**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра МО ЭВМ**

отчет

**по лабораторной работе №5**

**по дисциплине «Построение и анализ алгоритмов»**

Тема: «Алгоритм Ахо-Корасик»

|  |  |  |
| --- | --- | --- |
| Студент гр. 6381 |  | Шевелева А.М. |
| Преподаватель |  | Филатов А.Ю. |

Санкт-Петербург

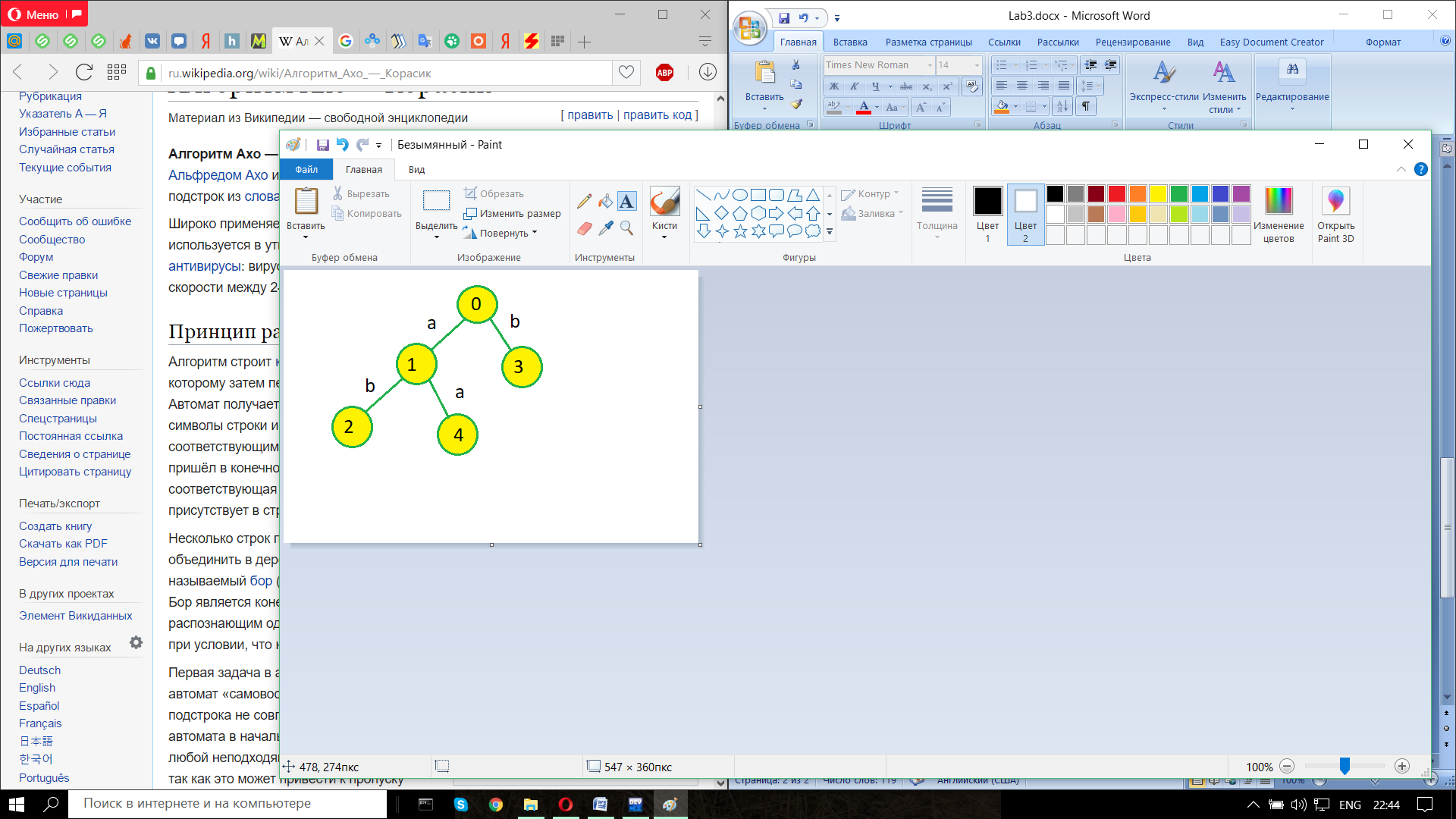
2018

**Цель работы**

Разработать программу, решающую задачу точного поиска набора образцов алгоритмом Ахо-Корасик и, используя реализацию точного множественного поиска, решить задачу точного поиска одного образца с джокером.

**Описание алгоритма Ахо-Корасик**

*Построение бора.*

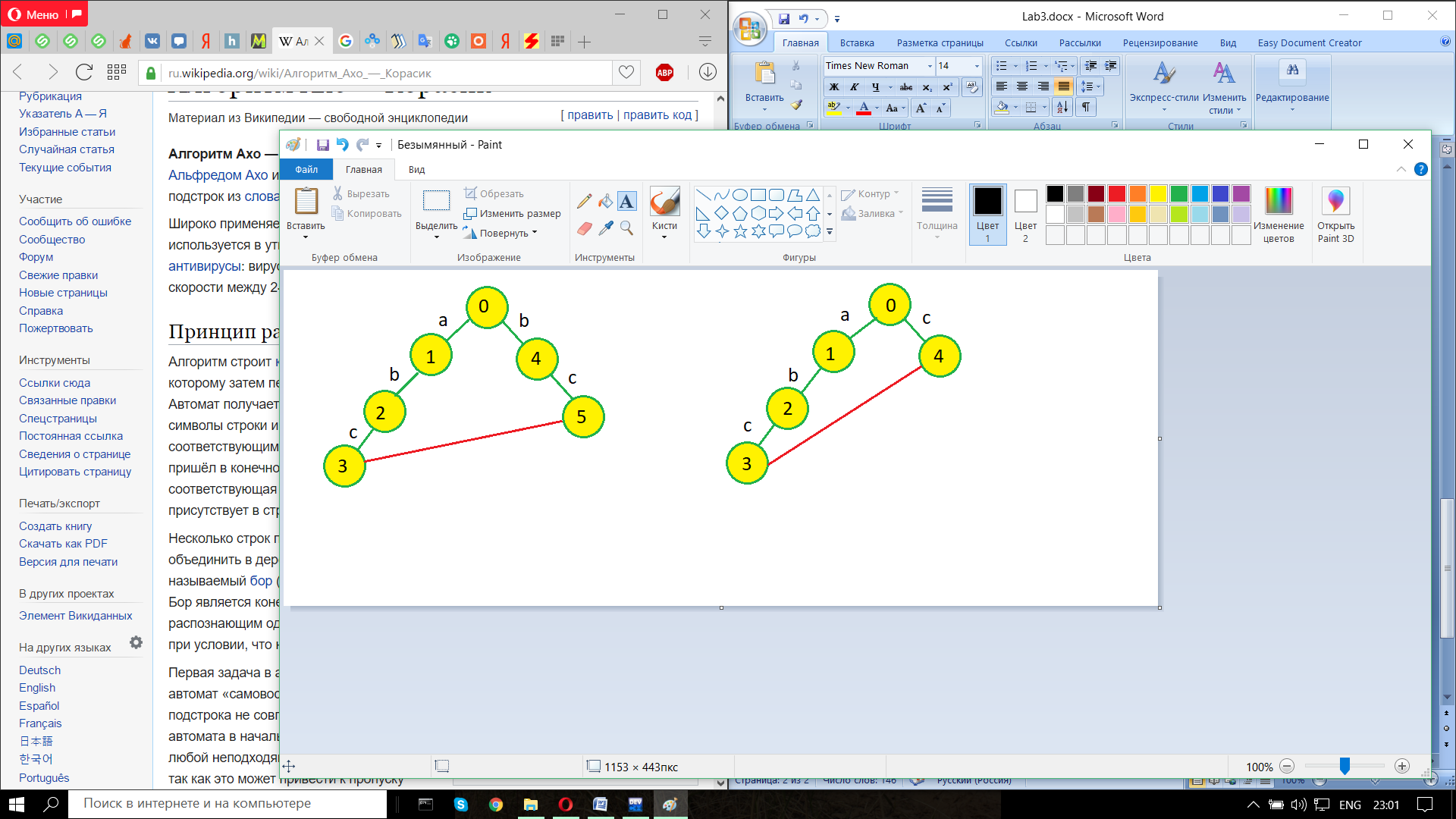
Корень дерева - это нулевая вершина. От нее исходят ветви дерева - образцы.

Образцы: ab, b, aa.

*Поиск суффиксных ссылок.*

Если текущая вершина является корнем или ее родитель корень, то суффиксная ссылка данной вершины указывает на корень.

Иначе мы ищем строку, начинающуюся с максимального суффикса данной ветви.



Первый пример:

1. Мы стоим в вершине 3. Ее родитель не корень, поэтому вызываем функцию поиска суффссылки из вершины 2 (родителя) после которой идет значение с.

2. Вершина 2 не корень, следовательно, ищем из вершины 1, после которой идет значение b.

3. Вершина 1 является сыном корня, поэтому мы возвращаем ноль и смотрим, есть ли значение b из корня. Оно есть, поэтому мы передаем 4.

4. Смотрим, есть ли значение с из вершины 4. Есть, мы передаем 5.

5. 5 - это суффссылка.

Второй пример:

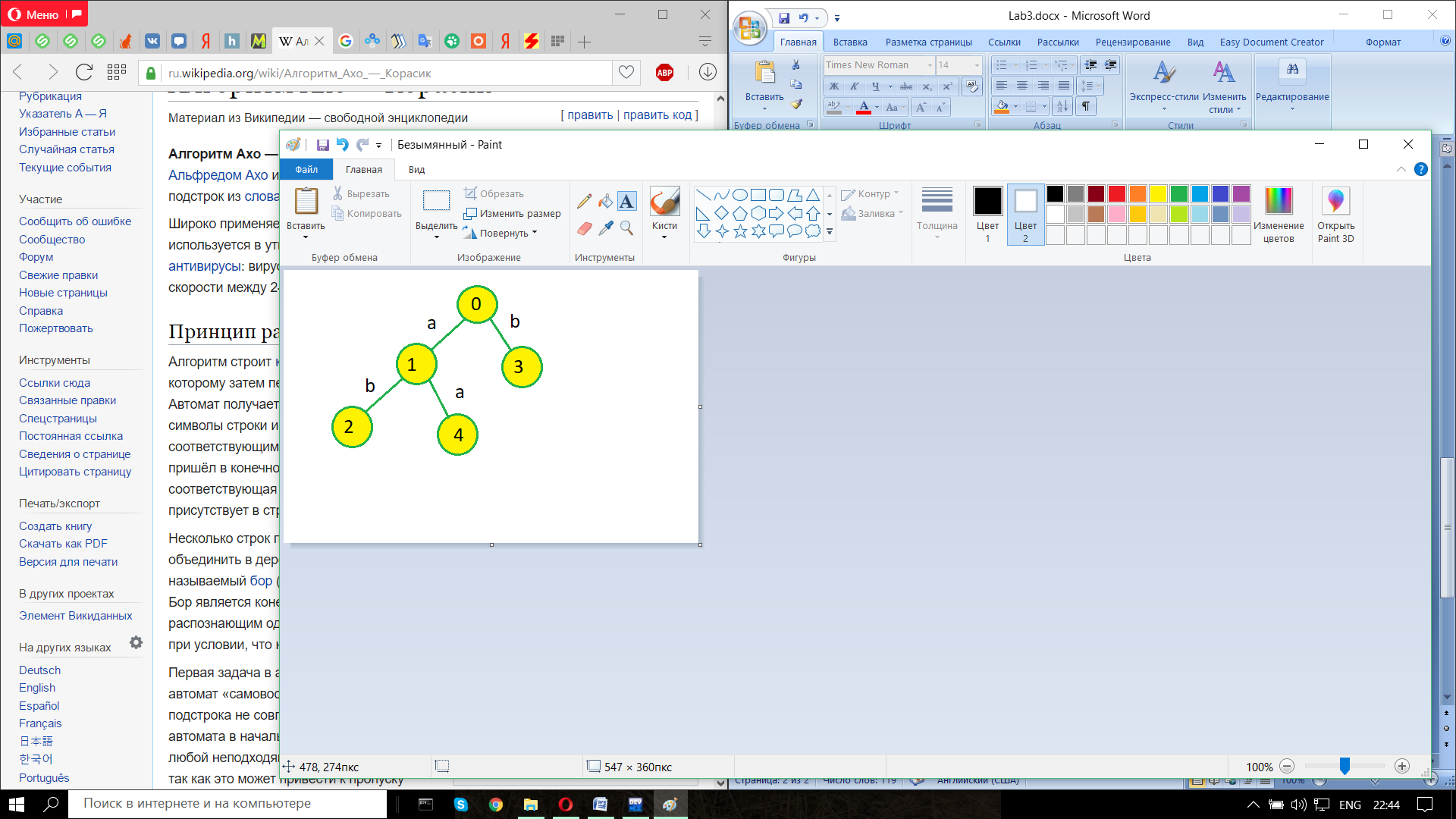
1-2 аналогичны предыдущему примеру.

3. Вершина 1 является сыном корня, поэтому мы возвращаем ноль и смотрим, есть ли значение b из корня. Его нет, поэтому мы передаем 0.

4. Смотрим, есть ли значение с из вершины 0. Есть, мы передаем 4.

5. 4 - это суффссылка.

*Алгоритм Ахо-Корасик*.

Идем по тексту abaabb.

1) i=0; текст[0]=a; суффссылка=0;

ищем суффссылку из вершины 0 в а.

суффссылка=1.

Из вершины 1 суффссылка в 0.

2) i=1; текст[1]=b; суффссылка=1;

ищем суффссылку из вершины 1 в b.

суффссылка=2. *Выводим значение.*

Из вершины 2 есть суффссылка в 3. *Выводим значение.*

3) i=2; текст[2]=a; суффссылка=2;

ищем суффссылку из вершины 2 в а.

суффссылка=1. Из вершины 1 суффссылка в 0.

4) i=3; текст[3]=а; суффссылка=1;

ищем суффссылку из вершины 1 в а.

суффссылка=4. *Выводим значение.* Из вершины 1 суффссылка в 0.

5) i=4; текст[4]=b; суффссылка=4;

ищем суффссылку из вершины 4 в b.

суффссылка=2. *Выводим значение.*

Из вершины 2 есть суффссылка в 3. *Выводим значение.*

6) i=5; текст[5]=b; суффссылка=2;

ищем суффссылку из вершины 2 в b.

суффссылка=3. *Выводим значение.* Из вершины 1 суффссылка в 0.

*Ответ*: каждая буква является началом одного из образцов.

*Поиск с джокером*.

Есть массив размером с длину строки, заполненный нулями.

Образцы - это максимальные подстроки, образца без джокеров.

Номера каждой подстроки-образца - это позиция в образце.

Применяем алгоритм и, если встречается подстрока, записываем в увеличиваем на единицу значения массива по адресу: положение подстроки в строке - позиция в образце + 1.

Потом пробегаемся по массиву, если количество подстрок совпадает с элементом массива, то это наш ответ.

**Вывод**

В процессе выполнения лабораторной работы был реализован алгоритм Ахо-Корасик - множественный поиск образцов в строке. Программа была протестирована на теоретических данных и дала верные ответы.