БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Кафедра системного анализа и компьютерного моделирования

КРИПТОАНАЛИЗ МНОГОАЛФАВИТНЫХ ШИФРОВ ЗАМЕНЫ

Методические материалы к лабораторной работе «Криптоанализ многоалфавитных шифров» по спецкурсу криптографический инжиниринг

для студентов специальности 1-98 01 01 Компьютерная безопасность

2023 г.

СОСТАВИТЕЛИ:

П.П. Коржуков, старший преподаватель кафедры системного анализа и компьютерного моделирования Белорусского государственного университета.

Текст представлен в авторской редакции.

РЕКОМЕНДОВАНА К РАЗМЕЩЕНИЮ В ЭЛЕКТРОННОЙ БИБЛИОТЕКЕ БГУ:

Кафедрой системного анализа и компьютерного моделирования Белорусского государственного университета
(протокол № 8 от 14 февраля 2023 г.);

Данная дисциплина предназначена для студентов специальности 1-98 01 01 Компьютерная безопасность (по направлениям) (направление 1-98 01 01-02 Компьютерная безопасность (Радиофизические методы и программно-технические средства)).

Криптографический инжиниринг ссылается на теорию и практику инжиниринга криптографических систем, т.е. устройства шифрования и расшифрования, аппаратно-программные системы цифровой сигнатуры и аутентификации, системы генерации, распределения и управления ключами. Криптографический инженер проектирует, реализовывает, тестирует и проверяет криптографические системы. В область интересов криптографического инжиниринга входит криптоанализ криптографических систем с целью проверки их устойчивости и надежности по отношению к атакам, а также построение контрмер против них, чтобы сорвать такие нападения или уменьшить вероятность их успешности.

В лабораторной работе исследуется метод частотного анализа для дешифрации многоалфавитных (полиалфавитных) шифров замены.

Криптографический инжиниринг
Павел Павлович
Коржуков
korzhukov@bsu.by 398-02-22
зима 2023

ЧАСТОТНЫЙ АНАЛИЗ ШИФРОТЕКСТА

Важный тезис: шифр замены (подстановки) не меняет частоты символов в тексте! Алфавит открытого текста и шифралфавит могут различаться, но частоты символов открытого текста точно такие же, как и частоты символов шифротекста. Остаётся лишь узнать соответствие одних с другими.

Примеры текстов

Текст № 1. В качестве примера текста на русском языке взят фрагмент из книги «Шлеер С., Меллор С. Объектно-ориентированный анализ: моделирование мира в состояниях», приведенный на рисунке 1. Этот исходный текст состоит из 1218 символов, записанных в одну строку, без переносов. В тексте применяются только прописные буквы. Поскольку строка очень длинная, на рисунке 1 она для наглядности разбита знаками переноса, которых в исходном тексте не существует.

Всего слов в этом отрывке 158, включая предлоги. Алфавит открытого текста содержит 40 символов вместе со знаками препинания. Собственно букв в тексте насчитывается 1031, остальные 187 символов – пробелы и знаки препинания. Тематика отрывка связана с системой управления железной дороги.

В ФОРМИРОВАНИИ ТИПИЧНОЙ БОЛЬШОЙ СИСТЕМЫ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ АНАЛИТИК, КАК ПРАВИЛО, ДОЛЖЕН РАССМАТРИВАТЬ РЯД ЧЕТКО ОПРЕДЕЛЕННЫХ ПРЕДМЕТНЫХ ОБЛАСТЕЙ, ИЛИ ДОМЕНОВ. КАЖДЫЙ ДОМЕН МОЖЕТ РАССМАТРИВАТЬСЯ КАК ОТДЕЛЬНЫЙ МИР, НАСЕЛЁННЫЙ СОБСТВЕННЫМИ КОНЦЕПТУАЛЬНЫМИ СУЩНОСТЯМИ, ИЛИ ОБЪЕКТАМИ. СЛЕДОВАТЕЛЬНО, В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ УПРАВЛЕНИЯ ЖЕЛЕЗНОЙ ДОРОГОЙ ДОМЕН РАБОТА ЖЕЛЕЗНОЙ ДОРОГИ ИМЕЕТ ОТНОШЕНИЕ К ПОЕЗДАМ, ДОРОГАМ И Т. П., В ТО ВРЕМЯ КАК ДОМЕН ПОЛЬЗОВАТЕЛЬСКИЙ ИНТЕРФЕЙС НАПОЛНЕН ОКНАМИ, ДИСПЛЕЯМИ И ПИКТОГРАММАМИ. КАЖДЫЙ ДОМЕН МОЖЕТ СУЩЕСТВОВАТЬ БОЛЕЕ ИЛИ МЕНЕЕ НЕЗАВИСИМО ОТ ДРУГИХ: ЖЕЛЕЗНАЯ ДОРОГА МОЖЕТ СУЩЕСТВОВАТЬ БЕЗ ЭКРАНОВ И ОКОН; ОКНА И ПИКТОГРАММЫ МОГУТ СУЩЕСТВОВАТЬ БЕЗ ПОЕЗДОВ. ДОМЕНЫ ИЗОБРАЖАЮТСЯ НА СХЕМЕ ДОМЕНОВ. НЕКОТОРЫЕ ДОМЕНЫ ДОВОЛЬНО МАЛЫ, ЧТОБЫ МОЖНО БЫЛО АНАЛИЗИРОВАТЬ ИХ КАК ЕДИНОЕ ЦЕЛОЕ, В ТО ВРЕМЯ КАК ДРУГИЕ СОДЕРЖАТ ТАК МНОГО ОБЪЕКТОВ, ЧТО ТРУДНО ПОДДАЮТСЯ КОНТРОЛЮ И ОБРАБОТКЕ. ПОЭТОМУ БОЛЬШИЕ ДОМЕНЫ РАЗБИВАЮТСЯ НА ПОДСИСТЕМЫ. ПОСЛЕ ТОГО КАК СИСТЕМА РАЗБИТА НА ДОМЕНЫ И ПОДСИСТЕМЫ, МОЖНО ПРОВОДИТЬ АНАЛИЗ. НЕЗАВИСИМО ОДНА ОТ ДРУГОЙ КАЖДАЯ ПОДСИСТЕМА (ИЛИ МАЛЕНЬКИЙ ДОМЕН) АНАЛИЗИРУЕТСЯ В ТРИ ЭТАПА: ИНФОРМАЦИОННОЕ МОДЕЛИРОВАНИЕ, МОДЕЛИРОВАНИЕ СОСТОЯНИЙ И МОДЕЛИРОВАНИЕ ПРОЦЕССОВ.

Рисунок 1 – Открытый текст № 1

На рисунке 2 приведена обзорная гистограмма частот встречаемости букв русского алфавита в приведенном тексте № 1 во всём диапазоне кодовой таблицы

символов 1251. Таблица 1 поясняет рисунок 2 и содержит исчерпывающую сводку численных значений частот текста № 1, включая процентное содержание, с полными кодами используемых символов.

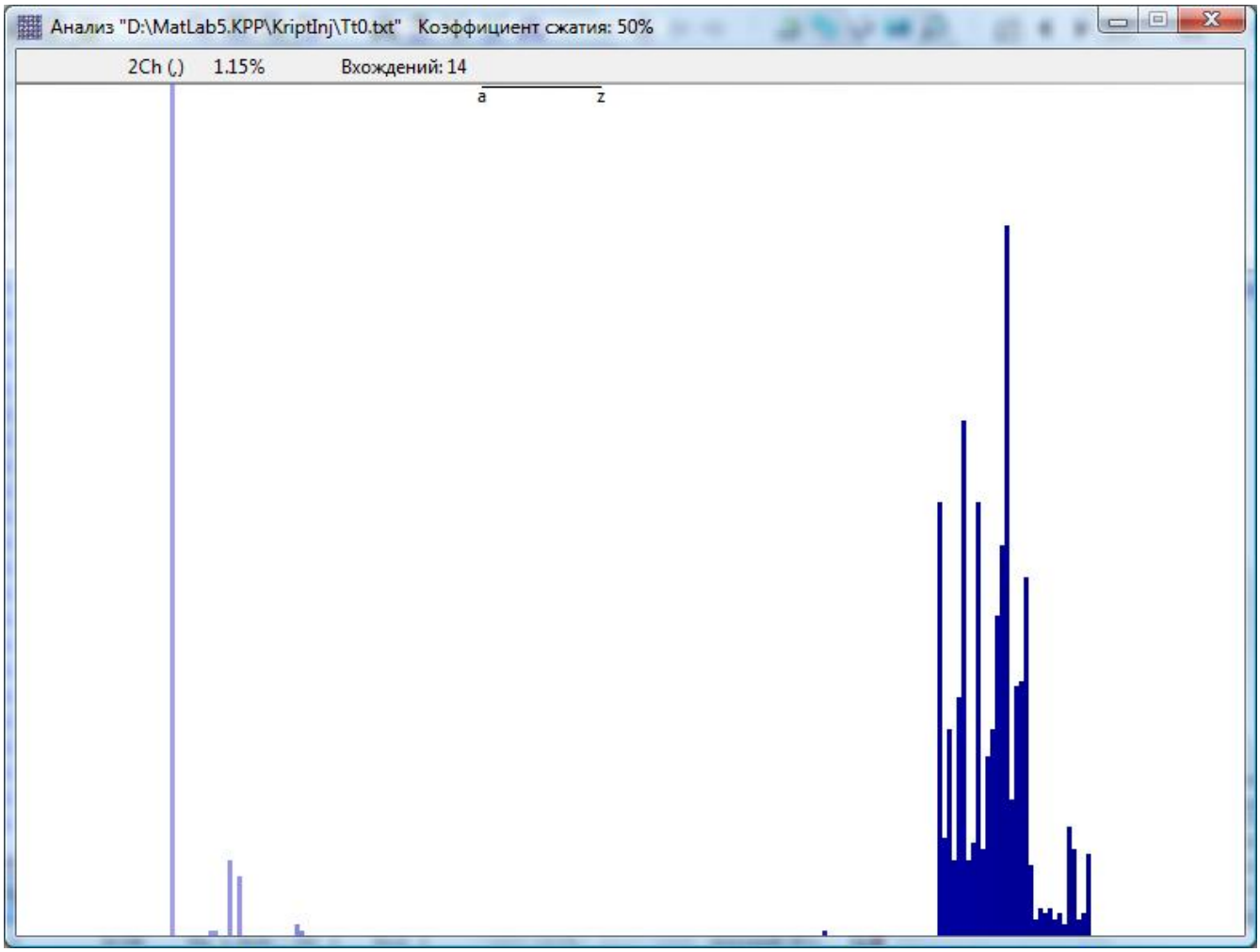


Рисунок 2 – Гистограмма частот букв в открытом тексте № 1

Таблица 1 – Полная таблица частот букв в открытом тексте № 1

Hex	Dec	Chr	No	%	Hex	Dec	Chr	No	%
20	32		157	12.89%	CC	204	М	59	4.84%
28	40	(	1	0.08%	CD	205	Н	72	5.91%
29	41	)	1	0.08%	CE	206	О	131	10.76%
2C	44	,	14	1.15%	CF	207	П	25	2.05%
2E	46	.	11	0.90%	D0	208	Р	46	3.78%
3A	58	:	2	0.16%	D1	209	С	47	3.86%
3B	59	;	1	0.08%	D2	210	Т	66	5.42%
A8	168	Ё	1	0.08%	D3	211	У	13	1.07%
C0	192	А	80	6.57%	D4	212	Ф	3	0.25%
C1	193	Б	18	1.48%	D5	213	Х	5	0.41%
C2	194	В	38	3.12%	D6	214	Ц	4	0.33%
C3	195	Г	14	1.15%	D7	215	Ч	5	0.41%
C4	196	Д	44	3.61%	D8	216	Ш	3	0.25%
C5	197	Е	95	7.80%	D9	217	Щ	4	0.33%
C6	198	Ж	14	1.15%	DA	218	Ъ	2	0.16%
C7	199	З	17	1.40%	DB	219	Ы	20	1.64%
C8	200	И	80	6.57%	DC	220	Ь	16	1.31%
C9	201	Й	16	1.31%	DD	221	Э	3	0.25%

CA	202	К	33	2.71%		DE	222	Ю	4	0.33%
CB	203	Л	38	3.12%		DF	223	Я	15	1.23%

Индекс совпадений Is = 0.060531296789807 в строке из 1031 букв русского алфавита без пробелов и знаков препинания.

Текст № 2. Прведен на рисунке 3. Без знаков препинания 1188 символов (-30) Алфавит 34 символа. Удалены с первого текста точки «.»=11, запятые «,»=14, двоеточия «:»=2, точка с запятой «;»=1, открывающая скобка «(»=1 и закрывающая скобка «)»=1.

В ФОРМИРОВАНИИ ТИПИЧНОЙ БОЛЬШОЙ СИСТЕМЫ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ АНАЛИТИК КАК ПРАВИЛО ДОЛЖЕН РАССМАТРИВАТЬ РЯД ЧЕТКО ОПРЕДЕЛЕННЫХ ПРЕДМЕТНЫХ ОБЛАСТЕЙ ИЛИ ДОМЕНОВ КАЖДЫЙ ДОМЕН МОЖЕТ РАССМАТРИВАТЬСЯ КАК ОТДЕЛЬНЫЙ МИР НАСЕЛЁННЫЙ СОБСТВЕННЫМИ КОНЦЕПТУАЛЬНЫМИ СУЩНОСТЯМИ ИЛИ ОБЪЕКТАМИ СЛЕДОВАТЕЛЬНО В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ УПРАВЛЕНИЯ ЖЕЛЕЗНОЙ ДОРОГОЙ ДОМЕН РАБОТА ЖЕЛЕЗНОЙ ДОРОГИ ИМЕЕТ ОТНОШЕНИЕ К ПОЕЗДАМ ДОРОГАМ И Т П В ТО ВРЕМЯ КАК ДОМЕН ПОЛЬЗОВАТЕЛЬСКИЙ ИНТЕРФЕЙС НАПОЛНЕН ОКНАМИ ДИСПЛЕЯМИ И ПИКТОГРАММАМИ КАЖДЫЙ ДОМЕН МОЖЕТ СУЩЕСТВОВАТЬ БОЛЕЕ ИЛИ МЕНЕЕ НЕЗАВИСИМО ОТ ДРУГИХ ЖЕЛЕЗНАЯ ДОРОГА МОЖЕТ СУЩЕСТВОВАТЬ БЕЗ ЭКРАНОВ И ОКОН ОКНА И ПИКТОГРАММЫ МОГУТ СУЩЕСТВОВАТЬ БЕЗ ПОЕЗДОВ ДОМЕНЫ ИЗОБРАЖАЮТСЯ НА СХЕМЕ ДОМЕНОВ НЕКОТОРЫЕ ДОМЕНЫ ДОВОЛЬНО МАЛЫ ЧТОБЫ МОЖНО БЫЛО АНАЛИЗИРОВАТЬ ИХ КАК ЕДИНОЕ ЦЕЛОЕ В ТО ВРЕМЯ КАК ДРУГИЕ СОДЕРЖАТ ТАК МНОГО ОБЪЕКТОВ ЧТО ТРУДНО ПОДДАЮТСЯ КОНТРОЛЮ И ОБРАБОТКЕ ПОЭТОМУ БОЛЬШИЕ ДОМЕНЫ РАЗБИВАЮТСЯ НА ПОДСИСТЕМЫ ПОСЛЕ ТОГО КАК СИСТЕМА РАЗБИТА НА ДОМЕНЫ И ПОДСИСТЕМЫ МОЖНО ПРОВОДИТЬ АНАЛИЗ НЕЗАВИСИМО ОДНА ОТ ДРУГОЙ КАЖДАЯ ПОДСИСТЕМА ИЛИ МАЛЕНЬКИЙ ДОМЕН АНАЛИЗИРУЕТСЯ В ТРИ ЭТАПА ИНФОРМАЦИОННОЕ МОДЕЛИРОВАНИЕ МОДЕЛИРОВАНИЕ СОСТОЯНИЙ И МОДЕЛИРОВАНИЕ ПРОЦЕССОВ

Рисунок 3 – Открытый текст № 2

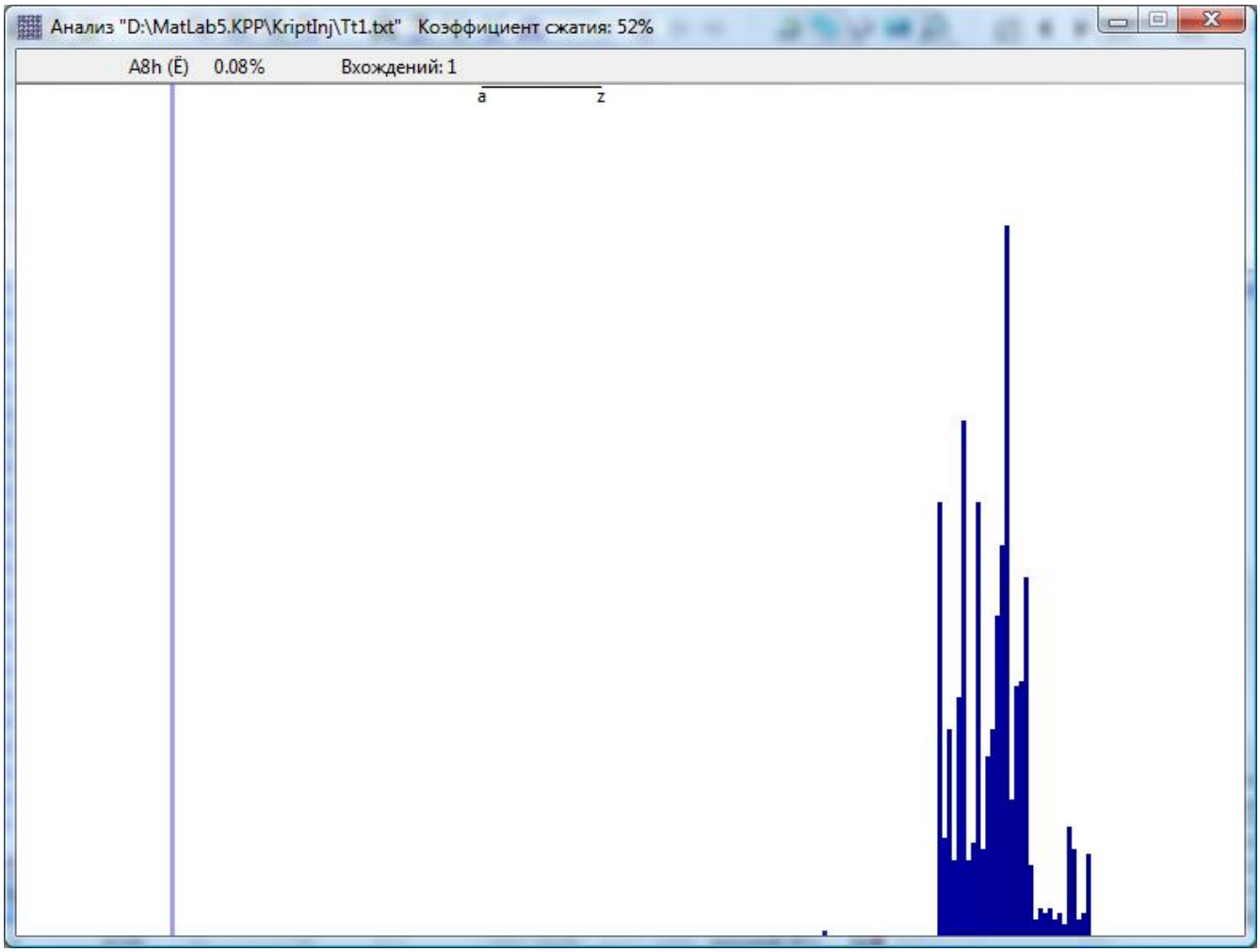


Рисунок 4 – Гистограмма частот букв в открытом тексте № 2

Таблица 2 – Полная таблица частот букв в открытом тексте № 2

Hex	Dec	Chr	No	%		Hex	Dec	Chr	No	%
20	32		157	13.22%		CF	207	П	25	2.10%
A8	168	Ё	1	0.08%		D0	208	Р	46	3.87%
C0	192	А	80	6.73%		D1	209	С	47	3.96%
C1	193	Б	18	1.52%		D2	210	Т	66	5.56%
C2	194	В	38	3.20%		D3	211	У	13	1.09%
C3	195	Г	14	1.18%		D4	212	Ф	3	0.25%
C4	196	Д	44	3.70%		D5	213	Х	5	0.42%
C5	197	Е	95	8.00%		D6	214	Ц	4	0.34%
C6	198	Ж	14	1.18%		D7	215	Ч	5	0.42%
C7	199	З	17	1.43%		D8	216	Ш	3	0.25%
C8	200	И	80	6.73%		D9	217	Щ	4	0.34%
C9	201	Й	16	1.35%		DA	218	Ъ	2	0.17%
CA	202	К	33	2.78%		DB	219	Ы	20	1.68%
CB	203	Л	38	3.20%		DC	220	Ь	16	1.35%
CC	204	М	59	4.97%		DD	221	Э	3	0.25%
CD	205	Н	72	6.06%		DE	222	Ю	4	0.34%
CE	206	О	131	11.03%		DF	223	Я	15	1.26%

Частоты незначительно изменились.

Текст № 3. Алфавит 32 символа. Ё = 1 → Е, Ъ = 2 → Ь, пробел =157 → Ъ

ВЪФОРМИРОВАНИИЪТИПИЧНОЙЪБОЛЬШОЙЪСИСТЕМЫЪПРОГРАММНОГОЪОБЕС-
ПЕЧЕНИЯЪАНАЛИТИКЪКАКЪПРАВИЛОЪДОЛЖЕНЪРАССМАТРИВАТЬЪРЯДЪЧЕТ-
КОЪОПРЕДЕЛЕННЫХЪПРЕДМЕТНЫХЪОБЛАСТЕЙЪИЛИЪДОМЕНОВЪКАЖДЫЙЪДО-
МЕНЪМОЖЕТЪРАССМАТРИВАТЬСЯЪКАКЪОТДЕЛЬНЫЙЪМИРЪНАСЕЛЕННЫЙЪСОБ-
СТВЕННЫМИЪКОНЦЕПТУАЛЬНЫМИЪСУЩНОСТЯМИЪИЛИЪОБЬЕКТАМИЪСЛЕДО-
ВАТЕЛЬНОЪВЪАВТОМАТИЗИРОВАННОЙЪСИСТЕМЕЪУПРАВЛЕНИЯЪЖЕЛЕЗ-
НОЙЪДОРОГОЙЪДОМЕНЪРАБОТАЪЖЕЛЕЗНОЙЪДОРОГИЪИМЕЕТЪОТНОШЕ-
НИЕЪКЪПОЕЗДАМЪДОРОГАМЪИЪТЪПЪВЪТОЪВРЕМЯЪКАКЪДОМЕНЪПОЛЬЗОВА-
ТЕЛЬСКИЙЪИНТЕРФЕЙСЪНАПОЛНЕНЪОКНАМИЪДИСПЛЕЯМИЪИЪПИКТОГРАММА-
МИЪКАЖДЫЙЪДОМЕНЪМОЖЕТЪСУЩЕСТВОВАТЬЪБОЛЕЕЪИЛИЪМЕНЕЕЪНЕЗАВИ-
СИМОЪОТЪДРУГИХЪЖЕЛЕЗНАЯЪДОРОГАЪМОЖЕТЪСУЩЕСТВОВАТЬЪБЕЗЪЭКРА-
НОВЪИЪОКОНЪОКНАЪИЪПИКТОГРАММЫЪМОГУТЪСУЩЕСТВОВАТЬЪБЕЗЪПОЕЗ-
ДОВЪДОМЕНЫЪИЗОБРАЖАЮТСЯЪНАЪСХЕМЕЪДОМЕНОВЪНЕКОТОРЫЕЪДОМЕ-
НЫЪДОВОЛЬНОЪМАЛЫЪЧТОБЫЪМОЖНОЪБЫЛОЪАНАЛИЗИРОВАТЬЪИХЪКАКЪ-
ЕДИНОЕЪЦЕЛОЕЪВЪТОЪВРЕМЯЪКАКЪДРУГИЕЪСОДЕРЖАТЪТАКЪМНОГОЪОБЬЕК-
ТОВЪЧТОЪТРУДНОЪПОДДАЮТСЯЪКОНТРОЛЮЪИЪОБРАБОТКЕЪПОЭТОМУЪБОЛЬ-
ШИЕЪДОМЕНЫЪРАЗБИВАЮТСЯЪНАЪПОДСИСТЕМЫЪПОСЛЕЪТОГОЪКАКЪСИСТЕ-
МАЪРАЗБИТАЪНАЪДОМЕНЫЪИЪПОДСИСТЕМЫЪМОЖНОЪПРОВОДИТЬЪАНА-
ЛИЗЪНЕЗАВИСИМОЪОДНАЪОТЪДРУГОЙЪКАЖДАЯЪПОДСИСТЕМАЪИЛИЪМАЛЕНЬ-
КИЙЪДОМЕНЪАНАЛИЗИРУЕТСЯЪВЪТРИЪЭТАПАЪИНФОРМАЦИОННОЕЪМОДЕЛИ-
РОВАНИЕЪМОДЕЛИРОВАНИЕЪСОСТОЯНИЙЪИЪМОДЕЛИРОВАНИЕЪПРОЦЕССОВ

Рисунок 5 – Открытый текст № 3

Открытый текст № 3 удобен для выполнения шифрования, но чисто русским его уже не назовёшь. Частота твёрдого знака принимает значение для знака пробела.

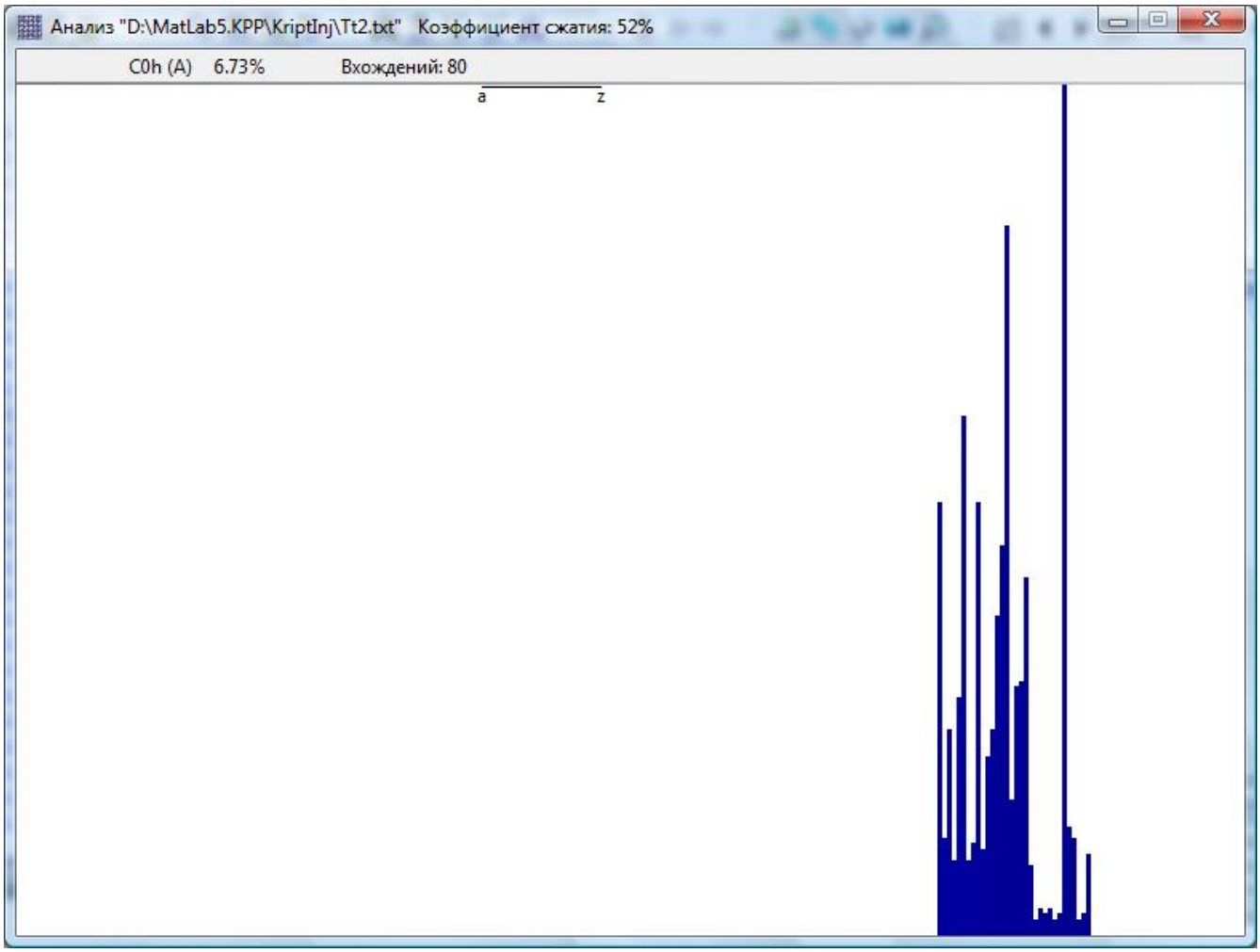


Рисунок 6 – Гистограмма частот букв в открытом тексте № 3

Таблица 3 – Полная таблица частот букв в открытом тексте № 3

Hex	Dec	Chr	No	%		Hex	Dec	Chr	No	%
C0	192	А	80	6.73%		D0	208	Р	46	3.87%
C1	193	Б	18	1.52%		D1	209	С	47	3.96%
C2	194	В	38	3.20%		D2	210	Т	66	5.56%
C3	195	Г	14	1.18%		D3	211	У	13	1.09%
C4	196	Д	44	3.70%		D4	212	Ф	3	0.25%
C5	197	Е	96	8.08%		D5	213	Х	5	0.42%
C6	198	Ж	14	1.18%		D6	214	Ц	4	0.34%
C7	199	З	17	1.43%		D7	215	Ч	5	0.42%
C8	200	И	80	6.73%		D8	216	Ш	3	0.25%
C9	201	Й	16	1.35%		D9	217	Щ	4	0.34%
CA	202	К	33	2.78%		DA	218	Ъ	157	13.22%
CB	203	Л	38	3.20%		DB	219	Ы	20	1.68%
CC	204	М	59	4.97%		DC	220	Ь	18	1.52%
CD	205	Н	72	6.06%		DD	221	Э	3	0.25%
CE	206	О	131	11.03%		DE	222	Ю	4	0.34%
CF	207	П	25	2.10%		DF	223	Я	15	1.26%

Частоты незначительно изменились. Перераспределились среди 4 букв: Ё, Е, Ъ и Ь.

Индекс совпадений Is = 0.06313202. Есть отклонение от общепризнанного значения 0.0553. Сказывается влияние пробела и перераспределения 4 букв.

Сдвиг Цезаря

Используемый алфавит шифротекста:
АБВГДЕЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ.

Шифрование осуществляем, работая с открытым текстом № 3 в качестве источника. Это сообщение длиной 1188 символов. Всего слов в исходном тексте 158, разделённых пробелами. Создадим две шифрограммы.

Вариант 1. Шифр Цезаря, сдвиг k = 3 (ROT3), на выходе получается шифротекст № 1. Эта шифрограмма приведена на рисунке 7. Шифрограмма представляет собой одну строку символов. Для наглядности, чтобы текст вместился в рамках страницы, на рисунке 7 присутствуют разбивка на строки и знаки переноса.

ЕЭЧСУПЛУСЕГРЛЛЭХЛТЛЪРСМЭДСОЯЫСМЭФЛФХИПЮЭТУСЖУГППРСЖСЭСДИФ-
ТИЪИРЛВЭГРГОЛХЛНЭНГНЭТУГЕЛОСЭЗСОЙИРЭУГФФПГХУЛЕГХЯЭУВЗЭЪИХНСЭС-
ТУИЗИОИРРЮШЭТУИЗПИХРЮШЭСДОГФХИМЭЛОЛЭЗСПИРСЕЭНГЙЗЮМЭЗСПИР-
ЭПСЙИХЭУГФФПГХУЛЕГХЯФВЭНГНЭСХЗИОЯРЮМЭПЛУЭРГФИОИРРЮМЭФСДФХЕ-
ИРРЮПЛЭНСРЩИТХЦГОЯРЮПЛЭФЦЬРСФХВПЛЭЛОЛЭСДЯИНХГПЛЭФОИЗСЕГХИО-
ЯРСЭЕЭГЕХСПГХЛКЛУСЕГРРСМЭФЛФХИПИЭЦТУГЕОИРЛВЭЙИОИКРСМЭЗСУ-
СЖСМЭЗСПИРЭУГДСХГЭЙИОИКРСМЭЗСУСЖЛЭЛПИИХЭСХРСЫИРЛИЭНЭТ-
СИКЗГПЭЗСУСЖГПЭЛЭХЭТЭЕЭХСЭЕУИПВЭНГНЭЗСПИРЭТСОЯКСЕГ-
ХИОЯФНЛМЭЛРХИУЧИМФЭРГТСОРИРЭСНРГПЛЭЗЛФТОИВПЛЭЛЭТЛНХСЖУ-
ГППГПЛЭНГЙЗЮМЭЗСПИРЭПСЙИХЭФЦЬИФХЕСЕГХЯЭДСОИИЭЛОЛЭПИРИИЭРИК-
ГЕЛФЛПСЭСХЭЗУЦЖЛШЭЙИОИКРГВЭЗСУСЖГЭПСЙИХЭФЦЬИФХЕ-
СЕГХЯЭДИКЭАНУГРСЕЭЛЭСНСРЭСНРГЭЛЭТЛНХСЖУГППЮЭПСЖЦХЭФЦЬИФХЕ-
СЕГХЯЭДИКЭТСИКЗСЕЭЗСПИРЮЭЛКСДУГЙГБХФВЭРГЭФШИПИЭЗСПИРСЕ-
ЭРИНСХСУЮИЭЗСПИРЮЭЗСЕСОЯРСЭПГОЮЭЪХСДЮЭПСЙРСЭДЮОСЭГРГОЛКЛУ-
СЕГХЯЭЛШЭНГНЭИЗЛРСИЭЩИОСИЭЕЭХСЭЕУИПВЭНГНЭЗУЦЖЛИЭФСЗИУЙГХЭХГ-
НЭПРСЖСЭСДЯИНХСЕЭЪХСЭХУЦЗРСЭТСЗЗГБХФВЭНСРХУСОБЭЛЭСДУГДСХНИЭТ-
САХСПЦЭДСОЯЫЛИЭЗСПИ-
РЮЭУГКДЛЕГБХФВЭРГЭТСЗФЛФХИПЮЭТСФОИЭХСЖСЭНГНЭФЛФХИПГ-
ЭУГКДЛХГЭРГЭЗСПИРЮЭЛЭТСЗФЛФХИПЮЭПСЙРСЭТУСЕСЗЛХЯЭГРГОЛКЭРИК-
ГЕЛФЛПСЭСЗРГЭСХЭЗУЦЖСМЭНГЙЗГВЭТСЗФЛФХИПГЭЛОЛЭПГОИРЯНЛМЭЗСПИР-
ЭГРГОЛКЛУЦИХФВЭЕЭХУЛЭАХГТГЭЛРЧСУПГЩЛСРРСИЭПСЗИОЛУСЕГРЛИЭПСЗИО-
ЛУСЕГРЛИЭФСФХСВРЛМЭЛЭПСЗИОЛУСЕГРЛИЭТУСЩИФФСЕ

Рисунок 7 – Шифротекст № 1 открытого текста № 3

Индекс совпадений Is = 0.06313202

Вариант 2. Шифр Цезаря, сдвиг k = 7 (ROT7), на выходе шифротекст № 2. Эта шифрограмма приведена на рисунке 8. Текст существенно отличается от первой шифрограммы на рисунке 7. На первый взгляд никакой связи в этих текстах не обнаруживается, только длины одинаковые.

ЙБЫХЧУПЧХЙЗФППБЩПЦПЮФХРБИХТГЯХРБШПШЩМУВБЦЧХКЧЗУУФХКХБХИМ-
ШЩМЮМФПЖБЗФЗТПЩПСБСЗСБЦЧЗЙПТХБЛХТ-
НМФБЧЗШШУЗЩЧПЙЗЩГБЧЖЛБЮМЩСХБХЦЧМЛМТМФФВЬБЦЧМЛУМЩФВЬБХИТ-
ЗШЩМРБПТПБЛХУМФХЙБСЗН-
ЛВРБЛХУМФБУХНМЩБЧЗШШУЗЩЧПЙЗЩГШЖБСЗСБХЩЛМТГФВРБУП-
ЧБФЗШМТМФФВРБШХИШЩЙМФФВУПБСХФЭМЦЩЪЗТГФВУ-
ПБШЪАФХШЩЖУПБПТПБХИГМСЩЗУПБШТМЛХЙЗЩМТГФХБЙБЗЙЩХУЗЩПОП-
ЧХЙЗФФХРБШПШЩМУМБЪЦЧЗЙТМФПЖБНМТМО-

ФХРБЛХЧХКХРБЛХУМФБЧЗИХЩЗБНМТМОФХРБЛХЧХКПБПУМ-
МЩБХЩФХЯМФПМБСБЦХМОЛ-
ЗУБЛХЧХКЗУБПБЩБЦБЙБЩХБЙЧМУЖБСЗСБЛХУМФБЦХТГО-
ХЙЗЩМТГШСПРБПФЩМЧЫМРШБФЗЦХТФМФБХСФЗУПБЛПШЦТМ-
ЖУПБПБЦПСЩХКЧЗУУЗУПБСЗНЛВРБЛХУМФБУХНМЩБШЪАМШЩЙХЙЗЩГБИ-
ХТММБПТПБУМФММБФМОЗЙПШПУХБХЩБЛЧЪКПЬБНМТМО-
ФЗЖБЛХЧХКЗБУХНМЩБШЪАМШЩЙХЙЗЩГБИ-
МОБДСЧЗФХЙБПБХСХФБХСФЗБПБЦПСЩХКЧЗУУВБУХКЪЩБШЪАМШЩЙХЙЗЩГБИ-
МОБЦХМОЛХЙБЛХУМФВБПОХИЧЗН-
ЗЕЩШЖБФЗБШЬМУМБЛХУМФХЙБФМСХЩХЧВМБЛХУМФВБЛХЙХТГФХБУЗТВБЮЩХ
ИВБУХНФХБИВТХБЗФЗТПОП-
ЧХЙЗЩГБПЬБСЗСБМЛПФХМБЭМТХМБЙБЩХБЙЧМУЖБСЗСБЛЧЪКПМБШХЛМЧНЗЩБ
ЩЗСБУФХКХБХИГМСЩХЙБЮЩХБЩЧЪЛФХБЦХЛЛ-
ЗЕЩШЖБСХФЩЧХТЕБПБХИЧЗИХЩСМБЦХДЩХУЪБИХТГЯПМБЛХУМФВБЧЗОИПЙ-
ЗЕЩШЖБФЗБЦХЛШПШЩМУВБЦХШТМБЩХКХБСЗСБШПШЩМУЗБЧЗОИ-
ПЩЗБФЗБЛХУМФВБПБЦХЛШПШЩМУВБУХНФХБЦЧХЙХЛПЩГБЗФЗТПОБФ-
МОЗЙПШПУХБХЛФЗБХЩБЛЧЪКХРБСЗН-
ЛЗЖБЦХЛШПШЩМУЗБПТПБУЗТМФГСПРБЛХУМФБЗФЗТПОП-
ЧЪМЩШЖБЙБЩЧПБДЩЗЦЗБПФЫХЧУЗЭПХФФХМБУХЛМТПЧХЙЗФПМБУХ-
ЛМТПЧХЙЗФПМБШХШЩХЖФПРБПБУХЛМТПЧХЙЗФПМБЦЧХЭМШШХЙ

Рисунок 8 – Шифротекст № 2 открытого текста № 3

Индекс совпадений Is = 0.06313202, совпадает с индексом шифротекста № 1.

Частотный анализ текста состоит в построении гистограммы частот встречаемости символов (букв) в шифротексте.

Гистограмма шифротекста № 1 приведена на рисунке 9. Численные значения частот сведены в таблицу 4.

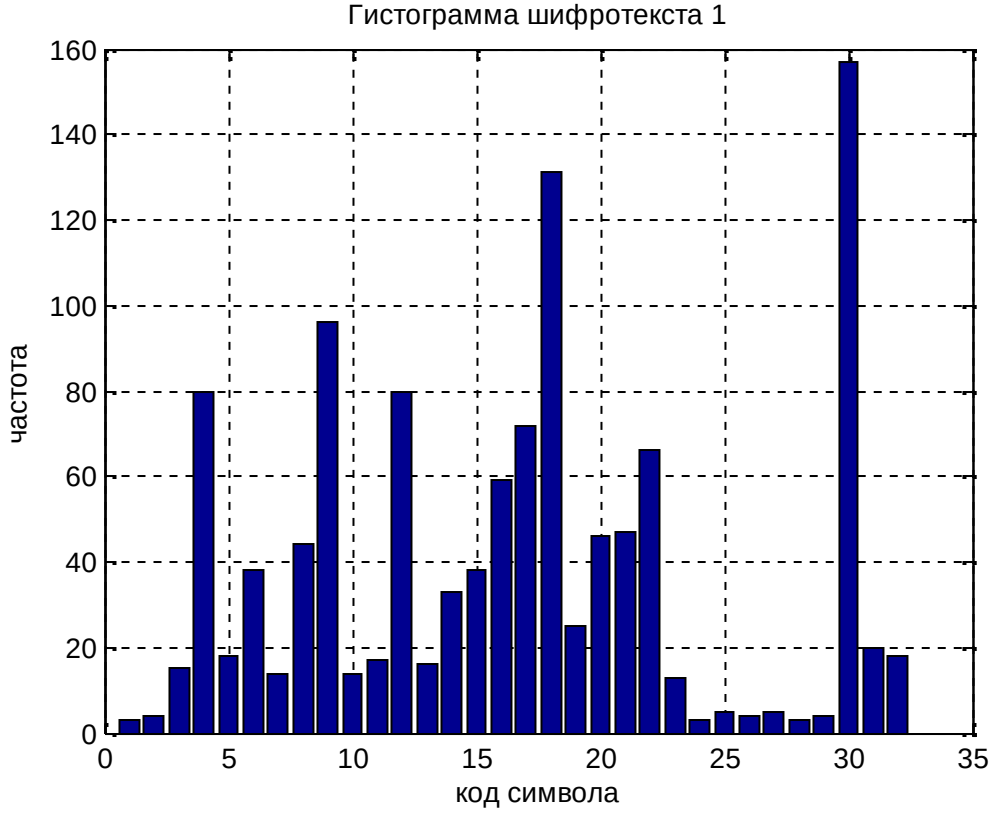


Рисунок 9 – Гистограмма частот букв в шифротексте № 1

Таблица 4 – Частоты букв в шифротексте № 1

А	3		Ж	14		М	16		Т	25		Ш	5		Ю	20
Б	4		З	44		Н	33		У	46		Щ	4		Я	18
В	15		И	96		О	38		Ф	47		Ъ	5			
Г	80		Й	14		П	59		Х	66		Ы	3			
Д	18		К	17		Р	72		Ц	13		Ь	4			
Е	38		Л	80		С	131		Ч	3		Э	157			

Гистограмма шифротекста № 2 приведена на рисунке 10. Численные значения частот сведены в таблицу 5.

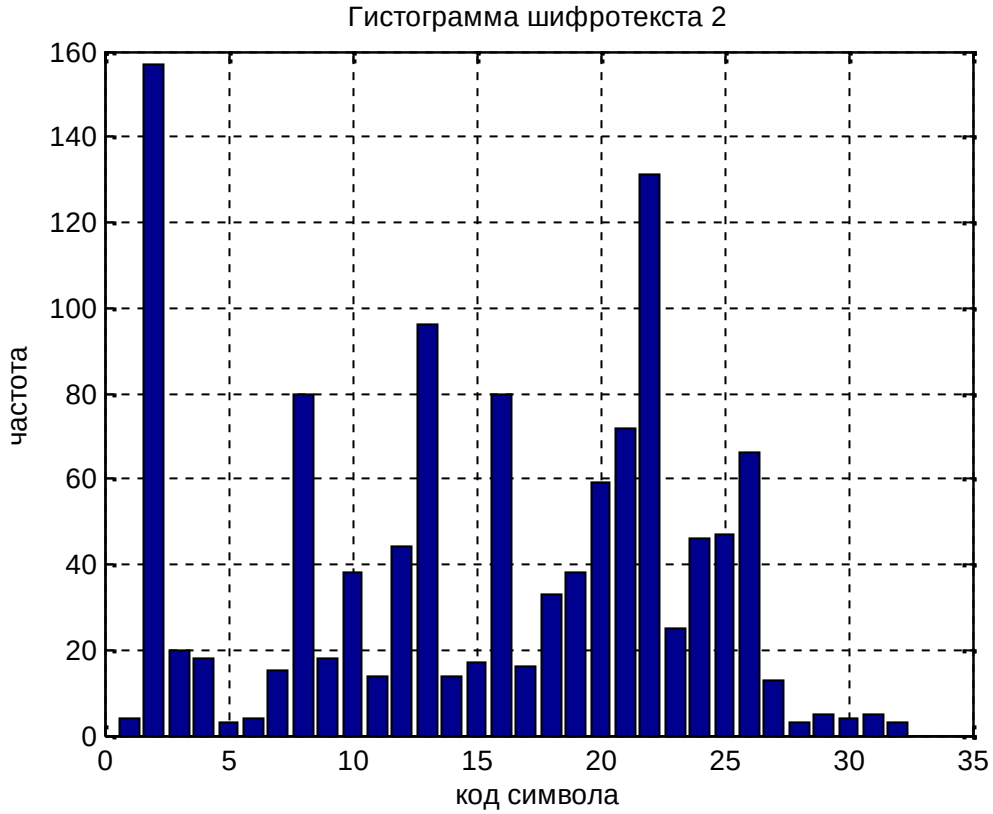


Рисунок 10 – Гистограмма частот букв в шифротексте № 2

Таблица 5 – Частоты букв в шифротексте № 2

А	4		Ж	15		М	96		Т	38		Ш	47		Ю	5
Б	157		З	80		Н	14		У	59		Щ	66		Я	3
В	20		И	18		О	17		Ф	72		Ъ	13			
Г	18		Й	38		П	80		Х	131		Ы	3			
Д	3		К	14		Р	16		Ц	25		Ь	5			
Е	4		Л	44		С	33		Ч	46		Э	4			

Гистограммы на рисунках 9 и 10 на первый взгляд различаются. Детали проясняются в таблицах 4 и 5 с частотами символов. В шифротексте № 1 символ «Э» встречается 157 раз, а в шифротексте № 2 символ «Б» встречается тоже 157 раз. Анализируя таблицы 4 и 5, приходим к выводу, что в них содержатся одни и те же частоты, только переставлены местами и закреплены за другими символами.

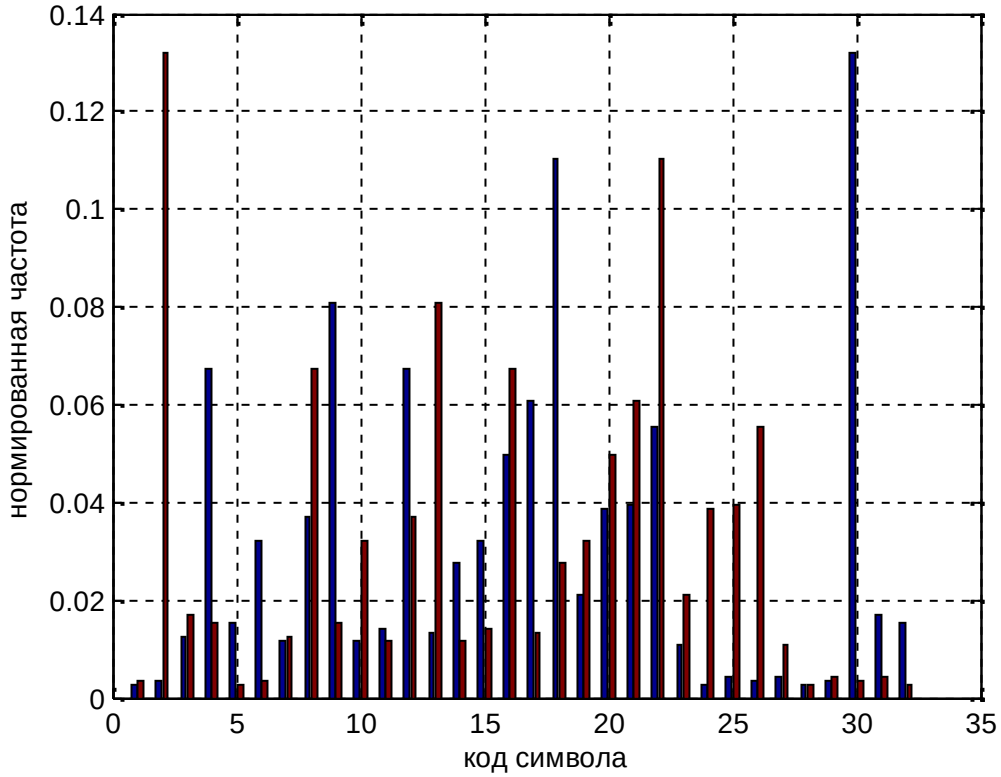
Поскольку две шифрограммы получены из одного открытого текста, у их символов в точности одинаковые частоты. Вывод 1 – шифры замены не меняют частоты символов.

По своей структуре гистограммы частот двух шифротекстов – разные. Это различие можно зафиксировать с помощью квадратичной статистики. Для сравнения гистограммы нормируются на число символов в шифрограмме, и вычисляется сумма квадратов разностей значений частот двух гистограмм. Пусть H и G – две нормированные гистограммы. Суммарная квадратичная разность Squ вычисляется по формуле

$$Squ = \sum_i (H_i - G_i)^2 . \qquad (1)$$

Здесь суммирование (i) по всем карманам.

Для наглядности сравнения гистограмм две гистограммы пронормированы и сведены воедино на рисунке 11. Видно, что каждый карман отличается или слабо, или существенно.



Синий - шифротекст № 1, коричневый – шифротекст № 2
Рисунок 11 – Нормированные гистограммы частот букв в двух шифрограммах № 1 и № 2

Для рисунка 11 Squ = 0.067324. Такое большое значение говорит, что два шифра совершенно разные. Это подтверждает их разное происхождение при выполнении шифрации двумя шифрами.

Сравнивая рисунки 9 и 10 между собой, тем не менее у них обнаруживается одно сходство. У них одинаковая конфигурация пиков, только они сдвинуты. Обнаруживать такие сдвиги позволяет циклическая взаимная корреляционная функция.

Циклическая взаимная корреляционная функция

Пусть имеются две гистограммы H и G, у которых m карманов. Проверяем запаздывание t гистограммы G относительно гистограммы H. Циклическая взаимная корреляционная функция Kf(t) определяется выражением

$$Kf(t) = \sum_k H(k) \cdot G((k + t))_m \, . \qquad (2)$$

Здесь суммирование (k) по всем карманам. Для гистограммы G номер кармана (k + t) вычисляется по модулю m. Функция Kf вычисляется для всех значений запаздывания от 0 до m.

На рисунке 12 представлена циклическая взаимная корреляционная функция гистограмм исследуемых шифротекстов № 1 и № 2. Наибольший пик соответствует запаздыванию 4. Значит шифралфавит № 2 сдвинут относительно шифралфавита № 1 на 4 буквы. Пик очень отчётливый, он в два раза превосходит все остальные значения. Это связано с тем, что две шифрограммы получены из одного и того же открытого текста.

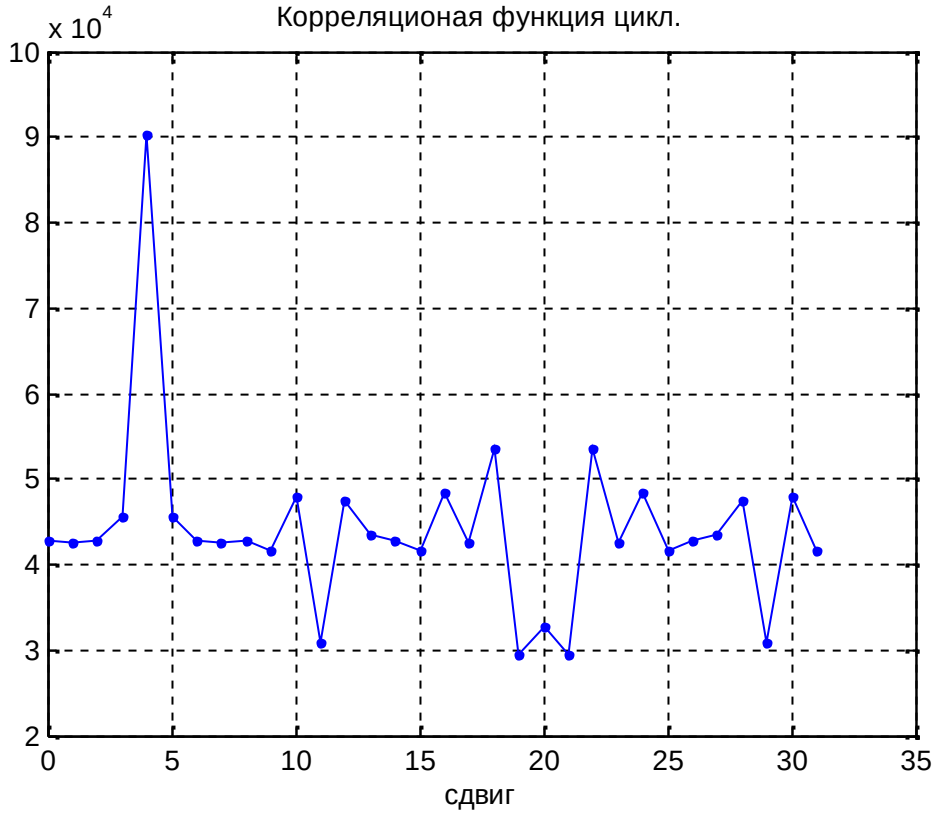


Рисунок 12 – Циклическая взаимная корреляционная функция из шифрограмм № 1 и № 2

Вывод 2. Чтобы определить ключ сдвига Цезаря достаточно посчитать взаимную корреляционную функцию гистограммы стандартного текста с гистограммой шифротекста.

Два открытых текста

Рассмотрим случай, когда шифрограммы получены из разных открытых текстов. В качестве первого используем прежний открытый текст № 3, исследования которого ранее проведены.

Второй текст

В качестве примера второго текста на русском языке взят фрагмент, приведенный на рисунке 13. Это стихотворная строфа И. Губермана:

Хотя и сладостен азарт
По сразу двум идти дорогам,
Нельзя с одной колодой карт
Играть и с дьяволом, и с Богом.

Здесь 110 символов с пробелами и знаками препинания. Для удобства избавились от знаков препинания и пробелы заменили на твёрдый знак Ъ. Алфавит согласован с открытым текстом № 3. В итоге второй текст содержит 107 символов, 21 слово.

ХОТЯЪИЪСЛАДОСТЕНЪАЗАРТЪПОЪСРАЗУЪДВУМЪИДТИЪДОРОГАМЪНЕЛЬ-
ЗЯЪСЪОДНОЙЪКОЛОДОЙЪКАРТЪИГРАТЬЪИЪСЪДЬЯВОЛОМЪИЪСЪБОГОМ
Рисунок 13 – Открытый текст № 4

Вариант 3. Шифрование открытого текста № 3. Шифр Цезаря, сдвиг k = 5 (ROT5), на выходе получается шифротекст № 3. Эта шифрограмма приведена на рисунке 14. Шифрограмма представляет собой одну строку символов. Для наглядности на рисунке 14, как и ранее, дополнительно присутствуют разбивка на строки и знаки переноса.

ЗЯЩУХСНХУЗЕТННЯЧНФНЬТУОЯЖУРБЭУОЯЦНЦЧКСАЯФХУИХЕС-
СТУИУЯУЖКЦФКЬКТНДЯЕТЕРНЧНПЯПЕПЯФХЕЗНРУЯЙУРЛКТЯХЕЦЦСЕЧХН-
ЗЕЧБЯХДЙЯЬКЧПУЯУФХКЙКРКТТАЪЯФХКЙСКЧТАЪЯУЖРЕЦЧКОЯНРНЯЙУСКТУЗЯ-
ПЕЛЙАОЯЙУСКТЯСУЛКЧЯХЕЦЦСЕЧХНЗЕЧБЦДЯПЕПЯУЧЙКРБТАОЯСНХЯТЕЦКРКТ-
ТАОЯЦУЖЦЧЗКТТАСНЯПУТЫКФЧШЕРБТАСНЯЦШЮТУЦЧДСНЯНРНЯУЖБКПЧЕС-
НЯЦРКЙУЗЕЧКРБТУЯЗЯЕЗЧУСЕЧНМНХУЗЕТТУОЯЦНЦЧКСКЯШФХЕЗРКТНДЯЛКРКМ-
ТУОЯЙУХУИУОЯЙУСКТЯХЕЖУЧЕЯЛКРКМТУОЯЙУХУИНЯНСККЧЯУЧТУЭКТНКЯ-
ПЯФУКМЙЕСЯЙУХУИЕСЯНЯЧЯФЯЗЯЧУЯЗХКСДЯПЕПЯЙУСКТЯФУРБМУ-
ЗЕЧКРБЦПНОЯНТЧКХЩКОЦЯТЕФУРТКТЯУПТЕСНЯЙНЦФРКДСНЯНЯФНПЧУИХЕС-
СЕСНЯПЕЛЙАОЯЙУСКТЯСУЛКЧЯЦШЮКЦЧЗУЗЕЧБЯЖУРККЯНРНЯСКТККЯТК-
МЕЗНЦНСУЯУЧЯЙХШИНЪЯЛКРКМТЕДЯЙУХУИЕЯСУЛКЧЯЦШЮКЦЧЗУЗЕЧБЯ-
ЖКМЯВПХЕТУЗЯНЯУПУТЯУПТЕЯНЯФНПЧУИХЕССАЯСУИШЧЯЦШЮКЦЧЗУЗЕЧБЯ-
ЖКМЯФУКМЙУЗЯЙУСКТАЯНМУЖХЕЛЕГЧЦДЯТЕЯЦЪКСКЯЙУСКТУЗЯТКПУЧУХ-
АКЯЙУСКТАЯЙУЗУРБТУЯСЕРАЯЬЧУЖАЯСУЛТУЯЖАРУЯЕТЕРНМНХУЗЕЧБЯНЪЯПЕ-
ПЯКЙНТУКЯЫКРУКЯЗЯЧУЯЗХКСДЯПЕПЯЙХШИНКЯЦУЙКХЛЕЧЯЧЕ-
ПЯСТУИУЯУЖБКПЧУЗЯЬЧУЯЧХШЙТУЯФУЙЙЕГЧЦДЯПУТЧХУР-
ГЯНЯУЖХЕЖУЧПКЯФУВЧУСШЯЖУРБЭНКЯЙУСКТАЯХЕМЖНЗЕГЧЦДЯТЕ-
ЯФУЙЦНЦЧКСАЯФУЦРКЯЧУИУЯПЕПЯЦНЦЧКСЕЯХЕМЖНЧЕЯТЕЯЙУСКТА-
ЯНЯФУЙЦНЦЧКСАЯСУЛТУЯФХУЗУЙНЧБЯЕТЕРНМЯТКМЕЗНЦНСУЯУЙТЕЯУЧЯЙХ-
ШИУОЯПЕЛЙЕДЯФУЙЦНЦЧКСЕЯНРНЯСЕРКТБПНОЯЙУСКТЯЕ-
ТЕРНМНХШКЧЦДЯЗЯЧХНЯВЧЕФЕЯНТЩУХСЕЫНУТТУКЯСУЙКРНХУЗЕТНКЯ-
СУЙКРНХУЗЕТНКЯЦУЦЧУДТНОЯНЯСУЙКРНХУЗЕТНКЯФХУЫКЦЦУЗ
Рисунок 14 – Шифротекст № 3 открытого текста № 3

Индекс совпадений Is = 0.06313202

Для полноты другой вариант шифрования открытого текста № 3. Шифр Цезаря, сдвиг k = 11 (ROT11), на выходе шифротекст № 4. Эта шифрограмма приведена на рисунке 15.

НЕЯЩЫЧУЫЩНЛШУУЕЭУЪУВШЩФЕМЩЦЗГЩФЕЬУЬЭРЧЖЕЪЫ-
ЩОЫЛЧЧШЩОЩЕЩМРЬЪРВРШУКЕЛШЛЦУЭУХЕХЛХЕЪЫЛНУЦЩЕПЩЦСР-
ШЕЫЛЬЬЧЛЭЫУНЛЭЗЕЫКПЕВРЭХЩЕЩЪЫРПРЦРШШЖАЕЪЫРПЧРЭШЖАЕЩМЦЛЬ-
ЭРФЕУЦУЕПЩЧРШЩНЕХЛСПЖФЕПЩЧРШЕЧЩСРЭЕЫЛЬЬЧЛЭЫУНЛЭЗЬ-
КЕХЛХЕЩЭПРЦЗШЖФЕЧУЫЕШЛЬРЦРШШЖФЕЬЩМЬ-
ЭНРШШЖЧУЕХЩШБРЪЭЮЛЦЗШЖЧУЕЬЮДШЩЬЭКЧУЕУЦУЕЩМЗРХЭЛ-
ЧУЕЬЦРПЩНЛЭРЦЗШЩЕНЕЛНЭЩЧЛЭУТУЫЩНЛШШЩФЕЬУЬЭРЧРЕЮЪЫЛНЦРШУ-
КЕСРЦРТШЩФЕПЩЫЩОЩФЕПЩЧРШЕЫЛМЩЭЛЕСРЦРТШЩФЕПЩЫЩОУЕУЧРР-
ЭЕЩЭШЩГРШУРЕХЕЪЩРТПЛЧЕПЩЫЩОЛЧЕУЕЭЕЪЕНЕЭЩЕНЫРЧКЕХЛХЕПЩЧР-
ШЕЪЩЦЗТЩНЛЭРЦЗЬХУФЕУШЭРЫЯРФЬЕШЛЪЩЦШРШЕЩХШЛЧУЕПУЬЪЦРК-
ЧУЕУЕЪУХЭЩОЫЛЧЧЛЧУЕХЛСПЖФЕПЩЧРШЕЧЩСРЭЕЬЮДРЬЭНЩНЛЭЗЕМЩЦРРЕ-
УЦУЕЧРШРРЕШРТЛНУЬУЧЩЕЩЭЕПЫЮОУАЕСРЦРТШЛКЕПЩЫЩОЛЕЧЩСРЭЕЬЮ-
ДРЬЭНЩНЛЭЗЕМРТЕИХЫЛШЩНЕУЕЩХЩШЕЩХШЛЕУЕЪУХ-
ЭЩОЫЛЧЧЖЕЧЩОЮЭЕЬЮДРЬЭНЩНЛЭЗЕМРТЕЪЩРТПЩНЕПЩЧРШЖЕ-
УТЩМЫЛСЛЙЭЬКЕШЛЕЬАРЧРЕПЩЧРШЩНЕШРХЩЭЩЫЖРЕПЩЧРШЖЕ-
ППЩНЩЦЗШЩЕЧЛЦЖЕВЭЩМЖЕЧЩСШЩЕМЖЦЩЕЛШЛЦУТУЫЩНЛЭЗЕУА-
ЕХЛХЕРПУШЩРЕБРЦЩРЕНЕЭЩЕНЫРЧКЕХЛХЕПЫЮО-
УРЕЬЩПРЫСЛЭЕЭЛХЕЧШЩОЩЕЩМЗРХЭЩНЕВЭЩЕЭЫЮПШЩЕЪЩППЛЙЭЬ-
КЕХЩШЭЫЩЦЙЕУЕЩМЫЛМЩЭХРЕЪЩИЭЩЧЮЕМЩЦЗГУРЕПЩЧРШЖЕЫЛТ-
МУНЛЙЭЬКЕШЛЕЪЩПЬУЬЭРЧЖЕЪЩЬЦРЕЭЩОЩЕХЛХЕЬУЬЭРЧЛЕЫЛТ-
МУЭЛЕШЛЕПЩЧРШЖЕУЕЪЩПЬУЬЭРЧЖЕЧЩСШЩЕЪЫЩНЩПУЭЗЕЛШЛЦУТЕШРТЛ-
НУЬУЧЩЕЩПШЛЕЩЭЕПЫЮОЩФЕХЛСПЛКЕЪЩПЬУЬЭРЧЛЕУЦУЕЧЛЦРШЗХУ-
ФЕПЩЧРШЕЛШЛЦУТУЫЮРЭЬКЕНЕЭЫУЕИЭЛЪЛЕУШЯЩЫЧЛБУЩШШЩРЕЧЩПР-
ЦУЫЩНЛШУРЕЧЩПРЦУЫЩНЛШУРЕЬЩЬЭЩКШУФЕУЕЧЩПРЦУЫЩНЛШУ-
РЕЪЫЩБРЬЬЩН

Рисунок 15 – Шифротекст № 4 открытого текста № 3

Индекс совпадений Is = 0.06313202, совпадает с индексом предыдущего шифротекста. Всегда один и тот же.

Вариант 4. Шифрование открытого текста № 4. Сообщение длиной 107 символов. Шифр Цезаря, сдвиг k = 11 (ROT11), на выходе шифротекст № 5. Эта шифрограмма приведена на рисунке 16.

АЩЭКЕУЕЬЦЛПЩЬЭРШЕЛТЛЫЭЕЪЩЕЬЫЛТЮЕПНЮЧЕУПЭУЕПЩЫЩОЛ-
ЧЕШРЦЗТКЕЬЕЩПШЩФЕХЩЦЩПЩФЕХЛЫЭЕУОЫЛЭЗЕУЕЬЕПЗКНЩЦЩЧЕУЕЬЕМЩО
ЩЧ

Рисунок 16 – Шифротекст № 5 открытого текста № 4

Индекс совпадений Is = 0.07229765, возрос по сравнению с шифротекстом № 4.

Для выполнения частотного анализа посчитаны две новые гистограммы. Гистограмма шифротекста № 3 приведена на рисунке 17. Численные значения частот сведены в таблицу 6. Эта гистограмма по конфигурации пиков совпадает с предыдущими гистограммами шифротекстов № 1 и № 2, но со своими сдвигами. Это естественно, поскольку шифротексты исходят из одного открытого текста.

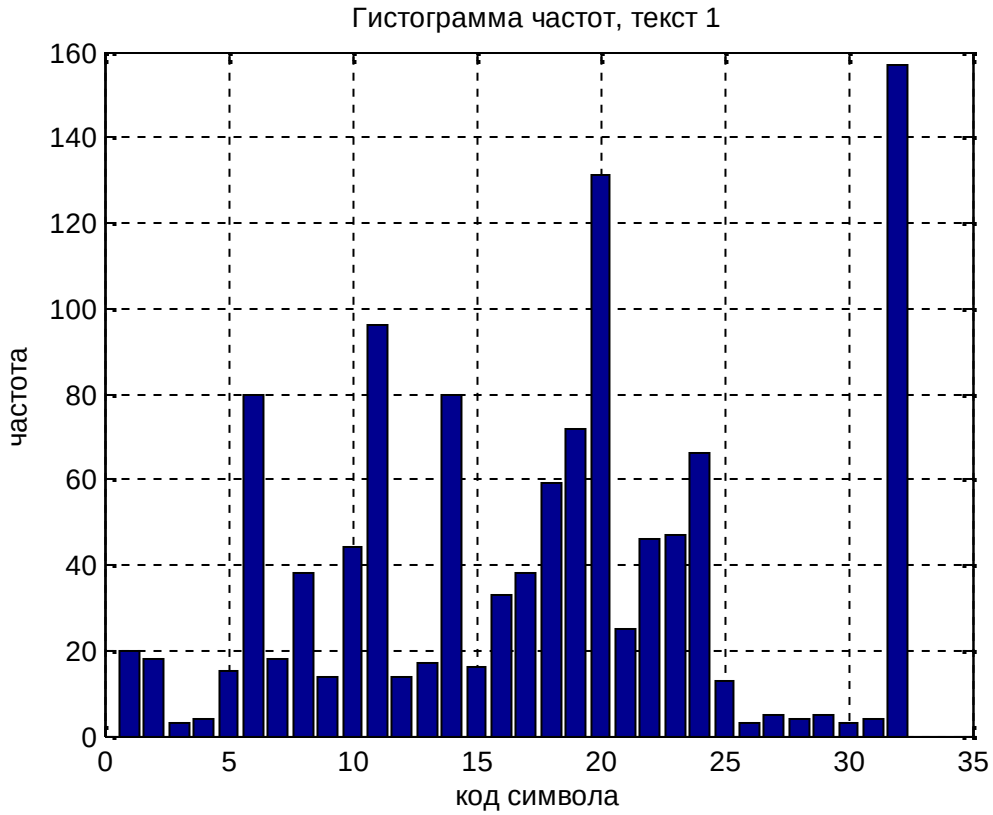


Рисунок 17 – Гистограмма частот букв в шифротексте № 3

Таблица 6 – Частоты букв в шифротексте № 3

А	20		Ж	18		М	17		Т	72		Ш	13		Ю	4
Б	18		З	38		Н	80		У	131		Щ	3		Я	157
В	3		И	14		О	16		Ф	25		Ъ	5			
Г	4		Й	44		П	33		Х	46		Ы	4			
Д	15		К	96		Р	38		Ц	47		Ь	5			
Е	80		Л	14		С	59		Ч	66		Э	3			

Гистограмма шифротекста № 5 приведена на рисунке 18. Численные значения частот сведены в таблицу 7. Конфигурации пиков существенно изменились в силу сокращения длины текста к приемлемому размеру. 9 букв выходного алфавита отсутствуют. Различие гистограмм шифротекстов № 3 и № 5 усилилось.

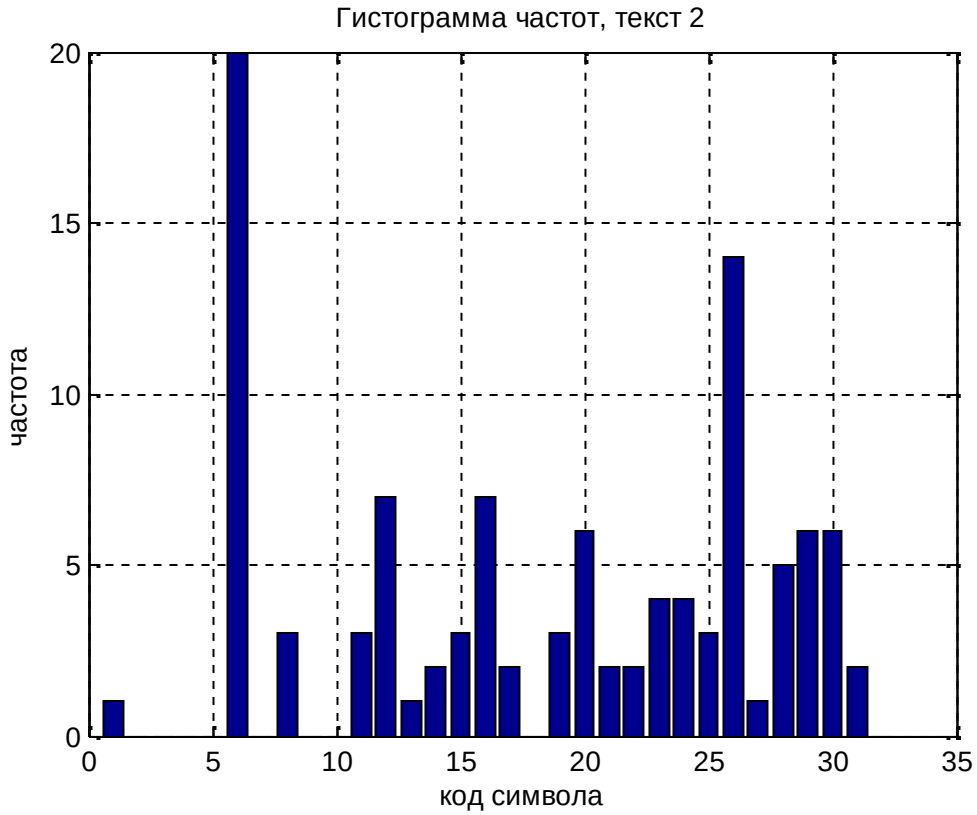
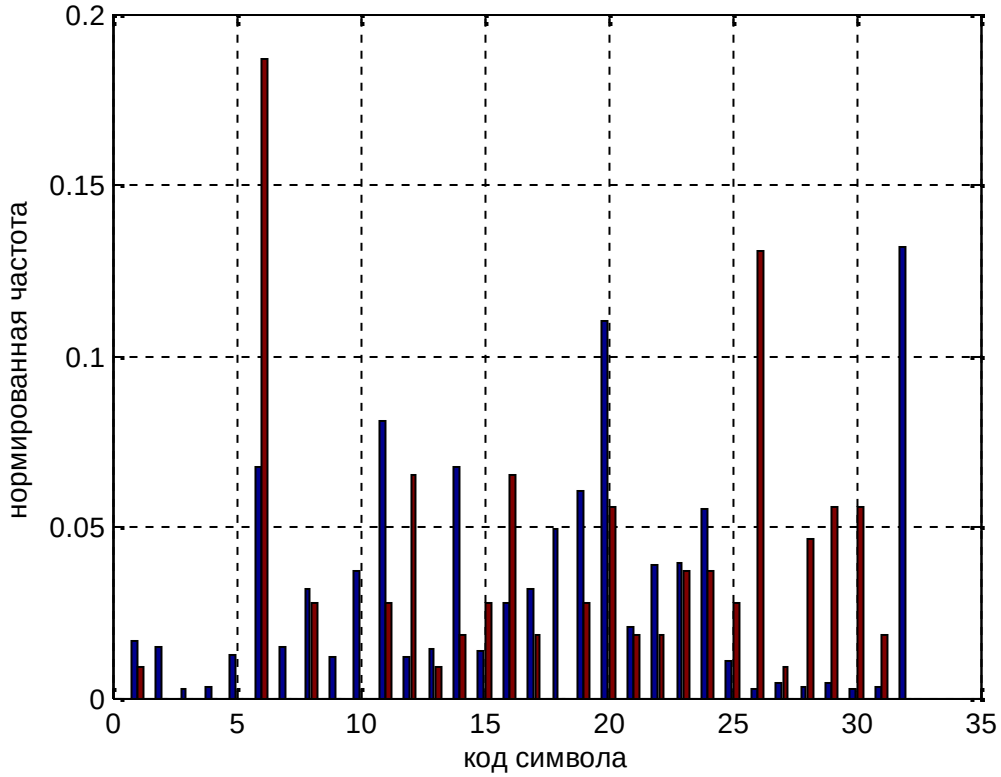


Рисунок 18 – Гистограмма частот букв в шифротексте № 5

Таблица 7 – Частоты букв в шифротексте № 5

А	1		Ж	0		М	1		Т	3		Ш	3		Ю	2
Б	0		З	3		Н	2		У	6		Щ	14		Я	0
В	0		И	0		О	3		Ф	2		Ъ	1			
Г	0		Й	0		П	7		Х	2		Ы	5			
Д	0		К	3		Р	2		Ц	4		Ь	6			
Е	20		Л	7		С	0		Ч	4		Э	6			

Две эти гистограммы для сравнения приведены на рисунке 19.



Синий - шифротекст № 3, коричневый – шифротекст № 5
Рисунок 19 – Нормированные гистограммы частот букв в двух шифрограммах № 3 и № 5

Для рисунка 19 Squ = 0.075503. Такое большое значение говорит, что два шифра совершенно разные. Эта характеристика возросла по сравнению с предыдущим примером на рисунке 11. При этом отметим, что корреляционная функция отчётливо обнаруживает сдвиг гистограммы шифротекста № 5 на 6 букв относительно гистограммы шифротекста № 3. Последнее демонстрирует рисунок 20.

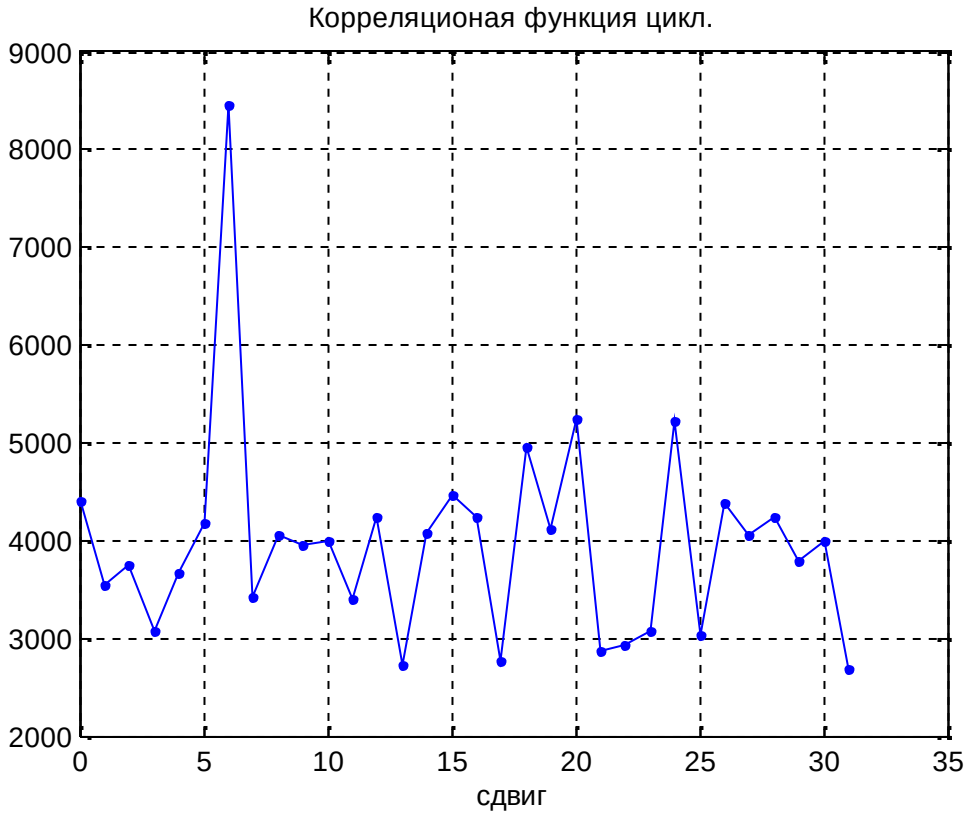


Рисунок 20 – Циклическая взаимная корреляционная функция из шифрограмм № 3 и № 5

Две шифрограммы с одним ключём

Два открытых текста № 3 и № 4 зашифруем с одним ключём, т.е. используем один шифр. Выбираем сдвиг алфавита равным 9.

Вариант 5. Шифрование открытого текста № 3. Шифр Цезаря, сдвиг k = 9 (ROT9), на выходе получается шифротекст № 6. Эта шифрограмма приведена на рисунке 21. Шифрограмма представляет собой одну строку символов. Для наглядности на рисунке 21, как и ранее, присутствуют знаки переноса.

ЛГЭЧЩХСЩЧЛЙЦССГЫСШСАЦЧТГКЧФЕБЧТГЪСЪЫОХДГШЩЧМЩЙХХЦЧМЧГЧ-
КОЪШОАОЦСИГЙЦЙФСЫСУГУЙУГШЩЙЛСФЧГНЧФПОЦГЩЙЪЪХЙЫЩСЛЙЫЕГ-
ЩИНГАОЫУЧГЧШЩОНОФОЦЩДЮГШЩОНХОЫЦДЮГЧКФЙЪЫОТ-
ГСФСГНЧХОЦЧЛГУЙПНДТГНЧХОЦГХЧПОЫГЩЙЪЪХЙЫЩСЛЙЫЕЪИГУЙУГЧЫНО-
ФЕЦДТГХСЩГЦЙЪОФОЦЦДТГЪЧКЪЫЛОЦЦДХСГУЧЦЯОШЫЬЙ-
ФЕЦДХСГЪЬВЦЧЪЫИХСГСФСГЧКЕОУЫЙХСГЪФОНЧЛЙЫОФЕЦЧГЛГЙЛЫЧХЙЫ-
СРСЩЧЛЙЦЦЧТГЪСЪЫОХОГЬШЩЙЛФОЦСИГПОФОР-
ЦЧТГНЧЩЧМЧТГНЧХОЦГЩЙКЧЫЙГПОФОРЦЧТГНЧЩЧМСГСХООЫГЧЫЦЧ-
БОЦСОГУГШЧОРНЙХГ-
НЧЩЧМЙХГСГЫГШГЛГЫЧГЛЩОХИГУЙУГНЧХОЦГШЧФЕРЧЛЙЫОФЕЪУСТГСЦЫО-
ЩЭОТЪГЦЙШЧФЦОЦГЧУЦЙХСГНСЪШФОИХСГСГШСУЫЧМЩЙХХЙХС-
ГУЙПНДТГНЧХОЦГХЧПОЫГЪЬВОЪЫЛЧЛЙЫЕГКЧФООГСФСГ-
ХОЦООГЦОРЙЛСЪСХЧГЧЫГНЩЬМСЮГПОФОР-
ЦЙИГНЧЩЧМЙГХЧПОЫГЪЬВОЪЫЛЧЛЙЫЕГКОРГЖУ-
ЩЙЦЧЛГСГЧУЧЦГЧУЦЙГСГШСУЫЧМЩЙХХДГХЧМЬЫГЪЬВОЪЫЛЧЛЙЫЕГКОРГ-
ШЧОРНЧЛГНЧХОЦДГСРЧКЩЙПЙЗЫЪИГЦЙГЪЮОХОГНЧХОЦЧЛГЦОУЧЫЧЩДО-
ГНЧХОЦДГНЧЛЧФЕЦЧГХЙФДГАЫЧ-
КДГХЧПЦЧГКДФЧГЙЦЙФСРСЩЧЛЙЫЕГСЮГУЙУГОНСЦЧОГЯОФЧО-
ГЛГЫЧГЛЩОХИГУЙУГНЩЬМСОГЪЧНОЩПЙЫГЫЙУГХЦЧМЧГЧ-

КЕОУЫЧЛГАЫЧГЫЩЬНЦЧГШЧННЙЗЫЪИГУЧЦЫЩЧФЗГСГЧКЩЙКЧЫУО-
ГШЧЖЫЧХЬГКЧФЕБСОГНЧХОЦДГЩЙРКСЛЙЗЫЪИГЦЙГШЧНЪСЪЫОХДГШЧЪФО-
ГЫЧМЧГУЙУГЪСЪЫОХЙГЩЙРК-
СЫЙГЦЙГНЧХОЦДГСГШЧНЪСЪЫОХДГХЧПЦЧГШЩЧЛЧН-
СЫЕГЙЦЙФСРГЦОРЙЛСЪСХЧГЧНЦЙГЧЫГНЩЬМЧТГУЙПНЙИГ-
ШЧНЪСЪЫОХЙГСФСГХЙФОЦЕ-
УСТГНЧХОЦГЙЦЙФСРСЩЬОЫЪИГЛГЫЩСГЖЫЙШЙГСЦЭЧЩХЙЯСЧЦЦЧОГХЧНОФ-
СЩЧЛЙЦСОГХЧНОФСЩЧЛЙЦСОГЪЧЪЫЧИЦСТГСГХЧНОФСЩЧЛЙЦСОГШЩЧЯОЪЪЧЛ
Рисунок 21 – Шифротекст № 6 открытого текста № 3

Индекс совпадений Is = 0.06313202, так как связан с открытым текстом № 3.

Вариант 6. Шифрование открытого текста № 4. Шифр Цезаря, сдвиг k = 9 (ROT9), на выходе получается шифротекст № 7. Эта шифрограмма приведена на рисунке 22.

ЮЧЫИГСГЪФЙНЧЪЫОЦГЙРЙЩЫГШЧГЪЩЙРЬГНЛЬХГСНЫСГНЧЩЧМЙХГЦОФЕ-
РИГЪГЧНЦЧТГУЧФЧНЧТГУЙЩЫГСМЩЙЫЕГСГЪГНЕИЛЧФЧХГСГЪГКЧМЧХ
Рисунок 22 – Шифротекст № 7 открытого текста № 4

Индекс совпадений Is = 0.07229765, так как связан с открытым текстом № 4.

Значения частот букв шифротекста № 6 совпадают с перечисленными частотами в таблице 3, но только переставлены. По аналогии значения частот букв шифротекста № 7 совпадают с перечисленными частотами в таблице 7, но опять же переставлены.

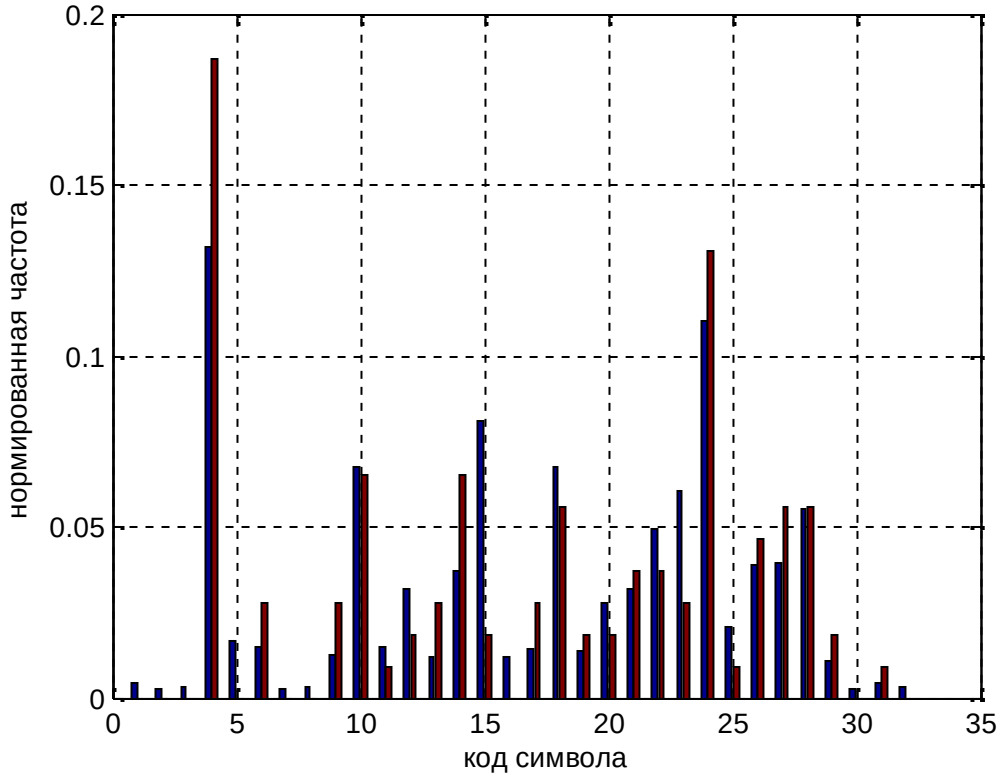
Таблица 8 – Частоты букв в шифротексте № 6

А	5		Ж	3		М	14		Т	16		Ш	25		Ю	5
Б	3		З	4		Н	44		У	33		Щ	46		Я	4
В	4		И	15		О	96		Ф	38		Ъ	47			
Г	157		Й	80		П	14		Х	59		Ы	66			
Д	20		К	18		Р	17		Ц	72		Ь	13			
Е	18		Л	38		С	80		Ч	131		Э	3			

Таблица 9 – Частоты букв в шифротексте № 7

А	0		Ж	0		М	3		Т	2		Ш	1		Ю	1
Б	0		З	0		Н	7		У	2		Щ	5		Я	0
В	0		И	3		О	2		Ф	4		Ъ	6			
Г	20		Й	7		П	0		Х	4		Ы	6			
Д	0		К	1		Р	3		Ц	3		Ь	2			
Е	3		Л	2		С	6		Ч	14		Э	0			

Две гистограммы шифротекстов № 6 и № 7 для сравнения приведены на рисунке 23. Значения частот заметно разнятся, но качественно гистограммы хорошо совпадают. Это следствие применения одного шифра.



Синий - шифротекст № 6, коричневый – шифротекст № 7
Рисунок 23 – Нормированные гистограммы частот букв в двух шифрограммах № 6 и № 7

В этом случае Squ = 0.011687. Такое малое значение говорит, что используются два одинаковых шифра, хотя исходные тексты сообщений разные.

Вывод 3. Суммарный квадрат разностей частот в двух гистограммах позволяет установить совпадение шифров одноалфавитной замены для разных сообщений.

Вывод 4. Суммарный квадрат разностей частот в двух отсортированных гистограммах можно использовать для сравнения шифров замены.

Корреляционная функция, приведенная на рисунке 24. уверенно подтверждает нулевой сдвиг шифралфавитов в шифротекстах № 6 и № 7.

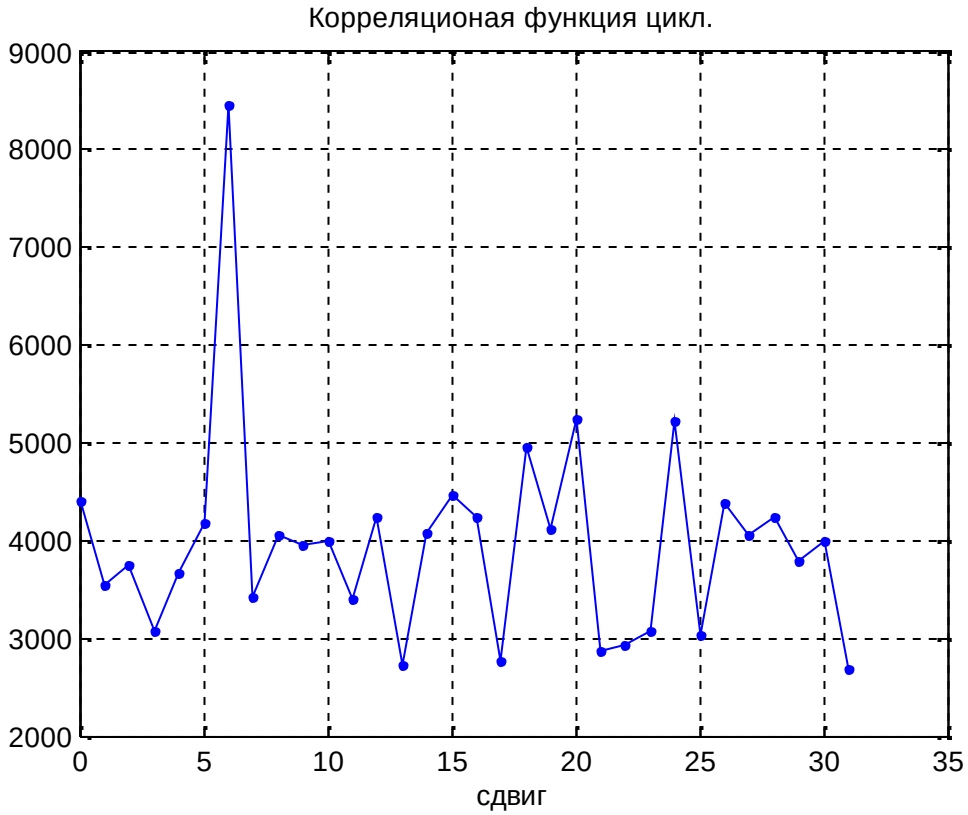


Рисунок 24 – Циклическая взаимная корреляционная функция из шифрограмм № 6 и № 7

ИНДЕКС СОВПАДЕНИЙ

Индекс совпадений используется в криптоанализе многоалфавитных шифров.

Определения и метод

Метод с использованием индекса совпадений находит применение в криптоанализе шифра Виженера. Метод разработан Уильямом Фридманом и опубликован в 1920 году.

Метод индекса совпадений предоставил новые возможности в криптоанализе шифра Виженера по сравнению с распространённым до него методом Касиски, Новый метод Фридмана работает с меньшей длиной шифротекста и чаще оказывается менее трудоёмким. Метод менее подвержен ошибкам и позволяет проводить анализ шифров с более длинными ключами. Использование индексов совпадений более эффективно с использованием компьютерной техники.

Индекс совпадений – вероятность того, что два случайных символа текста совпадут. Уильям Фридман показал, что эти вероятности существенно отличаются для текстов различной природы и языка в том числе. Вычисляя индексы совпадений сначала определяют длину ключа шифра, а затем и сам ключ.

Пусть имеем текст, написанный на некотором языке. Алфавит данного языка состоит из m символов. Строка текста достаточно длинная и состоит из n символов. Индекс совпадений – вероятность совпадения двух произвольных символов в этой строке. Если H – гистограмма частот символов исследуемого текста, то индекс совпадения вычисляется по формуле:

$$Is = {}_{i=1}\sum^{m} (H(i) \cdot (H(i) - 1)) /(n \cdot (n - 1)) . \quad (3)$$

Здесь суммирование (i) по всем карманам.

Индекс совпадений, вычисленный по формуле 3, не зависит от перестановки символов в строке и от их одноалфавитной замены (перестановки). Важна природа текста. Это показано на предыдущих примерах.

Индекс совпадений для открытого текста или его перестановки не зависит от его смыслового содержания, а зависит только от частот символов языка, на котором написан текст. Значения индексов совпадений для различных языков приведены в таблице 10.

Таблица 10 – Индексы совпадений для разных языков

Язык	Индекс совпадений
русский	0.0553
английский	0.0644 0.0667
итальянский	0.0738
испанский	0.0775
немецкий	0.0762
французский	0.0778
ведийский санскрит	0.021076696
пракрит	0.046635758
классический санскрит	0.045567736
хинди	0.041837864
урду	0.057535302

С другой стороны, если есть случайная строка с равномерным распределением используемых m символов, то индекс совпадений равен

$$Is = 1/m . \quad (4)$$

Многоалфавитный шифр Виженера с длинным ключём стремится создавать именно такие случайные строки шифротекста. Для английского языка индекс совпадений шифра Виженера будет стремиться к 0,03846 = 1/26, а для русского (без буквы «ё») – 0,03125 = 1/32.

Вывод 5. Посчитав индекс совпадений для текста, можно определить, получен ли этот текст из открытого простой перестановкой, или является многоалфавитным шифротекстом.

Пример

Для примера рассмотрим русский текст, из которого ранее получали открытый текст № 4. Это – открытый текст № 5. Он приведен на рисунке 25 и содержит 110 символов, но букв там 87. При этом не использовано 10 букв алфавита: Ж, Ф, Ц, Ч, Ш, Щ, Ъ, Ы, Э и Ю.

ХОТЯ И СЛАДОСТЕН АЗАРТ ПО СРАЗУ ДВУМ ИДТИ ДОРОГАМ, НЕЛЬЗЯ С ОДНОЙ КОЛОДОЙ КАРТ ИГРАТЬ И С ДЬЯВОЛОМ, И С БОГОМ.

Рисунок 25 – Открытый текст № 5

Численные значения частот букв этого текста сведены в таблицу 11. Этот набор значений частот немного отличается от значений в таблице 7.

Таблица 11 – Частоты букв в открытом тексте № 5

А	7		Ж	0		М	4		Т	6		Ш	0		Ю	0	

Б	1		З	3		Н	3		У	2		Щ	0		Я	3
В	2		И	6		О	14		Ф	0		Ъ	0			
Г	3		Й	2		П	1		Х	1		Ы	0			
Д	7		К	2		Р	5		Ц	0		Ь	3			
Е	2		Л	4		С	6		Ч	0		Э	0			

Индекс совпадений открытого текста № 5 Is = 0.05880781, что близко к стандартному значению 0.0553 из таблицы 10. Это явно подтверждает, что текст не является многоалфавитным шифротекстом. Более того, примеры, приведенные ранее, подтверждают это.

Шифр Виженера

Сообщение длиной 556 символов, содержащее 62 слова, зашифровано шифром Виженера. Это учебный пример под названием шифротекст № 8, шифрограмма которого представлена на рисунке 26.

ФЫЖЖЬЩБЗКЯЖОПЬИЯУЗЖДРУЛЯЪУЛЪЖЬПТХЛГЙЫКЪДКЗРЬШЫЖХЖУЪЖЪЛИЙ-
ТХСУТШДЯЧУОЯЧОЮУФЫЖЖЬЩБЗКЯЖОПЬИЯЯЗПЯЫЭГ-
ГЖЮПЙЪЩЗИЬНЮУВУТЗШНЮДТКЪЯЖЫЮИВУТЗШНЮДТКЪЧЩЪЮЗКЭЛЕЖЪОЕ-
НЫЮГЦШЩЬЖЬЖИЬРКТЖБЖЛЪЩАЕУЗПЯНШЮЙЭЫЩУТЗЮКЬРЛЙТЯ-
ЖБКБЖЯЖЬЖИЬРКТЖОГДПЫЮНТУЪЗКЬНЗППГВПШЖЦЖУЪКЩЫЮЩХР-
ЛЯЙЗИВИВЮГТЗИЗТЪРЕНЫЮЛТВГИФУЗУТШДЬЧРОУЩЫМЬФЭЖЗЭР-
РУЪРЮВТТМЩЕНЮЬЬЗРЬЫЭЖЗЭЕРРУТЗНЗШНГЗЙРРУФЫЖЖЬЩБЗКЯ-
ЖОПЬИЯПЗПЯЫЭГГЕЗАУШМЙЧЫЭШУТШРЬЪРПЕМЗИЗТЪРЕНЫЮЛТВГИФЩБЕ-
ЖУЛЭТШЖЗТШБЧЖНУЕОУРУФЫЖЖЬЩЮДКЦЖЮЖХОЯЩЭМЪЪЛ-
ТЯБРПБТАЪИТЬРЬЦЗПУАРЙСИЗНЗШНГЗФУЪЯЯЗСИЬЩЗОТНМИЬУЪЯЖШЮЫП-
СЛЕЫЭЖУЩЩЪЕЬШМППШЖХЖХЪЧЬЛИЧЦЗЮУЬЛИЭПЗНЕЫЭОЕ-
ПШЖЬЖХМДЬЫКЬЪЗНЗШЭЖЩЖШЖМ

Рисунок 26 – Шифротекст № 8

Отсортированная гистограмма шифротекста № 8 приведена на рисунке 27.

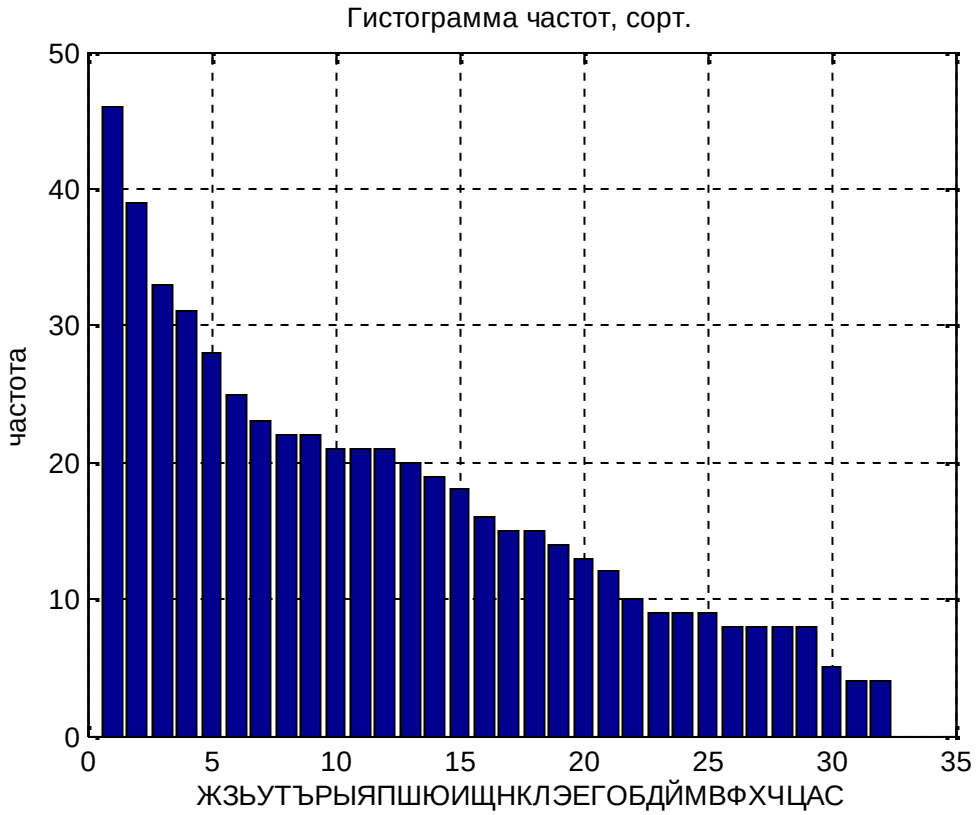


Рисунок 27 – Гистограмма частот букв в шифротексте № 8

Таблица 12 – Частоты букв в шифротексте № 8

А	4		Ж	46		М	9		Т	28		Ш	21		Ю	
Б	10		З	39		Н	18		У	31		Щ	19		Я	
В	8		И	20		О	12		Ф	8		Ъ	25			
Г	13		Й	9		П	21		Х	8		Ы	22			
Д	9		К	16		Р	23		Ц	5		Ь	33			
Е	14		Л	15		С	4		Ч	8		Э	15			

Индекс совпадений Is = 0.03991185

Алгоритм поиска длины ключа

Перебираем возможные длины ключей. Например, от 3 до 9.

1. Получаем прореженную строку из шифротекста. Для этого задаёмся длиной ключа и выбираем из шифротекста символы с шагом ключа, начиная с первого символа.

2. Вычисляем гистограмму частот символов в прореженной строке.

3. Вычисляем индекс совпадений из гистограммы.

Пункты 1, 2 и 3 повторяем в цикле. Где наибольший индекс совпадений, там и есть длина ключа.

Пример для шифротекста № 8

1. Пробная длина ключа = 3

Прореженная строка:

ФЖБЯПЯЖУЪЪПЛЫДРЫЖЖИХТЯООФЖБЯПЯПЭЖЙЗНВЗЮКЖИТНТЧЮЭЖЕЮШЖИК-
БЪЕПШЭУЮРТБЖЬЬТГЫТЗНППЦЪЫХЯИВТЗРЫТИЗШЧУМЭЭУЮТЕЬРЭЭУН-

НЙУЖЩКОИЗЫГАМЫУРРМЗРЫТИБУТЗБНОУЖЩКЮООЭЪЯПАТЬПРИЗГУЯИЗНЬЯЮ-
СЫУЪШПХЪЛЦУИЗЫЕЖХЬЬНЭЖМ

Индекс совпадений Is = 0.03894217 /дл.текста 186

2. Пробная длина ключа = 4

Прореженная строка:

ФЬКПУРЪЖХЫКШЖЪТТЧЧФЬКПЯЫЖЪЬВШТЖВШТЩКЖНЦЖЬЖЪУНЭТЬ-
ТКЖЬЖПТКППЖЩХЙИТТНТФТЧЩФЭЪТЕЬЫЭТШЙФЬКППЫЕШЫТЪМТНТФЖТТЖО-
ФЬКЖЩЪБТТЦАИШФЯЬТЬЖПЫЩЬПЖЬЦЬПЫПЖЬЪШЖ

Индекс совпадений Is = 0.06923157 /дл.текста 139

Длина ключа = 4 !!!

3. Пробная длина ключа = 5

Прореженная строка:

ФЩЖЯРУПЙКЫЪЙТУЮЖКЬПГЪНТДЖУЮЧКЪЮЬЬБАЯЭЗЛБЖРГНКПЖКХЗЮЗ-
НВЗЫЩЭРВЕЗЖУШРЖЗПЗГУЫШПЗНВБЭТНРЖКХМЯТЬПСШУ-
СОЬШЛУЬШЪЧЬЗОЬЬЗЖМ

Индекс совпадений Is = 0.04005792 /дл.текста 112

4. Пробная длина ключа = 6

Прореженная строка:

ФБПЖЪПЫРЖИТОФ-
БППЖЗВЮЖТТЮЖЮЖКЪПЭЮТЖЬГТНПЪХИТРТЗЧМЭЮЕРЭНЙЖ-
КИЫАЫРМРТБТБОЖКОЪПТПИГЯЗЬЮЫЪПЪЦИЫЖЬНЖ

Индекс совпадений Is = 0.05212716 /дл.текста 93

5. Пробная длина ключа = 7

Прореженная строка:

ФЗИУЖЙРУТЯЮЩПЯПНШЯТККЕЩРЪЯЩРКИГУПЦЮЗТЕГШЩЗЮНЫУГЫКЯГМ-
ТЕРВЖЗУЫКЯТАЦСГЗТЯЛЩПЧЮЗПДНШ

Индекс совпадений Is = 0.04050633 /дл.текста 80

6. Пробная длина ключа = 8

Прореженная строка:

ФКУЪХКЖТЧФКЯЖЬШЖШЩЖЦЬЪНТТЖЖТПЖХИТТТЩЭТЬ-
ЭШФКПЕЫЪТТЖТОЬЖЪТЦИФЬЬПЩПЬЬЫЖЪЖ

Индекс совпадений Is = 0.07784679 /дл.текста 70

7. Пробная длина ключа = 9

Прореженная строка:

ФЯЖЪЫЫИЯФЯПЙВКТЧЖШКЕЭРЖТТПЪЯТЫЗУЭТРУЙЩИГЫР-
РИТНЖЮЪАПЗЯНЮУПЛИЕЬЭ

Индекс совпадений Is = 0.03754627 /дл.текста 62

Взаимный индекс совпадений