Міністерство освіти і науки України
Національний технічний університет України
“Київський політехнічний інститут ім. Ігоря Сікорського”
Фізико-технічний інститут

КРИПТОГРАФІЯ
КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1
Криптоаналіз шифру Віженера

Варіант 11

Виконали:
Студенти групи ФБ-05
Алькова Аліна, Супрун Максим

Київ – 2022

Мета роботи

Засвоєння методів частотного криптоаналізу. Здобуття навичок роботи та аналізу потокових шифрів гамування адитивного типу на прикладі шифру Віженера.

Порядок виконання роботи

0. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
1. Самостійно підібрати текст для шифрування (2-3 кб) та ключі довжини r = 2, 3, 4, 5, а також довжини 10-20 знаків. Зашифрувати обраний відкритий текст шифром Віженера з цими ключами.
2. Підрахувати індекси відповідності для відкритого тексту та всіх одержаних шифртекстів і порівняти їх значення.
3. Використовуючи наведені теоретичні відомості, розшифрувати наданий шифртекст (згідно свого номеру варіанта).

Хід роботи

`Part 1-2.`

Для першої частини в якості експериментального тексту був взятий уривок з «лист Тетяни до Онєгіна» на російській мові. Оригінал тексту знаходиться у файлі *badtext.txt*. Функцією-редактором текст був очищений від зайвих символів, пробілів та літери ё. На виході отримали файл *goodtext.txt*, з яким працюємо надалі.

Попередньо були згенеровані ключі шифрування довжини 2-20. Функція шифрування Віженера схожа в цілому по роботі на шифрування Цезаря. Для відкритого тексту та кожного шифртексту був обрахований індекс відповідності за формулою:

$$I(Y) = \frac{1}{n(n-1)} \sum_{t \in Z_m} N_t(Y)(N_t(Y)-1),$$

У роботі також використовується функція обрахунку кількості зустріваності літер з попередньої роботи. Результати містяться у файлі *part1.xlsx*, а також наводимо їх нижче:

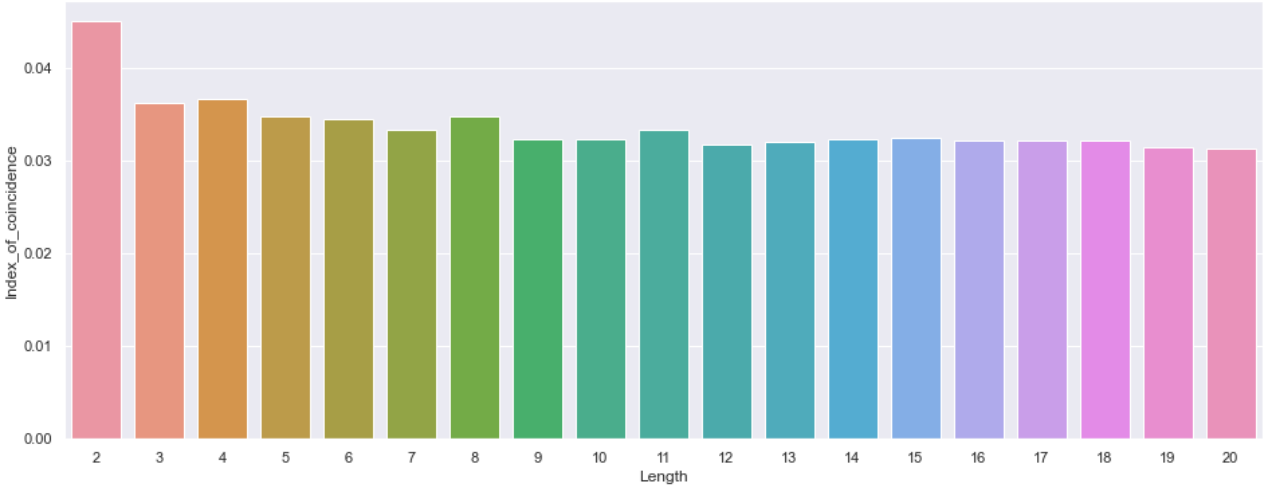
Length	Key	Index_of_coincidence
2	жг	0.04502014882570707
3	ишб	0.036282783740903234
4	зпеф	0.03666763500209806
5	юмыех	0.034721032493861
6	пжмзяг	0.03442804895308042
7	йогучмф	0.03328839263767123
8	сгряуцвш	0.03477565654383704
9	ахпатлдгп	0.032280330624477035
10	жячкьевщпю	0.03227784771311449
11	этбссндчмй	0.03334301668764727
12	рщалцйзьпохц	0.03168691480882824
13	ноинрсшдайупв	0.03207673189274816
14	ххдсынвдовзэгх	0.03225550151085156
15	йчшхвклябютмюйы	0.03249137809029355
16	чйцжоъоэнавгжвюб	0.03211149265182382
17	вевуюргначитвшдьж	0.03224060404267628
18	жзфкостеьлщйцэтсна	0.03218846290406279
19	рмыньквзжпюэчаднудс	0.0314535211407488
20	эбокрэижтжъебътчодиг	0.03127226861128284

Та для відкритого тексту `Coincidence index of all text: 0.05361102214012062`

Для порівняння індекс відповідності російської мови з Вікіпедії:

Язык	Индекс совпадений
русский	0.0553[1]

Якщо зобразимо у вигляді гістрограми то маємо таке:



Отже, відповідно до твердження з методичних вказівок «*Однак, якщо Y є шифртекстом, одержаним в результаті роботи шифру Віженера, то величина індексу відповідності та його математичне очікування буде стрімко падати із ростом довжини ключа r*», можемо спостерігати те саме.

Part 3.

У третій частині виконуємо варіант 11. Зашифрований текст можна знайти у файлі *task3.txt*. Для знаходження ключа шифрування скористуємось таким алгоритмом:

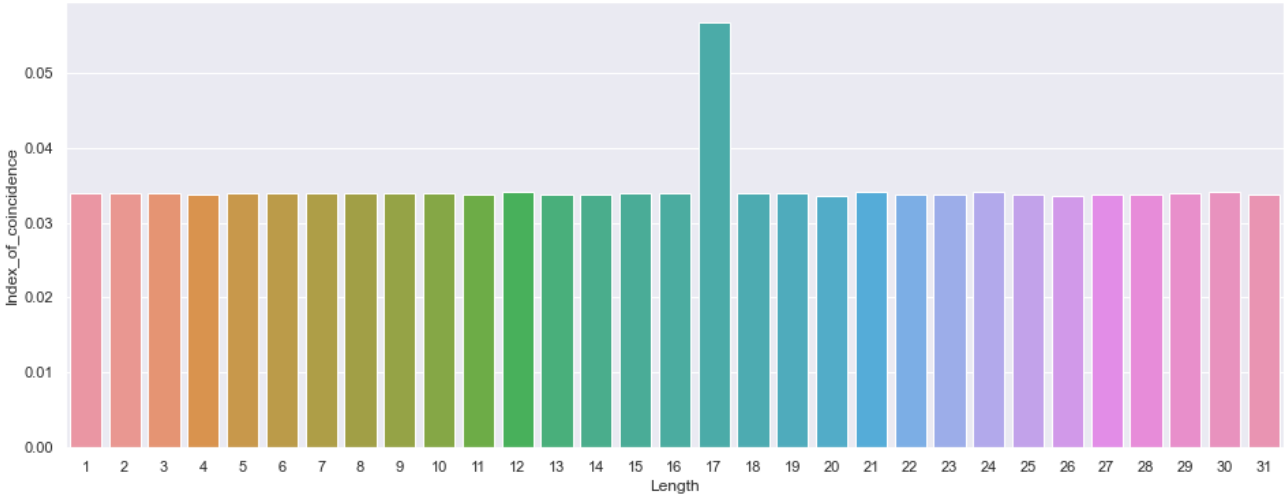
1) Для кожного кандидата $r = 2,3,...$ розбити шифртекст Y на блоки $Y_1, Y_2, ..., Y_r$.
2) Обчислити значення індексу відповідності для кожного блоку.
3) Якщо сукупність одержаних значень схиляється до теоретичного значення I для даної мови, то значення r вгадане вірне. Якщо сукупність значень схиляється до значення $I_0 = \frac{1}{m}$, що відповідає мові із рівноімовірним алфавітом, то значення r вгадане неправильно.

Будемо розбивати на блоки де r від 1 до 31 та обчислювати таким чином індекс відповідності. В результаті отримуємо значення, наведені у файлі *part3.xlsx*, а також нижче:

Key_Length	Index_of_coincidence
1	0.03389666144916544
2	0.03385469736120976
3	0.0338903572127828

4	0.03382856631255099
5	0.0339485646135541
6	0.03393201950461585
7	0.033887879672849
8	0.03387028521782783
9	0.033925669332816634
10	0.03389699079863339
11	0.03369705663706742
12	0.03402936176339339
13	0.033824971570802616
14	0.03374056874966326
15	0.03389000107283445
16	0.0340028030190302
17	0.056652871546688466
18	0.033914675695348
19	0.033895156796394856
20	0.033573302104999446
21	0.034174812268371026
22	0.03365396493285553
23	0.033783217934845344
24	0.03403767616107533
25	0.033792784827267594
26	0.033637618817285685
27	0.033791041307072986
28	0.033759700080605394
29	0.03399770139487165
30	0.03408868605919016
31	0.03379225398070093

Та гістрограма:



Бачимо, що найближче значення до значення індексу всієї мови у ключа довжини 17. Отже це наше шукане значення.

Спробуємо знайти за іншим алгоритмом:

Другий алгоритм використовує інший підхід.

1) Одержати оцінки індексу відповідності I_r для шифртекстів, що були зашифровані шифром Віженера із різними періодами r ($r \geq 2$).
2) Обчислити індекс відповідності даного шифтексту.
3) Порівнюючи обчислене значення із індексами I_r, зробити висновок щодо довжини ключа.

Для шифртексту отримали індекс відповідності: *0.03389666144916544*, але як і було сказано у методичних вказівках, дуже важко визначити однозначно довжину ключа для великого періоду, тому таких алгоритм не дуже ефективний.

Тепер знаючи довжину ключа, шукаємо сам ключ за допомогою функції у якій використовується формула $k = (y^* - x^*) \bmod m$,

Тобто можна сказати тут також застосовується частотний аналіз.

Попередньо наведемо найчастіші літери російської мови, які були визначені у минулій роботі.

о	0,110096
а	0,085067
е	0,08224
н	0,067052
и	0,066847
т	0,059604
с	0,053082
л	0,051947
р	0,045822
в	0,041994

Тоді на виході отримали такий ключ: венецианскийкужец

Враховуючи, що ключ має бути змістовний дуже легко здогадатись що фінальним результатом має бути «венецианскийкупец» як і назва твору. Проте якщо так сталося, що дуже важко зрозуміти, який же все таки змістовний ключ можна зробити так:

```
Finding the length of key ...
17
венецианскийкужец
One of possible keys was found, do you want to continue searching?
>>y
руыудцоыяшцчшбфуд
One of possible keys was found, do you want to continue searching?
>>y
лоцоясйцъустуьпря
One of possible keys was found, do you want to continue searching?
>>
```

Бачимо, що через декілька шагів функція вгадує і букву якої не вистачало.

Розшифруємо текст, з таким ключем і отримали уривок п’єси «Венеціанський купець», який знаходиться у файлі *task3_answer.txt*

Антонионезнаюотчегоятакпечаленмнеэтовтягостьвамяслышутоженогдеягрустьпоймал нашелильдобылчтосоставляетчтородитеехотелбызнатьбессмысленнаягрустьмоявиною чтосамогосебяузнатьмнетрудносаларионовыдухоммечетесьпоокеанугдевашивеличавыесу дакакбогатеишвельможиводильпышнаяпроцессияморскаяспрезреньемсмотрятнаторговц евмелкихчтокланяютсянизкоимспочтеньемкогдаонилетятнатканыхкрыльяхсаланиоповер ьтееслибятакрисковалпочтивсечувствабылибтаммоисмоейнадеждойябыпостоянносрыв алтравучтобзнатьоткудаветерискалнакартахгавани ……

Висновки:

Отже, під час роботи над цим практичним завданням, ми дослідили методи шифрування та розшифрування методом шифру Віженера. Було вивчене нове поняття індексу відповідності та алгоритми знаходження ключа за допомогою нього. Крім того побачили, що деякі алгоритми не завжди підходять для ключів більшого періоду. Також був застосований метод частотного аналізу для дешифрування повідомлення. Набуті навички знадобляться у подальшій роботі над практичними завданнями.