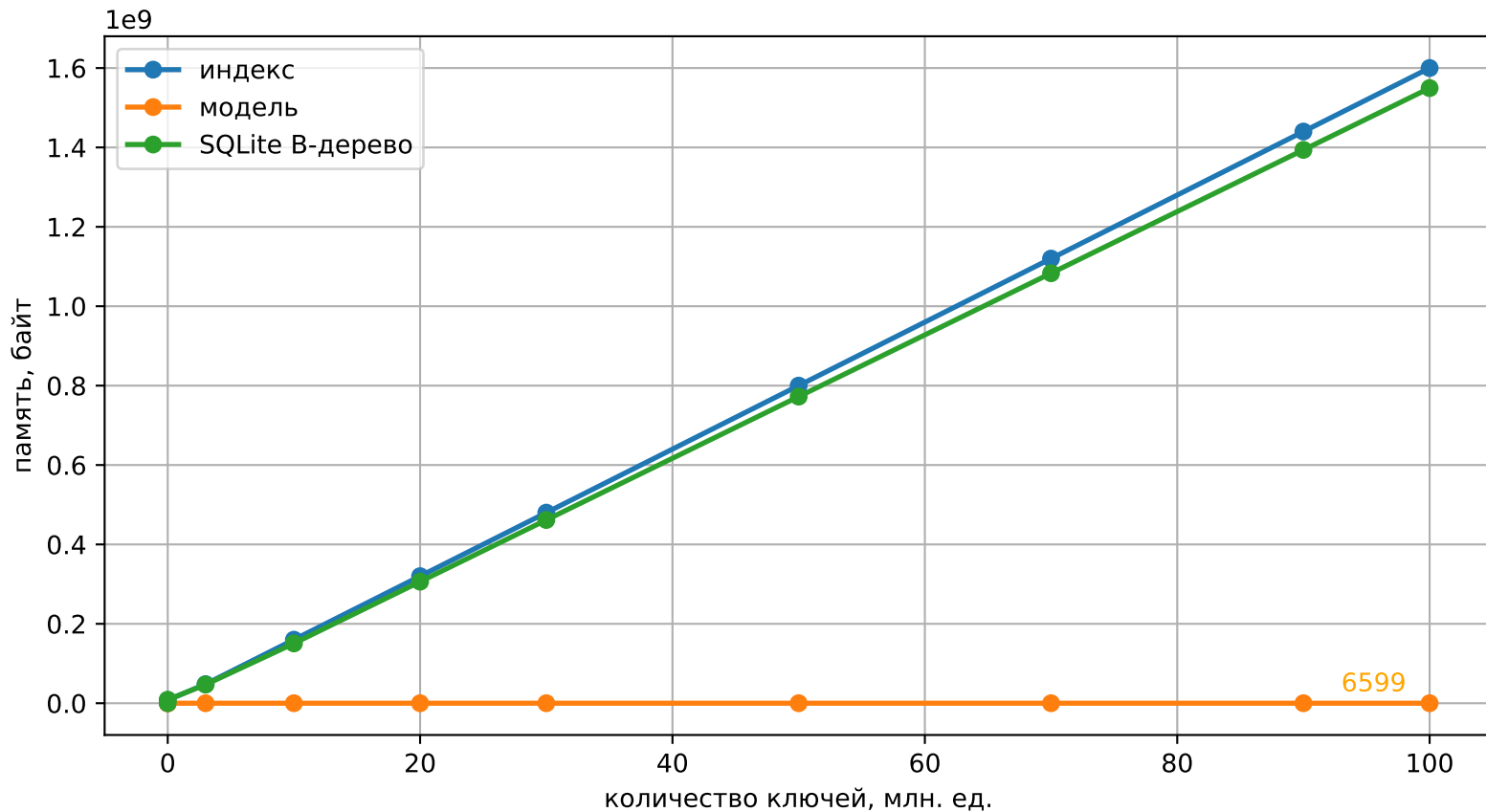


# Исследование времени вставки





# Исследование памяти, занимаемой индексом



# Заключение

В ходе данной работы:

- проанализированы известные методы построения индексов;
- приведено описание построения индексов с помощью нейронных сетей;
- разработан метод построения индексов в реляционной базе данных на основе глубоких нейронных сетей;
- разработано программное обеспечение, реализующее данный метод;
- проведено исследование (по времени и памяти) операций поиска и вставки с использованием индекса, построенного разработанным методом, при различных объемах данных.

Поставленная цель достигнута.

# Дальнейшее развитие

1. Оптимизация алгоритма вставки с учетом распределения ключей.
2. Добавление возможности построения индекса по ключам других типов данных.
3. Построение многомерных обученных индексов.