Лабораторная работа № 1 по дисциплине «Основы криптографии»
Исследование статистических свойств текстов

Тексты для исследований можно взять в каталоге D:\LAB\Crypto\LB1\texts, где размещены литературные произведения различных писателей.
В каталоге D:\LAB\Crypto\LB1\texts\big находятся объёмные тексты, которые можно использовать в качестве источника статистики для русского языка.

1 Определить энтропию и избыточность различных открытых текстов

1.1 В меню «Статистика» задайте алфавит из 32 символов.

1.2 Выберите 3 произвольных (не слишком коротких) текста. Найдите для каждого удельную энтропию и коэффициент избыточности, сведя результаты в таблицу.

	Прибл. длина	h_1^+	R_1	h_{10}^+	R_{10}
Текст 1					
Текст 2					
Текст 3					

1.3 Выбрав один текст, найдите эти же величины для разных алфавитов

	h_1^+	R_1	h_{10}^+	R_{10}
256 символов				
'А'-'Я' и 'пробел'				
'А'-'Я' без 'Ё'				

1.4 Выбрав 32-символьный алфавит и один длинный текст (из каталога big) проведите подробное исследование, заполняя 3 следующие таблицы

n	1	2	3	4	5	6	7	8	9	10
h_n^+										
R_n										

1.5 посимвольная модель текста

символ										
частость										

1.6 модель биграмм

биграмма										
частость										

2 Определить энтропию и избыточность текстов, зашифрованных различными шифрами

2.1 Для выполнения следующих пунктов выберите 2 произвольных ключевых фразы без пробелов (или слова). Первая длиной 4-5 символов, вторая длиной 9-10 символов

Ключевая фраза 1	
Ключевая фраза 2	

2.2 Определите удельную энтропию и избыточность для одного произвольного

текста, зашифрованного различными шифрами. Ключи для шифра Цезаря и гамма-шифра можно выбрать произвольно

	h_1^+	R_1	h_{10}^+	R_{10}
Шифр Цезаря				
Шифр Виженера (кл. фр. 2)				
Гамма-шифр				

2.3 Сравните распределение частот появления символов и биграмм для каждого из шифров с результатами п. 1.5 и 1.6

3 Определить индекс соответствия для 2 различных открытых текстов, используя меню «Криптоанализ», «Тест Казиски-Фридмана»

Длина ключа n	1	2	3	4	5	6	7	8	9	10
Текст 1, $I_c(n)$										
Текст 2, $I_c(n)$										

4 Определить индекс соответствия для одного текста, зашифрованного с применением различных шифров и ключей (шифры Цезаря, Виженера и гаммирования соответственно)

Длина ключа n	1	2	3	4	5	6	7	8	9	10
Шифр 1, $I_c(n)$										
Ш.2 (кл. фр. 1), $I_c(n)$										
Ш.2 (кл. фр. 2), $I_c(n)$										
Шифр 3, $I_c(n)$										

5 Определение истинного ключа

5.1 Используя один из длинных текстов получите распределение частот появления для отдельных символов (аналогично п. 1.5)

5.2 Загрузите произвольный текст, отличный от используемого в предыдущем пункте и зашифруйте его гамма-шифром с произвольно выбранным ключом.

5.3 Определить возможность однозначного определения ключа методом полного перебора всех ключей, с использованием критерия Байеса для текста, полученного в п. 5.2, задавая различную длину анализируемого текста (Меню «Криптоанализ», «Гамма-шифр Тест 1»). Для каждой из длин запишите в таблицу количество «верных» ключей, Байесовский порог, а также логарифмы вероятностей наиболее вероятного ключа и ключа следующего по вероятности («рейтинги»).

Длина	40	50	60	70	80	90	100	110	120
Принято ключей									
Порог									
$\log p(K_1)$									
$\log p(K_2)$									

По каждому из пунктов работы сделать выводы, а по таблице из п. 1.4 построить график.