

Практическая работа №3

Тема. Применение хеш-таблицы для поиска данных в двоичном файле с записями фиксированной длины.

Цель. Получить навыки по разработке хеш-таблиц и их применению при поиске данных в других структурах данных (файлах).

Задание 1.

1. Разработать приложение, которое использует хеш-таблицу для организации прямого доступа к записям двоичного файла, реализованного в практической работе 2.
2. Создать приложение и включить в него три заголовочных файла: управление хеш-таблицей, управление двоичным файлом (практическая работа 2), управление двоичным файлом посредством хеш-таблицы.
3. Для обеспечения прямого доступа к записи в файле элемент хеш-таблицы должен включать обязательные поля: ключ записи в файле, номер записи с этим ключом в файле. Элемент может содержать другие поля, требующиеся методу (указанному в вашем варианте), разрешающему коллизию.
4. Управление хеш-таблицей
 - 4.1. Определить структуру элемента хеш-таблицы и структуру хеш-таблицы в соответствии с методом разрешения коллизии, указанным в варианте. Определения разместить в соответствующем заголовочном файле. Все операции управления хеш-таблицей размещать в этом заголовочном файле.
 - 4.2. Тестирование операций выполнять в функции main приложения по мере их реализации.
 - 4.3. После тестирования всех операций, создать в заголовочном файле функцию с именем testHeshT переместить в нее содержание функции main, проверить, что приложение выполняется.
 - 4.4. Разработать операции по управлению хеш-таблицей.
 - 4.5. Разработать хеш-функцию (метод определить самостоятельно), выполнить ее тестирование, убедиться, что хеш (индекс элемента таблицы) формируется верно.
 - 4.6. Разработать операции: вставить ключ в таблицу, удалить ключ из таблицы, найти ключ в таблице, рехешировать таблицу. Каждую операцию тестируйте по мере ее реализации.
 - 4.7. Подготовить тесты (последовательность значений ключей), обеспечивающие:

- 4.7.1. вставку ключа без коллизии
- 4.7.2. вставку ключа и разрешение коллизии
- 4.7.3. вставку ключа с последующим рехешированием
- 4.7.4. удаление ключа из таблицы
- 4.7.5. поиск ключа в таблице. Для метода с открытым адресом подготовить тест для поиска ключа, который размещен в таблице после удаленного ключа, с одним значением хеша для этих ключей
- 4.8. Выполнить тестирование операций управления хеш-таблицей. При тестировании операции вставки ключа в таблицу предусмотрите вывод списка индексов, которые формируются при вставке элементов в таблицу.
5. Управление двоичным файлом
 - 5.1. Операции управления двоичным файлом: создание двоичного файла из текстового, добавить запись в двоичный файл, удалить запись с заданным ключом из файла, прочитать запись файла по заданному номеру записи.
 - 5.2. Структура записи двоичного файла и все операции по управлению файлом должны быть размещены в соответствующем заголовочном файле.
 - 5.3. Выполнить тестирование операций в main приложения, и содержание функции main переместить в соответствующую функцию заголовочного файла с именем testBinF.
6. Управление файлом посредством хеш-таблицы
7. В заголовочный файл управления файлом посредством хеш-таблицы подключить заголовочные файлы: управления хеш-таблицей, управления двоичным файлом. Реализовать поочередно все перечисленные ниже операции в этом заголовочном файле, выполняя их тестирование из функции main приложения. После разработки всех операций выполнить их комплексное тестирование.
8. Разработать и реализовать операции
 - 8.1. Прочитать запись из файла и вставить элемент в таблицу (элемент включает: ключ и номер записи с этим ключом в файле, и для метода с открытой адресацией возможны дополнительные поля).
 - 8.2. Удалить запись из таблицы при заданном значении ключа и соответственно из файла.
 - 8.3. Найти запись в файле по значению ключа (найти ключ в хеш-таблице, получить номер записи с этим ключом в файле, выполнить прямой доступ к записи по ее номеру).
 - 8.4. Подготовить тесты для тестирования приложения:
 - 8.5. Заполните файл небольшим количеством записей.
 - 8.6. Включите в файл записи как не приводящие к коллизиям, так и приводящие.

8.7. Обеспечьте включение в файл такого количества записей, чтобы потребовалось рехеширование.

8.8. Заполните файл большим количеством записей (до 1 000 000).

8.9. Определите время чтения записи с заданным ключом: для первой записи файла, для последней и где-нибудь в середине. Убедитесь (или нет), что время доступа для всех записей одинаково.

9. Составить отчет.

10. Ответы на вопросы

10.1. Расскажите о назначении хеш-функции.

10.2. Что такое коллизия?

10.3. Что такое «открытый адрес» по отношению к хеш-таблице?

10.4. Как в хеш-таблице с открытым адресом реализуется коллизия?

10.5. Какая проблема, может возникнуть после удаления элемента из хеш-таблицы с открытым адресом и как ее устранить?

10.6. Что определяет коэффициент нагрузки в хеш-таблице?

10.7. Что такое «первичный кластер» в таблице с открытым адресом?

10.8. Как реализуется двойное хеширование?

Таблица 1. Варианты заданий

№	Тип хеш-таблицы (метод разрешения коллизии)	Структура записи двоичного файла
1	С открытой адресацией (смещение на 1)	Читательский абонемент: <u>номер читательского билета</u> – целое пятизначное число, ФИО, адрес.
2	С открытой адресацией (смещение на номер выполняемого подбора)	Счет в банке: <u>номер счета</u> – 7 разрядное число, ФИО, адрес.
3	С открытой адресацией (двойное хеширование)	Владелец телефона: <u>номер телефона</u> – последовательность символов, адрес, ФИО.
4	Цепное хеширование	Владелец автомобиля: <u>номер машины</u> , марка, сведения о владельце, сведения об угоне (логического типа).
5	Цепное хеширование	Пациент поликлиники: <u>номер карточки</u> , код хронического заболевания, фамилия лечащего врача.

6	Цепное хеширование	Товар: название, <u>код</u> – шестизначное число, завод изготовитель, цена, страна (название).
7	Цепное хеширование	Специализация вуза: <u>код специальности</u> , название вуза, название специальности.
8	Открытый адрес(двойное хеширование)	Книга: <u>ISBN</u> – двенадцатизначное число, автор, название, год издания.
9	Цепное хеширование	Страховой полис: <u>номер</u> , компания, фамилия владельца.
10	Открытый адрес(смещение на 1)	Англо-русский словарь: <u>английское слово</u> , русское слово.
11	Открытый адрес(двойное хеширование)	Железнодорожная справка: <u>номер поезда</u> , пункт отправления, пункт назначения, время отправления.
12	Цепное хеширование	Регистрация малого предприятия: <u>номер лицензии</u> (текстовое значение), название, учредитель, признак действия лицензии (0 действует, 1 отозвана).
13	Открытый адрес(двойное хеширование)	Студент: <u>номер зачетной книжки</u> , номер группы, ФИО.
14	Цепное хеширование	Справочная межгорода: <u>код города</u> , название города, страна.
15	Открытый адрес (смещение на 1)	Найти и поздравить друга: <u>дата рождения</u> , имя
16	Цепное хеширование	Расписание занятий группы: <u>номер группы</u> , название дисциплины, номер пары, номер недели, номер дня недели, вид занятия, номер аудитории.
17	Открытый адрес (смещение на 1)	Частотный словарь: <u>слово</u> , количество вхождений в текст.
18	Открытый адрес (двойное хеширование)	Читательский билет: <u>номер, инвентарный</u> , номер книги, дата выдачи, дата возврата.

19	Цепное хеширование	Вызов такси: <u>номер</u> , фамилия водителя, время выезда, отметка о присутствии в гараже.
20	Открытый адрес(смещение на 1)	Продажи товаров: <u>код товара</u> , название, цена, дата продажи.
21	Открытый адрес(двойное хеширование)	Сотрудник: <u>табельный номер</u> , должность, оклад, количество детей.
22	Цепное хеширование	Расписание занятий группы: <u>номер группы</u> , название дисциплины, номер пары, номер недели, номер дня недели, вид занятия, номер аудитории.
23	Открытый адрес(двойное хеширование)	Нагрузка по дисциплине: <u>код дисциплины</u> , <u>код направления подготовки</u> , название дисциплины, номер семестра проведения дисциплины.
24	Цепное хеширование	Нагрузка по дисциплине: <u>код дисциплины</u> , <u>код направления подготовки</u> , название дисциплины, номера семестров проведения дисциплины (не более двух). <u>Подсказка</u> . Если только один семестр, то второе поле должно содержать 0.
25	Открытый адрес (смещение на 1)	Аэропорт (табло прибытия пассажирских авиарейсов сохраняет все данные в файле): пункт вылета, <u>номер рейса</u> , дата прилета, время прилета, информация о задержке прилета в часах.
26	Цепное хеширование	Учет заболеваний пациента. Структура записи о пациенте: <u>номер полиса</u> , фамилия, имя, отчество, код заболевания, дата установки диагноза, код врача.
27	Открытый адрес (смещение на 1)	Учет техосмотра автомобилей. Структура записи об автомобиле:

		<u>Номер</u> (код региона, цифровой код, буквенный код), Модель, Цвет, Сведения о владельце (Фамилия, Имя, Адрес), дата последнего техосмотра.
28	Открытый адрес (смещение на 1)	Учет нарушений ПДД. Структура записи о нарушении ПДД: <u>номер автомобиля</u> , фамилия и инициалы владельца, модель, дата нарушения, место нарушения (текстом), статья (КоАП), наказание (сумма штрафа).
29	Открытый адрес (двойное хеширование)	Справочник банков по городам страны. Об отдельном банке хранятся данные: наименование, <u>код банка</u> , адрес (город), форма собственности (коммерческий или государственный).
30	Цепное хеширование	Касса магазина. Структура записи операции по кассе: <u>номер кассы</u> , код товара, количество товара, цена товара, процентная скидка на товар, сумма за товар с учетом скидки.
31	Открытый адрес (смещение на 1)	Киноафиша города. Структура записи о сеансе: <u>название кинотеатра</u> , название фильма, дата, время начала, стоимость билета.