



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
"МИРЭА - Российский технологический университет"

РТУ МИРЭА

Институт информационных технологий (ИТ)
кафедра математического обеспечения и стандартизации информационных
технологий

ОТЧЕТ
ПО ПРАКТИЧЕСКОЙ РАБОТЕ №14
по дисциплине
«Структуры и алгоритмы
обработки данных»

Выполнил студент группы ИКБО-01-19

Рогов Д.В.

Принял старший преподаватель

Грушицын А.С.

«Зачтено»

«___»_____2020г.

Москва 2019

Оглавление

Задание 1.....	3
Задание 2.....	3
Задание 3.....	3
Задание 4.....	4
Задание 5.....	4
Задание 6.....	5
Выводы:	6

Задание 1. Построить таблицу частот встречаемости символов в исходной строке символов для чего сформировать алфавит исходной строки и посчитать количество вхождений (частот) символов и их вероятности появления, для строки **рогов дмитрий вадимович** такая таблица будет иметь вид:

Таблица частот

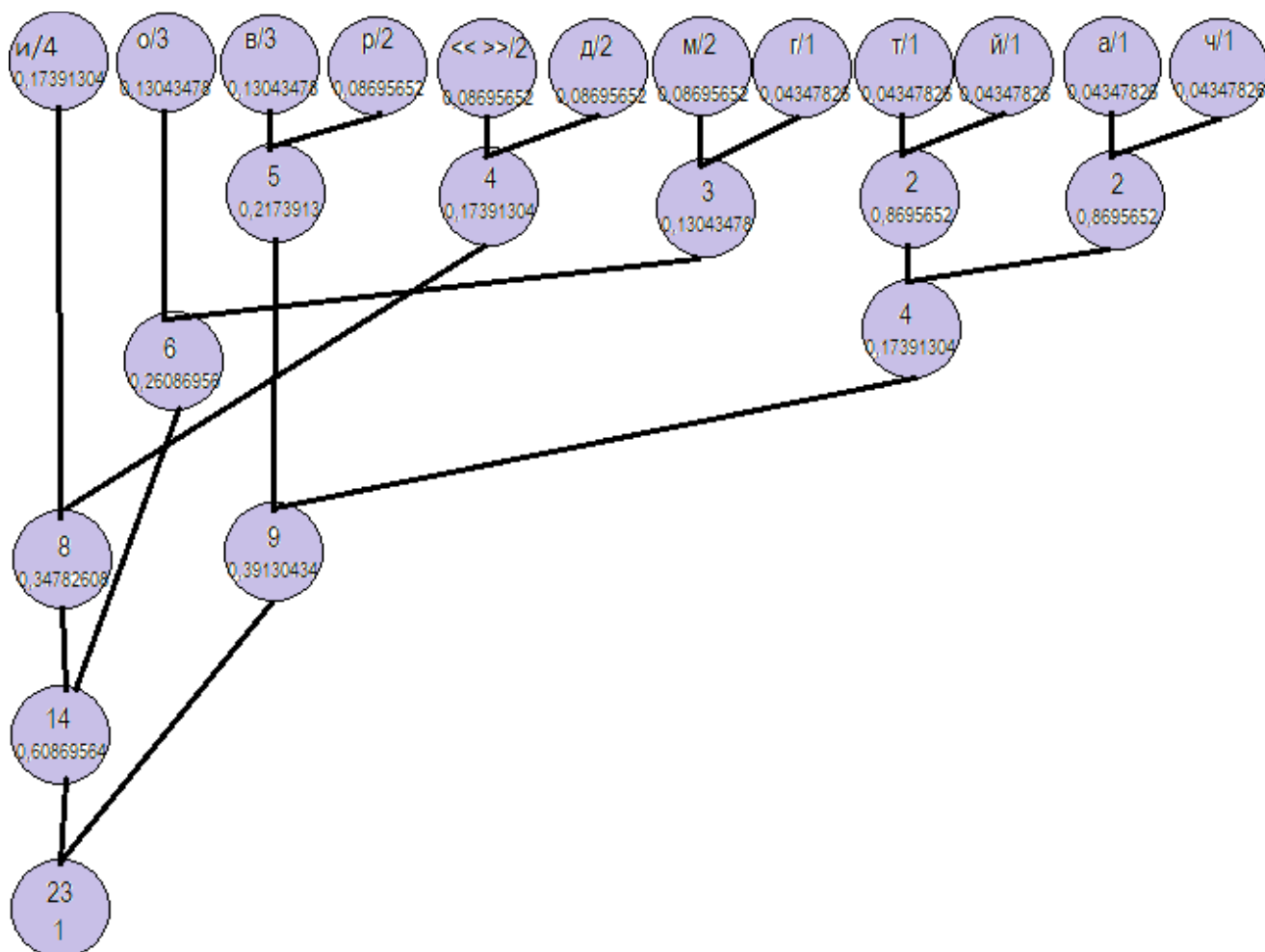
Алфавит	р	о	г	в	<<>>	д	м	и	т	й	а	ч
Кол. вх.	2	3	1	3	2	2	2	4	1	1	1	1
Вероятн.	0,08695652	0,13043478	0,04347826	0,13043478	0,08695652	0,08695652	0,08695652	0,17391304	0,04347826	0,04347826	0,04347826	0,04347826

Задание 2. Отсортировать алфавит в порядке убывания частот появления символов:

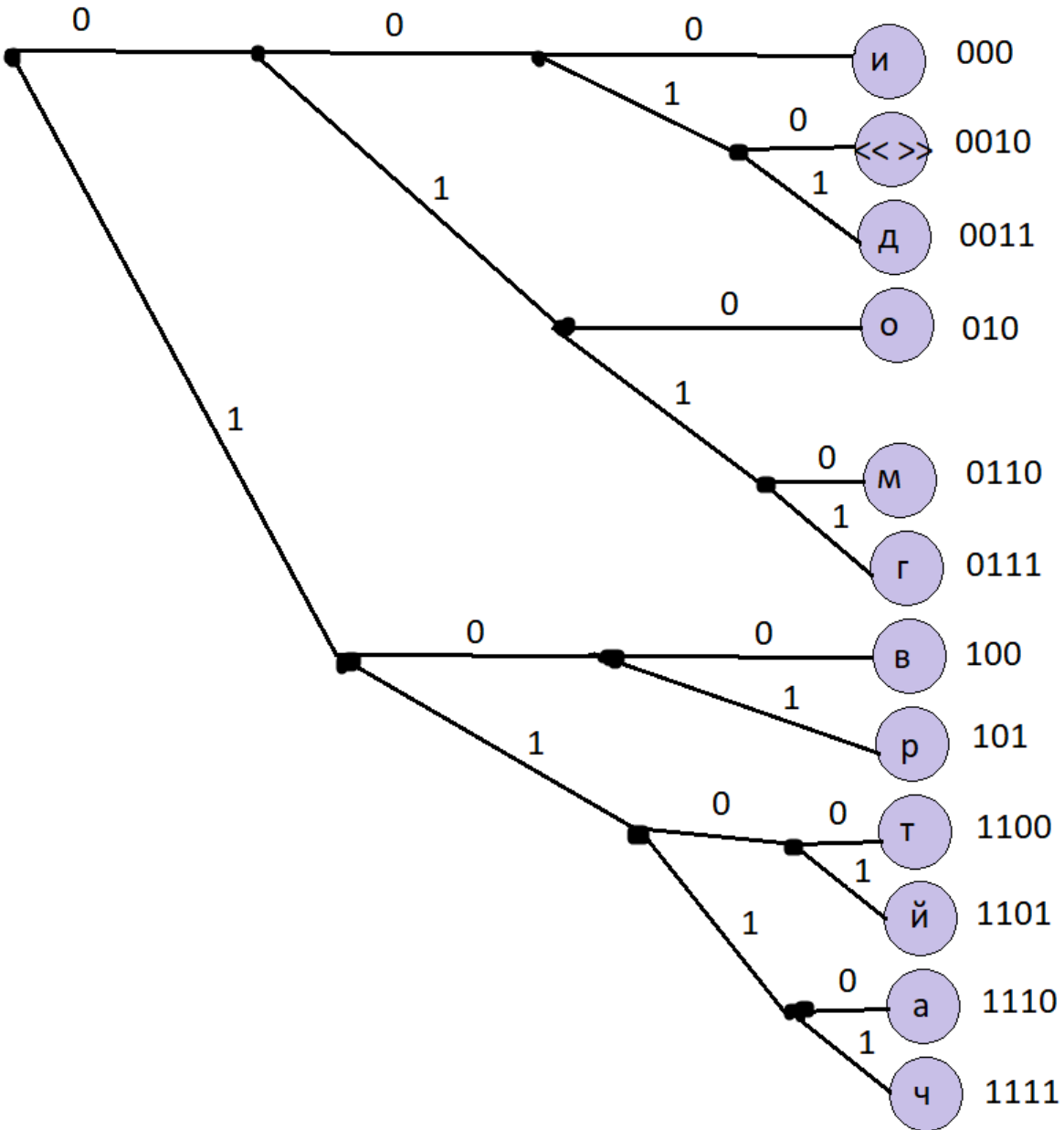
Таблица отсортированных частот

Алфавит	и	о	в	р	<<>>	д	м	г	т	й	а	ч
Кол. вх.	4	3	3	2	2	2	2	1	1	1	1	1
Вероятн.	0,17391304	0,13043478	0,13043478	0,08695652	0,08695652	0,08695652	0,08695652	0,04347826	0,04347826	0,04347826	0,04347826	0,04347826

Задание 3. Построить дерево кодирования Хаффмана:



Присвоить ветвям коды. Определить коды символов:



Задание 5. Провести кодирование исходной строки:

р	о	г	о	в	« »	д	м	и	т	р	и	й	« »
101	010	0111	010	100	0010	0011	0110	000	1100	101	000	1101	0010
в	а	д	и	м	о	в	и	ч					
100	1110	0011	000	0110	010	100	000	1111					

Рассчитать коэффициенты сжатия относительно кодировки ASCII и относительно равномерного кода.

Коэффициент сжатия относительно использования кодировки ASCII (8 бит/символ).

$$S_{ASCII} = 8 \cdot 23 = 184 \text{ бит}$$

$$S_{HUFF} = 4 \cdot 11 + 3 \cdot 12 = 80 \text{ бит}$$

Значит коэффициент сжатия относительно кодировки ASCII будет равен:

$$K_{ASCII} = \frac{S_{ASCII}}{S_{HUFF}} \approx 2.3$$

Коэффициент сжатия относительно равномерного кода (5 бит/символ, т. к. у нас всего 23 символа) будет равен:

$$K_{\text{равн.код}} = \frac{23 \cdot 5}{S_{HUFF}} = \frac{115}{80} \approx 1,4375$$

Задание 6. Рассчитать среднюю длину полученного кода и его дисперсию.

Рассчитаем среднюю длину полученного кода по формуле

$$l_{\text{ср.}} = \sum_S p_s \cdot l_s,$$

где s — множество символов алфавита; p_s — вероятность появления символа; l_s — количество бит в коде символа.

Для полученного кода средняя длина будет равна

$$l_{\text{ср.}} = 3 \cdot 0.17391304 + 6 \cdot 0.13043478 + 24 \cdot 0.08695652 + 20 \cdot 0.04347826 = 3,4782608 \text{ бит/символ}$$

Дисперсия рассчитывается по формуле:

$$\delta = \sum_S p_s (l_s - l_{\text{ср.}})^2$$

Для полученного кода дисперсия будет равна:

$$\delta = 0,17391304 \cdot (3 - 3,4782608)^2 + 2 \cdot 0,13043478 \cdot (3 - 3,4782608)^2 + 0,08695652 \cdot (3 - 3,4782608)^2 + 3 \cdot 0,08695652 \cdot (4 - 3,4782608)^2 + 5 \cdot 0,04347826 \cdot (4 - 3,4782608)^2 = 0,249527$$

Для уменьшения дисперсии кода существует правило: когда на дереве имеется более двух узлов с наименьшей вероятностью, следует объединять символы с наибольшей и наименьшей вероятностью; это сокращает общую дисперсию кода.

Выводы:

В ходе данной практической работы я изучил алгоритм оптимального префиксного кодирования Хаффмана и его использование для сжатия сообщений. Убедился в корректности данного алгоритма и его необходимости для сжатия данных.

Данный алгоритм позволяет значительно уменьшить объем необходимой информации путем измерения вероятности использования символа. Что было продемонстрировано в практической работе.