

**Міністерство освіти і науки України
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського"
Фізико-технічний інститут**

**КРИПТОГРАФІЯ
КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1**

**Виконала:
студентка
групи ФБ-13,
Буєва Христина.**

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Мета роботи : засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

Порядок виконання роботи

0. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H_1 та H_2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H_1 та H_2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилючено всі пробіли.
2. За допомогою програми CoolPinkProgram оцінити значення $H(10)$, $H(20)$, $H(30)$.
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Обраховані значення ентропії :

```
"D:\KPI\5 семестр\labs_crypto\venv\Scripts\python.exe" "D:\KPI\5 семестр\labs_crypto\lab1.py"
H1 для тексту з пробілами: 4.3982997701869255
H2 для тексту з пробілам: 7.966844867073893

Process finished with exit code 0
```

```
"D:\KPI\5 семестр\labs_crypto\venv\Scripts\python.exe" "D:\KPI\5 семестр\labs_crypto\lab1.py"
H1 для тексту без пробілів: 4.466116009991942
H2 для тексту без пробілів: 8.273234773195535

Process finished with exit code 0
```

```
"D:\KPI\5 семестр\labs_crypto\venv\Scripts\python.exe" "D:\KPI\5 семестр\labs_crypto\lab1.py"
H2 (пари букв не перетинаються) для тексту з пробілам: 7.968139204575967

Process finished with exit code 0
```

```
"D:\KPI\5 семестр\labs_crypto\venv\Scripts\python.exe" "D:\KPI\5 семестр\labs_crypto\lab1.py"
H2 (пари букв не перетинаються) для тексту без пробілів: 8.272650318626672

Process finished with exit code 0
```

Частоти букв (текст з пробілами) :

ъ	0.00015231836312475946	у	0.022500231989739952
ф	0.0015867728508045331	п	0.024354671090307412
щ	0.0029339770139759492	д	0.026232771295826292
э	0.0030167908813059155	м	0.027065346426304347
ц	0.003342131074387926	к	0.027763349022371205
ю	0.006066115781920032	л	0.035991498269227144
ш	0.006191815401974445	р	0.03636120303409306
х	0.007703168480746331	в	0.03829254072575264
ж	0.008596375192662396	с	0.04484666679729568
й	0.008662922050338261	н	0.05368113185853173
б	0.012927836217831527	т	0.05962006920133789
з	0.013757453710190654	а	0.06270044930220074
г	0.013841746396580083	и	0.06348126576559757
ч	0.015222963398119163	е	0.07379159224817837
ь	0.017824206123715783	о	0.09069893055502667
я	0.018048986620754263		
ы	0.018053423077932656		

Частоти букв (текст без пробілів) :

ъ: 0.00018017466446841913

ф: 0.0018769651939282888

щ: 0.0034705488767509084

э: 0.003568507917626942

ц: 0.003953347006782789

ю: 0.007175499744169469

ш: 0.007324187574070592

х: 0.009111940070058206

ж: 0.010168498296649713

й: 0.010247215383067955

б: 0.015292105988183691

з: 0.0162734456655311

г: 0.016373153974994205

ч: 0.01800697083531948

ь: 0.021083934279979184

я: 0.021349823105214133

ы: 0.021355070910975348

у: 0.026615121552300943

п: 0.02880870436048927

д: 0.031030275466070747

м: 0.0320151136805923

к: 0.0328407684536903

л: 0.04257369887215908

р: 0.04301101601892709

в: 0.04529556079364316

с: 0.0530483191715464

н: 0.06349844971071471

т: 0.07052351235639598

а: 0.07416723882326702

и: 0.07509085263724105

е: 0.08728675322630725

о: 0.10728614098230177

Частоти біграм (топ 10), що перетинаються (текст з пробілами):

ст: 0.01203130328515831

а_: 0.012645296259007387

_н: 0.013395403337902526

то: 0.013793389145048072

_в: 0.014535728842948735

_с: 0.015682712103133066

_п: 0.017893086808989746

и_: 0.01791232032383321

е_: 0.018440502231457597

о_: 0.018656509398161127

Частоти біграм (топ 10), що перетинаються (текст без пробілів):

по: 0.01012733961838351

ть: 0.010267292489254244

пр: 0.010428238290755587

но: 0.01106502385321743

ов: 0.011262707283322342

на: 0.011498877752916706

ни: 0.011812022301489974

ен: 0.012025450429567843

ст: 0.014390653947283252

то: 0.01688531387055409

Частоти біграм (топ 10), що не перетинаються (текст з пробілами):

ст: 0.011693977024826769

а_: 0.012655652767000022

_н: 0.013368772317350036

то: 0.013851089689578468

$$H(20) = 2,197; R = 0,123$$

Лабораторная работа №1

×

Произвольная часть текста:

е_внимание_на_два_п

Использованные буквы:

Порядок n-граммы:

5 символов

10 символов

15 символов

20 символов

25 символов

30 символов

35 символов

40 символов

45 символов

50 символов

Введенный символ:

Символ по счету:

Номер эксперимента:

51

Неравенство для энтропии:

1,88705974972532 < H < 2,50622721260073

Двоичная таблица угаданных символов:

00000000000100000000000000000000

00000000000000000000000000000000

00000000000000000000000000000000

10000000000000000000000000000000

10000000000000000000000000000000

00000000000000000000000000000000

Поле ввода символов:

Продолжить

Другой

Вероятности:

q[1] = 0,58

q[2] = 0,08

q[3] = 0,02

q[4] = 0

q[5] = 0

q[6] = 0,02

q[7] = 0,02

q[8] = 0,02

q[9] = 0

q[10] = 0

q[11] = 0,04

q[12] = 0,02

q[13] = 0,04

q[14] = 0,02

q[15] = 0

q[16] = 0

q[17] = 0,02

q[18] = 0,02

q[19] = 0

q[20] = 0

q[21] = 0,02

q[22] = 0

q[23] = 0

q[24] = 0

q[25] = 0,04

q[26] = 0,04

q[27] = 0

q[28] = 0

q[29] = 0

q[30] = 0

q[31] = 0

q[32] = 0

Строка состояния:

$$H(30) = 1,935; R = 0,109$$

Лабораторная работа №1

×

Произвольная часть текста:

ицей_однако_что_физическое_те

Использованные буквы:

Порядок n-граммы:

5 символов

10 символов

15 символов

20 символов

25 символов

30 символов

35 символов

40 символов

45 символов

50 символов

Введенный символ:

Символ по счету:

Номер эксперимента:

51

Неравенство для энтропии:

1,69993140313015 < H < 2,17243127287761

Двоичная таблица угаданных символов:

00000000000000000000000000000000

10000000000000000000000000000000

00000000000000000000000000000000

10000000000000000000000000000000

01000000000000000000000000000000

Поле ввода символов:

Продолжить

Другой

Вероятности:

q[1] = 0,6

q[2] = 0,1

q[3] = 0,02

q[4] = 0

q[5] = 0

q[6] = 0

q[7] = 0,02

q[8] = 0,02

q[9] = 0

q[10] = 0,06

q[11] = 0

q[12] = 0

q[13] = 0

q[14] = 0,02

q[15] = 0,02

q[16] = 0,08

q[17] = 0

q[18] = 0

q[19] = 0

q[20] = 0

q[21] = 0,04

q[22] = 0

q[23] = 0

q[24] = 0,02

q[25] = 0

q[26] = 0

q[27] = 0

q[28] = 0

q[29] = 0

q[30] = 0

q[31] = 0

q[32] = 0

Строка состояния:

Оцінка надлишковості російської мови:

$$R = 1 - \frac{H_{\infty}}{H_0} \text{ де } H_0 = \log_2 32 = 5$$

$$1 - 2,232469/5 < R < 1 - 2,837603/5 \rightarrow 0,553506 < R < 0,567521$$

$$1 - 1,887059/5 < R < 1 - 2,506227/5 \rightarrow 0,622588 < R < 0,501245$$

$$1 - 1,699931/5 < R < 1 - 2,172431/5 \rightarrow 0,660014 < R < 0,565514$$

Опис труднощів :

Для мене найскладніше було знайти таку кількість тексту та нормалізувати його. При копіюванні, напевно, деякі символи були не розпізнані і потім мали вигляд пробілу. Після очистки тексту від пробілів (я пробувала це робити як за допомогою python, так і використовуючи функції текстових редакторів) в пошуку жодних пробілів не знаходило, однак при розбитті тексту на біграми і розрахунку їх частот, можна побачити пари типу «буква, щось типу пробілу» (для тексту без пробілів), хоч їх дуже мало.

Висновки : під час виконання лабораторної роботи було засвоєно поняття ентропії на символ джерела та його надлишковості, порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуто практичні навички щодо оцінки ентропії на символ джерела.