Лабораторная работа № 4

Выполнила: студент группы 4309

Курбанов Д.И.

Основное назначение хеширования — проверка информации. Эта задача важна в огромном количестве случаев: от проверки паролей на сайте до сложных вычислений в блокчейне. Так как хеш — это уникальный код определенного набора данных, по нему можно понять, соответствует ли информация ожидаемой. Поэтому программа может хранить хеши вместо образца данных для сравнения. Это может быть нужно для защиты чувствительных сведений или экономии места.

Хеш-табли́ца — это структура данных, реализующая интерфейс ассоциативного массива, а именно, она позволяет хранить пары (ключ, значение) и выполнять три операции: операцию добавления новой пары, операцию поиска и операцию удаления пары по ключу. Существуют два основных варианта хеш-таблиц: с цепочками и открытой адресацией.

Хеш-функция— функция, осуществляющая преобразование массива входных данных произвольной длины в выходную битовую строку установленной длины, выполняемое определённым алгоритмом. Преобразование, производимое хеш-функцией, называется хешированием. Исходные данные называются входным массивом, «ключом» или «сообщением».

Под коллизией понимается ситуация, когда при добавлении разных объектов мы попадаем в одну и ту же ячейку массива. Для разрешения коллизий придумано 2 метода: метод цепочек и метод открытой адресации.

Коллизии существуют для большинства хеш-функций, но для «хороших» хеш-функций частота их возникновения близка к теоретическому минимуму. В некоторых частных случаях, когда множество различных входных данных [конечно](https://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D0%BD%D0%B5%D1%87%D0%BD%D0%BE%D0%B5_%D0%BC%D0%BD%D0%BE%D0%B6%D0%B5%D1%81%D1%82%D0%B2%D0%BE), можно задать [инъективную](https://ru.wikipedia.org/wiki/%D0%98%D0%BD%D1%8A%D0%B5%D0%BA%D1%82%D0%B8%D0%B2%D0%BD%D0%BE%D1%81%D1%82%D1%8C) хеш-функцию, по определению не имеющую коллизий. Однако для хеш-функций, принимающих вход переменной длины и возвращающих хеш постоянной длины, коллизии обязаны существовать, поскольку хотя бы для одного значения хеш-функции соответствующее ему множество входных данных будет бесконечно — и любые два набора данных из этого множества образуют коллизию.

Метод цепочек является наиболее простым методом разрешения коллизий. В ячейке массива мы будем хранить не элементы, а связанный список данных элементов. Потому как добавление в начало списка обладает асимптотикой , мы не испортим общую асимптотику, и она останется равной.

У данной реализации есть проблема: если списки будут очень сильно вырастать (в качестве крайнего случая можно рассмотреть хеш-функцию, которая возвращает константу для любого объекта), то мы получим асимптотику O(m), где m — число элементов во множестве, если размер массива фиксирован. Для избежания таких неприятностей вводится понятие коэффициент заполнения (он может быть равен, например, 1.5). Если при добавлении элемента оказывается, что доля числа элементов, находящихся в структуре данных по отношению к размеру массива, превосходит коэффициент заполнения, то происходит следующее: выделяется новый массив, размер которого превосходит размер старого массива (например в 2 раза), и структура данных перестраивается на новом массиве.

Метод открытой адресации. В данном методе в ячейках хранятся сами элементы, а в случае коллизии происходит последовательность проб, то есть мы начинаем по некоторому алгоритму перебирать ячейки в надежде найти свободную. Это можно делать разными алгоритмами (линейная / квадратичная последовательности проб, двойное хеширование), каждый из которых обладает своими проблемами (например, возникновение первичных или вторичных кластеров).

**Задание**. Синтаксический анализатор размещает распознанные с помощью лексического анализатора слова в таблицы. Количество таблиц соответствует числу типов слов. В нашем случае нужны три таблицы: таблица чисел (первый тип слова), таблица идентификаторов (второй тип слова), таблица составленных из специальных символов слов (третий тип слова). Таблицы организуются методом хеширования. Включить класс «Поиск слова в таблице методом хеширования» в разрабатываемый транслятор.

Код программы:

MyHashFunction.сs:

using System;

using System.Collections.Generic;

using System.Linq;

using System.Text;

using System.Threading.Tasks;

namespace WindowsFormsApp4

{

public class MyHashFunction

{

public int HashFunction(string word)

{

int hashValue = 0;

foreach (char c in word)

{

hashValue += (int)c;

}

return hashValue;

}

public void AddWord(Dictionary<int, List<string>> hashTable, string word)

{

int hashValue = HashFunction(word);

if (!hashTable.ContainsKey(hashValue))

{

hashTable[hashValue] = new List<string>();

}

else

{

if (hashTable[hashValue].Contains(word))

{

Console.WriteLine("Слово уже существует: " + word);

return;

}

}

hashTable[hashValue].Add(word);

}

public int SearchWord(Dictionary<int, List<string>> hashTable, string word)

{

int hashValue = HashFunction(word);

if (hashTable.ContainsKey(hashValue))

{

return 1;

}

else

{

return 0;

}

}

public bool RemoveWord(Dictionary<int, List<string>> hashTable, string word)

{

int hashValue = HashFunction(word);

if (hashTable.ContainsKey(hashValue))

{

List<string> words = hashTable[hashValue];

if (words.Contains(word))

{

words.Remove(word);

if (words.Count == 0)

{

hashTable.Remove(hashValue);

}

return true;

}

}

return false;

}

}

}

Form1.cs

using System;

using System.Collections.Generic;

using System.ComponentModel;

using System.Data;

using System.Drawing;

using System.Linq;

using System.Text;

using System.Threading.Tasks;

using System.Windows.Forms;

namespace WindowsFormsApp4

{

public partial class Form1 : Form

{

Dictionary<int, List<string>> hashTableIdentifier = new Dictionary<int, List<string>>();

Dictionary<int, List<string>> hashTableDigital = new Dictionary<int, List<string>>();

Dictionary<int, List<string>> hashTableRezerv = new Dictionary<int, List<string>>();

public MyHashFunction hashFunction = new MyHashFunction();

public Form1()

{

InitializeComponent();

//textBox2.AppendText("010000001 010000001 caaaa");

textBox2.AppendText("010000001 caaa cab 010000001 caaaa");

int n = textBox2.Lines.Length;

}

private void Form1\_Load(object sender, EventArgs e)

{

}

private void Form1\_Load\_1(object sender, EventArgs e)

{

}

private void label5\_Click(object sender, EventArgs e)

{

}

private void button1\_Click(object sender, EventArgs e)

{

textBox4.Clear();

uSyntAnalyzer Synt = new uSyntAnalyzer();

Synt.Lex.strPSource = textBox2.Lines;

Synt.Lex.strPMessage = textBox4.Lines;

Synt.Lex.enumPState = TState.Start;

Synt.Lex.hashTableRezerv = hashTableRezerv;

try

{

Synt.S();

Synt.Lex.NextToken();

throw new Exception("Текст верный");

}

catch (Exception exc)

{

textBox4.Text += exc.Message;

textBox2.Select();

textBox2.SelectionStart = 0;

int n = 0;

for (int i = 0; i < Synt.Lex.intPSourceRowSelection; i++) n += textBox2.Lines[i].Length + 2;

n += Synt.Lex.intPSourceColSelection;

textBox2.SelectionLength = n;

}

}

private void button2\_Click(object sender, EventArgs e)

{

CLex Lex = new CLex();

Lex.strPSource = textBox2.Lines;

Lex.strPMessage = textBox4.Lines;

Lex.intPSourceColSelection = 0;

Lex.intPSourceRowSelection = 0;

int x = textBox2.TextLength;

int y = textBox2.Lines.Length;

textBox4.Text = "";

try

{

while (Lex.enumPState != TState.Finish)

{

Lex.NextToken();

string s1 = "", s = "";

switch (Lex.enumPToken)

{

case TToken.lxmIdentifier:

{

hashFunction.AddWord(hashTableIdentifier, Lex.strPLexicalUnit);

listBox4.Items.Add(Lex.strPLexicalUnit);

s1 = "id " + Lex.strPLexicalUnit;

break;

}

case TToken.lxmNumber:

{

hashFunction.AddWord(hashTableDigital, Lex.strPLexicalUnit);

listBox5.Items.Add(Lex.strPLexicalUnit);

s1 = "num " + Lex.strPLexicalUnit;

break;

}

case (TToken.lxmtz):

{

hashFunction.AddWord(hashTableRezerv, "\_");

listBox6.Items.Add(Lex.strPLexicalUnit);

s1 = "rez \_" + Lex.strPLexicalUnit;

break;

}

}

String m = "(" + s + "" + s1 + ")";

textBox4.Text += m;

}

}

catch (Exception exc)

{

textBox4.Text += exc.Message;

textBox2.Select();

textBox2.SelectionStart = 0;

int n = 0;

for (int i = 0; i < Lex.intPSourceRowSelection; i++) n += textBox2.Lines[i].Length + 2;

n += Lex.intPSourceColSelection;

textBox2.SelectionLength = n;

}

}

private void button3\_Click(object sender, EventArgs e)

{

if (hashFunction.SearchWord(hashTableIdentifier, listBox4.SelectedItem.ToString()) == 1)

{

button3.BackColor = Color.Green;

}

else

{

button3.BackColor = Color.Red;

}

}

private void button4\_Click(object sender, EventArgs e)

{

hashFunction.AddWord(hashTableIdentifier, textBox3.Text.ToString());

}

private void button5\_Click(object sender, EventArgs e)

{

if (hashFunction.RemoveWord(hashTableIdentifier, listBox4.SelectedItem.ToString()))

{

button5.BackColor = Color.Green;

}

else

{

button5.BackColor = Color.Red;

}

}

private void button6\_Click(object sender, EventArgs e)

{

if (hashFunction.RemoveWord(hashTableIdentifier, listBox4.SelectedItem.ToString()))

{

hashFunction.AddWord(hashTableIdentifier, textBox3.Text.ToString());

button6.BackColor = Color.Green;

}

else

{

button6.BackColor = Color.Red;

}

}

private void button7\_Click(object sender, EventArgs e)

{

listBox4.Items.Clear();

listBox5.Items.Clear();

listBox6.Items.Clear();

foreach (var entry in hashTableIdentifier)

{

listBox4.Items.Add(string.Join(", ", entry.Value));

}

foreach (var entry in hashTableDigital)

{

listBox5.Items.Add(string.Join(", ", entry.Value));

}

foreach (var entry in hashTableRezerv)

{

listBox6.Items.Add(string.Join(", ", entry.Value));

}

}

private void label7\_Click(object sender, EventArgs e)

{

}

private void listBox6\_SelectedIndexChanged(object sender, EventArgs e)

{

}

}

}

Результаты программы:











