

Міністерство освіти і науки, молоді та спорту України Національний технічний університет України "Київський політехнічний інститут" Фізико-Технічний інститут

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1 за семестровий курс предмету «Симетрична криптографія»

Роботу виконали:

Студент групи ФІ-03 Гілевський Олександр

Приймав:

Чорний Олег Миколайович

КОМП′ЮТЕРНИЙ ПРАКТИКУМ №1

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 Н1 та Н2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення Н1 та Н2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення Н1 та Н2 на тому ж тексті, в якому вилучено 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.

Хід роботи

- Очистка вхідного текстового файлу та редагування;
- Написання функцій для підрахунку частот літер та частот біграм. Визначили частоти та значення ентропії;
- Обрахували оцінки надлишковості російської мови для даних умовних ентропій джерел;
- Оформлення звіту.
- 1. Результати частот букв, та Н1 і Н2:

З перетином, але без пробілів

0 95273 е 73163 а 65926 т 53931 и 53630 н 53037 с 43046 л 38362 в 37597 р 33528 к 26994 м 25946 д 25711 у 24780 п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520 ц 2319	Буква	Кількість
а 65926 т 53931 и 53630 н 53037 с 43046 л 38362 в 37597 р 33528 к 26994 м 25946 д 25711 у 24780 п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911	0	95273
т 53931 и 53630 н 53037 с 43046 л 38362 в 37597 р 33528 к 26994 м 25946 д 25711 у 24780 п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522	е	73163
и 53630 н 53037 с 43046 л 38362 в 37597 р 33528 к 26994 м 25946 д 25711 у 24780 п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	а	65926
Н 53037 С 43046 Л 38362 В 37597 Р 33528 К 26994 М 25946 Д 25711 У 24780 П 21133 Б 19375 Я 17406 Ч 14872 Б 14412 Г 14001 Ы 13920 З 12470 Ж 9574 Й 8441 X 7068 Ш 6911 Ю 4730 Э 2522 Щ 2520	Т	53931
с 43046 л 38362 в 37597 р 33528 к 26994 м 25946 д 25711 у 24780 п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 ц 2520	И	53630
л 38362 в 37597 р 33528 к 26994 м 25946 д 25711 у 24780 п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	н	53037
В 37597 р 33528 к 26994 м 25946 Д 25711 у 24780 п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	С	43046
р 33528 к 26994 м 25946 д 25711 у 24780 п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ь 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	Л	38362
к 26994 м 25946 д 25711 у 24780 п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ь 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	В	37597
м 25946 Д 25711 у 24780 п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	р	33528
Д 25711 У 24780 П 21133 Ь 19375 Я 17406 Ч 14872 Є 14412 Г 14001 Ы 13920 З 12470 Ж 9574 Й 8441 X 7068 Ш 6911 Ю 4730 Э 2522 Щ 2520	К	26994
у 24780 п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	M	25946
п 21133 ь 19375 я 17406 ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	Д	25711
Ь 19375 Я 17406 Ч 14872 Б 14412 Г 14001 Ы 13920 З 12470 Ж 9574 Й 8441 X 7068 Ш 6911 Ю 4730 Э 2522 Щ 2520	У	24780
я 17406 ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	П	21133
Ч 14872 б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	Ь	19375
б 14412 г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	Я	17406
г 14001 ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	Ч	14872
ы 13920 з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	б	14412
з 12470 ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	Γ	14001
ж 9574 й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	Ы	13920
й 8441 х 7068 ш 6911 ю 4730 э 2522 щ 2520	3	12470
х 7068 ш 6911 ю 4730 э 2522 щ 2520	ж	9574
ш 6911 ю 4730 э 2522 щ 2520	й	8441
ю 4730 э 2522 щ 2520	х	7068
э 2522 щ 2520	Ш	6911
щ 2520	ю	4730
	Э	2522
ц 2319	щ	2520
	ц	2319

ф 972

H1: 4,444373807624386

H2: 4,130122805459284

Без перетину та пробілів

	Des neperany ra npooring
Буква	Кількість
O	95273
е	73163
а	65926
т	53931
И	53630
н	53037
С	43046
Л	38362
В	37597
р	33528
К	26994
Μ	25946
Д	25711
У	24780
п	21133
Ь	19375
Я	17406
Ч	14872
б	14412
Г	14001
Ы	13920
3	12470
ж	9574
й	8441

Х	7068
Ш	6911
Ю	4730
Э	2522
щ	2520
ц	2319
ф	972

H1: 4,444373807624386 H2: 4,130127843664438

Без перетину, але з пробілами

Символ	Кількість
u u	167105
0	95273
е	73163
а	65926
Т	53931
И	53630
Н	53037
С	43046
Л	38362
В	37597
р	33528
К	26994
М	25946
Д	25711
У	24780
п	21133
Ь	19375
я	17406

ч	14872
б	14412
Γ	14001
Ы	13920
3	12470
ж	9574
й	8441
х	7068
Ш	6911
ю	4730
э	2522
щ	2520
ц	2319
ф	972

H1: 4,349377539781125 H2: 3,9499937912023224

3 перетином і пробілами

Символ	Кількість
<i>u u</i>	167105
0	95273
е	73163
a	65926
Т	53931
И	53630
Н	53037
С	43046
Л	38362
В	37597
р	33528

к	26994
М	25946
Д	25711
у	24780
П	21133
Ь	19375
Я	17406
ч	14872
б	14412
г	14001
Ы	13920
3	12470
ж	9574
й	8441
X	7068
Ш	6911
ю	4730
э	2522
щ	2520
ц	2319
ф	972

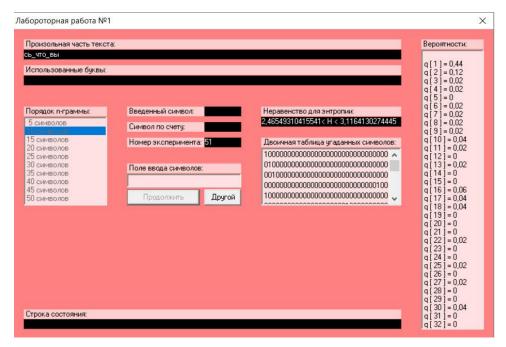
H1: 4,349377539781125

H2: 3,9501128865434714

Результати біграм в файлі results.txt

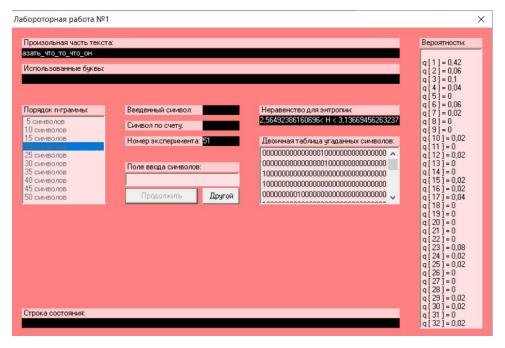
2. Оцінки для $H^{(10)}$, $H^{(20)}$, $H^{(30)}$

• $H^{(10)}$

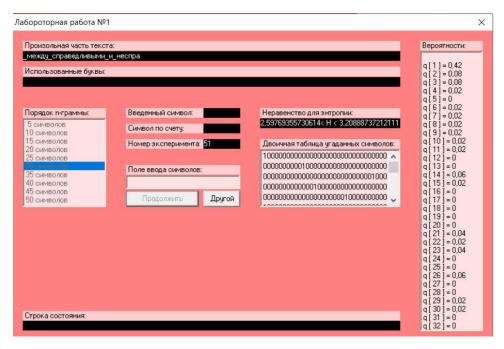


$$R = 1 - \frac{2.4655 + 3.1164}{2} \cdot \frac{1}{\log 2(32)} = 0.442$$

• $H^{(20)}$



$$R = 1 - \frac{2.5649 + 3.1367}{2} \cdot \frac{1}{\log 2(32)} = 0.43$$



$$R = 1 - \frac{2.5977 + 3.2089}{2} \cdot \frac{1}{\log 2(32)} = 0.42$$

Висновок: в цій лабі розглянули як рахувати частоту літер в тексті, біграм, ентропію та надлишковість. Перегоналися, що найчастішим символом є пробіл, а після нього голосні літери (а, о і тд).