



**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)**

Классификация методов построения индексов в базах данных

**Студент: Маслова Марина Дмитриевна ИУ7-73Б
Руководитель: Оленев Антон Александрович**

Цель и задачи

Цель: классификация методов построения индексов в базах данных.

Задачи:

- провести анализ предметной области: дать основные определение, описать свойства индексов и их типы;
- описать методы построения индексов в базах данных;
- предложить и обосновать критерии оценки качества описанных методов и сравнить методы по предложенным критериям оценки.

Основные определения

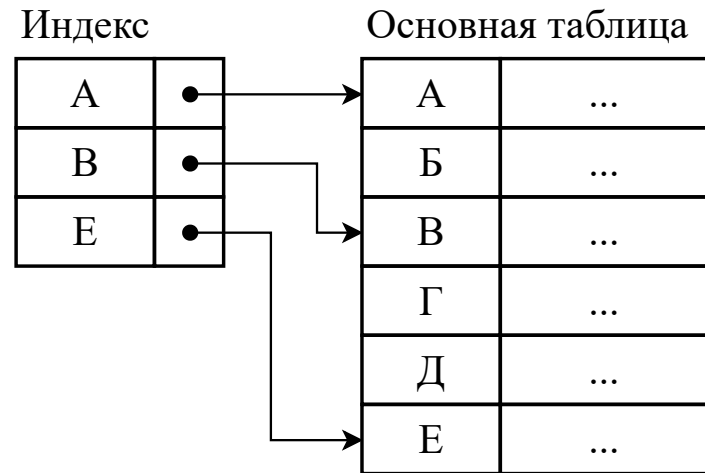
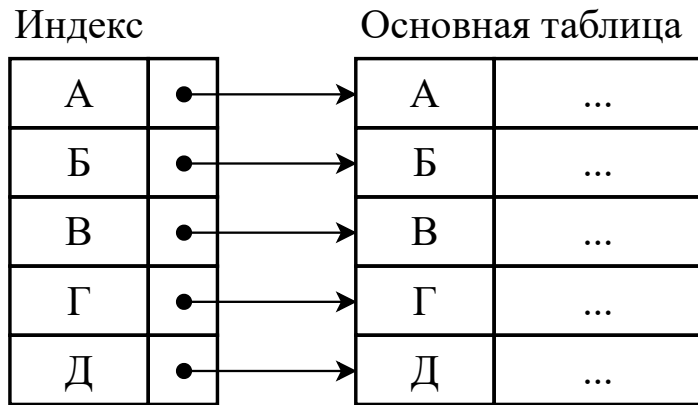
Индекс — это некоторая структура, обеспечивающая быстрый поиск записей в базе данных.

Индекс:

- определяет соответствие ключа поиска конкретной записи с положением этой записи;
- строится в дополнение к существующим данным;
- описывается:
 - типом и временем доступа;
 - временем вставки и удаления;
 - дополнительной памятью, занимаемая индексной структурой.

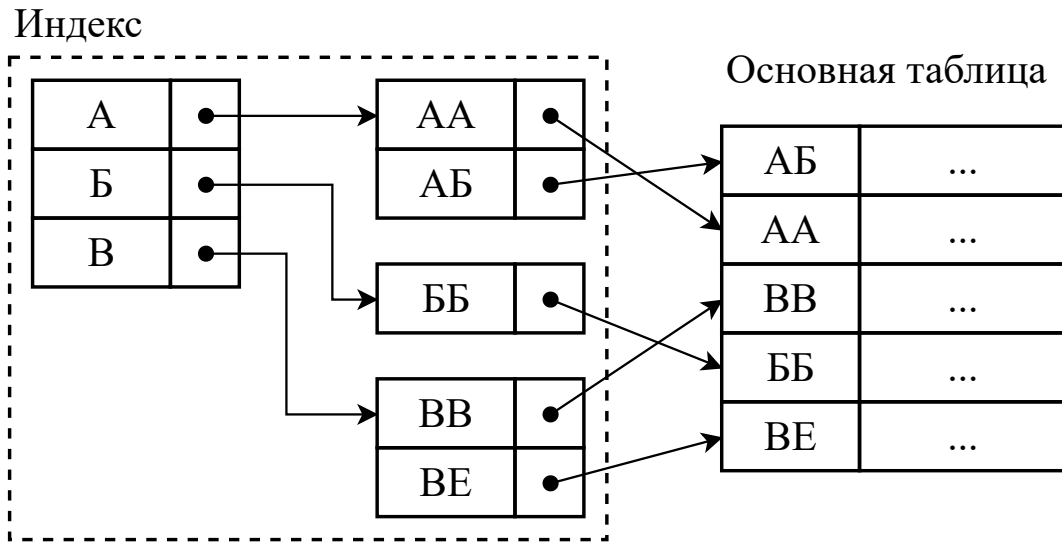
Типы индексов

- кластеризованные и некластеризованные;
- плотные и разреженные;



Типы индексов

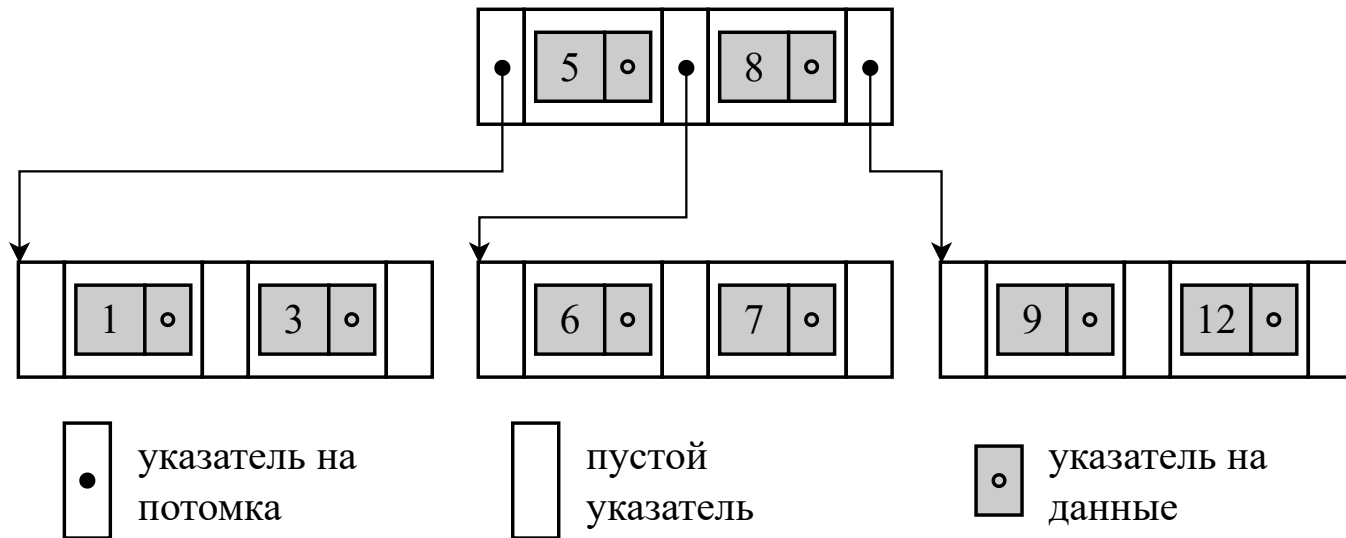
— одноуровневые и многоуровневые.



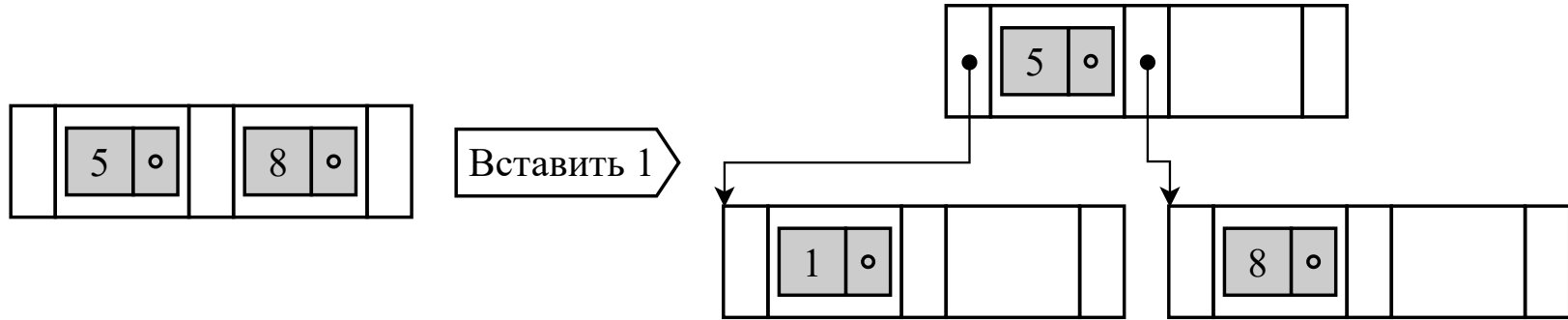
Типы индексов

- индексы на основе деревьев поиска;
- индексы на основе хеш-таблиц;
- индексы на основе битовых карт.

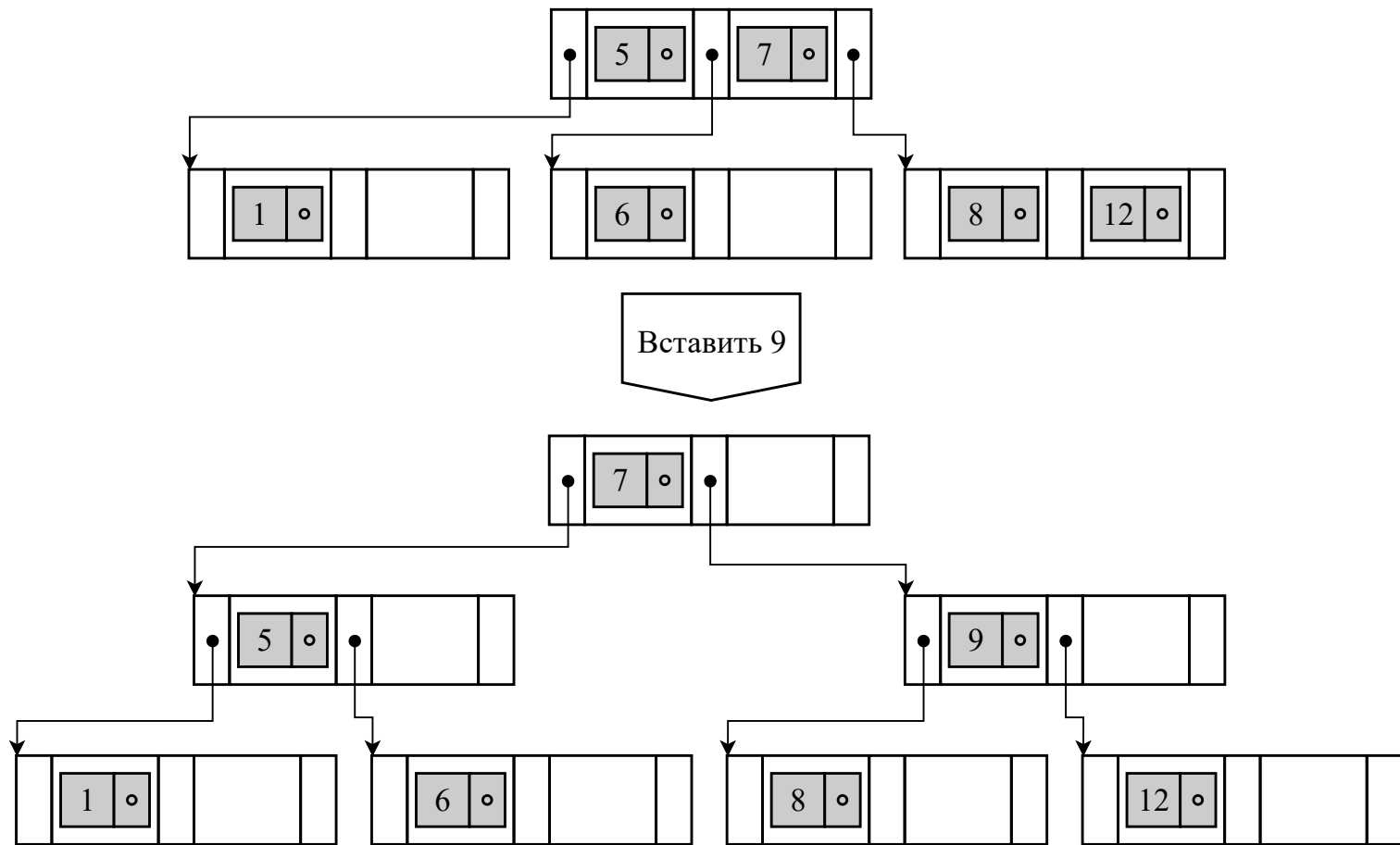
В-деревья



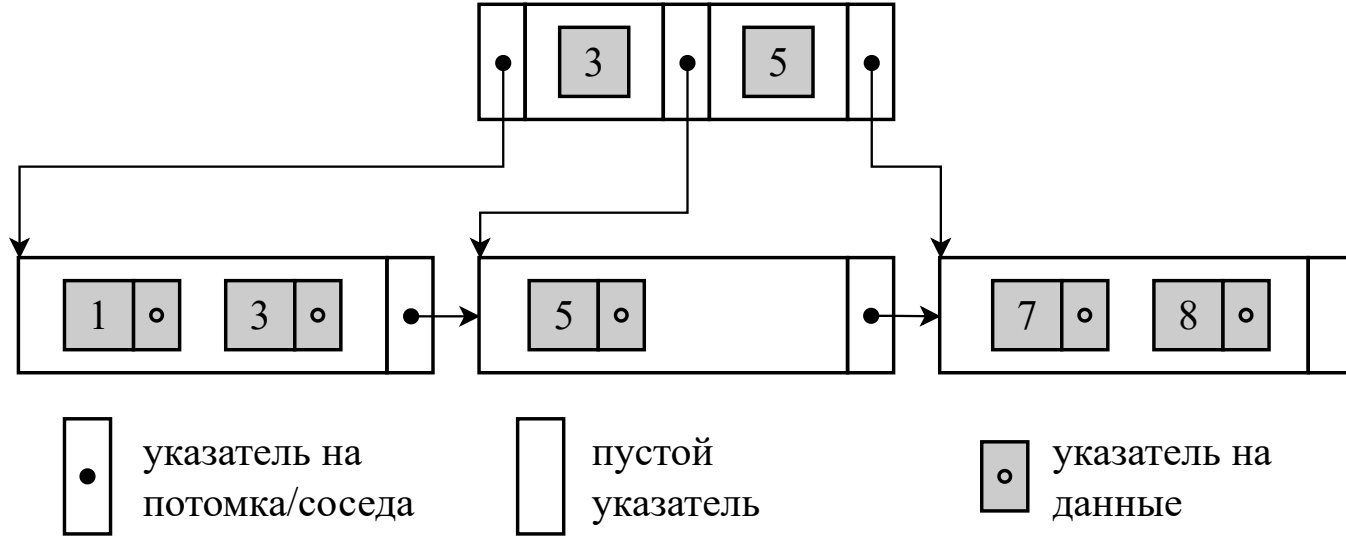
Построение В-деревьев



Построение В-деревьев

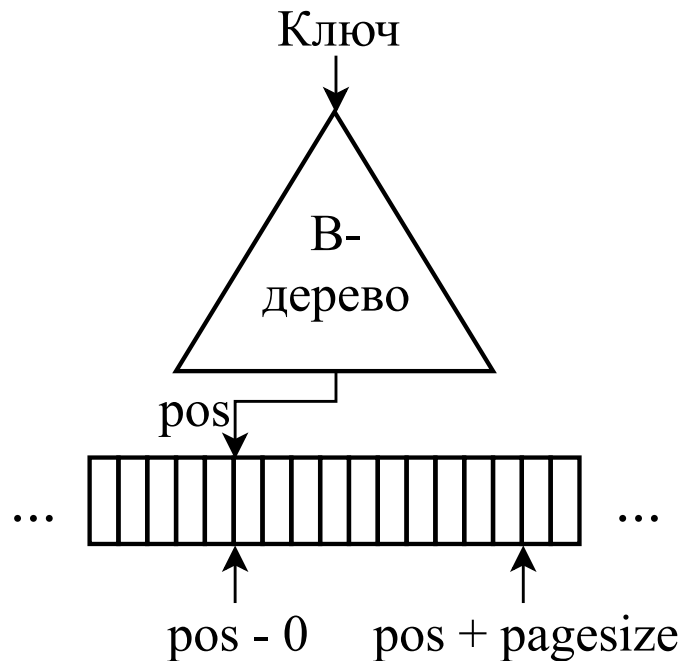


B^+ -деревья

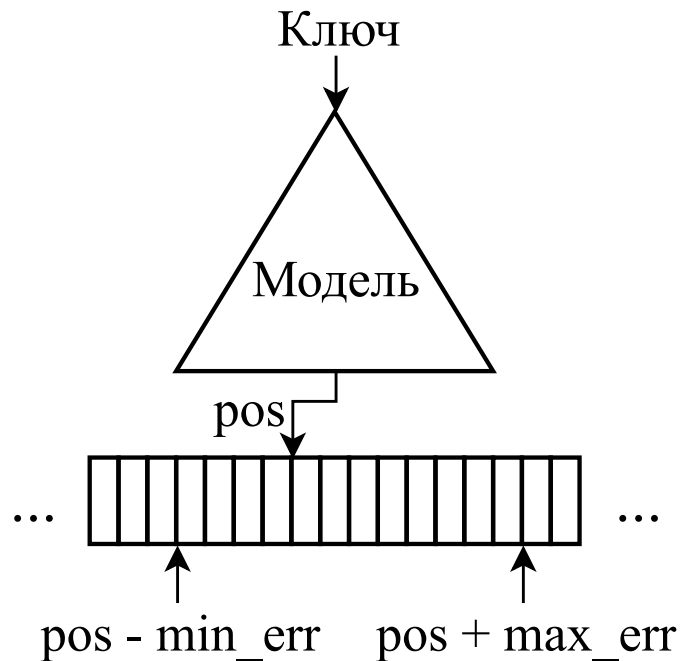


Обученные индексы

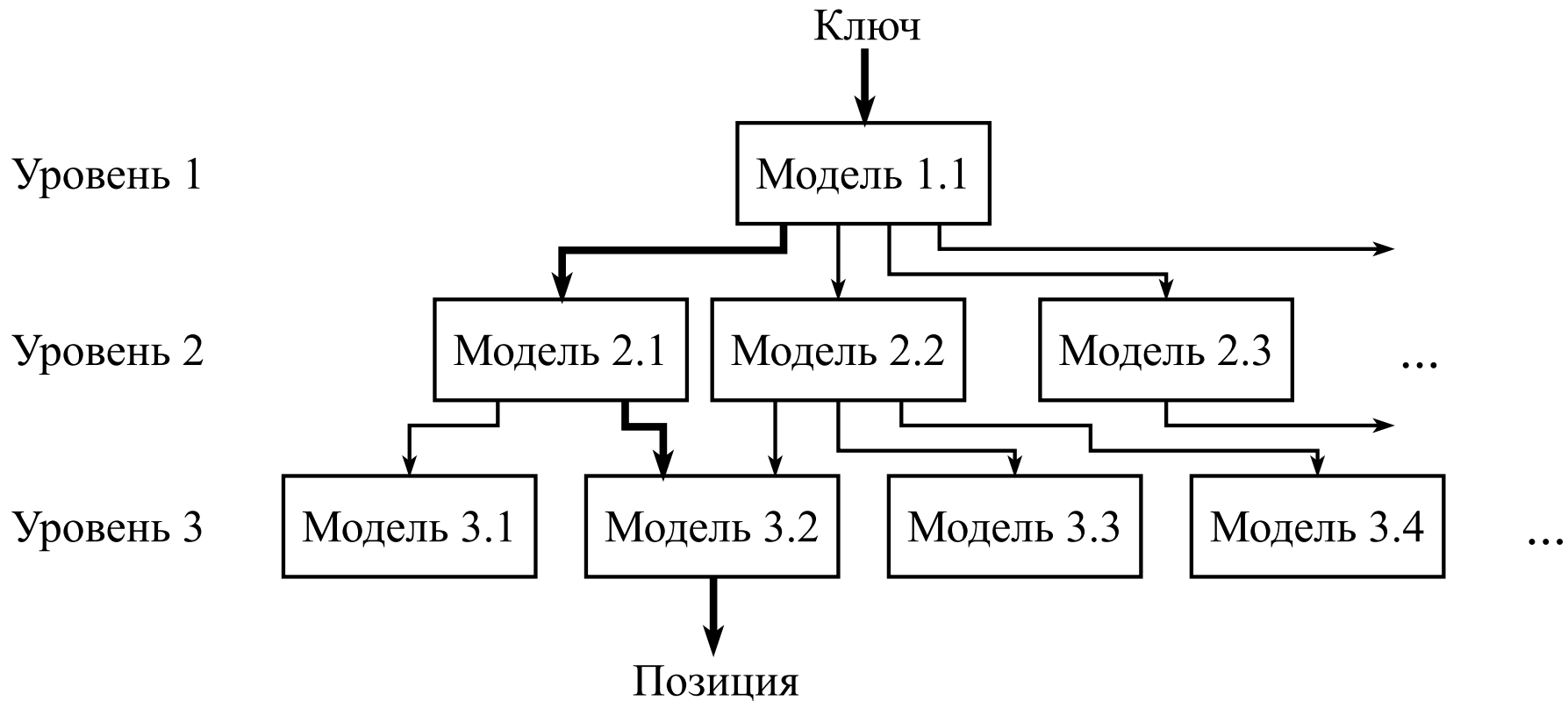
Индекс на основе В-дерева



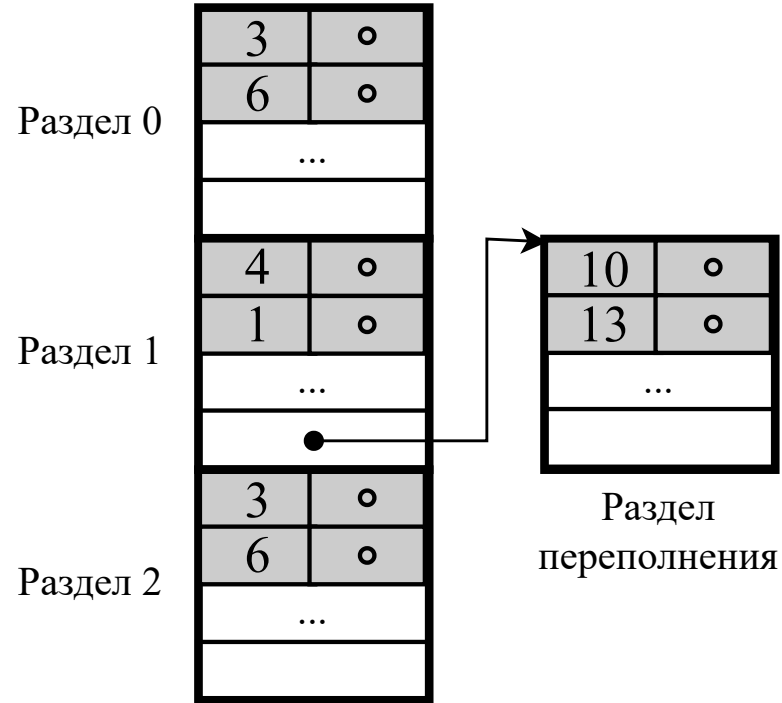
Обученный индекс



Рекурсивная модель

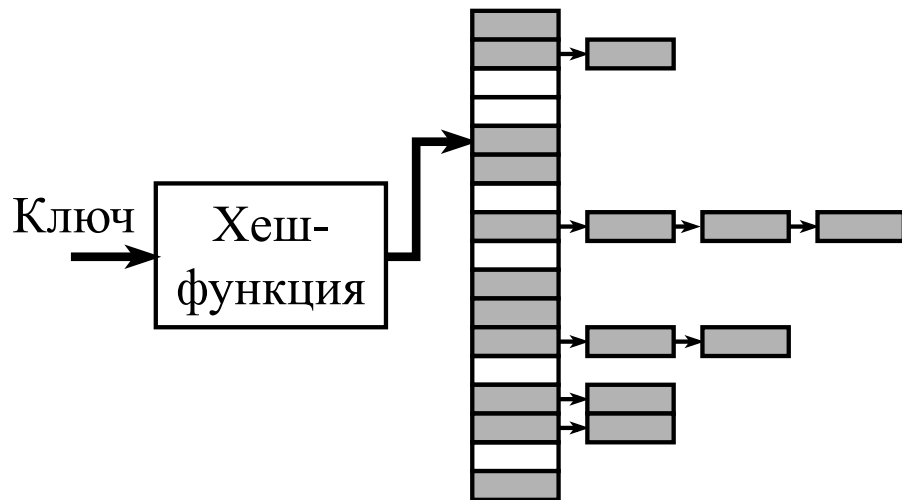


Хеш-индексы

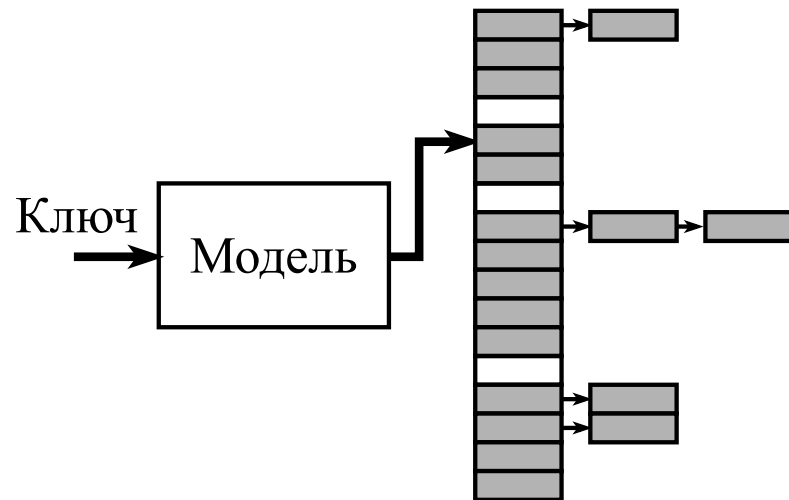


Обученные хеш-индексы

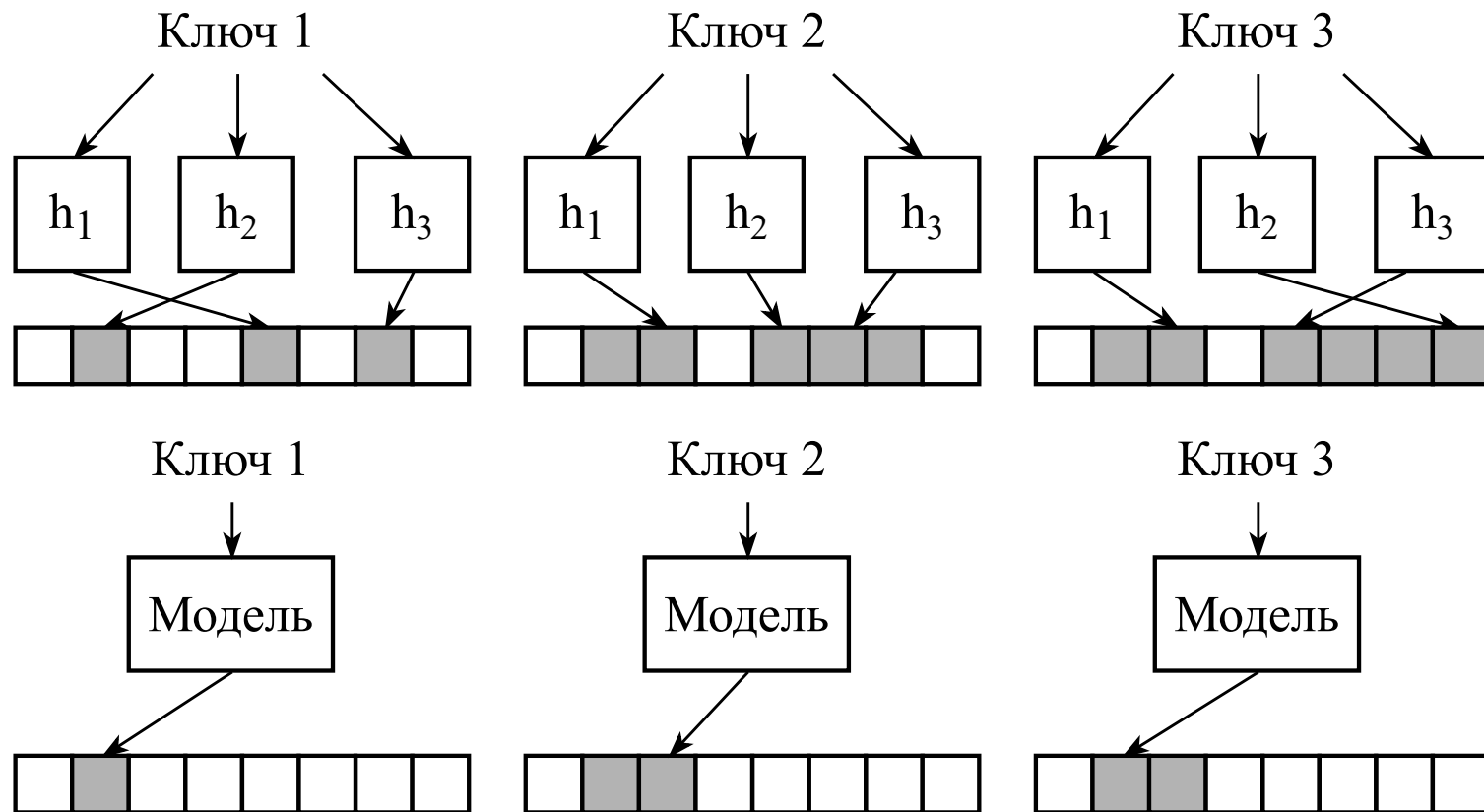
Традиционный хеш-индекс



Обученный хеш-индекс



Фильтр Блума и обученные индексы



Классификация

Классификация методов построения индексов для поиска ключей, принадлежащих некоторому диапазону (поиск).

Метод	Сложность	Время, нс	Обращения к памяти
В-деревья	$O(\log N)$	237.94	57.0
Обученные индексы	$O(\log N)$	139.09	12.6
LIPP	$O(\log N)$	24.23	3.1

Классификация

Классификация методов построения индексов для поиска ключей, принадлежащих некоторому диапазону (вставка)

Метод	Сложность	Время, нс	Обращения к памяти
В-деревья	$O(\log N)$	1114.19	57.8
Обученные индексы	—	—	—
LIPP	$O(\log^2 N)$	70.93	3.1

Классификация

Классификация методов построения индексов для поиска единичных ключей

Метод	Сложность		Процент коллизий
	Худший	Средний	
Хеш-индексы	$O(N)$	$O(1)$	35.3%
Обученные хеш-индексы	$O(N)$	$O(1)$	19.5%

Классификация

Классификация методов построения индексов для проверки существования ключа в наборе данных

Метод	Сложность	Размер, МБ
Фильтр Блума	$O(k)$	2.04
Обученные индексы	$O(1)$	1.31

Заключение

В ходе данной работы:

- проведен анализ предметной области;
- описаны методы построения индексов в базах данных;
- предложены и обоснованы критерии оценки качества описанных методов и проведено сравнение методов по предложенным критериям оценки.

Поставленная цель достигнута.