МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«ОРЛОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ   
ИМЕНИ И.С. ТУРГЕНЕВА»

Кафедра информационных систем

Отчет   
по лабораторной работе №1  
на тему: « Организация файла в виде кучи»  
по дисциплине «Базы данных»

Выполнили: Марочкин М.А. Шифр: 170584   
 Шорин В.Д. Шифр: 171406  
ИПАИТ  
Направление: 09.03.04 «Программная инженерия»  
Группа: 71-ПГ  
Проверил:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
Отметка о зачете:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дата «\_\_\_\_» \_\_\_\_\_\_ 2018 г.

Орел, 2018 г.

**Цель работы:**

Написать программу, которая работает с организованным в виде кучи файлом, хранящем информацию об отношении «студент».

В программе должны быть реализованы следующие функции:

* добавление информации о студенте;
* изменение информации о студенте;
* удаление информации о студенте;
* осуществление поиска информации о студенте.

Отношение студент должно содержать следующие атрибуты: номер зачетки (тип integer), фамилия (тип string (30)), имя (тип string (20)), отчество (тип string (30)), номер группы (тип integer).

**Контрольные вопросы**

1. **Что такое запись?**

Для N-арного отношения запись представляет собой упорядоченный набор из N значений, по одному значению для каждого поля отношения.

1. **Какие дополнительные байты может содержать запись?**

Запись включает столько байт, сколько требуется для хранения каждого поля в зависимости от его типа. Кроме того, запись может включать дополнительные байты, не относящиеся к полям. Это необходимо в следующих случаях:

1. Некоторые байты могут свидетельствовать о формате записи. Это необходимо, если совместно могут храниться записи различных форматов.

2. Один или несколько байт могут хранить информацию о длине записи. Это необходимо, если записи могут иметь переменную длину. В записях с постоянной длиной такая информация не храниться, а вычисляется исходя из типов полей.

3. Один бит может отводиться под информацию о том является ли запись удаленной или нет. В некоторых БД записи физически не удаляются, а соответствующий бит переводится в состояние «запись удалена». Такие записи продолжают храниться, но не рассматриваются при построении запросов. В этом случае СУБД должна иметь функцию сборки мусора. Этот процесс запускается в некоторый момент времени и перераспределяет записи, физически удаляя помеченные записи.

4. Бит использования. Он хранит информацию о наличии свободного места.

5. Неиспользуемое пространство. Байты, не хранящие никакой информации. Такие байты необходимы для более удобного чтения записей. Например, для достижения кратности разрядности процессора.

Если запись подразумевает такие типы данных как text, image, blob (двоичная информация), то такие данные не хранятся вместе с записью, а хранятся в отдельном блоке. В записи же хранится адрес.

1. **Что такое блок?**

Наиболее быстрая обработка данных была бы возможна, если размещать БД целиком в оперативной памяти. Однако практически размеры БД гораздо больше, поэтому возникает необходимость переноса фрагментов данных с диска (внешней памяти) в оперативную (первичную) память. Поэтому можно говорить, что данные на диске хранятся разделенными на части. Такие части называются ***блоками*** или ***страницами***. Размер блока зависит от количества байт, которые считываются с диска за один раз.

1. **В чем особенности организации файлов в виде кучи?**

Это наиболее тривиальный способ организации файлов. При этом записи записываются в блоки, без упорядочивания по какому-либо признаку. Сами блоки выстраиваются таким же образом. То есть, записи помещаются в один блок последовательно по мере появления. Когда блок заполняется, то выделяется новый блок, указатель на него помещается в предыдущий блок и аналогичным образом производится заполнение нового блока.

1. **Эффективность рабы с файлом, организованным в виде кучи.**

Пусть в файле существует ***N*** записей. В каждый блок может поместиться ***R*** записей. То есть в файле должно быть ***N/R*** блоков. Время выполнения операций измеряем количеством блоков, которые должны быть запрошены с диска или записаны на него.

При **поиске** записи в файле возможны два варианта:

1) N/R блоков придется просмотреть, если запись отсутствует в файле или, если ключевое поле, не является уникальным;

2) в противном случае потребуется считать ***N/2R*** блоков.

При **вставке** необходимо запросить последнюю запись кучи, так как именно после нее идет свободное место. То есть нужно считать последний блок. Если последний блок не имеет пустого места, то нужно начать новый блок. После вставки блок должен быть записан на диск. Таким образом, для вставки потребуется два обращения к диску – один по чтению, другой - по записи. Если необходимо контролировать отсутствие дублирующих записей, то время выполнения корректируется на время выполнения поиска.

**Удаление** требует найти запись, то есть выполнить операцию поиска, а затем выполнить операцию записи блока, в котором была найдена запись. Таким образом, для удаления записи потребуется в среднем ***((N/2R)+1)*** доступов, если запись существует в файле. Если ее нет в файле, то выполниться N/R доступов. Для дальнейшего использования освободившегося пространства последняя запись последнего блока может быть перемещена на место удаленной записи. Это немного замедлит процесс удаления, но уменьшит количество блоков, что в дальнейшем ускорит выполнение всех операций.

Время выполнения **модификации** аналогично времени выполнения операции удаления.

Таким образом, основное время при выполнении всех операций затрачивается на поиск записи. Следовательно, необходимо ускорить именно поиск. Это может быть достигнуто за счет другой организации файла.

**Выполнение работы**

**Код**