

# Реализация нового метода индексирования данных

Использование потоковых  
процессоров для анализа больших  
объемов слабо структурированной  
информации.

# Прошлогодняя работа

- Использование потоковых процессоров для анализа больших объемов слабо структурированной информации на базе NVidia CUDA.

# Turtle

- Инновационный метод индексирования данных. (Запатентован в США 2011 году)
- Единственный метод индексирования данных созданный в 21 веке
- Совмещает в себе достоинства индексирования на базе Hash Table и деревьев

# Достоинства.

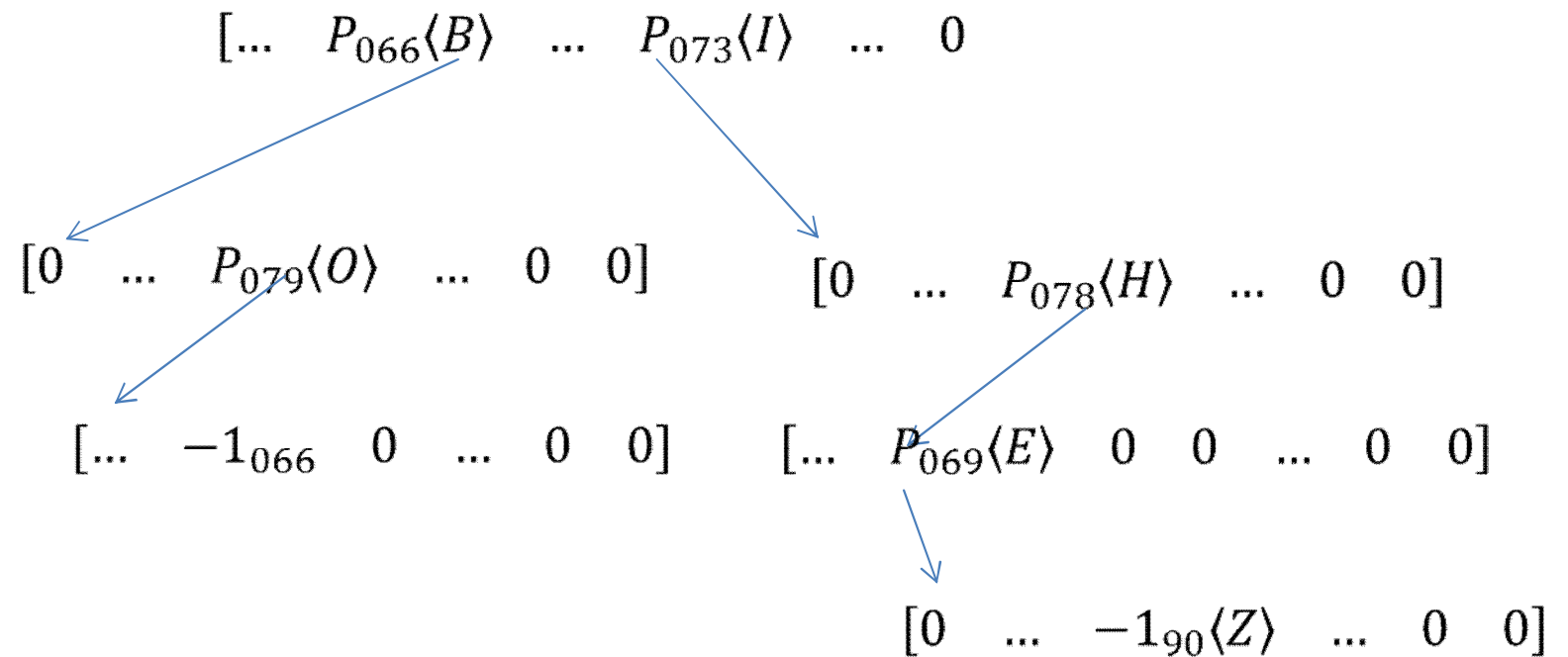
- Быстрый поиск по индексу (скорость не уступает лучшим методам на базе Hash tables)
- Возможность извлечения сортированных списков (как это получается в случае использования различных деревьев)

# Недостатки

- Алгоритм можно отнести к классу «жадных», хотя рост памяти идет логарифмически

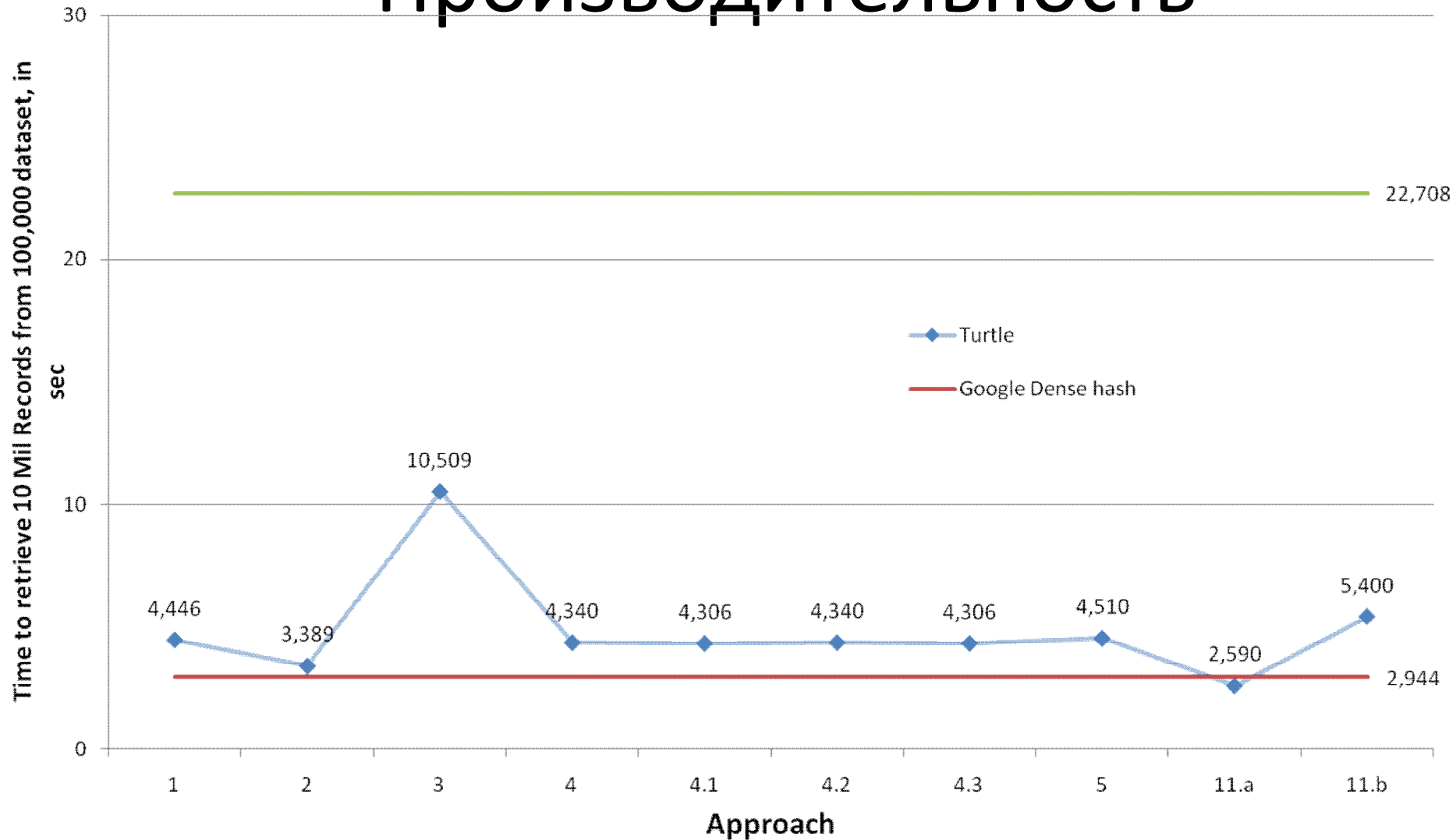
# Пример структуры дерева в простейшей имплементации

INEZ

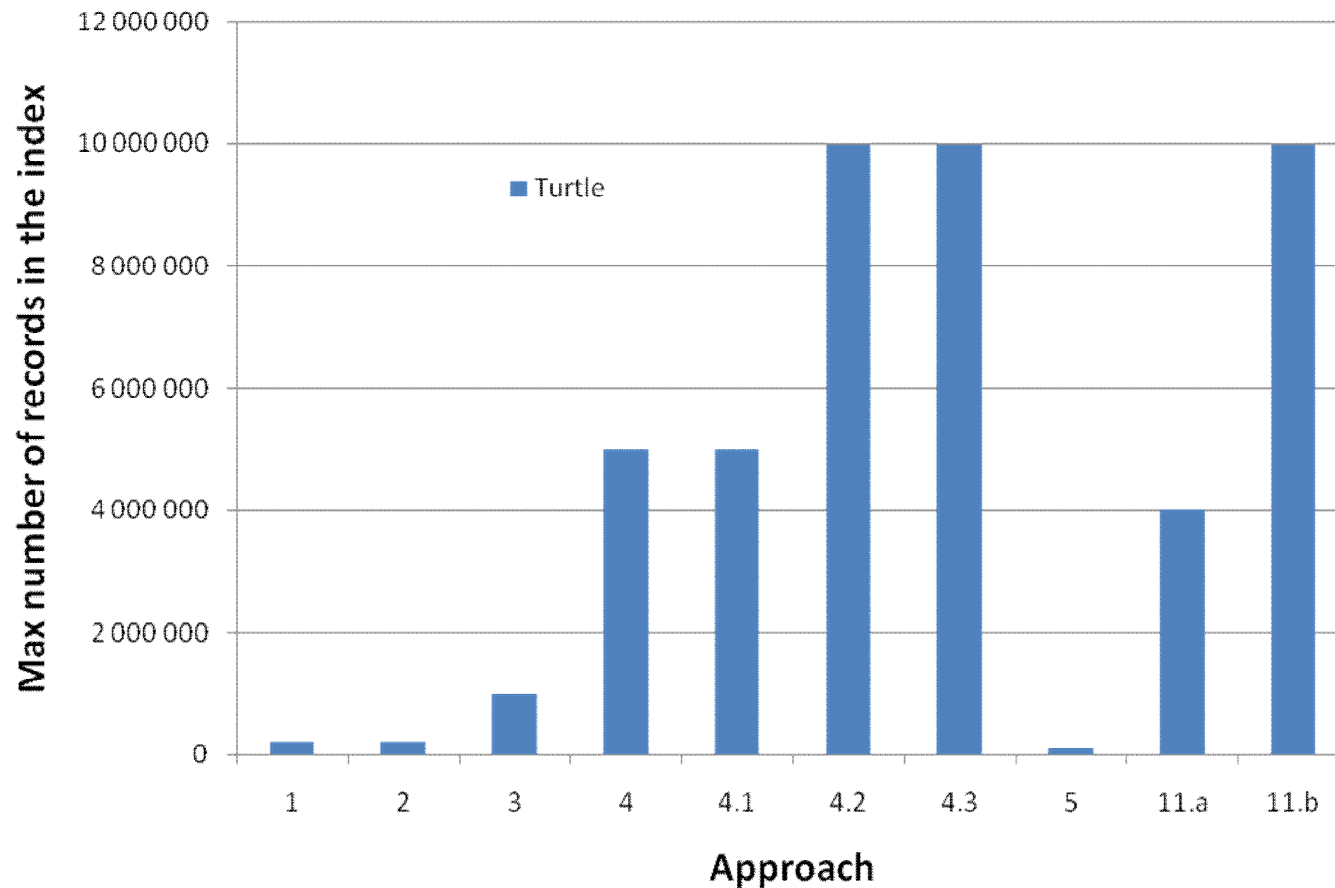


# Текущее состояние проекта

## Производительность



# Текущее состояние проекта (Память)





# Задачи, стоящие перед разработчиками

- Исследование производительности модификаций метода при индексировании различных данных.
- Реализация многопоточности для извлечения и добавления и удаления данных.
- Разработка механизма обеспечения целостности индекса при изменениях
- Разработка подходов для реализации условия поиска типа “like” для этого алгоритма.
- Реализация алгоритма, показавшего лучшие результаты на Asm