

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Сибирский государственный университет телекоммуникаций и
информатики»
(СибГУТИ)

02.03.02 Фундаментальная информатика
и информационные технологии
Профиль: Системное программное
обеспечение
(очная форма обучения)

ОТЧЕТ ПО УЧЕБНОЙ ПРАКТИКЕ

в/на _____
(наименование профильной организации/структурного подразделения СибГУТИ)

ТЕМА ИНДИВИДУАЛЬНОГО ЗАДАНИЯ

Выполнил: Григорьев Ю.В.
студент института ИВТ
гр. ИС-142
«27» мая 2023 г.

_____/_____
(подпись)

Проверил:
Руководитель от СибГУТИ
«27» мая 2023 г.

_____/_____
(подпись)

Новосибирск 2023

План-график проведения _____ учебной _____ практики

Вид практики

Григорьев Юрий Вадимович

Фамилия Имя Отчество студента

института Информатика и вычислительная техника, 2 курса, гр.
ИС-142

Направление: 02.03.02 Фундаментальная информатика и информационные технологии

Код – Наименование направления (специальности)

Профиль: Системное программное обеспечение

Место прохождения практики _____

Объем практики: **108/3** часов/ЗЕ

Вид практики **учебная**

Тип практики **научно-исследовательская работа (получение первичных навыков научно-исследовательской работы)**

Срок практики с **"30" января** 2023 г.
по **"27" мая** 2023 г.

Содержание практики*:

Наименование видов деятельности	Дата (начало – окончание)
1. Общее ознакомление со структурным подразделением предприятия, вводный инструктаж по технике безопасности	30.01.2023–01.02.2023
2. Выдача задания на практику, деление студентов на группы (если необходимо), определение конкретной индивидуальной темы, формирование плана работ	02.02.2023–04.02.2023
3. Работа с библиотечными фондами структурного подразделения или предприятия, сбор и анализ материалов по теме практики	06.02.2023–11.02.2023
4. Выполнение работ в соответствии с составленным планом: 1. Разработка заголовочного файла с прототипами методов работы со структурой данных 2. Разработка Makefile для автоматизации компиляции программы с библиотекой структуры данных 3. Разработка структуры данных в отдельном файле 4. Разработка программы для тестирования полученной библиотеки для работы со структурой данных 5. Отладка и форматирование кода программы	13.02.2023 – 20.05.2023
5. Анализ полученных результатов и произведенной работы. Составление отчета по практике, защита отчета	22.05.2023–27.05.2023

*В соответствии с программой практики

Руководитель от СибГУТИ

«28» 01 2023г.

_____/_____
(подпись)

ЗАДАНИЕ НА ПРАКТИКУ

Реализовать программно, исследовать эффективность и описать структуру данных “R-tree”.

ВВЕДЕНИЕ

R-дерево (R-Tree) - это индексная структура для доступа к пространственным данным, предложенная Антонином Гуттманом (Калифорнийский университет, Беркли) в 1984 году. R-дерево допускает произвольное выполнение операций добавления, удаления и поиска данных без периодической переиндексации. При этом дерево получается сбалансированным, что является одним из важных свойств любой иерархической структуры данных.

СТРУКТУРА R-ДЕРЕВА

R-дерево – это сбалансированное по высоте дерево, сходное с B+-деревом, листовые узлы которого содержат ссылки на конечные объекты. Если индексная структура находится на жестком диске, то каждый узел соответствует дисковой странице. Структура разработана так, чтобы для пространственного поиска требовалось посещение как можно меньшего числа узлов. Индексная структура полностью динамическая – добавление и удаление может выполняться одновременно с поиском, и никакой периодической реорганизации структуры производить не нужно. Для организации такой индексной структуры используют пространственную базу данных, состоящую из набора записей, каждой из которых соответствует некоторый уникальный идентификатор. Этот идентификатор используют как средство ссылки на запись из индекса. В качестве идентификатора может выступать некоторое уникальное число или номер записи в файле (второй вариант предпочтительнее, так как работает быстрее, однако для него присущи некоторые недостатки, связанные с удалением записей из файла).

Если принять описанные условия, то каждый листовой узел дерева будет состоять из элементов, имеющих вид:

$[MBR, \text{указатель_на_объект}]$,

где *указатель_на_объект* ссылается на объект в памяти устройства, а *MBR* – это *n*-мерный прямоугольник, который является минимальным охватывающим прямоугольником для пространственного объекта, со сторонами параллельными осям координат. Обычно *MBR* задают в виде интервала размерности *n* с закрытыми концами $[a, b]$, где *n* - число размерностей (измерений). Внутренние узлы дерева содержат элементы, имеющие похожую структуру:

$[MBR, \text{ссылка_на_потомка}]$,

где *ссылка_на_потомка* – это адрес узла низшего уровня в R-дереве (дочернего по отношению к данному), все записи внутри которого покрываются прямоугольником *MBR*.

И листовые узлы, и внутренние представляют собой набор из элементов описанной структуры, и даже в простейшей реализации таких элементов должно быть больше одного. Обозначим *M* как максимальное число элементов в любом узле, а *m* – минимальное. Для реализации основных алгоритмов необходимо выполнение условия $m \leq M/2$.

R-дерево должно удовлетворять следующим требованиям:

1. Каждый узел дерева содержит не меньше *m* и не больше *M* записей. Исключение может составлять только корень.
2. Корень, если он не является листом, содержит как минимум двух потомков. Максимальное количество элементов в корне также ограничивается значением *M*.

3. Для каждой индексной записи листового узла *MBR* является минимальным прямоугольником, который полностью вмещает в себя пространственный объект, на который ссылается запись.

4. Для каждой индексной записи внутреннего узла дерева *MBR* является минимальным прямоугольником, охватывающим все *MBR* дочерних узлов.

5. Все листовые узлы дерева расположены на одном уровне (дерево является сбалансированным).

6. Каждый объект упоминается в дереве ровно один раз.

На рис. 1 показан пример структуры *R*-дерева и проиллюстрированы отношения ограничения и перекрытия, которые могут существовать между его прямоугольниками.

Имея представление о свойствах *R*-дерева, можно оценить его высоту при числе элементов N . Из свойств, описанных выше, следует, что каждый узел дерева содержит как минимум m потомков. Поэтому наибольшая высота *R*-дерева, содержащего N индексных записей, будет не больше $\lceil \log_m N \rceil - 1$. При этом максимальное число узлов в таком дереве будет равно $\lceil N/m \rceil + \lceil N/m^2 \rceil + \dots + 1$.

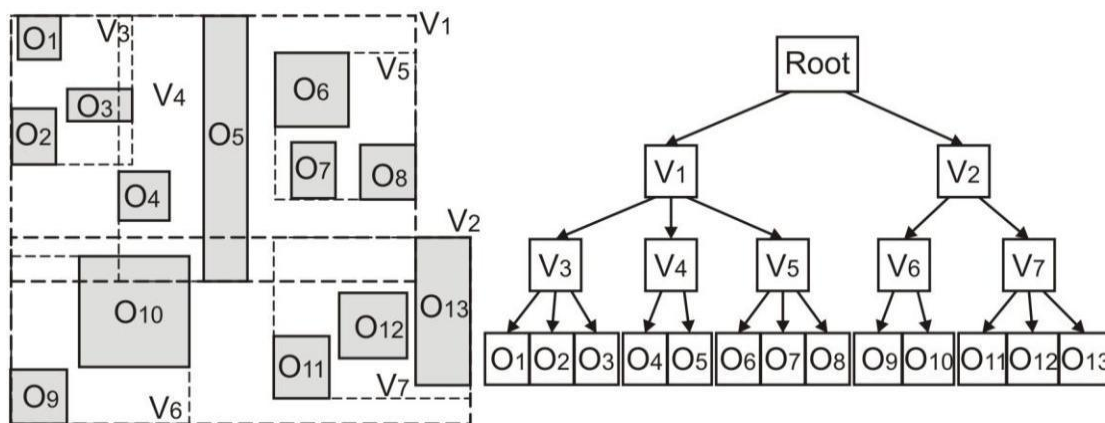


Рис. 1. Пример *R*-дерева

В худшем случае использование пространства памяти, в которой хранится индексная структура, будет m / M . Однако алгоритмы построения дерева разработаны таким образом, что структура будет стремиться содержать более m записей в узле. Это уменьшает высоту дерева и увеличивает полезное использование памяти.

АЛГОРИТМ РАБОТЫ

Поиск элемента

Алгоритм поиска в R -дереве похож на алгоритм поиска по B -дереву: он начинается в корне и опускается по нему к листовому узлу, выбирая в зависимости от заданных параметров поиска то или иное поддерево. Главное отличие состоит в том, что возможен вариант, при котором более одного поддерева текущего узла участвует в поиске: такая ситуация связана с применением метода размещения многомерных объектов, разрешающего пересекаться ограничивающим областям разных элементов. Это может привести к многократному уменьшению скорости поиска, однако алгоритмы построения и изменения дерева стараются поддерживать дерево в наиболее оптимальном виде. Пример рекурсивной процедуры поиска объектов, имеющих хотя бы одну общую точку с областью поиска S , в псевдокоде:

```

rtree_search (V, S, Res) // V - текущ. вершина, S - область поиска, Res - результаты поиска
  if V.kind != LEAF then
    for each V' in V // цикл по всем записям V' в узле V
      if V'.MBR * S != 0 then // MBR записи V' пересекается с S
        rtree_search (V', S, Res)
      end if
    end if
  end if
  if V.kind == LEAF then
    for each O in V
      if O.MBR * S != 0 then // MBR записи O пересекается с S
        Res += O
      end if
    end for
  end if
end
  
```

Рассмотрим описанный алгоритм на примере, показанном на рис. 2. Область поиска соответствует заданному прямоугольнику $ABCD$. Первоначально процедура поиска вызывается для корня. Так как корень является внутренней вершиной, то для него выполняется первая ветка алгоритма поиска. Она проверяет узлы V_1 и V_2 на пересечение с заданной областью. Как нетрудно заметить, оба этих узла имеют общие точки с областью поиска, и поэтому для обоих из этих узлов рекурсивно вызывается процедура **rtree_search**.

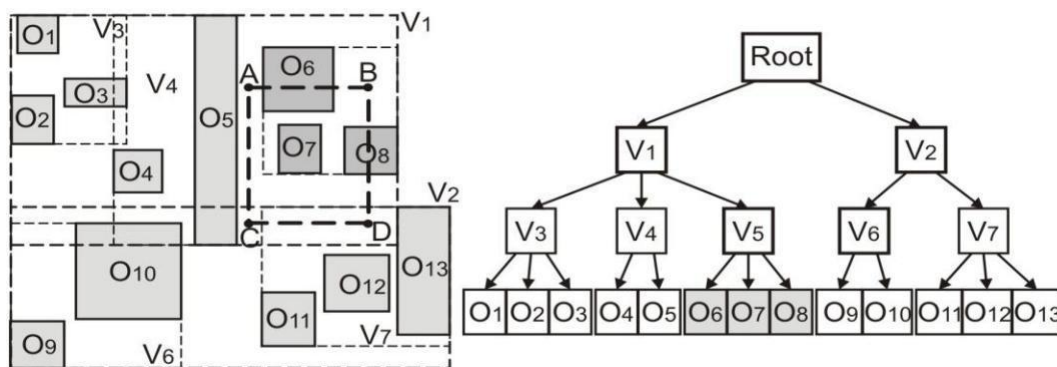


Рис. 2. Пример поиска в R -дереве

rtree_search для вершины V_1 перебирает элементы V_3, V_4, V_5 , причем только V_5 имеет пересечение с прямоугольником $ABCD$. Поэтому вершины V_3 и V_4 пропускаются и далее не рассматриваются. Дальнейший вызов процедуры для вершины V_5 выдаст в качестве результата

три элемента – O_6, O_7, O_8 , которые и будут добавлены в множество результата Res . Аналогичным образом будет просмотрена ветка V_2 . Из ее потомков только V_7 имеет общие точки с $ABCD$. Однако не один из элементов V_7 не пересекается с областью поиска. Данная ветка поиска оказалась ложной. В результате поиска получаем список элементов, удовлетворяющих заданному запросу:

$$Res = \{O_6, O_7, O_8\}.$$

Вставка элемента

Добавление нового объекта в R -дерево похоже на процедуру вставки в B^+ -дерево. Новая индексная запись добавляется в листовой узел, если узел переполняется, то происходит его деление, в результате которого у предка появляется еще один потомок. Если предок также оказывается переполненным, то и он делится дальше и так далее. Таким образом, вставка одного объекта может повлиять на структуру дерева в целом. Процедура вставки объекта представлена в следующем псевдокоде:

```
rtree_insert (O) // O - вставляемый объект
  L = choose_leaf (O)
  if L.count < M then
    L += O
    L'' = NULL
  else
    L'' = node_split (L, O)
  end if
  rtree_correct (L, L'')
end
```

Во-первых, процедура ищет листовой узел, в который необходимо поместить данный объект (шаг 1). Процедура поиска такого листа является важным шагом, так как неправильно выбранная позиция может привести к неэффективности структуры в целом. После того, как узел для вставки выбран, в нем размещается вставляемый объект (шаг 2). Если в листовом узле L есть место для новой записи, объект O помещается в него и процедура заканчивает свою работу. В противном случае, если узел L уже содержит максимально возможное число записей, то происходит деление узла на два новых L и L'' , которые содержат старые записи узла L и добавляемый объект O . После вставки объекта в дерево и возможного расщепления узла необходимо корректировать дерево (шаг 3). Эта процедура включает расширение границ MBR для текущего узла и всех его предков. Также эта процедура распространяется вверх по дереву при необходимости.

Рассмотрим алгоритмы упомянутых процедур подробнее в псевдокоде:

```
choose_leaf (O) // O - вставляемый объект
  V = root
  for each V' in V
    if V.kind == LEAF then
      return V
    end if
    Vnew = потомок, для которого MBR(V',O)-MBR(V') - минимальный
    V = Vnew
  end for
end
```

```
rtree_correct (L, L'') // L - корректируемая вершина, L'' - вершина от деления L
V=L, V''=L''
```

```

while V != tree->root then
    P = V.parent
    PV = запись в узле P о потомке V
    rtree_correct (PV)
    if V'' ≠ NULL then
        PV'' = новая запись о узле V''
        if P->count < M then
            P += PV''
        else
            P'' = node_split (P, PV'')
        end if
    end if
end if
V = P, V'' = P''
if V'' ≠ NULL then
    root = tree->root
    root.data[] = V, V''
end if
return
end while
end

```

Как было отмечено, процедура корректировки изменяет *MBR* всех вершин дерева, которые расположены выше листа с вставленным объектом. Второй и не менее важной функций процедуры корректировки является распространение деления вершин вверх по дереву, в случае, если будет происходить переполнение на внутренних узлах дерева. В качестве параметров в процедуру передаются два новых узла, которые получились при вставке объекта в дерево. Если разбиение не произошло, то первым параметром передается старый узел, а второй параметр приравнивается в *NULL*. На первом шаге процедура заносит переданные параметры в переменные *V* и *V''*. Эти переменные будут отвечать за текущие вершины в дереве, которые необходимо исправить. После этого происходит сравнение вершины *V* корня дерева. Если данная вершина является корнем, то это означает, что изменения уже распространились до верха дерева и необходимо просто завершить процедуру корректировки. Однако стоит учитывать один момент: если после предыдущих манипуляций произошло расщепление корня на два узла (переменная *V'' ≠ NULL*), то необходимо создать новый корень дерева, узлами-потомками которого будут *V* и *V''*. Если предыдущий пункт не выполнен, то происходит корректировка. Для этого определяется предок узла *V*, а также запись в нем об этом узле. После этого *MBR* найденной записи изменяется таким образом, чтобы включать в себя все *MBR* дочерних элементов узла *V*, но при этом не содержать лишних областей. Четвертый шаг алгоритма выполняется только в том случае, если предыдущие действия вызвали деление узла. В этом случае у нас в переменной *V''* будет находиться вершина с элементами, которые пока еще не помещены в дерево. Для этой вершины необходимо создать запись *P_{V''}*, которая будет содержать минимальный описывающий прямоугольник для данной вершины и ссылку на саму вершину. Эту запись и нужно разместить в предке узла *V*. Однако при помещении в узел *P* записи *P_{V''}* необходимо помнить, что данная операция может привести к переполнению и тогда придется разбивать узел *P* на два новых. После всех описанных операций в переменные *V* и *V''* заносятся новые значения *P* и *P''* соответственно, и алгоритм повторяется заново с шага 2.

Удаление элемента

Для того, чтобы структуру можно было считать динамической, необходима поддержка удаления уже существующих в дереве элементов, которая также должна корректировать дерево для сохранения его свойств.

```

rtree_delete (O) // O - удаляемый объект
V = root

```

```

L = rtree_search (V,O)
if L == NULL then
    return
end if
delete O
V = L, Q = NULL
if1 V == tree->root then goto if2 end if1
P = V->Parent
PV = запись в узле P о потомке V
if V->count < m then
    delete PV
    Q = V
    delete V
else
    rtree_correct (V)
end if
V = P
goto if1
if2 root->children->count == 1 then
    delete tree->root
    tree->root = root->child
end if2
tree->insert(Q[0], Q[1], ...)
end

```

Первое, что производит процедура удаления объекта O из R -дерева, это ищет листовой узел, в котором находится данный объект. Для этого используется процедура поиска `rtree_search`. В качестве параметров ей передается вершина, с которой нужно начать поиск (в нашем случае это корень) и объект поиска. Если объект не найден, то будет возвращён *NULL*. При этом необходимо завершить и процедуру удаления. На втором шаге удаляется объект O из узла L и подготавливаются временные переменные для коррекции дерева. В переменную V (текущая вершина для коррекции) заносится листовой узел L , а в переменную Q – пустое множество (это множество вершин, которые необходимо потом вставить в дерево заново). Далее необходимо проверить, является ли вершина V корнем. Если V – корневая вершина, то шаги 4–6 нужно пропустить и перейти сразу к седьмому пункту алгоритма. Иначе – находим предка для вершины V , и в нем определяем запись, ссылающуюся на V (P_V). Если в рассматриваемом узле число записей меньше минимально возможного (m), то необходимо удалить этот узел из дерева. При этом все элементы из V помещаются в множество Q (чтобы потом снова быть размещенными в дереве, но в других вершинах) и из вершины P удаляется элемент, ссылающийся на удаленную вершину (удаляется P_V). Если же записей в вершине V больше, чем заданный параметр m , то удалять вершину не нужно. При этом необходимо просто скорректировать *MBR* узла таким образом, чтобы он охватывал все прямоугольники дочерних узлов, но при этом не включал лишнего пространства (после удаления узлов вполне вероятно можно будет сузить *MBR*, который хранится в записи P_V). После проделанных операций необходимо распространить сделанные изменения вверх по дереву (скорректировать *MBR* узла предка или, возможно, даже удалить его, если он оказался не заполненным до предела m). Для этого в переменную V заносится предок текущей вершины и повторяется алгоритм с шага 3. После того, как все изменения дойдут до корня, алгоритм продолжится с шага 7. Исходя из свойств R -дерева, описанных в начале данного параграфа, корень должен иметь не меньше двух потомков. Поэтому необходимо просто проверить число дочерних узлов у корня и при нахождении там всего одного потомка сделать его новым корнем дерева. Последнее, что необходимо выполнить в процедуре удаления, это вставить временно удаленные узлы из множества Q обратно в дерево. Данная процедура выполняется полностью аналогично описанной ранее процедуре ВСТАВКА за одним лишь исключением: вершины из множества Q

необходимо разместить на тех же уровнях, на которых они были до процедуры удаления. Этого требования необходимо придерживаться для того, чтобы не нарушить сбалансированность дерева (одно из свойств *R*-дерева заключается в том, что все листовые узлы находятся в нем на одном уровне).

Разбиение узла

Изменение данных прикладной задачи требует частого изменения индексной структуры. Для добавления новой записи в уже заполненный узел *R*-дерева, содержащий M записей, необходимо распределить $M + 1$ элемент между двумя узлами. Процедура разбиения узла может быть вызвана не только при добавлении новых элементов в индекс, но и при перестройке дерева, при удалении ненужной записи, при обновлении данных или даже при его корректировке. Алгоритм, выполняющий деление узла, особенно важен, так как плохое разбиение может сильно затруднить операции поиска по дереву. Разбиение узла без учета критериев оптимальности построения дерева приводит к увеличению времени работы процедуры поиска конкретного объекта, а следовательно, к ухудшению работы индексной структуры в целом. При плохом разбиении узлы дерева разрастаются вдоль осей координат и захватывают много пространства, не содержащего ни одного объекта. Такой пример показан на рис. 3. С одной стороны вариант (а) обеспечивает нулевое перекрытие двух узлов дерева. Однако суммарная площадь этих узлов будет значительно больше самих узлов, что вызовет многократное ложное срабатывание процедуры поиска. При большой площади пространства, соответствующей узлам дерева, запросу поиска на промежуточных стадиях работы может удовлетворять большое число записей (более одной), хотя, в конечном счете, на каждом уровне интересует только одна. Следовательно, алгоритм будет ветвиться и обходить дерево неоптимальным путем, включая обход ненужных узлов, что может сильно отразиться на скорости работы индексной структуры. Кроме того, обход ненужных узлов потребует дополнительного расхода оперативной памяти. В условиях большого числа запросов это обстоятельство также может стать критичным.

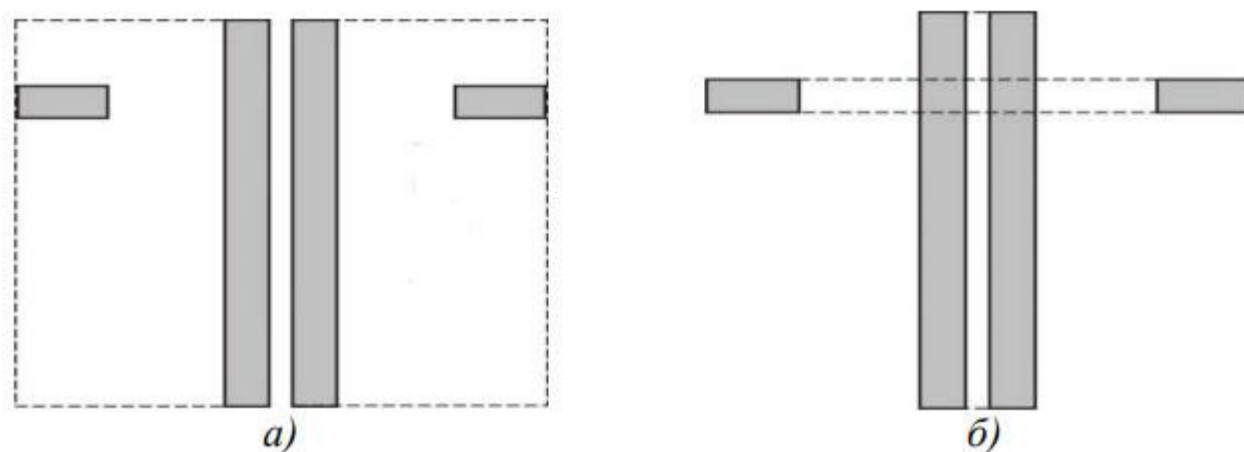


Рис. 3. Пример возможного разбиения узла на два новых:
а – “плохое” разбиение; б – “хорошее” разбиение с пересечением

Исчерпывающий перебор - самый простой из алгоритмов, при котором получается максимально оптимальное дерево: для нахождения минимальной площади покрытия и всех остальных параметров оптимальности деления необходимо произвести все возможные разбиения $M + 1$ записи на две группы и выбрать наилучшее. Задача разбиения в общем случае является NP-полной. Этот вариант применим без потери производительности при M не более 5, и именно поэтому данный подход применяется редко, а во всех вариантах построения *R*-деревьев используются эвристические подходы, дающие на реальных данных не всегда оптимальные разбиения (но практически всегда достаточно эффективные для решения конкретной задачи).

Квадратичный алгоритм разбиения был предложен основателем R-деревьев Антонином Гуттманом. В нём осуществлена попытка найти такое деление, при котором площадь охватывающих прямоугольников будет минимальна. Однако при этом не гарантируется, что это будет действительно наилучший вариант. Алгоритмическая сложность изменяется по квадратичному закону относительно M и по линейному – относительно числа измерений.

```

node_split (L,O) // O - вставляемый объект, L - разбиваемая вершина
  Q = L->children + O
  delete (L->children)
  L'' = NULL
  O1,O2 = choose_first (Q)
  L += O1, L'' += O2
  n = Q->size, n1 = L->size, n2 = L''->size
  if n == 0 then
    return L''
  end if
  if m - n1 ≥ n then
    L = Q
    return L''
  end if
  if m - n2 ≥ n then
    L'' = Q
    return L''
  end if
  O' = choose_next (L,L'',Q)
  d1 = MBR(L,O') - MBR(L)
  d2 = MBR(L'',O') - MBR(L'')
  if (d1 < d2) or (d1 == d2 && n1 < n2) then
    L += O'
  else
    L'' += O'
    node_split()
  end if
end

```

Алгоритм начинается с подготовительных операций (шаг 1). Все элементы, которые нужно будет распределить между двумя новыми вершинами, переносятся в множество Q . Создается еще одна пустая вершина L . Таким образом мы получаем две пустые вершины (L и L'') и множество элементов Q , которые необходимо распределить по этим вершинам. Затем алгоритм выбирает пару элементов, которые покрывают наибольшую площадь (соответствующая процедура представлена в листинге 3.8). Для этого рассчитывается коэффициент $A = MBR(O1,O2) - MBR(O1) - MBR(O2)$, где $MBR(O1,O2)$ – площадь прямоугольника, охватывающего обе записи, а $MBR(O1)$ и $MBR(O2)$ – площади прямоугольников соответствующих объектов. Этот коэффициент показывает неэффективность объединения двух данных объектов в одну группу. Элементы, на которых достигается максимум коэффициента A , становятся первыми элементами двух будущих групп.

```

choose_first (L,L'',Q) // Q - нераспред. элементы, L, L'' - вершины между которыми распредел.
  for each O1,O2 in Q
    A = MBR(O1,O2) - MBR(O1) - MBR(O2)
  end for
  O1,O2 = max(A)->nodes

```

```

Q := (O1 + O2)
return O1, O2
end

```

Оставшиеся записи распределяются в группы по одной (шаг 4 алгоритма **node_split**). Для этого вызывается процедура **choose_next**, на которую и возложена задача выбора элемента из множества Q, который будет распределен следующим. После выбора элемента для вставки он добавляется в вершину, ограничивающий прямоугольник которой потребует минимального увеличения площади (если d1 и d2 – увеличение площади ограничивающих прямоугольников при добавлении элемента O' в вершины L и L' соответственно, то вставку нужно произвести в ту вершину, для которой d меньше). При равенстве увеличения площадей (d1 = d2), для вставки выбирается та вершина, в которой меньше число записей. Описанное действие продолжается до тех пор, пока не будут выбраны все элементы из множества Q. Однако на каждом необходимо проверять выполнимость условия минимального наполнения узла (число элементов в любой вершине дерева, кроме корня, должно быть не меньше m). Если на каком-то шаге окажется, что для выполнения этого условия необходимо все оставшиеся в Q элементы переместить в одну из вершин, то необходимо сделать это и завершить процедуру деления узла (шаг 3).

```

choose_next (L,L'',Q) // Q - нераспред. элементы, L,L'' - вершины между которыми распредел.
  for each O' in Q
    d1 = MBR(L,O') - MBR(L)
    d2 = MBR(L'',O') - MBR(L'')
    a = |d1-d2|
  end for
  O = max(a)->node
  Q := O
  return O
end

```

Представленный алгоритм достаточно прост. Для каждого нераспределенного элемента рассчитывается площадь охватывающего прямоугольника, который получится после присоединения этого элемента к каждой группе (значения d1 и d2). Элемент с наибольшей разницей площадей обеих групп выбирается как следующий элемент. Существует несколько другая стратегия, предложенная позже: выбирается элемент, присоединение которого в какую либо группу минимально увеличивает площадь MBR группы.

```

choose_next (L,L'',Q) // Q - нераспред. элементы, L,L'' - вершины между которыми распредел.
  for each O' in Q
    d1 = MBR(L,O') - MBR(L)
    d2 = MBR(L'',O') - MBR(L'')
    d = min{d1,d2}
  end for
  O = min(d)->node
  Q := O
  return O
end

```

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ

R-дерево как структура индексирования пространственных объектов стало почти стандартом для промышленных СУБД, которые используют различные его варианты в качестве индексных структур. Однако при конкретной реализации R-деревьев в конечном приложении зачастую встает вопрос, какие параметры выбрать для лучшей её реализации и, соответственно, достижения лучших вычислительной/пространственной сложности. Временная сложность этих операций в большой нотации O зависит от количества записей данных в R-дереве (n), размерности данных, и максимального числа дочерних элементов в узле (M). Поиск, вставка и удаление (ключевые операции над структурой данных) имеют временную сложность в худшем случае $O(M \cdot \log M(n))$, в среднем - $O(\log M(n))$, в лучшем - $O(\log(n))$. Это связано с тем, что R-дерево представляет собой древовидную структуру данных, и поиск выполняется путем обхода дерева от корня до конечного узла, включая обход всех дочерних элементов в худшем случае. Высота дерева также логарифмично относится к количеству записей данных.

Как было отмечено ранее, одними из важнейших параметров для построения R-дерева являются минимально возможное (m) и максимально допустимое (M) количество элементов в узле. При выборе этих параметров необходимо руководствоваться следующими соображениями:

Максимальное число элементов в узле (M)

Чем больше значение M , тем сильнее будет ветвиться дерево, а следовательно, его глубина будет меньше. Если предположить, что индексная структура разрабатывается для внешней памяти, то уменьшение глубины дерева ведет к уменьшению обращений к диску (если учесть, что проверка узла дерева вызывает одно обращение к диску). Поэтому сильноветвящееся дерево (при большом M) будет более эффективным для внешней памяти. С другой стороны, процедура поиска вынуждена просматривать абсолютно все элементы вершины. Поэтому при очень большом M индексная структура может вырождаться просто к последовательному поиску. К тому же на сравнение с элементами вершины расходуется процессорная мощность. Поэтому чем больше M , тем больше нагрузка на процессор в процедурах поиска. Исходя из описанных фактов, можно сделать следующие выводы: если разрабатываемая индексная структура целиком размещается в оперативной памяти, то значение M стоит выбирать небольшим, порядка 4–10 элементов в вершине. Если же индексная структура хранится во внешней памяти, то значение M стоит вычислять по следующей формуле: $M = \text{Cluster} / \text{eSize}$, где Cluster – размер кластера жесткого диска (например, 512 или 1024 байт); eSize – размер одного элемента. Так, если один элемент занимает 16 байт, то в качестве верхней границы стоит взять $M = 32$ элементов в вершине.

Минимальное число элементов в узле (m)

Данный параметр зависит от M и, как было описано ранее, не может превышать $M / 2$. Минимальный же предел параметра m равен 2 (в узле не может быть меньше двух потомков, если это не корневой узел). Маленькое значение параметра m облегчает процедуру разделения узла, потому что исчезает необходимость повторной вставки элементов. В то же время маленькое значение нижней границы может привести к неэффективному использованию памяти. По исследованию А. Гуттмана, наименее плотные индексы могут потреблять приблизительно на 50% больше места, чем самые плотные. В практических применениях наиболее часто используемой операцией является процедура поиска элементов. Поэтому нижнюю границу заполнения узла стоит выбирать равной $M / 2$.

Центральным звеном при построении дерева также является процедура **разбиения узла пополам (node_split)**. От эффективности этой процедуры зависит оптимальность построения дерева в целом. При неоптимальной структуре дерева появляется неоднозначность поиска элементов. Возможны ситуации, когда уже на уровнях, близких к корню R-дерева, охватывающие прямоугольники пересекаются не по пустому множеству данных, что значительно усложняет процедуру поиска.

С проблемой качества изменения R-дерева можно бороться с помощью **«исчерпывающего»** алгоритма деления. Использование данного алгоритма для деления узла изменяет структуру R-дерева лучшим из возможных способов, что, конечно, отражается на дальнейшем поиске данных в лучшую сторону, но, в свою очередь, существенно замедляет работу индексной структуры. Применение данного алгоритма оправдано при малом числе записей в узле, а также в ситуациях, когда структура дерева редко меняется, т. е. при индексировании неподвижных (например, жилых домов, складов и т. д.) или малоподвижных пространственных объектов (например, небесной карты звезд).

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

Для выполнения моделирования обозначенной структуры данных были созданы файлы `rtree.c`, `rtree.h` как модуль для работы с R-деревьями и файл `main.c` как тестовая программа. В заголовочном файле `rtree.h` описаны прототипы функций и глобальные директивы препроцессора, задающие тип данных, с которым будет работать R-дерево, минимальное и максимальное количество элементов в узле дерева, количество измерений пространства, с которыми будет взаимодействовать пользователь при работе со структурой данных, а в `rtree.c` - реализация этих и скрытых от пользователя функций, которые необходимы для корректной работы и построения R-деревьев. Имея псевдокод описанных выше функций, было несложно реализовать их на языке программирования C. В качестве функции разбиения узла был выбран оригинальный квадратичный алгоритм Антонина Гуттмана, в качестве значений M - 64 элемента, m - 6 элементов (10% от M (относительная реализация) соответственно). Также был создан `Makefile` для автоматизации компиляции программы. В файле программы для тестирования реализации полученной структуры данных (**`main.c`**), была создана новая структура **`city`**, хранящая в себе название города и его широту-долготу, и несколько экземпляров этой структуры в качестве известных городов мира, располагающихся в разных частях света: Торонто, Новосибирск, Токио, Буэнос Айрес, Рио де Жанейро и Сидней. Эти экземпляры были переданы в новосозданное R-дерево, с которым далее были выполнены операции поиска по разным широте-долготе (северо-западные, северо-восточные, юго-западные и юго-восточные города относительно нулевого меридиана и экватора, где -180 - 0 градусы - это юг и запад в понимании широты и долготы, 0 - 180 градусы - север и восток соответственно) и удаления элементов из дерева (для примера был удалён Новосибирск, находящийся в северо-восточной четверть сфере Земли). R-дерево умеет находить всю область с заданными параметрами и выводить все объекты, находящиеся в ней. Чтобы работать R-дереву с новой структурой **`city`**, было достаточно передавать в функции работы с R-деревом указатель на нашу новую структуру, так как в **`rtree.h`** в качестве рабочего формата данных используется простой указатель - **`void *`**. Исходный код всего проекта можно увидеть в приложении 1.

РЕЗУЛЬТАТЫ РЕАЛИЗАЦИИ

Компиляция проекта:

```
1  ► Run make
4  gcc -Wall -Wextra -o main main.c rtree.c
```

Запуск проекта:

```
1  ► Run ./main
4
5  Northwestern cities:
6  Toronto
7
8  Northeastern cities:
9  Novosibirsk
10 Tokyo
11
12 Southwestern cities:
13 Buenos Aires
14 Rio de Janeiro
15
16 Southeastern cities:
17 Sydney
18
19 Northeastern cities after element deletion:
20 Tokyo
```

ЗАКЛЮЧЕНИЕ

В ходе проведения работы была изучена и смоделирована структура данных «R-дерево» и тестовая программа для работы с ней. Подводя итоги, можно сказать, что главные свойства R-дерева следующие:

1. Оно состоит из внутренних узлов, листовых узлов и единственного корня
2. Корень R-дерева содержит указатель на самую большую область в пространстве
3. Родительские узлы содержат указатели на свои дочерние узлы, чья совокупная область (их сумма) покрывает область родительского узла
4. MBR — важнейший параметр, обозначающий минимальную ограничивающую область (рамку/прямоугольник), окружающую рассматриваемую область/объект в пространстве
5. Листовые узлы содержат данные об MBR объектов, на которые они ссылаются

Преимущества R-деревьев над B+деревьями заключаются в том, что для построения B+дерева необходимы данные, которые можно составить в одном последовательном порядке. Это не всегда возможно, поскольку некоторые типы данных (например, географические координаты) не предполагают единого порядка, который можно было бы использовать для эффективного сканирования диапазона (например, всех точек в заданной области) по индексам, построенным с использованием B+Tree.

Обобщённо, преимущества R-деревьев заключаются в том, что эта структура данных эффективна для задач, включающих пространственную индексацию и поиск в двух или более измерениях. Примерами таких задач могут являться:

1. Поиск ближайшего соседа: поиск ближайшей точки к заданной точке в наборе данных.
2. Запросы диапазона: поиск всех точек в пределах заданного расстояния или площади от заданной точки.
3. Пространственное соединение: объединение двух наборов данных на основе их пространственной близости.
4. Кластеризация: группировка похожих точек на основе их пространственной близости.
5. Маршрутизация: поиск кратчайшего пути между двумя точками на карте.
6. Обработка изображений: Обнаружение и отслеживание объектов на изображениях на основе их пространственного положения.

В целом, любая задача, требующая быстрой и эффективной пространственной индексации и поиска, может выиграть от использования R-деревьев.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. *Томас Х. Кормен, Чарльз И. Лейзерсон, Рональд Л. Ривест, Клиффорд Штайн.* Алгоритмы: построение и анализ, 3-е издание = Introduction to Algorithms, Third Edition. — М.: «Вильямс», 2013. — 1328 с. — ISBN 978-5-8459-1794-2
2. *Курносков М.Г., Берлизов Д.М.* Алгоритмы и структуры обработки информации. — Новосибирск: Параллель, 2019. — 211 с. — ISBN 978-5-98901-230-5
3. *Гулаков В.К., Трубаков А.О., Трубаков Е.О.* Структуры и алгоритмы обработки многомерных данных: монография. - 2-е изд. - СПб., М., Краснодар: Лань, 2021. - 355 с.
4. R-tree // Wikipedia URL: <https://en.wikipedia.org/wiki/R-tree> (дата обращения: 23.05.2023).
5. *Guttman A.* R-trees: A dynamic index structure for spatial searching // ACM SIGMOD. - 1984. - №14(2). - С. 47-57.
6. *Samet H.* The design and analysis of spatial data structures. - 2 изд. - Addison-Wesley Publishing Co., 1990
7. *Garcia-Molina, H., Salem, K.* R-trees: A dynamic index structure for spatial searching // IEEE Transactions on Knowledge and Data Engineering. - 1987. - №1(1). - С. 25-39.

ПРИЛОЖЕНИЯ

Приложение 1. Исходный код проекта

1. Файл main.c

```
1  #include <stdio.h>
2  #include <string.h>
3  #include <math.h>
4  #include "rtree.h"
5
6  struct city {
7      char *name;
8      double latitude;
9      double longitude;
10 };
11
12 struct city nsk = { .name = "Novosibirsk", .latitude = 55.0333, .longitude = 82.9167 };
13 struct city bai = { .name = "Buenos Aires", .latitude = -34.5997, .longitude = -58.3819 };
14 struct city rio = { .name = "Rio de Janeiro", .latitude = -22.9111, .longitude = -43.2056 };
15 struct city tok = { .name = "Tokyo", .latitude = 35.6897, .longitude = 139.6922 };
16 struct city tor = { .name = "Toronto", .latitude = 43.6532, .longitude = -79.3832 };
17 struct city syd = { .name = "Sydney", .latitude = -33.8688, .longitude = 151.2093 };
18
19 bool city_iter(const double *min, const double *max, const void *item, void *udata) {
20     const struct city *city = item;
21     printf("%s\n", city->name);
22     return true;
23 }
24
25 int main() {
26     struct rtree *tr = rtree_new();
27     rtree_insert(tr, (double[2]){nsk.longitude, nsk.latitude}, NULL, &nsk);
28     rtree_insert(tr, (double[2]){tor.longitude, tor.latitude}, NULL, &tor);
29     rtree_insert(tr, (double[2]){bai.longitude, bai.latitude}, NULL, &bai);
30     rtree_insert(tr, (double[2]){rio.longitude, rio.latitude}, NULL, &rio);
31     rtree_insert(tr, (double[2]){tok.longitude, tok.latitude}, NULL, &tok);
32     rtree_insert(tr, (double[2]){syd.longitude, syd.latitude}, NULL, &syd);
33
34     printf("\nNorthwestern cities:\n");
35     rtree_search(tr, (double[2]){-180, 0}, (double[2]){0, 90}, city_iter, NULL);
36     printf("\nNortheastern cities:\n");
37     rtree_search(tr, (double[2]){0, 0}, (double[2]){180, 90}, city_iter, NULL);
38     printf("\nSouthwestern cities:\n");
39     rtree_search(tr, (double[2]){-180, -90}, (double[2]){0, 0}, city_iter, NULL);
40     printf("\nSoutheastern cities:\n");
41     rtree_search(tr, (double[2]){0, -90}, (double[2]){180, 0}, city_iter, NULL);
42
43     rtree_delete(tr, (double[2]){nsk.longitude, nsk.latitude}, NULL, &nsk);
44     printf("\nNortheastern cities after element deletion:\n");
45     rtree_search(tr, (double[2]){0, 0}, (double[2]){180, 90}, city_iter, NULL);
46     rtree_free(tr);
47 }
```

2. Файл rtree.c


```

1  #include <stdlib.h>
2  #include <stdio.h>
3  #include <string.h>
4  #include <stdbool.h>
5  #include "rtree.h"
6
7  struct node *node_new(struct rtree *tr, enum kind kind) {
8      struct node *node = (struct node *)tr->malloc(sizeof(struct node));
9      memset(node, 0, sizeof(struct node));
10     node->kind = kind;
11     return node;
12 }
13
14 void node_free(struct rtree *tr, struct node *node) {
15     if (node->kind == BRANCH) {
16         for (int i = 0; i < node->count; i++) {
17             node_free(tr, node->children[i]);
18         }
19     }
20     tr->free(node);
21 }
22
23 void rect_expand(struct rect *rect, struct rect *other) {
24     for (int i = 0; i < DIMS; i++) {
25         if (other->min[i] < rect->min[i]) { rect->min[i] = other->min[i]; }
26         if (other->max[i] > rect->max[i]) { rect->max[i] = other->max[i]; }
27     }
28 }
29
30 double rect_area(struct rect *rect) {
31     double area = (double)(rect->max[0]) - (double)(rect->min[0]);
32     for (int i = 1; i < DIMS; i++) {
33         area *= (double)(rect->max[i]) - (double)(rect->min[i]);
34     }
35     return area;
36 }
37
38 bool rect_contains(struct rect *rect, struct rect *other) {
39     for (int i = 0; i < DIMS; i++) {
40         if (other->min[i] < rect->min[i] || other->max[i] > rect->max[i]) {
41             return false;
42         }
43     }
44     return true;
45 }
46
47 bool rect_intersects(struct rect *rect, struct rect *other) {
48     for (int i = 0; i < DIMS; i++) {
49         if (other->min[i] > rect->max[i] || other->max[i] < rect->min[i]) {
50             return false;
51         }
52     }
53     return true;
54 }
55
56 bool nums_equal(NUMTYPE a, NUMTYPE b) {

```

```

58     return !(a < b || a > b);
59 }
60
61 bool rect_onedge(struct rect *rect, struct rect *other) {
62     for (int i = 0; i < DIMS; i++) {
63         if (nums_equal(rect->min[i], other->min[i])) {
64             return true;
65         }
66         if (nums_equal(rect->max[i], other->max[i])) {
67             return true;
68         }
69     }
70     return false;
71 }
72
73 bool rect_equals(struct rect *rect, struct rect *other) {
74     for (int i = 0; i < DIMS; i++) {
75         if (!nums_equal(rect->min[i], other->min[i])) {
76             return false;
77         }
78         if (!nums_equal(rect->max[i], other->max[i])) {
79             return false;
80         }
81     }
82     return true;
83 }
84
85 void node_swap(struct node *node, int i, int j) {
86     struct rect tmp = node->rects[i];
87     node->rects[i] = node->rects[j];
88     node->rects[j] = tmp;
89     if (node->kind == LEAF) {
90         struct item tmp = node->items[i];
91         node->items[i] = node->items[j];
92         node->items[j] = tmp;
93     } else {
94         struct node *tmp = node->children[i];
95         node->children[i] = node->children[j];
96         node->children[j] = tmp;
97     }
98 }
99
100 void node_qsort(struct node *node, int s, int e, int axis, bool rev, bool max) {
101     int nrects = e - s, left = 0, right = nrects - 1, pivot = nrects / 2;
102     if (nrects < 2) { return; }
103     node_swap(node, s + pivot, s + right);
104     struct rect *rects = &node->rects[s];
105     if (!rev) {
106         if (!max) {
107             for (int i = 0; i < nrects; i++) {
108                 if (rects[i].min[axis] < rects[right].min[axis]) {
109                     node_swap(node, s + i, s + left);
110                     left++;
111                 }
112             }
113         } else {
114             for (int i = 0; i < nrects; i++) {

```

```

115         if (rects[i].max[axis] < rects[right].max[axis]) {
116             node_swap(node, s + i, s + left);
117             left++;
118         }
119     }
120 }
121 } else {
122     if (!max) {
123         for (int i = 0; i < nrects; i++) {
124             if (rects[right].min[axis] < rects[i].min[axis]) {
125                 node_swap(node, s + i, s + left);
126                 left++;
127             }
128         }
129     } else {
130         for (int i = 0; i < nrects; i++) {
131             if (rects[right].max[axis] < rects[i].max[axis]) {
132                 node_swap(node, s + i, s + left);
133                 left++;
134             }
135         }
136     }
137 }
138 node_swap(node, s + left, s + right);
139 node_qsort(node, s, s + left, axis, rev, max);
140 node_qsort(node, s + left + 1, e, axis, rev, max);
141 }
142
143 void node_sort(struct node *node) {
144     node_qsort(node, 0, node->count, 0, false, false);
145 }
146
147 void node_sort_by_axis(struct node *node, int axis, bool rev, bool max) {
148     node_qsort(node, 0, node->count, axis, rev, max);
149 }
150
151 int rect_largest_axis(struct rect *rect) {
152     int axis = 0;
153     double nlength = (double)rect->max[0] - (double)rect->min[0];
154     for (int i = 1; i < DIMS; i++) {
155         double length = (double)rect->max[i] - (double)rect->min[i];
156         if (length > nlength) {
157             nlength = length;
158             axis = i;
159         }
160     }
161     return axis;
162 }
163
164 void node_move_rect_at_index_into(struct node *from, int index, struct node *into) {
165     into->rects[int->count] = from->rects[index];
166     from->rects[index] = from->rects[from->count - 1];
167     if (from->kind == LEAF) {
168         into->items[int->count] = from->items[index];
169         from->items[index] = from->items[from->count - 1];
170     } else {
171         into->children[int->count] = from->children[index];

```

```

172         from->children[index] = from->children[from->count - 1];
173     }
174     from->count--;
175     into->count++;
176 }
177
178 struct node *node_split_largest_axis_edge_snap
179 (struct rtree *tr, struct rect *rect, struct node *left) {
180     int axis = rect_largest_axis(rect);
181     struct node *right = node_new(tr, left->kind);
182     if (!right) return NULL;
183     for (int i = 0; i < left->count; i++) {
184         double min_dist = (double)left->rects[i].min[axis] - (double)rect->min[axis];
185         double max_dist = (double)rect->max[axis] - (double)left->rects[i].max[axis];
186         if (min_dist < max_dist) { // stay left
187             if (else { // move right
188                 node_move_rect_at_index_into(left, i, right);
189                 i--;
190             }
191         }
192         // make sure that both left and right nodes have at least min_entries by moving items into
193         underflowed nodes
194         if (left->count < MIN_ENTRIES) { // reverse sort by min axis
195             node_sort_by_axis(right, axis, true, false);
196             while (left->count < 2) {
197                 node_move_rect_at_index_into(right, right->count-1, left);
198             }
199         } else if (right->count < 2) { // reverse sort by max axis
200             node_sort_by_axis(left, axis, true, true);
201             while (right->count < 2) {
202                 node_move_rect_at_index_into(left, left->count-1, right);
203             }
204         }
205         node_sort(right);
206         node_sort(left);
207         return right;
208     }
209 }
210
211 struct node *node_split(struct rtree *tr, struct rect *r, struct node *left) {
212     return node_split_largest_axis_edge_snap(tr, r, left);
213 }
214
215 int node_rsearch(struct node *node, NUMTYPE key) {
216     for (int i = 0; i < node->count; i++) {
217         if (!(node->rects[i].min[0] < key)) {
218             return i;
219         }
220     }
221     return node->count;
222 }
223
224 double rect_unioned_area(struct rect *rect, struct rect *other) {
225     double area = (double)MAX(rect->max[0], other->max[0]) - (double)MIN(rect->min[0],
226     other->min[0]);
227     for (int i = 1; i < DIMS; i++) {
228         area *= (double)MAX(rect->max[i], other->max[i]) - (double)MIN(rect->min[i],
229         other->min[i]);
230     }
231     return area;
232 }

```

```

225     }
226     return area; // returns the area of two rects expanded
227 }
228
229 int node_choose_least_enlargement(struct node *node, struct rect *ir) {
230     int j = -1;
231     double jenlargement = 0, jarea = 0;
232     for (int i = 0; i < node->count; i++) {
233         // calculate the enlarged area
234         double uarea = rect_unioned_area(&node->rects[i], ir);
235         double area = rect_area(&node->rects[i]);
236         double enlargement = uarea - area;
237         if (j == -1 || enlargement < jenlargement || (!(enlargement > jenlargement) && area < jarea)){
238             j = i;
239             jenlargement = enlargement;
240             jarea = area;
241         }
242     }
243     return j;
244 }
245
246 int node_choose_subtree(struct node *node, struct rect *ir) {
247     // take a quick look for the first node that contain the rect.
248     if (FAST_CHOOSER == 1) {
249         int index = -1;
250         double narea;
251         for (int i = 0; i < node->count; i++) {
252             if (rect_contains(&node->rects[i], ir)) {
253                 double area = rect_area(&node->rects[i]);
254                 if (index == -1 || area < narea) {
255                     narea = area;
256                     index = i;
257                 }
258             }
259         }
260     } else if (FAST_CHOOSER == 2) {
261         for (int i = 0; i < node->count; i++) {
262             if (rect_contains(&node->rects[i], ir)) {
263                 return i;
264             }
265         }
266     }
267     // fallback to using the choose-least-enlargment algorithm
268     return node_choose_least_enlargement(node, ir);
269 }
270
271 struct rect node_rect_calc(struct node *node) {
272     struct rect rect = node->rects[0];
273     for (int i = 1; i < node->count; i++) {
274         rect_expand(&rect, &node->rects[i]);
275     }
276     return rect;
277 }
278
279 int node_order_to_right(struct node *node, int index) {
280     while (index < node->count - 1 && node->rects[index + 1].min[0] < node->rects[index].min[0]) {
281         node_swap(node, index + 1, index);

```

```

282         index++;
283     }
284     return index;
285 }
286
287 int node_order_to_left(struct node *node, int index) {
288     while (index > 0 && node->rects[index].min[0] < node->rects[index - 1].min[0]) {
289         node_swap(node, index, index - 1);
290         index--;
291     }
292     return index;
293 }
294
295 // performs a copy of the data from args[1] & args[2], expects a rectangle (double[] double[])
296 // first N values are min corner, next N values - max corner, N - num of dimensions (max coords are
optional)
297 bool node_insert(struct rtree *tr, struct rect *nr, struct node *node, struct rect *ir, struct item
item, bool *split, bool *grown) {
298     *split = false;
299     *grown = false;
300     if (node->kind == LEAF) {
301         if (node->count == MAX_ENTRIES) {
302             *split = true;
303             return true;
304         }
305     }
306     int index = node_rsearch(node, ir->min[0]);
307     memmove(&node->rects[index + 1], &node->rects[index], (node->count-index) * sizeof(struct
rect));
308     memmove(&node->items[index + 1], &node->items[index], (node->count-index) *
sizeof(struct item));
309     node->rects[index] = *ir;
310     node->items[index] = item;
311     node->count++;
312     *grown = !rect_contains(nr, ir);
313     return true;
314 }
315 int index = node_choose_subtree(node, ir); // choose a subtree for inserting the rectangle
316 if (!node_insert(tr, &node->rects[index], node->children[index], ir, item, split, grown)) {
317     return false;
318 }
319 if (*split) {
320     if (node->count == MAX_ENTRIES) {
321         return true;
322     }
323     struct node *left = node->children[index];
324     struct node *right = node_split(tr, &node->rects[index], left); // split child node
325     if (!right) {
326         return false;
327     }
328     node->rects[index] = node_rect_calc(left);
329     memmove(&node->rects[index + 2], &node->rects[index + 1], (node->count - (index + 1)) *
sizeof(struct rect));
330     memmove(&node->children[index + 2], &node->children[index + 1], (node->count - (index +
1)) * sizeof(struct node*));
331     node->rects[index + 1] = node_rect_calc(right);
332     node->children[index + 1] = right;
333     node->count++;

```

```

334     if (node->rects[index].min[0] > node->rects[index + 1].min[0]) {
335         node_swap(node, index + 1, index);
336     }
337     index++;
338     node_order_to_right(node, index);
339     return node_insert(tr, nr, node, ir, item, split, grown);
340 }
341 if (*grown) { // child rectangle must expand to accomadate new item
342     rect_expand(&node->rects[index], ir);
343     node_order_to_left(node, index);
344     *grown = !rect_contains(nr, ir);
345 }
346 return true;
347 }
348
349 struct rtree *rtree_new_with_allocator(void *(*cust_malloc)(size_t), void (*cust_free)(void*)) {
350     if (!cust_malloc) cust_malloc = malloc;
351     if (!cust_free) cust_free = free;
352     struct rtree *tr = (struct rtree *)cust_malloc(sizeof(struct rtree));
353     if (!tr) { return NULL; }
354     memset(tr, 0, sizeof(struct rtree));
355     tr->malloc = cust_malloc;
356     tr->free = cust_free;
357     return tr;
358 }
359
360 struct rtree *rtree_new() { return rtree_new_with_allocator(NULL, NULL); }
361
362 bool rtree_insert
363 (struct rtree *tr, const NUMTYPE *min, const NUMTYPE *max, const DATATYPE data) {
364     struct rect rect;
365     memcpy(&rect.min[0], min, sizeof(NUMTYPE) * DIMS);
366     memcpy(&rect.max[0], max ? max : min, sizeof(NUMTYPE) * DIMS);
367     struct item item;
368     memcpy(&item.data, &data, sizeof(DATATYPE));
369     if (!tr->root) {
370         struct node *new_root = node_new(tr, LEAF);
371         if (!new_root) return false;
372         tr->root = new_root;
373         tr->rect = rect;
374     }
375     bool split = false, grown = false;
376     if (!node_insert(tr, &tr->rect, tr->root, &rect, item, &split, &grown)) { return false; }
377     if (split) {
378         struct node *new_root = node_new(tr, BRANCH);
379         if (!new_root) return false;
380         struct node *left = tr->root;
381         struct node *right = node_split(tr, &tr->rect, left);
382         tr->root = new_root;
383         tr->root->rects[0] = node_rect_calc(left);
384         tr->root->rects[1] = node_rect_calc(right);
385         tr->root->children[0] = left;
386         tr->root->children[1] = right;
387         tr->root->count = 2;
388         tr->height++;
389         node_sort(tr->root);
390         return rtree_insert(tr, min, max, data);

```

```

390     }
391     if (grown) {
392         rect_expand(&tr->rect, &rect);
393         node_sort(tr->root);
394     }
395     tr->count++;
396     return true;
397 }
398
399 void rtree_free(struct rtree *tr) {
400     if (tr->root) { node_free(tr, tr->root); }
401     tr->free(tr);
402 }
403
404 bool node_search(struct node *node, struct rect *rect, bool (*iter)
405 (const NUMTYPE *min, const NUMTYPE *max, const DATATYPE data, void *udata), void *udata) {
406     if (node->kind == LEAF) {
407         for (int i = 0; i < node->count; i++) {
408             if (rect_intersects(&node->rects[i], rect)) {
409                 if (!iter(node->rects[i].min, node->rects[i].max, node->items[i].data, udata)) {
410                     return false;
411                 }
412             }
413         }
414         return true;
415     }
416     for (int i = 0; i < node->count; i++) {
417         if (rect_intersects(&node->rects[i], rect)) {
418             if (!node_search(node->children[i], rect, iter, udata)) {
419                 return false;
420             }
421         }
422     }
423     return true;
424 }
425
426 void rtree_search(struct rtree *tr, const NUMTYPE *min, const NUMTYPE *max, bool (*iter)
427 (const NUMTYPE *min, const NUMTYPE *max, const DATATYPE data, void *udata), void *udata) {
428     struct rect rect;
429     memcpy(&rect.min[0], min, sizeof(NUMTYPE) * DIMS);
430     memcpy(&rect.max[0], max ? max : min, sizeof(NUMTYPE) * DIMS);
431     if (tr->root && rect_intersects(&tr->rect, &rect)) {
432         node_search(tr->root, &rect, iter, udata);
433     }
434 }
435
436 size_t rtree_count(struct rtree *tr) { return tr->count; }
437
438 void node_delete(struct rtree *tr, struct rect *nr, struct node *node, struct rect *ir, struct item
439 item, bool *removed, bool *shrunk, int (*compare)(const DATATYPE a, const DATATYPE b, void
440 *udata), void *udata) {
441     *removed = false;
442     *shrunk = false;
443     if (node->kind == LEAF) {
444         for (int i = 0; i < node->count; i++) {
445             if (!rect_contains(ir, &node->rects[i])) {
446                 continue;

```



```

443     }
444     int cmp = compare ?
445         compare(node->items[i].data, item.data, udata) :
446         memcmp(&node->items[i].data, &item.data, sizeof(DATATYPE));
447     if (cmp != 0) {
448         continue;
449     }
450     // found the target item to delete
451     memmove(&node->rects[i], &node->rects[i + 1], (node->count - (i + 1)) * sizeof(struct
rect));
452     memmove(&node->items[i], &node->items[i + 1], (node->count - (i + 1)) * sizeof(struct
item));
453     node->count--;
454     if (rect_onedge(ir, nr)) { // item was on the edge of node rect
455         *nr = node_rect_calc(node); // recalculation of node rect
456         *shrunk = true; // notify the caller that rect is shrunk
457     }
458     *removed = true;
459     return;
460 }
461 return;
462 }
463 for (int i = 0; i < node->count; i++) {
464     if (!rect_contains(&node->rects[i], ir)) {
465         continue;
466     }
467     struct rect crect = node->rects[i];
468     node_delete(tr, &node->rects[i], node->children[i], ir, item, removed, shrunk, compare, udata);
469     if (!*removed) {
470         continue;
471     }
472     if (node->children[i]->count == 0) { // underflow
473         node_free(tr, node->children[i]);
474         memmove(&node->rects[i], &node->rects[i + 1], (node->count - (i + 1)) * sizeof(struct
rect));
475         memmove(&node->children[i], &node->children[i + 1], (node->count - (i + 1)) *
sizeof(struct node *));
476         node->count--;
477         *nr = node_rect_calc(node);
478         *shrunk = true;
479         return;
480     }
481     if (*shrunk) {
482         *shrunk = !rect_equals(&node->rects[i], &crect);
483         if (*shrunk) {
484             *nr = node_rect_calc(node);
485         }
486         node_order_to_right(node, i);
487     }
488     return;
489 }
490 return;
491 }
492
493 // search the tree for an item contained within provided rect, perform a binary comparison of its data to
provided, first item found is deleted
494 void rtree_delete

```

```

495  (struct rtree *tr, const NUMTYPE *min, const NUMTYPE *max, const DATATYPE data) {
496      struct rect rect;
497      memcpy(&rect.min[0], min, sizeof(NUMTYPE) * DIMS);
498      memcpy(&rect.max[0], max ? max : min, sizeof(NUMTYPE) * DIMS);
499      struct item item;
500      memcpy(&item.data, &data, sizeof(DATATYPE));
501      if (!tr->root) { return; }
502      bool removed = false, shrunk = false;
503      node_delete(tr, &tr->rect, tr->root, &rect, item, &removed, &shrunk, NULL, NULL);
504      if (!removed) {
505          return;
506      }
507      tr->count--;
508      if (tr->count == 0) {
509          node_free(tr, tr->root);
510          tr->root = NULL;
511          memset(&tr->rect, 0, sizeof(struct rect));
512      } else {
513          while (tr->root->kind == BRANCH && tr->root->count == 1) {
514              struct node *prev = tr->root;
515              tr->root = tr->root->children[0];
516              prev->count = 0;
517              node_free(tr, prev);
518          }
519          if (shrunk) {
520              tr->rect = node_rect_calc(tr->root);
521          }
522      }
523  }

```

3. Файл rtree.h

```

1  #pragma once
2
3  #include <stdlib.h>
4  #include <stdbool.h>
5
6  #define DATATYPE void *
7  #define NUMTYPE double
8  #define DIMS 2
9  #define MAX_ENTRIES 64
10 #define MIN_ENTRIES_PERCENTAGE 10
11 #define FAST_CHOOSER 2 // 0 - off , 1 - fast, 2 - faster
12 #define panic(_msg_) { \
13     fprintf(stderr, "panic: %s (%s:%d)\n", (_msg_), __FILE__, __LINE__); \
14     exit(1); \
15 }
16 #define MIN(a,b) ((a) < (b) ? (a) : (b))
17 #define MAX(a,b) ((a) > (b) ? (a) : (b))
18 #define MIN_ENTRIES ((MAX_ENTRIES) * (MIN_ENTRIES_PERCENTAGE) / 100 + 1)
19
20 enum kind {
21     LEAF = 1,
22     BRANCH = 2,
23 };
24
25 struct rect {
26     NUMTYPE min[DIMS];
27     NUMTYPE max[DIMS];
28 };
29
30 struct item {
31     DATATYPE data;
32 };
33
34 struct node {
35     enum kind kind; // LEAF or BRANCH
36     int count; // number of rects
37     struct rect rects[MAX_ENTRIES];
38     union { struct node *children[MAX_ENTRIES]; struct item items[MAX_ENTRIES]; };
39 };
40
41 struct rtree {
42     size_t count;
43     int height;
44     struct rect rect;
45     struct node *root;
46     void *(*malloc)(size_t);
47     void (*free)(void *);
48 };
49
50 struct rtree *rtree_new();
51 bool rtree_insert(struct rtree *tr, const double *min, const double *max, const void *data);
52 void rtree_free(struct rtree *tr);
53 void rtree_search(struct rtree *tr, const double *min, const double *max, bool (*iter)
54 (const double *min, const double *max, const void *data, void *udata), void *udata);
55 size_t rtree_count(struct rtree *tr);
56 void rtree_delete(struct rtree *tr, const double *min, const double *max, const void *data);

```

4. Makefile

```
1 all: main
2 .PHONY: main
3 main: main.c rtree.c
4     gcc -Wall -Wextra -o $@ $^
5
6 .PHONY: clean
7 clean:
8     rm -rf main
```

Отзыв о работе студента

Григорьев Юрий Вадимович

(ФИО студента)

This image shows a full page of blank, lined paper. It features approximately 28 horizontal black lines spaced evenly across the page, typical of notebook paper. The lines are thin and extend from the left edge to the right edge. There are no margins, text, or other markings on the page.

Уровень освоения компетенций

Григорьев Юрий Вадимович

(ФИО студента)

Компетенции	Уровень сформированности компетенций
<i>ОПК-1 - Способен применять фундаментальные знания, полученные в области математических и (или) естественных наук, и использовать их в профессиональной деятельности</i>	

отметка о зачете _____

Руководитель практики от СибГУТИ:

Должность руководителя

подпись

ФИО руководителя

" ____ " _____ 20 ____ г.