

Сравнительный анализ алгоритмов расщепления вершин в R-дереве

Выполнил: Чередник К.Е.

Научный руководитель: Чернышев Г.А.

Рецензент: Новиков Б.А.

СПбГУ

2014 г.

Многомерное индексирование

- OLAP
- Геоинформационные системы

Древовидные структуры

- B⁺-дерево
- Семейство R-деревьев

GiST (Generalized Search Tree)

Постановка задачи

Сравнение поведения алгоритмов расщепления вершин в структурах, основанных на R-дереве

- Провести обзор модификаций R-дерева
- Проанализировать возможности встраивания дополнительных структур в GiST
- При необходимости, модифицировать GiST
- Провести экспериментальное сравнение изученных структур

Семейство R-деревьев

- R-дерево (1984)
 - Al-Badarneh's Split (2010)
 - Ang-Tan linear Split (1997)
 - Corner-based Split (2014)
 - Double sort Split (2012)
 - Greene's Split (1989)
 - Guttman's linear Split (1984)
 - Guttman's quadratic Split (1984)
 - K-means Split (2002)

- Гильбертово R-дерево (1994)

- Revised R*-дерево (2009)

Шаблон для реализации поисковых деревьев

- Структура Predicate
- 6 методов:
 - Consistent
 - Union
 - Compress
 - Decompress
 - Penalty
 - PickSplit

- Структура `NodeInfo`
- Методы:
 - Добавлен `UpdateNodeInfo` — сохранение информации о состоянии вершины
 - Изменена сигнатура `PickSplit` — добавление возможности оперировать историей изменения вершины
 - `Penalty` → `FindOptimalEntry` — обобщение алгоритма выбора поддерева

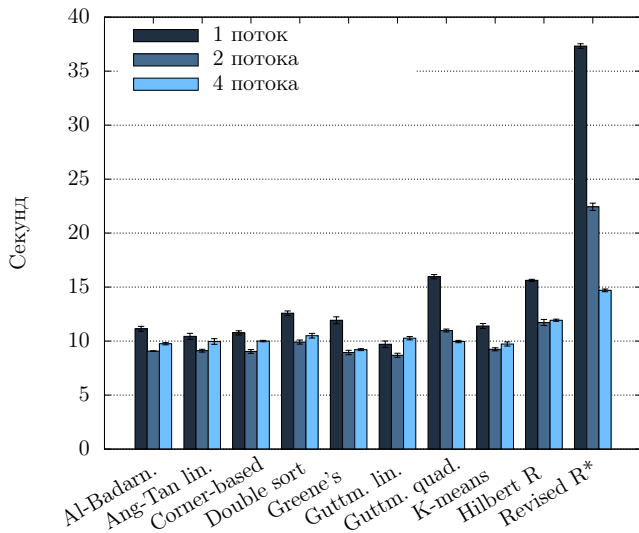
Варьируемые параметры:

- Размерность пространства (2, 4, 6, 8)
- Распределение (равномерное, нормальное, Парето)
- Количество потоков (1, 2, 4)
- Тип запросов (на диапазон, точечные)

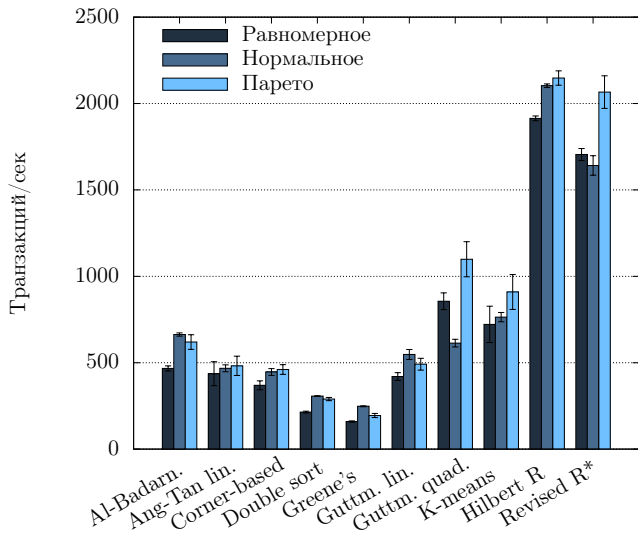
Конфигурация стенда:

Процессор	Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz
ОЗУ	8 GB
ОС	x86_64 GNU/Linux, ядро 3.8.0-26-generic

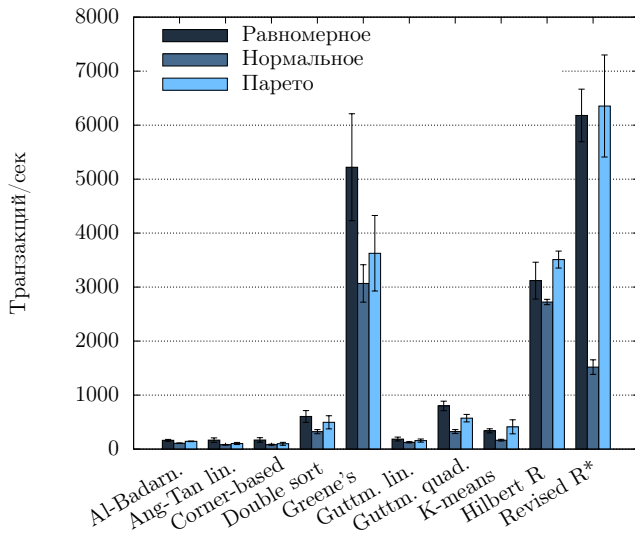
Время построения



Запросы на диапазон



Точечные запросы



- Проведен обзор основных вариаций R-дерева
- Проведен анализ возможности встраивания изученных деревьев в GiST
- Проведена модификация GiST для поддержки дополнительных деревьев поиска
- Реализованы все изученные деревья и проведено экспериментальное сравнение их производительности