Лабораторная работа № 1

ШИФРЫ И КРИПТОАНАЛИЗ

ЦЕЛЬ РАБОТЫ : Изучение подстановочного шифра и метода частотного криптоанализа.

Введение

Наиболее простой тип криптограмм – это так называемые подстановочные криптограммы. Составляя их, каждой букве алфавита сопоставляют определенный символ (чаще тоже букву) и при кодировании всякую букву текста заменяют на соответствующий ей символ.

Расшифровка (криптоанализ) подобных криптограмм не составляет большой проблемы. Все основывается на том, что различные буквы естественного языка – русского, английского или какого-либо другого встречаются в осмысленных текстах неодинаково часто. Следовательно, тоже самое верно и для сопоставляемых им знаков. В еще большей мере это относится к буквосочетаниям из двух или нескольких букв. Лишь некоторые из них часто употребляются, многие же вообще не употребляются.

Анализируя частоту появления тех или иных знаков и их сочетаний можно с большой уверенностью восстановить буквы зашифрованного текста. Этот метод называется частотным анализом. Он основывается на подсчете частоты появления зашифрованных знаков. В таблице 1 указаны относительные частоты букв русского языка. Буквы *Е* и *Ё* , а также *Ь* и *Ъ* кодируются обычно одинаково, поэтому в таблице они не различаются. Как следует из таблицы наиболее часто встречающаяся буква русского алфавита – это О. Ее относительная частота, равная 0,090, означает, что на 1000 букв русского текста приходится в среднем 90 букв О. В таком же смысле понимаются относительные частоты и остальных букв. В таблицу 1 не включен символ пробел. Его относительная частота наибольшая и равна 0,175.

Таблица 1

№	буква	Отн. частота	№	буква	Отн. частота	№	буква	Отн. частота
0	А	0,062	10	К	0,028	20	Ф	0,002
1	Б	0,014	11	Л	0,035	21	Х	0,009
2	В	0,038	12	М	0,026	22	Ц	0,004
3	Г	0,013	13	Н	0,053	23	Ч	0,012
4	Д	0,025	14	О	0,090	24	Ш	0,006
5	Е	0,072	15	П	0,023	25	Щ	0,003
6	Ж	0,007	16	Р	0,040	26	Ы	0,016
7	З	0,016	17	С	0,045	27	Ь ,Ъ	0,014
8	И	0,062	18	Т	0,053	28	Э	0,003
9	Й	0,010	19	У	0,021	29	Ю	0,006
						30	Я	0,018

Рассмотрим криптограмму :

ЦЯРСНСМЩИ ЯМЯКЗЖ ОНКДЖДМ МД СНКЫЙН ГКЮ ОНГРСЯМНБНЦМЩФ ЙПЗОСНВПЯЛЛ МН Б ГПТВЗФ РКТЦЯЮФ НМ РКНЕМДД

Для расшифровки подсчитаем сколько раз в криптограмме встречается каждая буква. Результаты подсчета приведены в таблице 2.

Таблица 2

Н	М	Я	К	Д	С	Р	Г	О	П	З	Ф	Ц	Б	В	Ж	Й	Л	Т	Щ	Ю	Е	И	Ы
11	9	6	6	5	5	4	3	3	3	3	3	3	2	2	2	2	2	2	2	2	1	1	1

Наиболее часто встречающийся символ Н скорее всего означает букву О. Сделав такое предположение , рассмотрим следующий по частоте символ М. В криптограмме имеется двубуквенное сочетание МН. Так как Н – это О, то символ М соответствует согласной. Среди согласных в русском языке выделяются по частоте буквы Т и Н. Разберем случай, когда М означает Н.

Если М – это Н, то в сочетании МД, встречающемся в криптограмме , Д скорее всего означает гласную. Из наиболее вероятных для Д вариантов А, Е, И выбираем Е, потому что лишь в этом случае имеющееся в криптограмме слово РКНЕМДД допускает осмысленную расшифровку. Теперь обратимся к сочетанию ЯМЯКЗЖ. В нем Я может означать лишь гласную А или И. Любые другие возможности

заведомо не допускают разумного прочтения слова ЯМЯКЗЖ. Испытаем букву А. Подставляя вместо Я букву А, вместо М – Н, вместо других знаков точки, получим недописанное слово АНА… . В словаре имеется всего лишь несколько слов из 6 букв с таким началом: АНАЛИЗ, АНАЛОГ, АНАНАС, АНАТОМ. Из них годится лишь первое. Если вместо Я подставить букву И, то получится шестибуквенное сочетание с началом ИНИ, но в словаре нет ни одного такого слова. Расшифрованы еще четыре буквы: Я, К, З, Ж. Они означают соответственно А, Л, И, З.

В слове ОНКЖДМ известны все символы кроме первого. Заменяя их буквами, получаем: . ОЛЕЗЕН. Ясно, что неизвестная буква – это П. Значит О расшифровывается как П.

Рассмотрим сочетание РКНЕМДД, означающее .ЛО.НЕЕ. Имеется несколько вариантов его прочтения, один из них – СЛОЖНЕЕ. Следовательно, скорее всего Р – это С, Е - это Ж.

Из нерасшифрованных знаков чаще всего встречается С. В соответствии с таблицей 1 среди оставшихся согласных наибольшую частоту имеет Т. Естественно предположить, что С означает Т.

Попытаемся восстановить зашифрованный текст, подставляя вместо разгаданных знаков соответствующие им буквы:

.АСТОТН.. АНАЛИЗ ПОЛЕЗЕН НЕ ТОЛ..О .Л. ПО.СТАНО.О.Н.. ..ИПТО..А.. НО . ….И. СЛ..А.. ОН СЛОЖНЕЕ

Ясны по контексту, по крайней мере три слова: .АСТОТН.. означает ЧАСТОТНЫЙ, ТОЛ..О – ТОЛЬКО, .Л. – ДЛЯ. С учетом новой информации текст примет следующую форму:

ЧАСТОТНЫЙ АНАЛИЗ ПОЛЕЗЕН НЕ ТОЛЬКО ДЛЯ ПОДСТАНО.ОЧНЫ. К.ИПТО..А.. НО . Д…И. СЛ.ЧАЯ. ОН СЛОЖНЕЕ

Окончательная расшифровка не представляет труда. Текст таков:

ЧАСТОТНЫЙ АНАЛИЗ ПОЛЕЗЕН НЕ ТОЛЬКО ДЛЯ ПОДСТАНОВОЧНЫХ КРИПТОГРАММ, НО В ДРУГИХ СЛУЧАЯХ ОН СЛОЖНЕЕ.

ЗАДАНИЯ

1. Зашифровать любой текст с помощью подстановочного шифра Цезаря (Он состоит в том, что весь алфавит циклически сдвигается на определенное число букв.) Предложить метод расшифровки более простой, чем частотный анализ.
2. Расшифровать заданный преподавателем текст зашифрованный шифром Цезаря.
3. Зашифровать любой текст с помощью подстановочного шифра. Расшифровать текст методом частотного анализа. Для сбора статистики использовать файл test.txt
4. Расшифровать заданный преподавателем шифртекст методом частотного анализа.

ИСХОДНЫЕ ДАННЫЕ: Файлы с шифртекстом (шифр Цезаря и подстановочный шифр).

СОДЕРЖАНИЕ ОТЧЕТА: Отчет по лабораторной работе должен содержать

1. Описание алгоритма шифрования
2. Описание алгоритма криптоанализа
3. Программы шифрования и дешифрования
4. Расшифрованные тексты
5. Выводы по работе