КРИПТОГРАФІЯ

КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

Екпериментальна оцінка ентропії на символ джерела відкритого тексту

ФБ-23 Моісеєнко Дмитро

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:

- Уважно прочитати методичні вказівки до виконання комп’ютерного практикуму.

- Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1 та H2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H1 та H2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H1 та H2 на тому ж тексті, в якому вилучено всі пробіли.

- За допомогою програми CoolPinkProgram оцінити значення H(10), H(20), H(30).

- Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи:

Спочатку пишемо програму мовою Python, що буде задовольняти поставлену задачу. Окрім того програма підрахує H1, H2, R1, R2 для текстів з пробілами і без нього. Для частотного аналізу біграм в тексті беремо їх з перетинами і без. Текст довільний, а також виконую на віртуальній машині Kali Linux.

Результати виконання програми:

```
┌──(kali㉿kali)-[~]
└─$ python3 entropy_assessment.py

[*] Counting the frequency for characters one by one...
[!] Frequency for each char:
| ъ:0.000142186833 | э:0.000924214418 | щ:0.001137494668 | ф:0.001208588085 | ц:0.002274989336
| ю:0.005332006256 | ш:0.006967154841 | х:0.010806199346 | ч:0.010877292763 | ж:0.010877292763
| й:0.011446040097 | з:0.015640551685 | г:0.015711645102 | ь:0.016138205602 | б:0.016706952936
| ы:0.017488980520 | я:0.018768662022 | п:0.023958481445 | д:0.025664723447 | у:0.026091283947
| м:0.026233470781 | к:0.032631878288 | в:0.034053746623 | р:0.036755296460 | с:0.043793544718
| т:0.043935731551 | л:0.044504478885 | н:0.049978671975 | и:0.051898194227 | е:0.061993459406
| а:0.066685624911 | о:0.095051898194 |  :0.174321057870 |
[!] Entropy H1: 4.369143103722749393540888281743688294265213541498526187001813623755886156069250247924173891078680 75
[!] Redundancy R1: 0.141194100073475760103059906406219883556995724263318129599398541885774611624976931234827179014 31
|
[*] Same text without spaces and special chars...
| ъ:0.000172205958 | э:0.001119338729 | щ:0.001377647667 | ф:0.001463750646 | ц:0.002755295333
| ю:0.006457723437 | ш:0.008438091958 | х:0.013087652833 | ч:0.013173755812 | ж:0.013173755812
| й:0.013862579645 | з:0.018942655416 | г:0.019028758395 | ь:0.019545376270 | б:0.020234200103
| ы:0.021181332874 | я:0.022731186499 | п:0.029016703978 | д:0.031083175478 | у:0.031599793353
| м:0.031771999311 | к:0.039521267436 | в:0.041243327019 | р:0.044515240227 | с:0.053039435164
| т:0.053211641123 | л:0.053900464956 | н:0.060530394352 | и:0.062855174789 | е:0.075081797830
| а:0.080764594455 | о:0.115119683141 |
[!] H1 w/o spaces: 4.48315707889327003725625312816547698114864781403782434557417529933379595971643105656312400242313
[!] R1 w/o spaces: 0.111259554108068292750594335076999540928767304708746035650802891300639596291918615729768397066 12
|
[*] Counting frequency for bigrams with overlay...
[!] Frequency for bigrams:
> Show bigram frequencies table? [y/n] n
[!] Entropy H2: 3.95022903163947448547819418719443569267440070189392428692093264810204561107709331224668858340010046 2
[!] Redundancy R2: 0.22353653384746258765906794539247759033277519271748022124687029387965024236534961769052211068605 07160
[!] Entropy H2` for text without overlapping: 3.94032981269105878910028057696335051413450684244332673772373161749098 97373984657420464827737305313 35
```

```
| м:0.026233470781 | к:0.032631878288 | в:0.034053746623 | р:0.036755296460 | с:0.043793544718
| т:0.043935731551 | л:0.044504478885 | н:0.049978671975 | и:0.051898194227 | е:0.061993459406
| а:0.066685624911 | о:0.095051898194 |  :0.174321057870 |
[!] Entropy H1: 4.369143103722749393540888281743688294265213541498526187001813623755886156069250247924173891078680 75
[!] Redundancy R1: 0.141194100073475760103059906406219883556995724263318129599398541885774611624976931234827179014 31
|
[*] Same text without spaces and special chars...
| ъ:0.000172205958 | э:0.001119338729 | щ:0.001377647667 | ф:0.001463750646 | ц:0.002755295333
| ю:0.006457723437 | ш:0.008438091958 | х:0.013087652833 | ч:0.013173755812 | ж:0.013173755812
| й:0.013862579645 | з:0.018942655416 | г:0.019028758395 | ь:0.019545376270 | б:0.020234200103
| ы:0.021181332874 | я:0.022731186499 | п:0.029016703978 | д:0.031083175478 | у:0.031599793353
| м:0.031771999311 | к:0.039521267436 | в:0.041243327019 | р:0.044515240227 | с:0.053039435164
| т:0.053211641123 | л:0.053900464956 | н:0.060530394352 | и:0.062855174789 | е:0.075081797830
| а:0.080764594455 | о:0.115119683141 |
[!] H1 w/o spaces: 4.48315707889327003725625312816547698114864781403782434557417529933379595971643105656312400242313
[!] R1 w/o spaces: 0.111259554108068292750594335076999540928767304708746035650802891300639596291918615729768397066 12
|
[*] Counting frequency for bigrams with overlay...
[!] Frequency for bigrams:
> Show bigram frequencies table? [y/n] n
[!] Entropy H2: 3.95022903163947448547819418719443569267440070189392428692093264810204561107709331224668858340010046 2
[!] Redundancy R2: 0.22353653384746258765906794539247759033277519271748022124687029387965024236534961769052211068605 07160
[!] Entropy H2` for text without overlapping: 3.94032981269105878910028057696335051413450684244332673772373161749098 97373984657420464827737305313 35
[!] Redundancy H2` for text without overlapping: 0.22548234048179326284281210416872118516176168372478777673434628344 17464849768567205023892705333826674
|
[*] Counting frequency for bigrams without spaces...
[!] Frequency for bigrams:
> Show bigram frequencies table? [y/n] n
[!] H2 w/o spaces: 4.12452491848532083816938376498347432791508790564775115642057658370928538925891115862754077170393 4302
[!] R2 w/o spaces: 0.18235474451590265872208173771193366166000762579162663036430656142093239409958246186084025196158 81848
[!] H2` w/o spaces, w/o overpalling: 4.09908405999473913311025426820564080940737055290761796647128119892533018760183 427531273991917259991 2
[!] R2` w/o spaces, w/o overlapping: 0.18739813682201716912187187619106286949698546460428290222069588640527701141374 0825234994451790762509 5
> Show bigram frequencies matrix? [y/n] n

┌──(kali㉿kali)-[~]
└─$
```

Створюю таблицю частот для монограми та біграми:

	H	R
Монограми з пробілом	4.369	0.141
Монограми без пробілу	4.483	0.111
Біграми з перетином з пробілом	3.951	0.224
Біграми з перетином без пробілу	3.950	0.223
Біграми без перетину з пробілом	4.124	0.182
Біграми без перетину без пробілу	4.123	0.182

Записи частоти біграм в матриці вказані в відсотках(%)

Частота символів з пробілом:

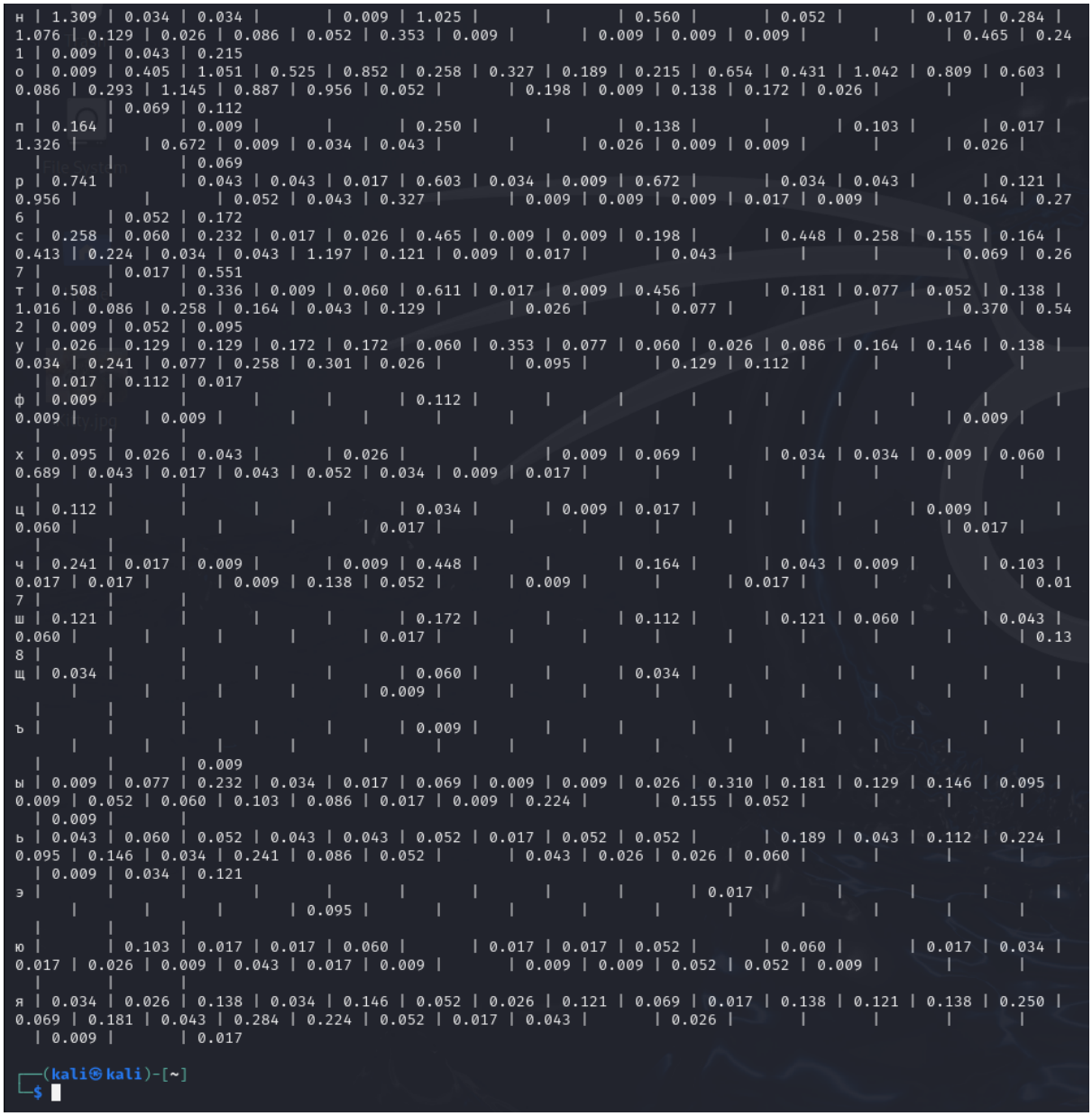
```
[*] Counting the frequency for characters one by one ...
[!] Frequency for each char:
| ъ:0.000142186833 | э:0.000924214418 | щ:0.001137494668 | ф:0.001208588085 | ц:0.002274989336
| ю:0.005332006256 | ш:0.006967154841 | х:0.010806199346 | ч:0.010877292763 | ж:0.010877292763
| й:0.011446040097 | з:0.015640551685 | г:0.015711645102 | ь:0.016138205602 | б:0.016706952936
| ы:0.017488980520 | я:0.018768662022 | п:0.023958481445 | д:0.025664723447 | у:0.026091283947
| м:0.026233470781 | к:0.032631878288 | в:0.034053746623 | р:0.036755296460 | с:0.043793544718
| т:0.043935731551 | л:0.044504478885 | н:0.049978671975 | и:0.051898194227 | е:0.061993459406
| а:0.066685624911 | о:0.095051898194 |  :0.174321057870 |
```

Частота символів без пробілу:

```
[*] Same text without spaces and special chars ...
| ъ:0.000172205958 | э:0.001119338729 | щ:0.001377647667 | ф:0.001463750646 | ц:0.002755295333
| ю:0.006457723437 | ш:0.008438091958 | х:0.013087652833 | ч:0.013173755812 | ж:0.013173755812
| й:0.013862579645 | з:0.018942655416 | г:0.019028758395 | ь:0.019545376270 | б:0.020234200103
| ы:0.021181332874 | я:0.022731186499 | п:0.029016703978 | д:0.031083175478 | у:0.031599793353
| м:0.031771999311 | к:0.039521267436 | в:0.041243327019 | р:0.044515240227 | с:0.053039435164
| т:0.053211641123 | л:0.053900464956 | н:0.060530394352 | и:0.062855174789 | е:0.075081797830
| а:0.080764594455 | о:0.115119683141 |
```

Матриця частот в % для біграм з перетином:

```
> Show bigram frequencies matrix? [y/n] y
  |   а   |   б   |   в   |   г   |   д   |   е   |   ж   |   з   |   и   |   й   |   к   |   л   |   м   |   н   |
  о   |   п   |   р   |   с   |   т   |   у   |   ф   |   х   |   ц   |   ч   |   ш   |   щ   |   ъ   |   ы   |   ь
  |   э   |   ю   |   я
а | 0.009 | 0.319 | 0.362 | 0.138 | 0.422 | 0.112 | 0.146 | 0.474 | 0.172 | 0.077 | 0.637 | 1.197 | 0.482 | 0.439 |
0.121 | 0.327 | 0.499 | 0.758 | 0.577 | 0.060 | 0.026 | 0.164 | 0.009 | 0.121 | 0.060 | 0.017 |       |       |
  | 0.009 | 0.086 | 0.258
б | 0.431 |       |       |       | 0.319 |       |       | 0.095 |       | 0.017 | 0.086 |       |       |
0.431 |       | 0.215 | 0.009 | 0.009 | 0.121 |       |       | 0.009 |       |       |       | 0.009 | 0.207 | 0.00
9 |       |       | 0.060
в | 0.680 | 0.034 | 0.052 | 0.069 | 0.043 | 0.646 | 0.009 | 0.043 | 0.301 |       | 0.077 | 0.069 | 0.034 | 0.095 |
0.878 | 0.086 | 0.043 | 0.396 | 0.043 | 0.103 |       | 0.009 | 0.009 | 0.009 | 0.009 |       | 0.009 | 0.327 | 0.01
7 | 0.017 |       | 0.017
г | 0.129 |       |       | 0.009 | 0.095 | 0.017 |       | 0.026 | 0.052 |       | 0.034 | 0.181 | 0.009 | 0.155 |
1.016 | 0.017 | 0.043 |       | 0.017 | 0.103 |       |       |       |       |       |       |       |       |
  |       |       |
д | 0.439 | 0.009 | 0.181 | 0.026 | 0.017 | 0.482 |       | 0.017 | 0.362 |       | 0.069 | 0.069 | 0.017 | 0.198 |
0.448 | 0.034 | 0.189 | 0.043 | 0.052 | 0.172 |       | 0.009 | 0.034 |       |       |       |       | 0.069 | 0.05
2 |       | 0.026 | 0.095
е | 0.034 | 0.319 | 0.370 | 0.499 | 0.465 | 0.052 | 0.189 | 0.284 | 0.086 | 0.155 | 0.207 | 0.784 | 0.405 | 0.861 |
0.112 | 0.276 | 0.663 | 0.654 | 0.551 | 0.043 | 0.026 | 0.112 | 0.026 | 0.138 | 0.095 | 0.052 |       |       |
  | 0.009 | 0.009 | 0.034
ж | 0.189 |       |       | 0.009 | 0.069 | 0.362 | 0.017 |       | 0.344 |       | 0.069 | 0.009 |       | 0.077 |
0.017 |       |       | 0.052 | 0.034 | 0.043 |       |       |       |       |       |       |       | 0.00
9 |       |       | 0.017
з | 0.689 | 0.172 | 0.103 | 0.043 | 0.146 | 0.077 | 0.017 | 0.017 | 0.017 |       | 0.026 | 0.009 | 0.017 | 0.241 |
0.052 | 0.026 | 0.043 | 0.009 |       | 0.043 |       |       |       | 0.009 |       |       |       | 0.077 | 0.03
4 |       |       | 0.026
и | 0.009 | 0.103 | 0.422 | 0.043 | 0.207 | 0.121 | 0.034 | 0.396 | 0.146 | 0.129 | 0.474 | 0.680 | 0.387 | 0.921 |
0.138 | 0.301 | 0.086 | 0.431 | 0.525 | 0.043 | 0.017 | 0.146 | 0.086 | 0.164 | 0.155 | 0.026 |       |       |
  |       |       | 0.095
й | 0.034 | 0.043 | 0.077 | 0.043 | 0.103 |       | 0.034 | 0.069 | 0.095 |       | 0.086 | 0.043 | 0.069 | 0.155 |
0.043 | 0.138 | 0.009 | 0.129 | 0.052 | 0.017 | 0.009 | 0.026 |       | 0.077 | 0.009 |       |       |       |
  | 0.009 |       | 0.017
к | 1.051 | 0.034 | 0.060 |       | 0.034 | 0.043 |       | 0.009 | 0.534 |       | 0.017 | 0.103 | 0.009 | 0.077 |
0.990 | 0.060 | 0.267 | 0.069 | 0.052 | 0.456 |       | 0.052 |       | 0.017 | 0.009 |       |       |       |
  |       |       |
л | 0.474 | 0.017 | 0.086 | 0.103 | 0.017 | 0.551 | 0.043 | 0.026 | 0.784 |       | 0.155 | 0.026 | 0.112 | 0.215 |
0.896 | 0.060 |       | 0.413 | 0.034 | 0.431 | 0.017 | 0.043 |       | 0.052 | 0.009 |       |       | 0.103 | 0.34
4 |       | 0.146 | 0.232
м | 0.198 | 0.034 | 0.086 | 0.026 | 0.052 | 0.448 | 0.017 | 0.026 | 0.405 |       | 0.077 | 0.052 | 0.026 | 0.344 |
0.439 | 0.138 | 0.009 | 0.121 | 0.052 | 0.215 |       | 0.060 | 0.017 | 0.060 |       |       |       | 0.215 | 0.00
9 | 0.009 |       | 0.043
н | 1.309 | 0.034 | 0.034 |       | 0.009 | 1.025 |       |       | 0.560 |       | 0.052 |       | 0.017 | 0.284 |
1.076 | 0.129 | 0.026 | 0.086 | 0.052 | 0.353 | 0.009 |       | 0.009 | 0.009 | 0.009 |       |       | 0.465 | 0.24
1 | 0.009 | 0.043 | 0.215
о | 0.009 | 0.405 | 1.051 | 0.525 | 0.852 | 0.258 | 0.327 | 0.189 | 0.215 | 0.654 | 0.431 | 1.042 | 0.809 | 0.603 |
0.086 | 0.293 | 1.145 | 0.887 | 0.956 | 0.052 |       | 0.198 | 0.009 | 0.138 | 0.172 | 0.026 |       |       |
  |       | 0.069 | 0.112
п | 0.164 |       | 0.009 |       |       | 0.250 |       |       | 0.138 |       |       | 0.103 |       | 0.017 |
1.326 |       | 0.672 | 0.009 | 0.034 | 0.043 |       |       | 0.026 | 0.009 | 0.009 |       |       | 0.026 |
  |       |       | 0.069
р | 0.741 |       | 0.043 | 0.043 | 0.017 | 0.603 | 0.034 | 0.009 | 0.672 |       | 0.034 | 0.043 |       | 0.121 |
0.956 |       |       | 0.052 | 0.043 | 0.327 |       | 0.009 | 0.009 | 0.009 | 0.017 | 0.009 |       | 0.164 | 0.27
6 |       | 0.052 | 0.172
с | 0.258 | 0.060 | 0.232 | 0.017 | 0.026 | 0.465 | 0.009 | 0.009 | 0.198 |       | 0.448 | 0.258 | 0.155 | 0.164 |
0.413 | 0.224 | 0.034 | 0.043 | 1.197 | 0.121 | 0.009 | 0.017 |       | 0.043 |       |       |       | 0.069 | 0.26
```



Оцінка значень H(10), H(20), H(30) за допомогою програми CoolPinkProgram:


Произольная часть текста:
тая_сделк
Использованные буквы:
Порядок n-граммы:
5
10
15
20
25
30
35
40
45
50
Введенный символ:
Символ по счету:
Номер эксперимента: 51
Поле ввода символов:
Неравенство для энтропии:
1.68450478586586< H < 2.48014416732061
Двоичная таблица угаданных символов:
Строка состояния:
Вероятности:
q[1] = 0.52
q[2] = 0.1
q[3] = 0.1
q[4] = 0.06
q[5] = 0.02
q[6] = 0.02
q[7] = 0
q[8] = 0.08
q[9] = 0
q[10] = 0
q[11] = 0.02
q[12] = 0.02
q[13] = 0.02
q[14] = 0.02
q[15] = 0
q[16] = 0
q[17] = 0
q[18] = 0
q[19] = 0
q[20] = 0.02
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0


Произольная часть текста:
ываем_на_себе_такое_сильное_давлениеэтого_закона_или_правила_что_не_в_состо
Использованные буквы:
Порядок n-граммы:
5
10
15
20
25
30
35
40
45
50
Введенный символ: _ (пробел)
Символ по счету: 1
Номер эксперимента: 50
Поле ввода символов:
Неравенство для энтропии:
1.83195837413134< H < 2.75489025085716
Двоичная таблица угаданных символов:
Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка
Вероятности:
q[1] = 0.46
q[2] = 0.14
q[3] = 0.08
q[4] = 0.06
q[5] = 0.04
q[6] = 0.04
q[7] = 0.04
q[8] = 0.04
q[9] = 0.02
q[10] = 0.02
q[11] = 0
q[12] = 0.02
q[13] = 0.02
q[14] = 0
q[15] = 0
q[16] = 0
q[17] = 0
q[18] = 0.02
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0
q[25] = 0
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0

Висновки: Під час виконання комп’ютерного практикуму отримав навички аналізу частот монограм на біграм. У свою чергу закріпив настійно на практиці формули обрахування надлишковості та ентропії. Порівняв та оцінив при цьому отриманні значення для різних наборів вхідних алфавітів. При програмі CoolPinkProgram застосував отримані дані для передбачення наступних символів.