

Лабораторная работа №2

Частотный анализ. Основы криптоанализа

Требования к сдаче работы: в ходе лабораторной работы необходимо сохранить расшифрованные тексты в отдельные файлы, заполнить предлагаемую таблицу и ответов письменно на вопросы.

Основным средством обеспечения конфиденциальности являются различные криптографические алгоритмы. История развития криптографии насчитывает около четырёх тысяч лет. Одними из средств криптографии являются шифры.

Первые шифры были **моноалфавитными**. Например, шифр Атбаш, шифр квадрат Полибия, шифр Цезаря и т.д.

Моноалфавитные шифры подвержены частотному анализу, так как при шифровании выполняется взаимно-однозначное преобразование символов одного алфавита в символы другого (или этого же) алфавита. То есть, зная частотное распределение букв алфавита, можно дешифровать крипто-текст.

Задание 1. Используя самостоятельно разработанную программу для частотного анализа, расшифруйте предлагаемые тексты: tis_text1.txt, tis_text2.txt, tis_text3.txt.

Для выполнения частотного анализа необходимо наличие распределения частот появления букв, которое зависит от языка, стиля, автора и прочих свойств текста. Поэтому распределения частот букв для разных текстов могут отличаться. Чем больше длина текстов, тем более сходны распределения частот букв этих текстов. В общем случае криптоанализ на основе частот появления букв представляет подбор из нескольких вариантов соответствия букв закрытого и открытого текста. Для сокращения времени расшифровки предлагается воспользоваться распределениями частот букв, которые содержатся в файлах freq_text1.txt, freq_text2.txt и freq_text3.txt, соответственно для каждого текста.

(!) Расшифрованные тексты сохраните для отчета.

Заполните следующую таблицу:

Документ	Файл частот букв	Источник (откуда текст)	Предполагаемый шифр
tis_text1.txt	freq_text1.txt		
tis_text2.txt	freq_text2.txt		
tis_text3.txt	freq_text3.txt		

Проверьте совпадение распределений частот букв расшифрованных текстов с данными для выполнения задания, используя самостоятельно разработанную программу подсчета частот появления букв.

Пример демонстрации программы. Расшифруем текст из первого файла (tis_text1.txt), используя распределение частот из соответствующего файла freq_text1.txt. Заметим, что в криптотексте сохранены знаки пунктуации, что также может помочь при дешифровании.

криптотекст

расшифровываемый текст

Наиболее часто встречающейся буквой по частотному распределению является буква О, а в криптотексте чаще всех встречается буква З. Поэтому можно сделать замену З->О.

И так далее...

С некоторой итерации можно будет угадывать слова, что подтверждает правильность замен.

freq_text1.txt

1	О	0,113
2	Е	0,085
3	А	0,076
4	Н	0,075
5	И	0,059
6	Т	0,059
7	Л	0,054
8	С	0,052
9	Р	0,047
10	М	0,037
11	В	0,036
12	К	0,032
13	Д	0,031
14	П	0,024
15	Б	0,023
16	У	0,022
17	Б	0,021
18	Я	0,021
19	Ь	0,020
20	Э	0,017
21	Г	0,016
22	Ч	0,013
23	Э	0,012
24	Ю	0,012
25	Й	0,011
26	Ж	0,010
27	Х	0,008
28	Ш	0,006
29	Ц	0,004
30	Щ	0,003
31	Ф	0,002
32	Ъ	0,000

Frequency Analysis

russian bl B Substitute Run

ЪДЩЪЗЭЩЙЧ ЫЩК, - КГЩАЩД ЦЖЭЙЧ ЕЩЙЛБЖ
Б КЮД ЖЩ ИЙЮЭДЗЯЮЖЖФВ КЛМД. ЗЖ ЖЮ
ЫФЬДШЭЮД ГЩГ РЮДЗЫЮГ, ЭЗЫЮЭЮЖЖФВ
ЭЗ ИЗКДЮЭЖЮВ РЮЙЛФ, ЖЗ ЭЮДЗ
ЗЪКЛЗЩДЗ БЕЮЖЖЗ ЛЩГ. ЖЩ ЮЪЗ ДБПЮ ЖЮ

ЪЛАЬОЭАРЧ ВАС, - СГАААЛ ЦНЭРЧ МАТИН И
СЕЛ НА ИРЕЭЛОЯЕННФВ СТМЛ. ОН НЕ
ВФЬЛШЭЕЛ ГАГ РЕЛОВЕГ, ЭОВЕЭЕННФВ ЭО
ИОСЛЕЭНЕВ РЕРТФ, НО ЭЕЛО ОЪСТОШЛО
ИМЕННО ТАГ. НА ЕЬО ЛИПЕ НЕ РИТАЛОСХ
НИГАГИО ОСОЪЕННФО ЦМОПИВ, ЛИСХ

З->593
Ю->447
Ш->397
Ж->392
Б->309
Л->308
Д->285
К->273
Й->245
Е->194
Ы->189
Г->167
Э->162
И->127
Ф->121
М->114
Ь->112
Ш->108
Х->106
А->91
Б->86
Р->66
Ц->63
Ч->62
В->58
Я->55
О->41
С->31
П->23
Т->14
Н->9
У->0

З->О
Ю->Е
Ш->А
Ж->Н
Б->И
Л->Т
Д->П
К->С
Й->Р
Е->М
Ы->В
Г->К
Э->Д

В окне расшифрованного текста большие буквы относятся к открытому тексту, а маленькие к криптотексту. После ряда замен можно делать дополнительные предположения о замене букв. Например, третье слово предположительно «сказал», последнее слово в первом предложении – «стол» или «стул». Но слово «стол» не подходит, так как буква О уже была расшифрована.

Продолжая делать замены, получим

freq_text1.txt

1	О	0,113
2	Е	0,085
3	А	0,076
4	Н	0,075
5	И	0,059
6	Т	0,059
7	Л	0,054
8	С	0,052
9	Р	0,047
10	М	0,037
11	В	0,036
12	К	0,032
13	Д	0,031
14	П	0,024
15	Б	0,023
16	У	0,022
17	Б	0,021
18	Я	0,021
19	Ь	0,020
20	Э	0,017
21	Г	0,016
22	Ч	0,013
23	Э	0,012
24	Ю	0,012
25	Й	0,011
26	Ж	0,010
27	Х	0,008
28	Ш	0,006
29	Ц	0,004
30	Щ	0,003
31	Ф	0,002
32	Ъ	0,000

Frequency Analysis

russian З Д Substitute Run

ЪДЩЪЗЭЩЙЧ ЫЩК, - КГЩАЩД ЦЖЭЙЧ ЕЩЙЛБЖ
Б КЮД ЖЩ ИЙЮЭДЗЯЮЖЖФВ КЛМД. ЗЖ ЖЮ
ЫФЬДШЭЮД ГЩГ РЮДЗЫЮГ, ЭЗЫЮЭЮЖЖФВ
ЭЗ ИЗКДЮЭЖЮВ РЮЙЛФ, ЖЗ ЭЮДЗ
ЗЪКЛЗЩДЗ БЕЮЖЖЗ ЛЩГ. ЖЩ ЮЪЗ ДБПЮ ЖЮ

ЪЛАЬОДАРЧ ВАС, - СКАААЛ ЦНДРЧ МАТИН И
СЕЛ НА ИРЕДЛОЯЕННФВ СТМЛ. ОН НЕ
ВФЬЛШДЕЛ КАК РЕЛОВЕК, ДОВЕДЕННФВ ДО
ИОСЛЕДНЕВ РЕРТФ, НО ДЕЛО ОЪСТОШЛО
ИМЕННО ТАК. НА ЕЬО ЛИПЕ НЕ РИТАЛОСХ
НИКАКИО ОСОЪЕННФО ЦМОПИВ, ЛИСХ

З->593
Ю->447
Ш->397
Ж->392
Б->309
Л->308
Д->285
К->273
Й->245
Е->194
Ы->189
Г->167
Э->162
И->127
Ф->121
М->114
Ь->112
Ш->108
Х->106
А->91
Б->86
Р->66
Ц->63
Ч->62
В->58
Я->55
О->41
С->31
П->23
Т->14
Н->9
У->0

З->О
Ю->Е
Ш->А
Ж->Н
Б->И
Л->Т
Д->П
К->С
Й->Р
Е->М
Ы->В
Г->К
Э->Д

Предполагая первое слово «благодарю», можно сделать замены Ъ->Б, Ъ->Г и Ч->Ю. Скорее всего третье слово – «сказал», поэтому добавим замену А->З. После слов «мартин и сел на» скорее всего следует «предложенный стул». Поэтому можно добавить замену И->П, Я->Ж, Ф->Ы, В->Й и М->У. Получим

The 'Frequency Analysis' window displays the following mappings:

З -> 593	э -> О
Ю -> 447	ю -> Е
Щ -> 397	щ -> А
Ж -> 392	ж -> Н
Б -> 309	б -> И
Л -> 308	л -> Т
Д -> 285	д -> Л
К -> 273	к -> С
Й -> 245	й -> Р
Е -> 194	е -> М
Ы -> 189	ы -> В
Г -> 167	г -> К
Э -> 162	э -> Д
И -> 127	и -> Б
Ф -> 121	ф -> Г
М -> 114	ч -> Ю
Ъ -> 112	а -> З
Ш -> 108	и -> П
Х -> 106	я -> Ж
А -> 91	ф -> Ы
Ь -> 86	в -> Й
Р -> 66	м -> У
Ц -> 63	
Ч -> 62	
В -> 58	
Я -> 55	
О -> 41	
С -> 31	
П -> 23	
Т -> 14	
Н -> 9	
У -> 0	

Таким образом, можно расшифровать текст полностью.

Задание 2. Ответьте письменно на следующие вопросы:

1. Приведите пример шифра, который не подвержен частотному анализу. Объясните, благодаря какому свойству достигается стойкость шифра к частотному анализу.
2. Какие части слова «угадываются» лучше при расшифровке? Как вы считаете с чем это связано?