**Лабораторная работа №6**

**Тема:** Моделирование хеш-таблиц;

**Цель работы:** освоение навыков построения и проверки таблиц хеширования.

**Краткие теоретические сведения**

Теоретические сведения , необходимые для выполнения лабораторной работы №6, включающие сведения об основных особенностях программного способа адресации по содержанию(хешированию), методам и функциям хеширования, способам перевода ключевых слов в числовые значения (V), способам преобразования числовых значений V в хеш-адрес(h), о понятии “Коллизия”, основным методам(способам) обработки коллизии, о понятии “пробинг” и видах пробинга, структуре хеш-таблиц, приведены в теоретической части ЭРУД “АОИС”(тема 11, подразделы 11.1-11.3).

*Назначение и применение таблиц хеширования* Таблицы хеширования применяются для обеспечения программного поиска информации по содержанию специального ключевого слова или фрагмента (части) самих данных.

*Основные понятия*  
 Для формирования таблиц хеширования используются следующие компоненты и понятия:  
 ID – ключевое слово (любое имя, слово, сочетание любых символов и т.д.) или его идентификатор;  
 V – числовое значение ключевого слова;  
 h (V) – хеш-адрес, вычисленный по числовому значению V ключевого слова;  
 В – начальный адрес таблицы (величина кратная размеру строки таблицы);  
 Н – общее количество строк (ячеек) в таблице.  
 Способы перевода ключевых слов в числовые значения и преобразование их в хеш-адреса рассмотрены в [9] и на лекции по теме 11.  
 Напомним, что при совпадении хеш-адресов, вычисленных по разным ключевым словам, возникает коллизия (конфликт). В этом случае для размещения второй и последующих записей необходимо использовать резервные ячейки памяти, которые размещаются либо в самой таблице хеширования (внутренняя адресация), либо в специальной области памяти (области переполнения).  
 Для поиска свободных резервных ячеек при внутренней адресации применяется процедура пробинга (линейного, квадратичного или случайного).  
 *Структура ячейки (строки) хеш-таблицы*  
 В состав ячейки (строки) хеш-таблицы должны входить:  
 – ключевое слово (ID) или его идентификатор;  
 – соответствующие этому ключу данные или указатели;  
 – различные флажки (признаки).

Формат ячейки (строки) хеш-таблицы приведен на рисунке 1.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ID | C | U | T | L | D | Po | Данные или указатели (Pi) |

Рисунок 1. Формат ячейки (строки) хеш-таблицы

Обозначения на рисунке 1[9]:  
ID – идентификатор ключевого слова или само ключевое слово (К),  
C – флажок коллизий,  
U – флажок «занято»,  
Т – терминальный флажок,  
L – флажок связи, указывающий, что находится в поле Pi – данные или указатель(адрес), где находятся данные;  
D – флажок вычеркивания,  
Ро – указатель области переполнения (или следующей записи в цепочке),  
Pi – данные или указатель области данных.

**Рассмотрим пример построения хеш-таблицы.**

Построим хеш-таблицу в соответствии с приведенными ниже исходными данными.  
 Справочные данные о группе или подгруппе студентов (не менее 10 человек), включающие:  
1) фамилии студентов  
2) любые данные (например, имя, увлечения(спорт,искусство)).  
 Ключевое слово – фамилия студента.  
 *Примечание:* могут быть использованы другие ключевые слова, например, набор предметов, перечень специальностей и т. д.

**Решение**

1. Применим седующий способ определения значения ключевого слова (V) по его первым двум буквам:  
    - используем русский алфавит;  
    - буквам присваиваются соответствующие номера:  
   А=0, Б=1, ... , Я=32, т. е. всего 33 номера. Будем эту величину считать основанием позиционной системы счисления, тогда сочетание двух первых букв фамилии (ключевого слова), можно записать, например, следующим образом:  
    а) Вяткин V[Вя] = 2∙  
    б) Третьяк V[Tp] =   
   Аналогично может быть использован латинский алфавит.
2. Структура ячейки (строки) хеш-таблицы и ее размер определяется в соответствии с исходными данными и форматом ячейки (см. Рисунок 1).
3. Спасоб формирования хеш-адреса определяется размером таблицы и вычисленным значением V (например, h(V)=V ∙ mod H + B, где V ∙ mod H – остаток от деления V на H).  
    Способы обработки коллизий:  
    - размещение резервных ячеек в хеш-таблице (внутренняя адресация);  
    - использование области переполнения.  
    Способ (вид) пробинга при выборе резервных ячеек при использовании внутренней адресации – линейный.  
    Вариант хеш-таблицы некоторой условной группы студентов приведен в таблице 1, эта таблица состоит из двух частей:  
    а) правая часть – это собственно хеш-таблица;  
    б) левая часть, вкючающая фамилии студентов, расположенные в порядке записи строк в хеш-таблицу, вычисленные значения V и h для каждого ключевого слова (фамилии студента).  
    Размер таблицы Н=20 строк.  
    Начальный адрес таблицы В=0 (можно взять любое другое значение, но оно не вляет на постороение хеш-таблицы, а только на расположение таблицы в памяти).

*Таблица 1.*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Левая часть | | | Правая часть (ТХ) | | | | | | | | |
| Фамилия студента | V | h | № строки ТХ | Значение полей и флажков ТХ | | | | | | | |
| ID | C | U | T | L | D |  | (данные) |
| 1.Абаев (ID1) | 1 | 1 | 0 | ID6 | 1 | 1 | 1 | 0 | 0 | 2 | Сергей |
| 2.Бобков (ID2) | 48 | 18 | 1 | ID1 |  | 1 | 1 | 0 | 0 |  | Тимур |
| 3.Видерт (ID3) | 75 | 15 | 2 | ID7 | 0 | 1 | 0 | 0 | 0 | 3 | Евгений |
| 4.Гракова (ID4) | 116 | 16 | 3 | ID8 |  | 1 | 0 | 0 | 0 | 12 | Иван |
| 5.Кожевников (ID5) | 388 | 18 | 4 | ID12 |  | 1 | 1 | 0 | 0 |  | Максим |
| 6.Ковалев (ID6) | 378 | 18 | 5 |  |  |  |  |  |  |  |  |
| 7.Крикунов (ID7) | 380 | 0 | 6 |  |  |  |  |  |  |  |  |
| 8.Кот (ID8) | 378 | 18 | 7 |  |  |  |  |  |  |  |  |
| 9.Давыденко (ID9) | 132 | 12 | 8 | ID13 | 0 | 1 | 1 | 0 | 0 |  | Александр |
| 10.Горбань (ID10) | 114 | 14 | 9 |  |  |  |  |  |  |  |  |
| 11.Данилов (ID11) | 132 | 12 | 10 |  |  |  |  |  |  |  |  |
| 12.Козлов (ID12) | 378 | 18 | 11 |  |  |  |  |  |  |  |  |
| 13.Азимов (ID13) | 8 | 8 | 12 | ID9 | 1 | 1 | 0 | 0 | 0 | 13 | Ирина |
|  |  |  | 13 | ID11 |  | 1 | 1 | 0 | 0 |  | Павел |
|  |  |  | 14 | ID10 |  | 1 |  | 0 | 0 |  | Петр |
|  |  |  | 15 | ID3 |  | 1 |  | 0 | 0 |  | Руслан |
|  |  |  | 16 | ID4 |  | 1 |  | 0 | 0 |  | Наталья |
|  |  |  | 17 |  |  |  |  |  |  |  |  |
|  |  |  | 18 | ID2 | 1 | 1 | 0 | 0 | 0 | 19 | Андрей |
|  |  |  | 19 | ID5 | 0 | 1 | 1 | 0 | 0 | 00 | Константин |

*Примечание:* В графе «ID» правой части таблицы записываются полное ключевое слово (фамилия) соответствующая индексу (1-20).

Напомним, что с хеш-таблицей можно выполнять различные операции, в частности, поиск информации по ключевому слову, запись в таблицу новых строк, удаление отдельных строк из таблицы. Правила и выполнения таких операций подробно рассмотрели в части «Теория ЭРУД АОИС» (Тема 11, подразделы 11.1-11.3)

*Кратко напомним только операцию удаления.*  
 Удаление строк из хеш-таблицы производиться следующим образом:   
 1. По ключевому слову удаляемой строки определяется ее хеш-адрес(h), отыскивается в ТХ эта строка и в ней устанавливается флажок D в «1».

2. Анализируется состояние флажков Т и С, содержимое поля и определяется какая это строка – одиночная (т.е не является частью цепочки строк(ячеек), сформированных из-за возникших при формировании таблицы коллизий), последняя строка цепочки или любая другая строка в цепочке. Выполнить это можно следующим образом:   
 а) если состояние флажка Т=1 и в поле записан адрес этой же строки (h), то эта строка является одиночной и ее можно просто удалить, сделать свободным, т.е. установить флажок U в «0»  
 б) если состояние флажка Т=1 и в поле записан какой-то код, указывающий, что эта строка в ходит в цепочку резервных ячеек, то эта строка является конечной в какой-то цепочке. В этом случае отслеживается предыдущая относительно удаляемой строки в цепочке путем сравнения содержимого поля всех предыдущих строк с адресом удаляемой строки h и далее:  
 - удаляемая строка делается свободной (флажок U устанавливается в 0, значение остальных полей и флажков может быть произвольным)  
 - в найденной предыдущей строке, у коротой значения поля = h удаляемой строки, флажок Т устанавливается в «1».  
 в) Если состояние флажка Т=0 и содержание поля , то удаляемая строка является другой строкой (не последней) в цепочке резервных ячеек. В этом случае находится последующая за удаляемой строкой в этой цепочке по содержимому поля и эта строка переписывается на место удаляемой и делается свободной, т.е. в ней устанавливается флажок U в «0», фактически удаляется не вычисление по хеш-адресу строки, а следующее за ней, при этом содержимое остальных строк этой цепочки не изменяется.  
 г) если состояние флажка Т=0, а С=1, то эта строка является первой в цепочке, начинающейся с хеш-адреса удаляемой строки. В этом случае по содержимому поля остлеживается следующая за удаляемой строка, её содержимое переписывается в удаляемую, в ней устанавливается С=1, и эта строка делается свободной (U=0).

**Контрольные вопросы:**

1. Для чего спользуются таблицы хеширования?
2. Что используется для программного поиска информации по содержанию?
3. Назовите основные функции хештрования.
4. Приведите структуру хеш-таблицы.
5. Приведите способы перевода ключевых слов в числовые значения.
6. Приведите способы преобразования числового значения ключевого слова в хеш-адрес.
7. Что такое коллизия? В каких случаях она возникает?
8. Назовите способы обработки коллизий.
9. В каких случаях производится процедура пробинга?

**Индивидуальные задания:**

1. Разработать и проверить программу, обеспечивающую формирование хеш-таблицы, по ключевым словам и выполнение различных операций с этой таблицей – включение в таблицу новых строк, поиск информации в таблице по ключевым словам, удаление строк из таблицы.
2. Хеш-таблица должна быть разработана по образцу примера, рассмотренного выше.
3. Хеш-таблица должна быть тематической, т.е. все ключевые слова и данные должны быть из определенной отрасли науки, техники или деятельности людей. Перечень вариантов тематической направленности хеш-таблицы приведен в таблице 2.  
   *Таблица 2.*

|  |  |  |  |
| --- | --- | --- | --- |
| № варианта | Тематическая направленность (ТХ) | Ключевые слова (ID) | Данные () |
| 1 | Математика | Названия, термины, понятия из соответствующей отрасли науки, техники, деятельности. | Определение, что это или из какого раздела данной науки, техники и т.д. это ключевое слово. |
| 2 | Физика |
| 3 | География |
| 4 | Биология |
| 5 | Грамматика |
| 6 | Литература |
| 7 | Медицина |
| 8 | Транспорт |
| 9 | Спорт |
| 0 | Армия, вооружение |

1. Каждый студент выполняет вариант хеш-таблицы, номер которого совпадает с последней цифрой номера записи его фамилии в списке группы.  
    *Примечание:* если у студента доставшийся ему вариант хеш-таблицы вызывает затруднее, например, из-за недостатка знаний в данной тематике, он может предложить и по согласованию с преподавателем выполнить свой другой вариант, отсутствующий в таблице 2 с соблюдением всех требований к построению хеш-таблиц, приведенных в данном методическом материале.
2. При выполнение любого варианта ЛР №6 должны быть выполнены слудующие требовония:  
    - размер таблицы не менее 20 строк;  
    - количество записанных в таблице строк – не менее 10-ти;  
    - количество коллизий не менее 2-х;  
    - количество строк, объединенных в цепочки резервных ячеек, сформированных из-за коллизий – не менее 3-х;  
    - размер и структура каждой строки хеш-таблицы должны быть одинаковыми и постоянными.  
    *Примечание:* если размер поля в хеш-таблице не позволяет записать все данные по конкретному ключевому слову, может быть использована отдельная область памяти для записи этих данных, при этом в поле записывается указатель, где в памяти находятся данные, и флажок «L» (флажок связи) устанавливается в «1».

**Требовония к программе:**

1. Разработаная программа должна уметь выполнять следующие функции:  
    - определение числовых значений ключевых слов (V);  
    - формирование хеш-адресов [h(v)];  
    - формирование самой хеш-таблицы, в соответствии со структурой, приведенной на рисунке 1;  
    - обработку коллизии (т.е. обеспечение нахождения резервных ячеек и запись в них);  
    - выборку из хеш-таблицы необходимой информации по соответствующему ключевому слову;  
    - занесение в хеш-таблицу новых новых записей по новым ключевым словам;  
    - удаление из хеш-таблицы отдельных записей;  
    - контроль за потыткой записи в хеш-таблицу информации по ключевому слову, уже имеющемуся в таблице;  
    - расчет коэффициентов заполнения хеш-таблицы.
2. Программа должна обеспечивать индикацию (вывод на экран) содержимого всех полей и указателей всех строк хеш-таблицы, вычисленных значений V(K) и h(V) для каждой записи, коэффициента заполнения хеш-таблицы.

**Методика выполнения:**

1. Определяется перечень ключевых слов и данных для записи в таблицу.
2. Определяется структура и размер строк таблицы.
3. Строится программа выполняющая:  
    - формирование хеш-таблицы;  
    - формирование хеш-адресов;  
    - занесение (запись) в таблицу записей (строк);  
    - обработку коллизий;  
    - поиск записей в хеш-таблице по задонному ключевому слову;  
    - удаление записей (строк) из хеш-таблицы;  
    - контроль за коррекностью использования ключевых слов (каждое ключевое слово может быть использовано только для одной записи);  
    - индикацию вычисленных для каждого ключевого слова числового значения и хеш-адреса.

Стандартная процедура проверки разработанной программы заключается в анализе результатов выполнения предъявленных к ней требований.