

# **Алгоритмы предварительной подгрузки данных для рабочей нагрузки с большой долей случайных запросов в СХД**

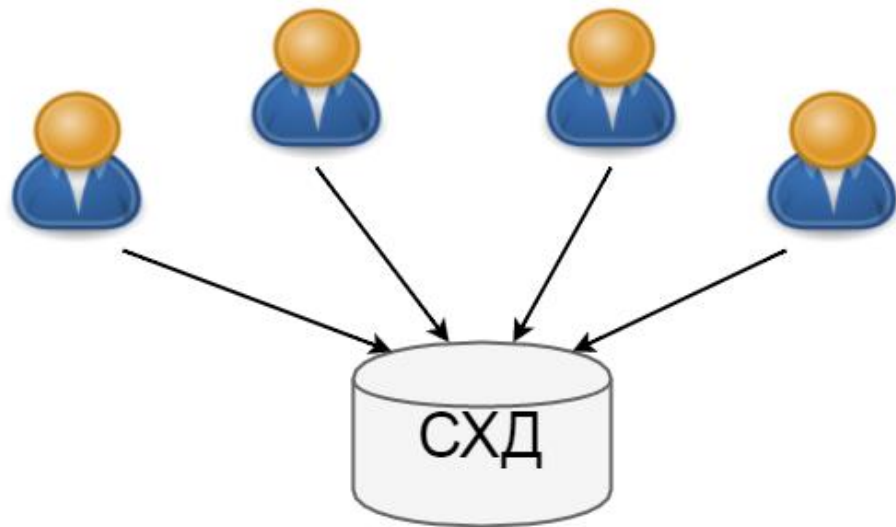
Бутрова А.С., 344

Научный руководитель: к.т.н. С.В.Лазарева

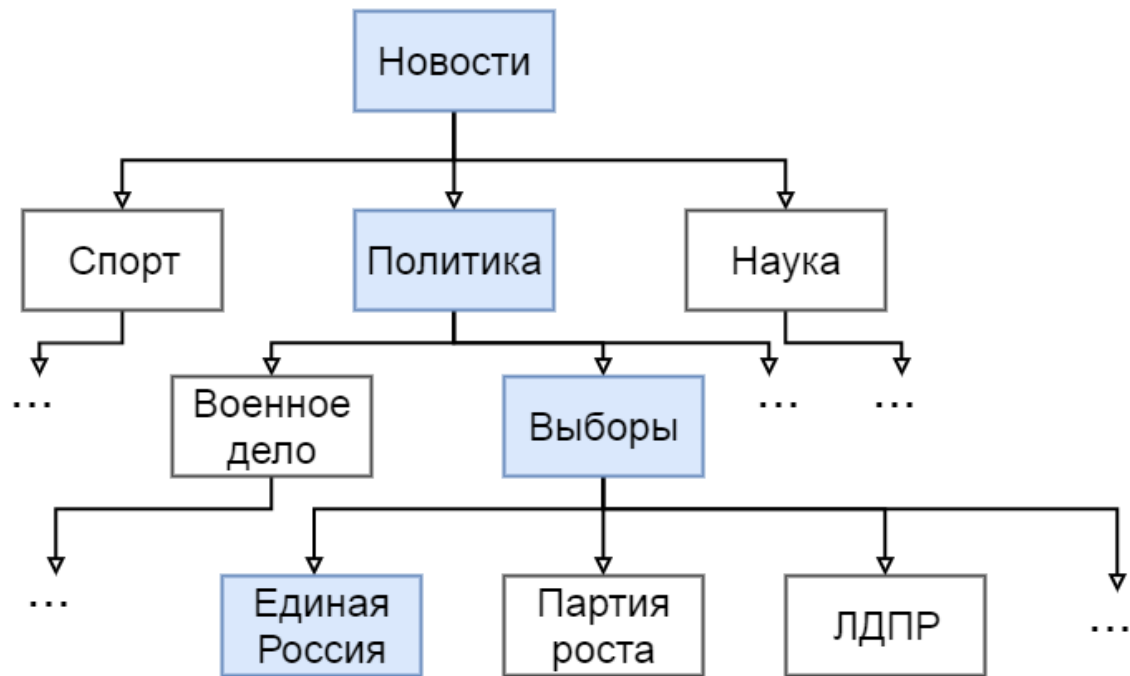
# Производительность СХД

Количество информации  
Количество пользователей

Полоса пропускания  
Время доступа



# Технология предварительной подгрузки



## Цели:

- Исследовать существующие алгоритмы предварительной подгрузки
- Разработать алгоритм для СХД с блочным доступом
- Реализовать модель алгоритма
- Произвести тестирование

# Обзор решений

- На основе теории вероятности:

Предварительное чтение на основе модели Маркова

- На основе анализа данных:

C-Miner

# C-Miner

Основывается на алгоритме CloSpan.

Семейство алгоритмов анализа данных:

- Apriori
- PrefixSpan
- CloSpan
- GSP

# Ассоциативный анализ

Правила:  $a, b \rightarrow c$

История запросов: 1, 2, 4, 2, 162, 32

Поддержка последовательности – количество раз, которое встретилась данная последовательность.

Частая последовательность - последовательность, которая имеет поддержку не меньше установленной.

# Алгоритм

1. Заполнение истории.
2. Нахождение частых единичных последовательностей и их поддержек.
3. Генерация кандидатов и подсчёт их поддержек.
4. Повтор шага 3, пока появляются новые кандидаты.



# Классификации ассоциативных алгоритмов

- Многомерные шаблоны
- Шаблоны на основе ограничений
- Шаблоны с временными интервалами
- Замкнутые шаблоны

# Классификации ассоциативных алгоритмов

## Apriori Based

- Обход в ширину
- Генерация и тестирование
- Многократный обход истории

## Pattern Growth Base

- Обход в глубину
- Проектирование дерева
- Отсечение кандидатов
- Разбиение истории

# Классификации ассоциативных алгоритмов

- Антимонотонность.

У любой частой последовательности любая подпоследовательность частая.

- Префикс-монотонность.

Префикс – любая подпоследовательность, с которой начинается данная последовательность.

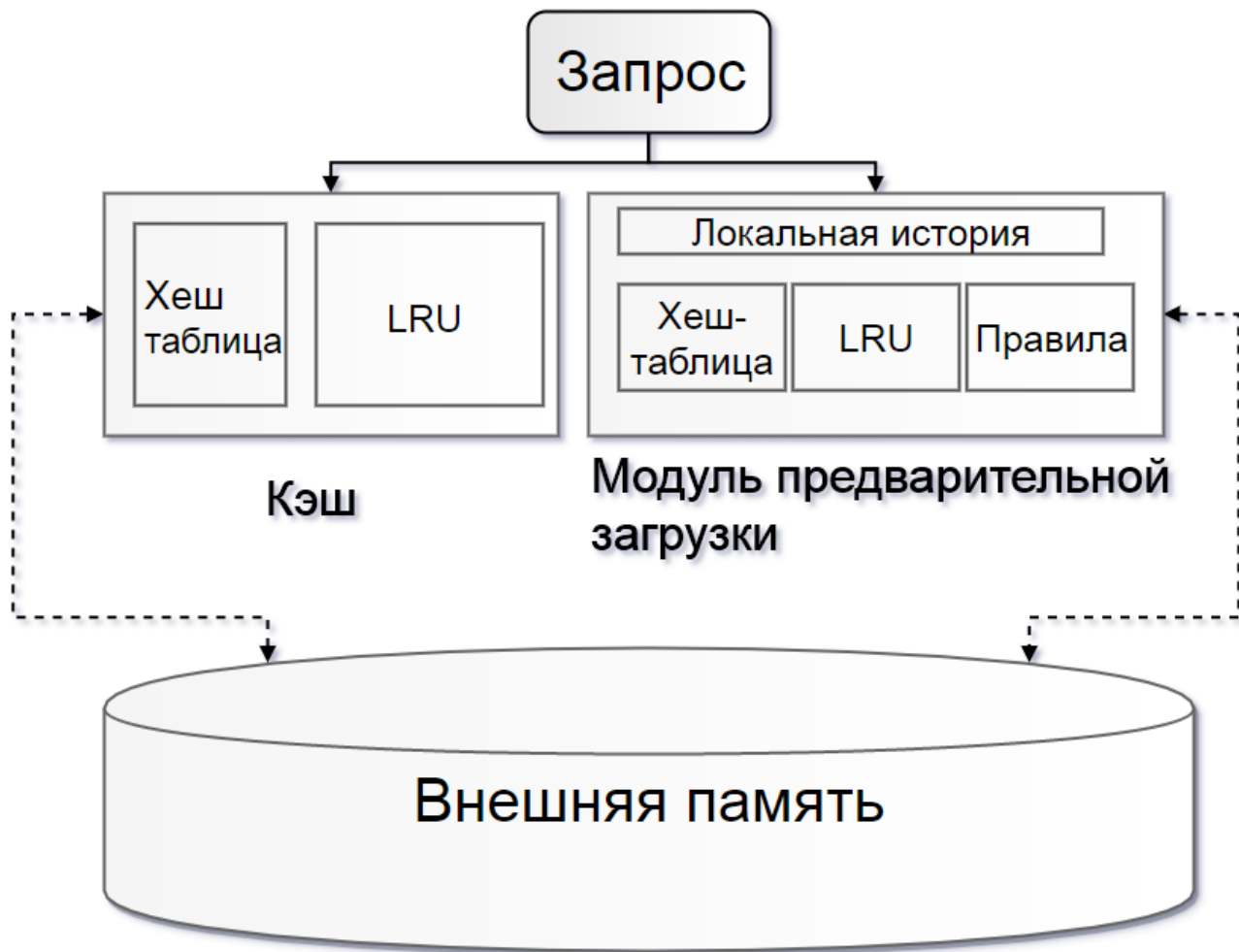
У любой частой последовательности любой префикс является частой последовательностью.

# Выбор алгоритма

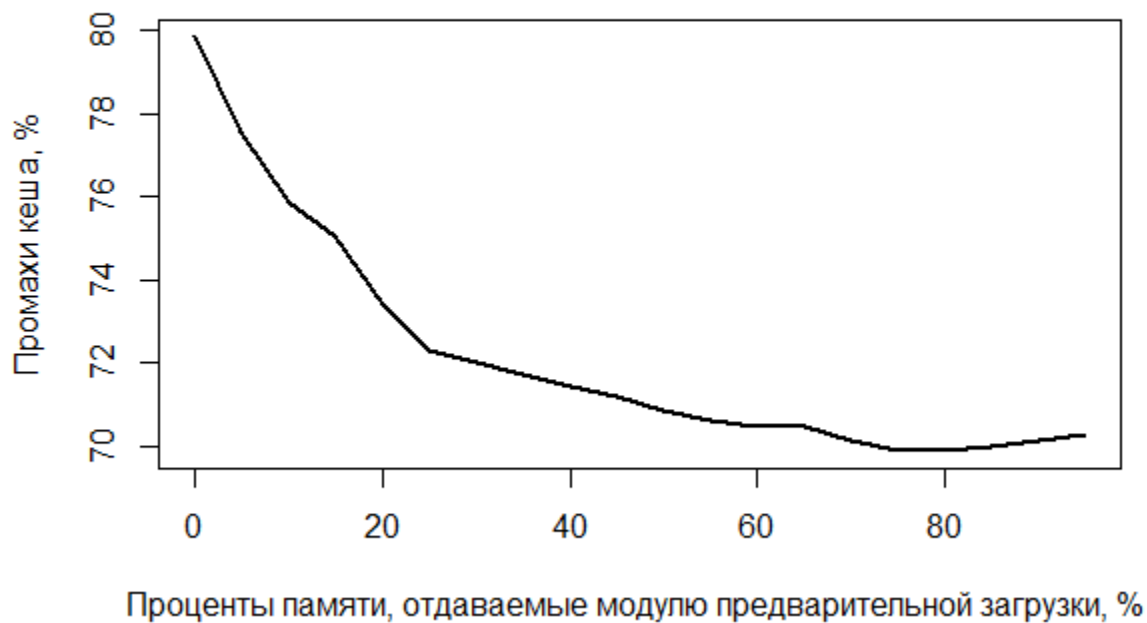
- Шаблоны с временными интервалами и замкнутые шаблоны
- Отсечение кандидатов.
- Разбиение истории.
- Антимонотонность, префикс-монотонность.
- Обход в ширину.

# Применяемый алгоритм

1. Создать историю и разбить на окна.
2. Найти единичные частые последовательности и их поддержку. Удалить из истории нечастые элементы.
3. Генерация кандидатов: составление кандидатов из имеющихся частых последовательностей. Проверить их на антимонотонность.
4. Посчитать поддержку кандидатов и удалить нечастые.
5. Повторять 3 и 4, пока появляются частые кандидаты.



# Результаты



# Результаты





# Итоги

- Изучены существующие алгоритмы предварительной подгрузки
- Разработан алгоритм для СХД с блочным доступом
- Реализована модель алгоритма
- Произведено тестирование