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

ЛАБОРАТОРНА РОБОТА №1

з дисципліни «Криптографія»

«Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконали:
студенти 3 курсу ФТІ
групи ФБ-81
Близнюк Микола та Мишкін Артем
Перевірили:
Чорний О.

Київ – 2020

Завдання

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.

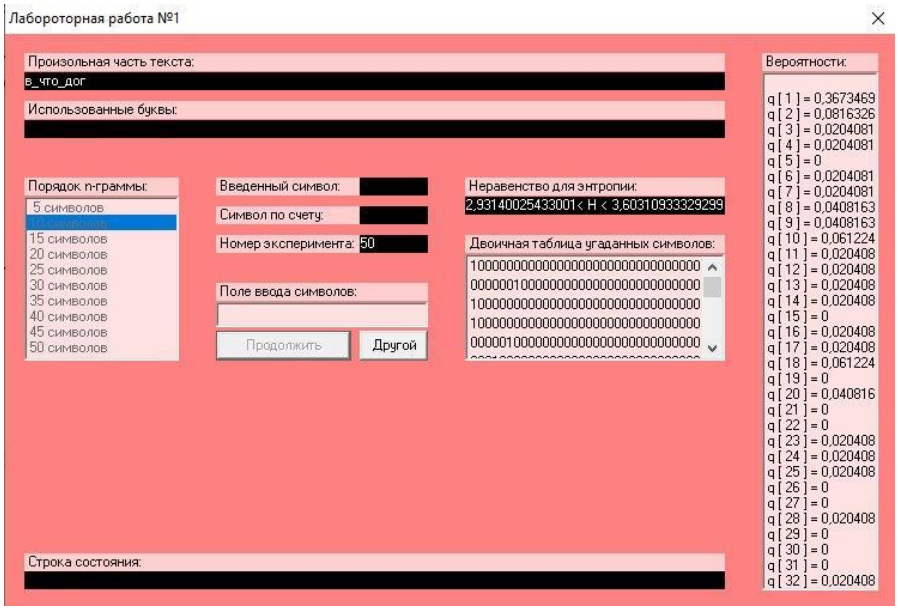
1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1 та H2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H1 та H2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення 1 H та 2 H на тому ж тексті, в якому вилучено всі пробіли.

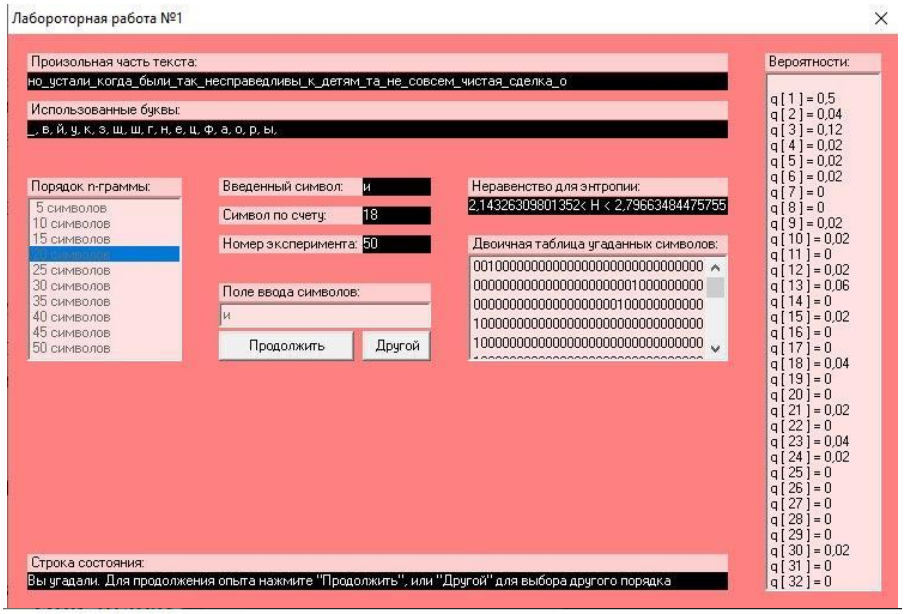
2. За допомогою програми CoolPinkProgram оцінити значення H(10) , H (20), H(30) .

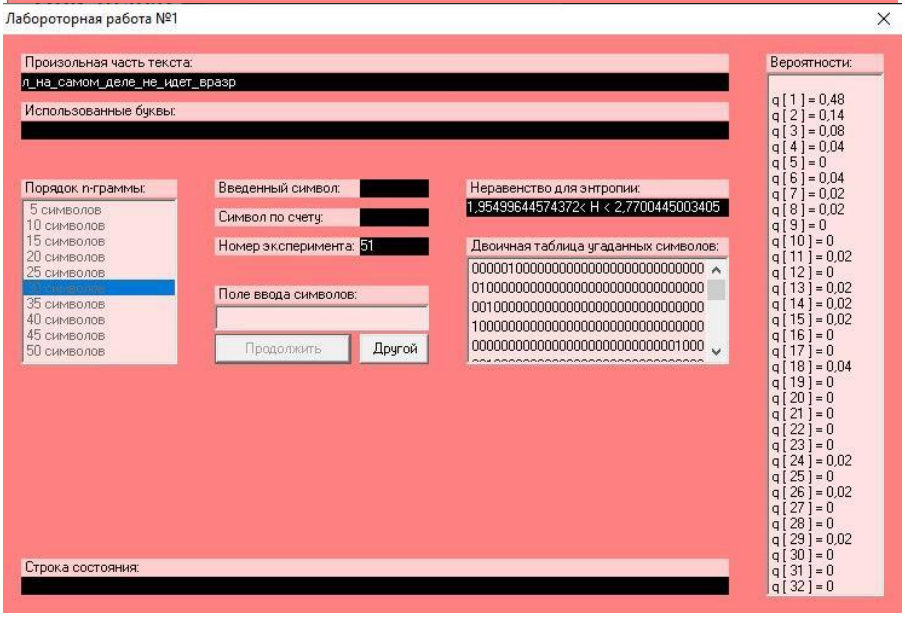
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела

Хід роботи

Файл	Коментар
funcs.py	Включає в себе імпорт потрібних бібліотек та функції
main.py	Головний виконуваний файл
text.txt	Файл з текстом для аналізу
values.txt	Файл з результатами аналізу


Лабороторная работа №1
Произольная часть текста:
в_что_дог
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: 50
Поле ввода символов:
Продолжить
Другой
Неравенство для энтропии:
2,93140025433001< H < 3,60310933329299
Двоичная таблица угаданных символов:
Строка состояния:
Вероятности:
q [1] = 0,3673469
q [2] = 0,0816326
q [3] = 0,0204081
q [4] = 0,0204081
q [5] = 0
q [6] = 0,0204081
q [7] = 0,0204081
q [8] = 0,0408163
q [9] = 0,0408163
q [10] = 0,061224
q [11] = 0,020408
q [12] = 0,020408
q [13] = 0,020408
q [14] = 0,020408
q [15] = 0
q [16] = 0,020408
q [17] = 0,020408
q [18] = 0,061224
q [19] = 0
q [20] = 0,040816
q [21] = 0
q [22] = 0
q [23] = 0,020408
q [24] = 0,020408
q [25] = 0,020408
q [26] = 0
q [27] = 0
q [28] = 0,020408
q [29] = 0
q [30] = 0
q [31] = 0
q [32] = 0,020408


Лабороторная работа №1
Произольная часть текста:
но_устали_когда_были_так_несправедливы_к_детям_та_не_совсем_чистая_сделка_о
Использованные буквы:
_, в, й, у, к, з, щ, ш, г, н, е, ц, ф, а, о, р, ы,
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: и
Символ по счету: 18
Номер эксперимента: 50
Поле ввода символов:
и
Продолжить
Другой
Неравенство для энтропии:
2,14326309801352< H < 2,79663484475755
Двоичная таблица угаданных символов:
Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка
Вероятности:
q [1] = 0,5
q [2] = 0,04
q [3] = 0,12
q [4] = 0,02
q [5] = 0,02
q [6] = 0,02
q [7] = 0
q [8] = 0
q [9] = 0,02
q [10] = 0,02
q [11] = 0
q [12] = 0,02
q [13] = 0,06
q [14] = 0
q [15] = 0,02
q [16] = 0
q [17] = 0
q [18] = 0,04
q [19] = 0
q [20] = 0
q [21] = 0,02
q [22] = 0
q [23] = 0,04
q [24] = 0,02
q [25] = 0
q [26] = 0
q [27] = 0
q [28] = 0
q [29] = 0
q [30] = 0,02
q [31] = 0
q [32] = 0


Лабороторная работа №1
Произольная часть текста:
л_на_самом_деле_не_идет_вразр
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
Поле ввода символов:
Продолжить
Другой
Неравенство для энтропии:
1,95499644574372< H < 2,77004450034 05
Двоичная таблица угаданных символов:
Строка состояния:
Вероятности:
q [1] = 0,48
q [2] = 0,14
q [3] = 0,08
q [4] = 0,04
q [5] = 0
q [6] = 0,04
q [7] = 0,02
q [8] = 0,02
q [9] = 0
q [10] = 0
q [11] = 0,02
q [12] = 0
q [13] = 0,02
q [14] = 0,02
q [15] = 0,02
q [16] = 0
q [17] = 0
q [18] = 0,04
q [19] = 0
q [20] = 0
q [21] = 0
q [22] = 0
q [23] = 0
q [24] = 0,02
q [25] = 0
q [26] = 0,02
q [27] = 0
q [28] = 0
q [29] = 0,02
q [30] = 0
q [31] = 0
q [32] = 0

Вид тесту	Entropy	Redundancy
Монограма (whitespace)	4.33270764082674	0.13345847183465198
Біграма (whitespace)	3.966574582247116	0.2066850835505768
Монограма (no whitespace)	3.884654462910488	0.22306910741790242
Біграма (no whitespace)	4.154389978943852	0.16912200421122958

Висновки

Під час виконання комп’ютерного практикуму № 1. Ми експериментально оцінили ентропію на символ джерела та надлишковість на прикладі російського алфавіту. Ми порівняли ентропію та надлишковість для монограм і для біграм з пробілом та без нього.