МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ

імені ІГОРЯ СІКОРСЬКОГО»

Фізико-технічний інститут

КРИПТОГРАФІЯ

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №2

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Варіант №11

Виконали:

ФБ-32 Пінькас Б. О.

ФБ-32 Драчук О. І.

Київ 2025

Мета роботи

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

Порядок виконання роботи

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

У якості довільного тексту російською мовою достатньої довжини, було обрано “Зелений шум” Миколи Олексійовича Некрасова (Nekrasov_Zelenyu_shum.txt, довжини 3 Кб)

Для підрахунку частот індексів відповідності ВТ та для всіх одержаних ШТ, а також для знаходження довжини ключа, самого ключа, та розшифрування наданого шифротексту згідно з номером варіанту, було написано скрипт lab2.py. Всі вхідні файли зберігаються в папці input (довільний текст, ШТ з варіанта), а вихідні файли - в папці output (весь вміст формується скриптом lab2.py). Перед безпосереднім запуском скрипта, рекомендується встановити необхідні бібліотеки (pip install -r requirements.txt)

На початку, скрипт фільтрує ВТ від зайвих символів. Всі символи, окрім літер російського алфавіту вилучено, а літеру “ё” замінено на “е”.

Зашифруємо відфільтрований ВТ (Nekrasov_Zelenyu_shum.txt) ключами зі списку різної довжини

```
ВТ: николайнекрасовзеленыйшумидетгудетзеленыйшумзеленыйшумвесеннийшумиграючирасходит...

К: он
ШТ: ыхшыщнчъучюняырфушуъйцжаъхттарбсуяхтщтыичебщхтщтыичебщртятыъццжаъхсэолехюнявьсця...

К: код
ШТ: чцошщдуыйфюдыьжсуппыяужчцципазэтйьхйхусечьэълпщйчйнвбрмухпыстчьэъмнюдиемъохяьита...

К: свет
ШТ: юкпаьвояцмхтврзщцнкямлэеэкйчгешццфмчьзтнъъшюшзрчюэокдозчвзтящлэеэкивсаъъбвцзяжнд...
```

...

```
К: ловичудесныймомент
ШТ: шцмцвунтцчлйэьомтэрыэспжрнхтнмятсчфчцупгалчсштжощйхэаюнуундамойазспюмгдъыоуэечмч...

К: пустьденьбудетярким
ШТ: ыыыаздоъблгдцабчпусьоъкпрнсбуцчйчсчпусьоъкпрмтзжаяоктьмнэфаюъеьшщддгдгйзакщбэчщд...

К: пустьденьбудеттеплым
ШТ: ыыыаздоъблгдцафмфцащкьйеимйтоджикдщкъризшлдюгйртйьььшюфкарищчьйеимиэьякмхтгъэпгю...
```

Всі ключі та відповідні ШТ можна переглянути у папці output в файлі encrypted_Nekrasov_Zelenyu_shum.txt

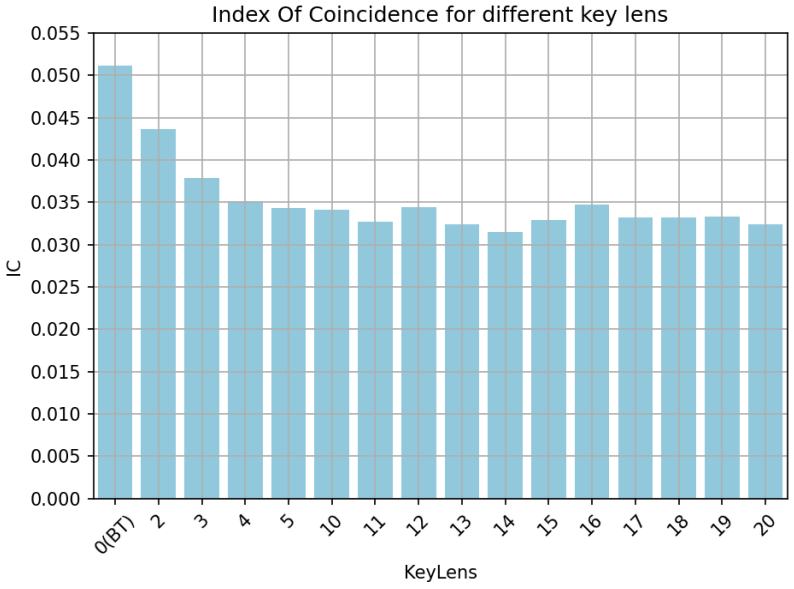
Потім, для ВТ і отриманих ШТ, ми знаходимо індекси відповідності за формулою

$I(Y) = \frac{1}{n(n-1)} \sum_{t \in Z_m} N_t(Y)(N_t(Y) - 1)$, де – $N_t(Y)$ к-сть появ букви t у шифртексті Y

```
KeyLens        IC
  0(BT) 0.051149
      2 0.043606
      3 0.037908
      4 0.035117
      5 0.034285
     10 0.034172
     11 0.032741
     12 0.034378
     13 0.032413
     14 0.031517
     15 0.032913
     16 0.034756
     17 0.033259
     18 0.033238
     19 0.033291
     20 0.032449
```

KeyLens	IC
0(BT)	0,051149
2	0,043606
3	0,037908
4	0,035117
5	0,034285
10	0,034172
11	0,032741
12	0,034378
13	0,032413
14	0,031517
15	0,032913
16	0,034756
17	0,033259
18	0,033238
19	0,033291
20	0,032449

Для порівняння отриманих значень, побудуємо діаграму на їх основі



Отримані значення зберігаються у вигляді таблиць в файл ICByKeyLens.xlsx та у вигляді діаграми в файл ICByKeyLens.png

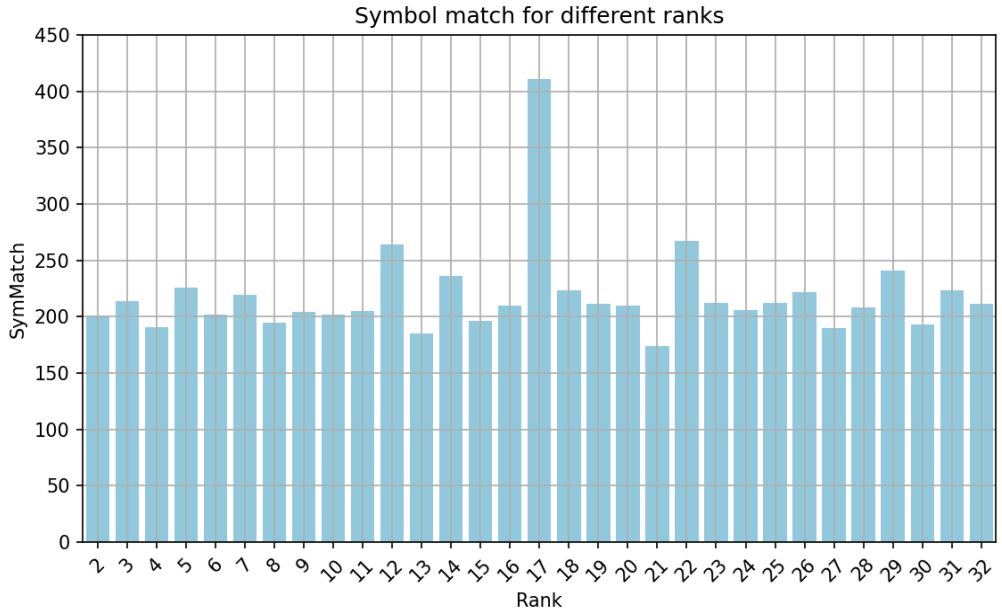
Для знаходження довжини ключа ми обчислювали статистику співпадінь символів:

$D_r = \sum_{i=1}^{n-r} \delta(y_i, y_{i+r})$, де $\delta(a, b)$ - символ Кронекера

Так, для r, що кратні істинному періоду, значення D_r будуть істотно більшими за інші одержані значення

Rank	SymMatch
2	201
3	214
4	191
5	226
6	202
7	219
8	195
9	204
10	202
11	205
12	264
13	185
14	236
15	196
16	210
17	411
18	223
19	211
20	210
21	174
22	267
23	212
24	206
25	212
26	222
27	190
28	208
29	241
30	193
31	223
32	211

Для візуального порівняння, побудуємо діаграму:



Таблиці - SymMatchByRank.xlsx, діаграма - SymMatchByRank.png

Бачимо, що найбільше значення D_r маємо при $r = 17$, отже довжина нашого ключа - 17 символів

```
Знайдено довжину ключа: 17
```

Далі, знайдемо наш ключ методом частотного аналізу. Розбиваємо наш текст на 17 фрагментів (відповідає довжині ключа). Кожен фрагмент, відповідно, зашифрований одним символом (частиною нашого ключа шифру Віженера). Задача зводиться до пошуку ключа шифру Цезаря, знайдемо k_i , $i = \overline{1,r}$; за формулою $k = (y^* - x^*) mod\, m$, де y^* – буква, що частіше за всіх зустрічається у фрагменті Y_i, а – x^* найімовірніша буква у мові, якою написано відкритий текст.

Так як фрагменти у нас невеликого розміру, найчастішою літерою не завжди буде найчастіша літера в мові, тому, для повноти експерименту, візьмемо 3 найчастіші літери мови і для них знайдемо ключ шифру Віженера

```
Ключ (найчастіша літера - о): венецианскийкужьц
Ключ (найчастіша літера - е): лоцоясйцъустуьпея
Ключ (найчастіша літера - а): руыудцоыяшцчшбфкд
Отримали ключ: венецианскийкупец
```

Бачимо, що для перших літер ключа, в нас найчастіша літера в фрагментах дійсно буде “о”, однак, для 15 і 16 фрагменту в нас найчастішою виявилась літера “е”. Таке припущення ми робимо на основі того факту, що наш ключ змістовний, і отримуємо ключ “венецианскийкупец”

Для знайденого ключа розшифруємо ШТ за варіантом:

```
ШТ: втяугроъцсхйиббыеумчцтптикуочяькуфупчхлоюгжкйцтарсъшяуьннъфонингвциюфыовилъсвнфтюлйдгашьицсывьилхтфчнфуэуърт
тцяцыпюраэпеябчнсюэещфпаьехехацидмырмрцшсжчдуещущцсттйырчубббвпкяхймнывкуйъыьушэйаъдфмтипъоыпюудмкнтйлдтукасмшьннв
зикзыдныкткшцпчыкнпкбдмычткчоыьбеээъехчрызпщьттыужупндзчртшьнцжшыцврчэдихаяяьлчмйфзвзрчнлятыыхицйсбцхпнфпдрмюашяы
палквмурйццнхьпъиьапчавтиъашышнйэъкюптюрфызышьяцпщфтфочцмххцацвньщцаыысцыщшпцикаомхркьуысдкцщуысннхпоншьожссуючдз
нъяышдмуъчжвзаьицбфюкъешещшьвзтчышюыкуцкэпхивърешинхщлыюьоъгчроьхыммтгбъчцбтжспкайцяущюпчщпчскпвчйсыхяомчнъшяькг
пупижысянщцлпгтебуешежрнывыынйяэозхфсалинйццзлхыдужвйчкчгдэярифшеыазнндчдфоуцькхшгфшжвинтгидтькъечшыущгапнъянтйрби
ъшхюкзрьъалхепвщцхчысэюрстрхэиыбтьйявякъучнзюубиышшйлюлзезцчкэивмшврхнюпзйупшугрвещцхсршжквгученъоозпучмуббздулсд
лишдмюоъэснзоуяхххачсцхссчптюбцпдицгыыктхшцрахпкпцецмыщьдъфуъуевцъалятыжъышфышсдлпыхцйлйцокйьбъпгхзпцычрмюшщытгпц
зэфнрюйыпушмътхэргэуорытлхтмфчтлфравтацбцвыэбъчцбфждееяцикоюгкуччыжквксыибрбмялеышяушввчйтымущсйчщте...
ВТ: антонионезнаютчегоятакпечаленмнеэтовтягостьвамяслышутоженогдеягрустьпоймалнашелильдобылчтосоставляетчтородит
еехотелбызнатьбессмысленнаягрустьмоявиноючтосамогосебяузнатьмнетрудносалариновыдухоммечетесьпоокеанугдевашивелича
выесудакакбогатеиивельможиводильпышнаяпроцессияморскаяспрезреньемсмотрятнаторговцевмелкихчтокланяютсянизкоимспочт
еньемкогдаонилетятнаткаыхкрыльяхсаланиоповерьтееслибятакрисковалпочтивсечувствабылибтаммоисмоейнадеждойябыпостоя
нносрывалтравучтобзнатьоткудаветерискалнакартахгаваниибухтылюбойпредметчтомогбынеудачумнепредвещатьменябынесомнен
новгрустьповергалсалариностудямойсупдыханьемявлихорадкебыдрожалотмысличтоможетвмореураганнаделатьнемогбывидетьяча
совпесочныхневспомнившиомеляхиорифахпредставилбыкорабльвпескезавязшимглавусклонившимнижечембокачтобцеловатьсвоюмо
гилувцерквисмотрянакамнизданиясвятогокакмогбыяневспомнитьскалопасныхчтохрупкиймойкорабльедватолкнуввсепряностирас
сыпалибывводуиволныоблеклибвмоишелканусловомчтомоебогатствосталоничемимоглибяобэтомдуматьнедумаяприт...
```

Повний текст збережено в файл decrypted_var11.txt

Висновки:

У ході даного комп’ютерного практикуму, було експериментально досліджено знаходження ключа шифру Віженера на ШТ. Для цього, ми шукали індекси відповідності для різних довжин ключа, визначати довжину ключа, базуючись тільки на індексах відповідності - погана ідея, тому що для великих довжин ключа, точність обрахунків значно падає.

Тому ми обчислювали статистику співпадінь символів з різними періодами на ШТ, для якого ми хочемо встановити ключ. Період означає, що ми значення y_i порівнюємо значенням y_{i+r}, і якщо вони однакові, то D_r зростає на 1. Так, значення D_r будуть найбільші, для періодів r, що кратні довжині ключа. Порівнявши значення D_r для різних періодів, ми встановили, що довжина нашого ключа - 17 символів.

Далі, щоб знайти ключ, ми розбили текст на 17 фрагментів, кожен фрагмент відповідає одному символу ключа. Тобто, фактично, ми отримали 17 фрагментів, зашифрованих шифром Цезаря. Для знаходження ключа для кожного шифру Цезаря ми використали метод частотного аналізу, який полягає в тому, щоб порівняти найчастіший символ в шифротексті y* з найчастішим символом у мові x*, і для цих значень обчислити k.

Оскільки фрагменти невеликого розміру, у відповідному ВТ для цих фрагментів не завжди найчастішим символом буде найчастіший символ у мові. Тому, ми знайшли ключі для трьох найчастіших символів у мові, і, співставивши їх, враховуючи природу мови, визначили ключ, який має бути змістовним, тобто відповідати словам або словосполученням з мови. Отриманий ключ “венецианскийкупец” відповідає словосполученню “венецианский купец” в російській мові.

Отже, порівняння статистики співпадінь символів для знаходження довжини ключа, а також розбиття ШТ на фрагменти, к-сть яких відповідає довжині ключа, тобто, для кожного фрагменту ми знаходимо один символ ключа, методом частотного аналізу шифру Цезаря, є ефективними методами, за допомогою яких вдалось розшифрувати ШТ, зашифрований невідомим ключем шифру Віженера.