**Министерство образования и науки Российской Федерации**

*Федеральное государственное бюджетное образовательное учреждение   
высшего профессионального образования*   
**«АЛТАЙСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. И.И. ПОЛЗУНОВА» (АлтГТУ)**

Факультет информационных технологий

Кафедра прикладной математики

Отчёт по дисциплине «Алгоритмы и структуры данных»

Лабораторная работа №3

«Суффиксные автоматы»

Выполнил:

Студент гр. ПИ-32

Глушков Г.Г.

Проверил:

ст. преподаватель Н.Д. Бубнова

Барнаул 2015

* Формулировка задачи:
* Даны строки S и T и текст Х. Найти наидлиннейшую общую подстроку S и T, не содержащуюся в тексте Х.
* Алгоритм решения:
* Считываем из файла input.txt исходный текст (S и T и текст Х);
* Реализуем алгоритм построения суффиксных автоматов:
* Вначале опишем структуру данных, которая будет хранить всю информацию о конкретном переходе (, , список переходов). При необходимости сюда можно добавить флаг терминальности, а также другую требуемую информацию. Список переходов мы храним в виде стандартного контейнера  , что позволяет достичь суммарно  памяти и  времени на обработку всей строки;
* Сам суффиксный автомат будем хранить в виде массива этих структур . Как доказывается в следующем разделе, если  — это максимально возможная в программе длина строки, то достаточно завести память под  состояний. Также мы храним переменную  — состояние, соответствующее всей строке на данный момент;
* Приведём функцию, инициализирующую суффиксный автомат (создающую автомат с единственным начальным состоянием);
* Наконец, приведём реализацию основной функции — которая добавляет очередной символ в конец текущей строки, перестраивая соответствующим образом автомат;
* Найдем наибольшую общую строку и поищем её же в тексте();
* Построив суффиксные автоматы выведем в файл тот, который для теста;
* Результат работы программы пишем в файл output.txt.
* Текст программы с комментариями:

/\*

Даны строки S и T и текст Х.

Найти наидлиннейшую общую подстроку S и T, не содержащуюся в тексте Х

Вывести суффиксный автомат для текста Х

\*/

#include "stdafx.h"

#include <cstdio>

#include <map>

#include <string>

#include <tchar.h>

#include <vector>

const int MAXLEN = 10000; //максимальная длина строки

char buf[MAXLEN];

bool eof = false;

using namespace std;

//структура, описывающая состояние суффиксного автомата

struct state {

int len, link; //длина строки, соответствующей текущему состоянию автомата и суффиксная ссылка

map<char, int> next; //переходы по символам

state() :

len(0), link(-1){}//начальные значения

};

struct SuffixAutomaton{

vector<state> states;

int last; //состояние, соответствующее всей текущей строке до добавления символа c

int stateId; //состояние суффиксного автомата

SuffixAutomaton() :

last(0), stateId(1){}

void build(string &str); //построить суффиксный автомат на строке str

void extend(char c); //функция добавления очередного символа строки и преобразование автомата затем

bool find(string &str); //функция поиска строки (показывает наличие)

};

//построить суффиксный автомат на строке str

void SuffixAutomaton::build(string &str){

states.assign(2 \* (int)str.size(), state());

for (int i = 0; i < (int)str.size(); i++)

extend(str[i]);

}

//функция добавления очередного символа строки и преобразование автомата затем

void SuffixAutomaton::extend(char c){

int cur = stateId++;

states[cur].len = states[last].len + 1;

int p; //состояние, которое связано суффиксными ссылками с last и которое содержит переход по С

//проходим по суффиксным ссылкам начиная с last и если нет перехода по символу С в текущее состояние, то добавить его

for (p = last; p != -1 && !states[p].next.count(c); p = states[p].link)

states[p].next[c] = cur;

//если не было таких состояний, связанных суффиксными ссылками с last,

//то суффиксная ссылка текущего состояния указывает на начальное

if (p == -1)

states[cur].link = 0;

else {

//а если такой переход уже существовал, то рассмотрим два случая

int q = states[p].next[c]; //запоминаем состояние, в которое ведет тот самый переход

//если len(p) + 1 = len(q), то мы можем просто присвоить link(cur) = q и выйти

if (states[p].len + 1 == states[q].len)

states[cur].link = q;

else {

//иначе нужно создать клона состояния q (вместе с переходами и суффиксной ссылкой)

int clone = stateId++;

//длина строки, соответствующей состоянию будет равна len(p)+1

states[clone].len = states[p].len + 1;

states[clone].next = states[q].next;

states[clone].link = states[q].link;

//перенаправление переходов по символу С со всех состояний в q, соединенных с состоянием p, на состояние clone

for (; p != -1 && states[p].next[c] == q; p = states[p].link)

states[p].next[c] = clone;

//назначение суффиксных ссылок: ссылки с q и с текущего состояния указывают на clone

states[q].link = states[cur].link = clone;

}

}

last = cur;

}

//функция поиска строки (показывает наличие)

bool SuffixAutomaton::find(string &str){

int curState = 0;

bool found = true;

//последовательно для каждого символа строки пытаемся

//сделать соответствующий переход в суффиксном автомате

for (int i = 0; i < (int)str.size() && found; i++){

if (states[curState].next.count(str[i]) == 0)

found = false;

else

curState = states[curState].next[str[i]];

}

return found;

}

//функция считывания строки

string readLine(){

if (scanf("%s", buf) == -1){

eof = true;

return string();

}

return string(buf);

}

void printSA(SuffixAutomaton &sa){//функция вывода суффиксного автомата

for (int i = 0; i < sa.stateId; i++){

printf("состояние №%d, суффиксная ссылка: %d\n", i, sa.states[i].link);

for (map<char, int>::iterator j = sa.states[i].next.begin(); j != sa.states[i].next.end(); j++){

printf(" переход по символу '%c' в состояние %d\n", j->first, j->second);

}

}

}

int \_tmain(int argc, \_TCHAR\* argv[])

{

freopen("input.txt", "rt", stdin);

freopen("output.txt", "wt", stdout);

setlocale(LC\_CTYPE, "Russian");

SuffixAutomaton text\_sa, t\_sa;

string s, t, text;

//ввод данных

s = readLine();

t = readLine();

//строка s - строка с минимальной длиной

if (s.size() > t.size())

swap(s, t);

//ввод текста

do{

text = text + readLine();

} while (!eof);

//построение суффиксных автоматов

text\_sa.build(text);

t\_sa.build(t);

//нахождение наибольшей общей подстроки и поиск этой подстроки в тексте

bool generalSubStr = false;

bool found = true;

for (int len = (int)s.size(); len > 0 && found; len--){

for (int first = 0; first <= (int)s.size() - len && found; first++){

//поиск текущей подстроки в другой строке

string substr = s.substr(first, len);

if (t\_sa.find(substr)){

//если подстрока общая, то попытаться найти ее в тексте

generalSubStr = true;

if (!text\_sa.find(substr)){

//нашли искомую подстроку

found = false;

printf("Искомая строка: %s\n", substr.c\_str());

}

}

}

}

if (!generalSubStr){

printf("Общих подстрок нет.\n");

}

else if (found){

printf("Все общие подстроки содержатся в тексте.\n");

}

printf("\nСуффиксный автомат будет выглядеть следующим образом:\n");

printSA(text\_sa);

return 0;

}

* Тесты:
* Содержимое исходного файла:

Привет

Пока

Меня зовут Артур. Что делаешь? Я пишу программу, которая реализует алгоритм написания суффиксного автомата.

Содержимое итогового файла:

Искомая строка: П

Суффиксный автомат будет выглядеть следующим образом:

состояние №0, суффиксная ссылка: -1

переход по символу 'А' в состояние 10

переход по символу 'М' в состояние 1

переход по символу 'Ч' в состояние 19

переход по символу 'Я' в состояние 32

переход по символу 'а' в состояние 45

переход по символу 'в' в состояние 113

переход по символу 'г' в состояние 77

переход по символу 'д' в состояние 23

переход по символу 'е' в состояние 25

переход по символу 'з' в состояние 69

переход по символу 'и' в состояние 67

переход по символу 'к' в состояние 103

переход по символу 'л' в состояние 65

переход по символу 'м' в состояние 48

переход по символу 'н' в состояние 85

переход по символу 'о' в состояние 22

переход по символу 'п' в состояние 39

переход по символу 'р' в состояние 17

переход по символу 'с' в состояние 96

переход по символу 'т' в состояние 13

переход по символу 'у' в состояние 15

переход по символу 'ф' в состояние 100

переход по символу 'ш' в состояние 36

переход по символу 'ь' в состояние 30

переход по символу 'я' в состояние 60

переход по символу ',' в состояние 50

переход по символу '.' в состояние 122

переход по символу '?' в состояние 31

состояние №1, суффиксная ссылка: 0

переход по символу 'е' в состояние 2

состояние №2, суффиксная ссылка: 25

переход по символу 'н' в состояние 3

состояние №3, суффиксная ссылка: 85

переход по символу 'я' в состояние 4

состояние №4, суффиксная ссылка: 60

переход по символу 'з' в состояние 5

состояние №5, суффиксная ссылка: 69

переход по символу 'о' в состояние 6

состояние №6, суффиксная ссылка: 22

переход по символу 'в' в состояние 7

состояние №7, суффиксная ссылка: 113

переход по символу 'у' в состояние 8

состояние №8, суффиксная ссылка: 15

переход по символу 'т' в состояние 9

состояние №9, суффиксная ссылка: 13

переход по символу 'А' в состояние 10

состояние №10, суффиксная ссылка: 0

переход по символу 'р' в состояние 11

состояние №11, суффиксная ссылка: 17

переход по символу 'т' в состояние 12

состояние №12, суффиксная ссылка: 13

переход по символу 'у' в состояние 14

состояние №13, суффиксная ссылка: 0

переход по символу 'А' в состояние 10

переход по символу 'а' в состояние 120

переход по символу 'м' в состояние 83

переход по символу 'о' в состояние 55

переход по символу 'у' в состояние 14

состояние №14, суффиксная ссылка: 15

переход по символу 'р' в состояние 16

состояние №15, суффиксная ссылка: 0

переход по символу 'е' в состояние 71

переход по символу 'п' в состояние 38

переход по символу 'р' в состояние 16

переход по символу 'т' в состояние 9

переход по символу 'ф' в состояние 98

переход по символу ',' в состояние 50

состояние №16, суффиксная ссылка: 17

переход по символу '.' в состояние 18

состояние №17, суффиксная ссылка: 0

переход по символу 'а' в состояние 58

переход по символу 'е' в состояние 62

переход по символу 'и' в состояние 81

переход по символу 'о' в состояние 41

переход по символу 'т' в состояние 12

переход по символу '.' в состояние 18

состояние №18, суффиксная ссылка: 122

переход по символу 'Ч' в состояние 19

состояние №19, суффиксная ссылка: 0

переход по символу 'т' в состояние 20

состояние №20, суффиксная ссылка: 13

переход по символу 'о' в состояние 21

состояние №21, суффиксная ссылка: 55

переход по символу 'д' в состояние 23

состояние №22, суффиксная ссылка: 0

переход по символу 'а' в состояние 111

переход по символу 'в' в состояние 7

переход по символу 'г' в состояние 108

переход по символу 'д' в состояние 23

переход по символу 'м' в состояние 116

переход по символу 'р' в состояние 80

переход по символу 'т' в состояние 53

состояние №23, суффиксная ссылка: 0

переход по символу 'е' в состояние 24

состояние №24, суффиксная ссылка: 25

переход по символу 'л' в состояние 26

состояние №25, суффиксная ссылка: 0

переход по символу 'а' в состояние 63

переход по символу 'л' в состояние 26

переход по символу 'н' в состояние 3

переход по символу 'т' в состояние 72

переход по символу 'ш' в состояние 29

состояние №26, суффиксная ссылка: 65

переход по символу 'а' в состояние 27

состояние №27, суффиксная ссылка: 45

переход по символу 'е' в состояние 28

состояние №28, суффиксная ссылка: 25

переход по символу 'ш' в состояние 29

состояние №29, суффиксная ссылка: 36

переход по символу 'ь' в состояние 30

состояние №30, суффиксная ссылка: 0

переход по символу '?' в состояние 31

состояние №31, суффиксная ссылка: 0

переход по символу 'Я' в состояние 32

состояние №32, суффиксная ссылка: 0

переход по символу 'п' в состояние 33

состояние №33, суффиксная ссылка: 39

переход по символу 'и' в состояние 34

состояние №34, суффиксная ссылка: 89

переход по символу 'ш' в состояние 35

состояние №35, суффиксная ссылка: 36

переход по символу 'у' в состояние 37

состояние №36, суффиксная ссылка: 0

переход по символу 'у' в состояние 37

переход по символу 'ь' в состояние 30

состояние №37, суффиксная ссылка: 15

переход по символу 'п' в состояние 38

состояние №38, суффиксная ссылка: 39

переход по символу 'р' в состояние 40

состояние №39, суффиксная ссылка: 0

переход по символу 'и' в состояние 89

переход по символу 'р' в состояние 40

состояние №40, суффиксная ссылка: 17

переход по символу 'о' в состояние 41

состояние №41, суффиксная ссылка: 22

переход по символу 'г' в состояние 42

состояние №42, суффиксная ссылка: 108

переход по символу 'р' в состояние 43

состояние №43, суффиксная ссылка: 17

переход по символу 'а' в состояние 44

состояние №44, суффиксная ссылка: 58

переход по символу 'м' в состояние 46

состояние №45, суффиксная ссылка: 0

переход по символу 'в' в состояние 112

переход по символу 'е' в состояние 28

переход по символу 'л' в состояние 75

переход по символу 'м' в состояние 46

переход по символу 'н' в состояние 92

переход по символу 'п' в состояние 87

переход по символу 'т' в состояние 118

переход по символу 'я' в состояние 59

переход по символу '.' в состояние 121

состояние №46, суффиксная ссылка: 48

переход по символу 'м' в состояние 47

состояние №47, суффиксная ссылка: 48

переход по символу 'у' в состояние 49

состояние №48, суффиксная ссылка: 0

переход по символу 'а' в состояние 117

переход по символу 'м' в состояние 47

переход по символу 'н' в состояние 84

переход по символу 'у' в состояние 49

состояние №49, суффиксная ссылка: 15

переход по символу ',' в состояние 50

состояние №50, суффиксная ссылка: 0

переход по символу 'к' в состояние 51

состояние №51, суффиксная ссылка: 103

переход по символу 'о' в состояние 52

состояние №52, суффиксная ссылка: 22

переход по символу 'т' в состояние 53

состояние №53, суффиксная ссылка: 13

переход по символу 'о' в состояние 54

состояние №54, суффиксная ссылка: 55

переход по символу 'р' в состояние 56

состояние №55, суффиксная ссылка: 22

переход по символу 'д' в состояние 23

переход по символу 'м' в состояние 116

переход по символу 'р' в состояние 56

состояние №56, суффиксная ссылка: 80

переход по символу 'а' в состояние 57

состояние №57, суффиксная ссылка: 58

переход по символу 'я' в состояние 59

состояние №58, суффиксная ссылка: 45

переход по символу 'м' в состояние 46

переход по символу 'я' в состояние 59

состояние №59, суффиксная ссылка: 60

переход по символу 'р' в состояние 61

состояние №60, суффиксная ссылка: 0

переход по символу 'з' в состояние 5

переход по символу 'р' в состояние 61

переход по символу 'с' в состояние 95

состояние №61, суффиксная ссылка: 17

переход по символу 'е' в состояние 62

состояние №62, суффиксная ссылка: 25

переход по символу 'а' в состояние 63

состояние №63, суффиксная ссылка: 45

переход по символу 'л' в состояние 64

состояние №64, суффиксная ссылка: 75

переход по символу 'и' в состояние 66

состояние №65, суффиксная ссылка: 0

переход по символу 'а' в состояние 27

переход по символу 'г' в состояние 76

переход по символу 'и' в состояние 66

состояние №66, суффиксная ссылка: 67

переход по символу 'з' в состояние 68

состояние №67, суффиксная ссылка: 0

переход по символу 'з' в состояние 68

переход по символу 'к' в состояние 102

переход по символу 'с' в состояние 90

переход по символу 'т' в состояние 82

переход по символу 'ш' в состояние 35

переход по символу 'я' в состояние 94

состояние №68, суффиксная ссылка: 69

переход по символу 'у' в состояние 70

состояние №69, суффиксная ссылка: 0

переход по символу 'о' в состояние 6

переход по символу 'у' в состояние 70

состояние №70, суффиксная ссылка: 15

переход по символу 'е' в состояние 71

состояние №71, суффиксная ссылка: 25

переход по символу 'т' в состояние 72

состояние №72, суффиксная ссылка: 13

переход по символу 'а' в состояние 73

состояние №73, суффиксная ссылка: 120

переход по символу 'л' в состояние 74

состояние №74, суффиксная ссылка: 75

переход по символу 'г' в состояние 76

состояние №75, суффиксная ссылка: 65

переход по символу 'г' в состояние 76

переход по символу 'и' в состояние 66

состояние №76, суффиксная ссылка: 77

переход по символу 'о' в состояние 78

состояние №77, суффиксная ссылка: 0

переход по символу 'о' в состояние 110

переход по символу 'р' в состояние 43

состояние №78, суффиксная ссылка: 110

переход по символу 'р' в состояние 79

состояние №79, суффиксная ссылка: 80

переход по символу 'и' в состояние 81

состояние №80, суффиксная ссылка: 17

переход по символу 'а' в состояние 57

переход по символу 'и' в состояние 81

состояние №81, суффиксная ссылка: 67

переход по символу 'т' в состояние 82

состояние №82, суффиксная ссылка: 13

переход по символу 'м' в состояние 83

состояние №83, суффиксная ссылка: 48

переход по символу 'н' в состояние 84

состояние №84, суффиксная ссылка: 85

переход по символу 'а' в состояние 86

состояние №85, суффиксная ссылка: 0

переход по символу 'а' в состояние 86

переход по символу 'и' в состояние 93

переход по символу 'о' в состояние 106

переход по символу 'я' в состояние 4

состояние №86, суффиксная ссылка: 45

переход по символу 'п' в состояние 87

состояние №87, суффиксная ссылка: 39

переход по символу 'и' в состояние 88

состояние №88, суффиксная ссылка: 89

переход по символу 'с' в состояние 90

состояние №89, суффиксная ссылка: 67

переход по символу 'с' в состояние 90

переход по символу 'ш' в состояние 35

состояние №90, суффиксная ссылка: 96

переход по символу 'а' в состояние 91

состояние №91, суффиксная ссылка: 45

переход по символу 'н' в состояние 92

состояние №92, суффиксная ссылка: 85

переход по символу 'и' в состояние 93

состояние №93, суффиксная ссылка: 67

переход по символу 'я' в состояние 94

состояние №94, суффиксная ссылка: 60

переход по символу 'с' в состояние 95

состояние №95, суффиксная ссылка: 96

переход по символу 'у' в состояние 97

состояние №96, суффиксная ссылка: 0

переход по символу 'а' в состояние 91

переход по символу 'н' в состояние 105

переход по символу 'у' в состояние 97

состояние №97, суффиксная ссылка: 15

переход по символу 'ф' в состояние 98

состояние №98, суффиксная ссылка: 100

переход по символу 'ф' в состояние 99

состояние №99, суффиксная ссылка: 100

переход по символу 'и' в состояние 101

состояние №100, суффиксная ссылка: 0

переход по символу 'и' в состояние 101

переход по символу 'ф' в состояние 99

состояние №101, суффиксная ссылка: 67

переход по символу 'к' в состояние 102

состояние №102, суффиксная ссылка: 103

переход по символу 'с' в состояние 104

состояние №103, суффиксная ссылка: 0

переход по символу 'о' в состояние 52

переход по символу 'с' в состояние 104

состояние №104, суффиксная ссылка: 96

переход по символу 'н' в состояние 105

состояние №105, суффиксная ссылка: 85

переход по символу 'о' в состояние 106

состояние №106, суффиксная ссылка: 22

переход по символу 'г' в состояние 107

состояние №107, суффиксная ссылка: 108

переход по символу 'о' в состояние 109

состояние №108, суффиксная ссылка: 77

переход по символу 'о' в состояние 109

переход по символу 'р' в состояние 43

состояние №109, суффиксная ссылка: 110

переход по символу 'а' в состояние 111

состояние №110, суффиксная ссылка: 22

переход по символу 'а' в состояние 111

переход по символу 'р' в состояние 79

состояние №111, суффиксная ссылка: 45

переход по символу 'в' в состояние 112

состояние №112, суффиксная ссылка: 113

переход по символу 'т' в состояние 114

состояние №113, суффиксная ссылка: 0

переход по символу 'т' в состояние 114

переход по символу 'у' в состояние 8

состояние №114, суффиксная ссылка: 13

переход по символу 'о' в состояние 115

состояние №115, суффиксная ссылка: 55

переход по символу 'м' в состояние 116

состояние №116, суффиксная ссылка: 48

переход по символу 'а' в состояние 117

состояние №117, суффиксная ссылка: 45

переход по символу 'т' в состояние 118

состояние №118, суффиксная ссылка: 13

переход по символу 'а' в состояние 119

состояние №119, суффиксная ссылка: 120

переход по символу '.' в состояние 121

состояние №120, суффиксная ссылка: 45

переход по символу 'л' в состояние 74

переход по символу '.' в состояние 121

состояние №121, суффиксная ссылка: 122

состояние №122, суффиксная ссылка: 0

переход по символу 'Ч' в состояние 19

* Содержимое исходного файла:

Привет

Поквет

Abracadabra

Содержимое итогового файла:

Искомая строка: вет

Суффиксный автомат будет выглядеть следующим образом:

состояние №0, суффиксная ссылка: -1

переход по символу 'a' в состояние 1

переход по символу 'b' в состояние 2

переход по символу 'c' в состояние 5

переход по символу 'd' в состояние 7

переход по символу 'r' в состояние 3

состояние №1, суффиксная ссылка: 0

переход по символу 'b' в состояние 2

переход по символу 'c' в состояние 5

переход по символу 'd' в состояние 7

состояние №2, суффиксная ссылка: 0

переход по символу 'r' в состояние 3

состояние №3, суффиксная ссылка: 0

переход по символу 'a' в состояние 4

состояние №4, суффиксная ссылка: 1

переход по символу 'c' в состояние 5

состояние №5, суффиксная ссылка: 0

переход по символу 'a' в состояние 6

состояние №6, суффиксная ссылка: 1

переход по символу 'd' в состояние 7

состояние №7, суффиксная ссылка: 0

переход по символу 'a' в состояние 8

состояние №8, суффиксная ссылка: 1

переход по символу 'b' в состояние 9

состояние №9, суффиксная ссылка: 2

переход по символу 'r' в состояние 10

состояние №10, суффиксная ссылка: 3

переход по символу 'a' в состояние 11

состояние №11, суффиксная ссылка: 4