



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Факультет компьютерных наук
Департамент программной инженерии

КОМПОНЕНТ-РАСШИРЕНИЕ РСУБД SQLITE ДЛЯ ИНДЕКСИРОВАНИЯ ДАННЫХ МОДИФИКАЦИЯМИ В-ДЕРЕВЬЕВ

SQLITE RDBMS EXTENSION FOR DATA INDEXING USING B-TREE MODIFICATIONS

Выпускная квалификационная работа

Исполнитель: Ригин А.М., студент группы БПИ153

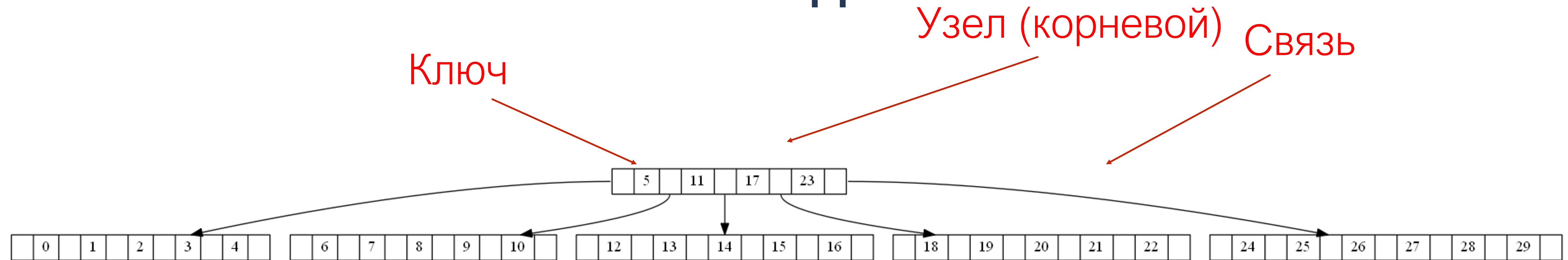
Научный руководитель: Шершаков С.А.,
ст. преп. ДПИ ФКН, н.с. НУЛ ПОИС ФКН

Москва, 2018

ОПИСАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ

- Сильно ветвящиеся деревья: В-деревья, V^+ -деревья, V^* -деревья и V^{*+} -деревья
- Эмпирическая оценка сложности основных операций и объёма занимаемой памяти
- Индексы в СУБД
- SQLite – встраиваемая реляционная СУБД
- Использование сильно ветвящихся деревьев в качестве индекса в SQLite

ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ



Пример B-дерева порядка 6, высоты 2

Узел (листовой)

ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

- **Сильно ветвящееся дерево** (структура данных) – такое дерево, которое содержит в одном узле более одного элемента с ключом и более одной связи с дочерними узлами.
- **В-дерево** – Сильно ветвящееся дерево. В-дерево построено так, что если некоторый узел содержит k ключей, то у данного узла $k + 1$ потомков, и для любого i , такого, что $1 \leq i \leq k + 1$, верно, что все ключи в i -м потомке данного узла не меньше, чем i -й ключ данного узла, и не больше, чем $i + 1$ -й ключ данного узла [6].
- **Порядок В-дерева** – такое число t , что для любого некорневого узла дерева верно неравенство: $t - 1 \leq k \leq 2t - 1$, где k – число ключей в узле. Корневой узел для непустого В-дерева содержит $1 \leq k \leq 2t - 1$ ключей, для пустого В-дерева – 0 ключей [6].
- В-дерево является **сбалансированным деревом**, поэтому его высота будет равна $O(\log_t n)$, где n – число ключей в дереве [6].

[6] Т. Кормен, Ч. Лейзерсон, Р. Ривест, К. Штайн. Алгоритмы: построение и анализ. 3-е изд.

ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

- **B^+ -дерево** – модификация B -дерева. В B^+ -дереве настоящие ключи хранятся лишь в листьях дерева, а во внутренних узлах хранятся лишь ключи-маршрутизаторы, необходимые для поиска по дереву. Листья в B^+ -дереве содержат $t \leq n \leq 2t$ ключей, где t – порядок дерева, ограничения для внутренних узлов такие же, как и в B -дереве. Деление листьев происходит поровну на две части, крайний ключ из левой половины делимого узла копируется в родительскую вершину в качестве ключа-маршрутизатора аналогично перемещению медианы для обычного деления, деление внутренних узлов происходит так же, как и в B -дереве [3].
- **B^* -дерево** – модификация B -дерева. Каждый узел заполняется не менее, чем на $2/3$, а не $1/2$. По этой причине, вместо традиционного разбиения узла, происходит

перераспределение ключей между соседними узлами-потомками, либо, если нет незаполненных соседей, то узел разбивается на три (а не на две) части [1] [5].

- **B^{*+} -дерево** – модификация B -дерева, **разработанная в рамках выполнения курсовой работы за 3 курс**. Представляет собой совмещение B^+ -дерева и B^* -дерева: модификация заполненных вершин при вставке выполняется как в B^* -дереве (переливание либо разбиение на три части), при этом, как в B^+ -дереве, реальные данные хранятся только в листьях, в остальных вершинах находятся лишь ключи-маршрутизаторы.

[1] <https://xlinux.nist.gov/dads/HTML/bstartree.html>

[3] Kerttu Pollari-Malmi. B^+ -trees:

<https://www.cs.helsinki.fi/u/mluukkai/tirak2010/B-tree.pdf>

[5] Д. Кнут. Искусство программирования. Том 3. 2-е изд.

ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

- **СУБД** – система управления базами данных
- **РСУБД** – реляционная СУБД
- **SQLite** – РСУБД с открытым исходным кодом (написана на языке C)
- **Расширение SQLite** – библиотека динамического подключения, расширяющая функционал РСУБД SQLite и предоставляющая новые функции [4]

АКТУАЛЬНОСТЬ ТЕМЫ И СУЩЕСТВУЮЩИЕ РЕШЕНИЯ

- **Актуальность темы**

- ✓ В настоящее время растут объёмы обрабатываемых данных
- ✓ Необходимо разрабатывать новые эффективные подходы к индексации данных в СУБД
- ✓ СУБД SQLite содержит небольшое число способов индексирования данных – актуально добавление новых

- **Существующие решения**

- ✓ В-дерево является способом индексации по умолчанию в SQLite
- ✓ Существует ряд расширений для SQLite, добавляющих, например, индексирование при помощи R-дерева
- ✓ Расширений с B^+ -деревом, B^* -деревом и B^{*+} -деревом не обнаружено

ЦЕЛЬ И ЗАДАЧИ РАБОТЫ

- **Цель работы**
 - ✓ Разработать расширение для SQLite, позволяющее использовать в качестве индекса в данной СУБД модификации В-дерева: B^+ -дерево, B^* -дерево и B^{*+} -дерево
- **Задачи работы**
 - ✓ Реализовать API на C для имеющейся C++-библиотеки алгоритмов над сильно ветвящимися деревьями
 - ✓ Разработать интерфейс подключения данной библиотеки в качестве расширения к SQLite
 - ✓ Разработать алгоритм выбора лучшей структуры данных в качестве индекса
 - ✓ Написать расширение для SQLite-менеджера для визуализации В-деревьев и их модификаций, а также рассчитанных для них метрик



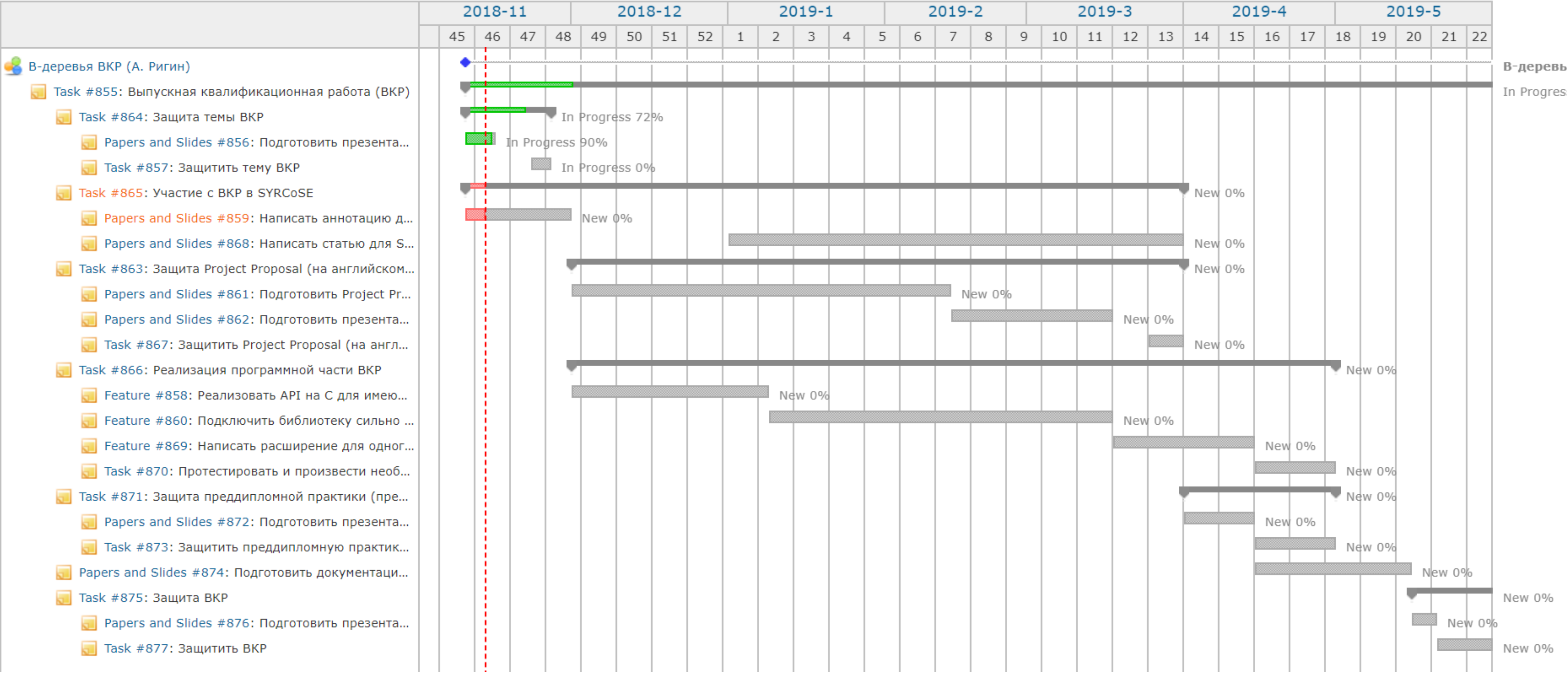
ЦЕЛЬ И ЗАДАЧИ РАБОТЫ

Департамент программной инженерии ФКН

Subtasks			
Task #863: Защита Project Proposal (на английском языке) для ВКР	New	Anton Rigin	
▶ Papers and Slides #861: Подготовить Project Proposal (на английском языке) для ВКР	New	Anton Rigin	
▶ Papers and Slides #862: Подготовить презентацию к защите Project Proposal (на английском языке) для ВКР	New	Anton Rigin	
▶ Task #867: Защитить Project Proposal (на английском языке) для ВКР	New	Anton Rigin	
Task #864: Защита темы ВКР	In Progress	Anton Rigin	
▶ Papers and Slides #856: Подготовить презентацию к защите темы ВКР	In Progress	Anton Rigin	
▶ Task #857: Защитить тему ВКР	In Progress	Anton Rigin	
Task #865: Участие с ВКР в SYRCoSE	New	Anton Rigin	
▶ Papers and Slides #859: Написать аннотацию для SYRCoSE	New	Anton Rigin	
▶ Papers and Slides #868: Написать статью для SYRCoSE	New	Anton Rigin	
Task #866: Реализация программной части ВКР	New	Anton Rigin	
▶ Feature #858: Реализовать API на C для имеющейся C++-библиотеки сильно ветвящихся деревьев	New	Anton Rigin	
▶ Feature #860: Подключить библиотеку сильно ветвящихся деревьев в качестве расширения к SQLite	New	Anton Rigin	
▶ Feature #869: Написать расширение для одного из менеджеров SQLite для более удобного просмотра различных метрик, связанных с B-деревьями, и самих деревьев	New	Anton Rigin	
▶ Task #870: Протестировать и произвести необходимые доработки	New	Anton Rigin	
Task #871: Защита преддипломной практики (предзащита ВКР)	New	Anton Rigin	
▶ Papers and Slides #872: Подготовить презентацию к защите преддипломной практики (предзащите ВКР)	New	Anton Rigin	
▶ Task #873: Защитить преддипломную практику (сдать предзащиту ВКР)	New	Anton Rigin	
Papers and Slides #874: Подготовить документацию к ВКР	New	Anton Rigin	
Task #875: Защита ВКР	New	Anton Rigin	
▶ Papers and Slides #876: Подготовить презентацию к защите ВКР	New	Anton Rigin	
▶ Task #877: Защитить ВКР	New	Anton Rigin	



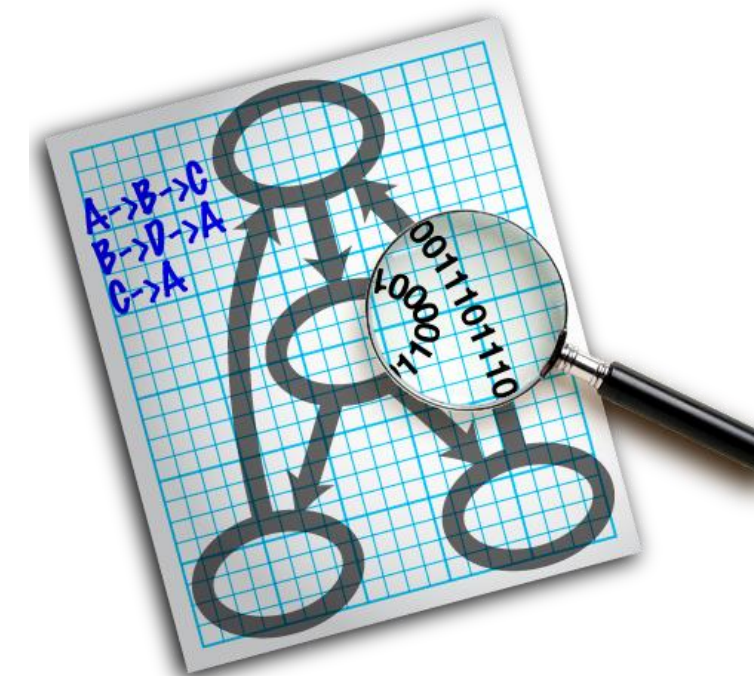
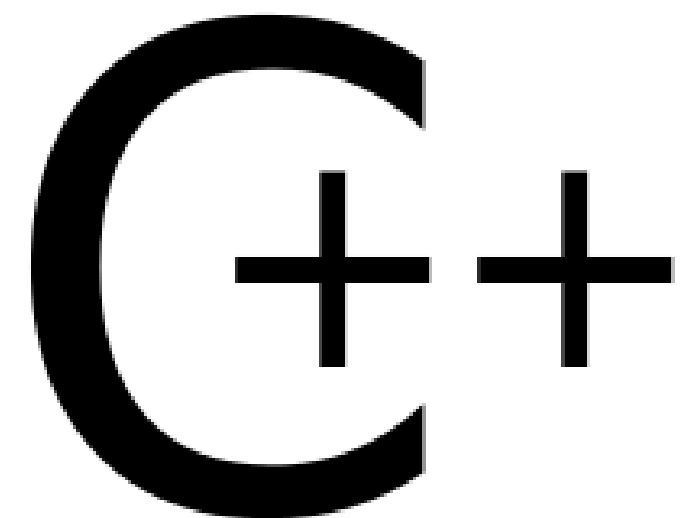
Департамент программной инженерии ФКН



МЕТОДЫ И ТЕХНОЛОГИИ

- Реализация межъязыкового адаптера для организации взаимодействия между компонентами на С и С++
- Описанные в документации SQLite способы подключения расширений к данной СУБД
- Визуализация В-деревьев и их модификаций

МЕТОДЫ И ТЕХНОЛОГИИ



ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ

- Разработанное расширение для СУБД SQLite, позволяющее использовать в качестве индекса модификации В-дерева: B^+ -дерево, B^* -дерево и B^{*+} -дерево
- Разработанное расширение для SQLite-менеджера для визуализации В-деревьев и их модификаций, а также рассчитанных для них метрик

ХОД ВЫПОЛНЕНИЯ

- **Выполнено**
 - ✓ Реализовано API на C для имеющейся C++-библиотеки алгоритмов над сильно ветвящимися деревьями
- **Осталось выполнить**
 - ✓ Разработать интерфейс подключения данной библиотеки в качестве расширения к SQLite (**в процессе** – до 17.03.2019 г.)
 - ✓ Разработать алгоритм выбора лучшей структуры данных в качестве индекса (до 17.03.2019 г.)
 - ✓ Написать расширение для SQLite-менеджера для визуализации B-деревьев и их модификаций, а также рассчитанных для них метрик (до 14.04.2019 г.)



СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. B*-tree [Электронный ресурс] // NIST Dictionary of Algorithms and Data Structures. Режим доступа: <https://xlinux.nist.gov/dads/HTML/bstartree.html>, свободный. (дата обращения: 23.12.2017)
2. B-Star Trees [Электронный ресурс] // College of Engineering, Computer Science & Construction Management, Chico, California State University. Режим доступа: http://www.ecst.csuchico.edu/~mjstapleton/courses/Fall2007/CSCI311/ProgTwo_Bstar.htm, свободный. (дата обращения: 15.01.2018)
3. Kerttu Pollari-Malmi. B+-trees // [Электронный ресурс]: Computer Science | University of Helsinki. Режим доступа: <https://www.cs.helsinki.fi/u/mluukkai/tirak2010/B-tree.pdf>, свободный. (дата обращения: 07.12.2017)
4. Run-Time Loadable Extensions [Электронный ресурс] // SQLite. Режим доступа: <https://www.sqlite.org/loadext.html>, свободный (дата обращения: 04.11.2018)
5. Д. Кнут. Искусство программирования. Том 3. 2-е изд. — М.: ИД «Вильямс». — 2017. — 824 с.
6. Т. Кормен. Алгоритмы: построение и анализ. 3-е изд. / Т. Кормен, Ч. Лейзерсон, Р. Ривест, К. Штайн. — М.: ИД «Вильямс». — 2013. — 1324 с.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Спасибо за внимание!

amrigin@edu.hse.ru

anton19979@yandex.ru

anton19979@yandex-team.ru