



Міністерство освіти і науки України
НТУУ «Київський політехнічний інститут»
Фізико-технічний інститут

Лабораторні роботи № 1
з предмету «Криптографія»
на тему: «Експериментальна оцінка ентропії на символ джерела
відкритого тексту»

Виконала:
Студентка III
курсу
ФТІ групи ФБ-84
Матвієнко В. С.
Перевірив:
Чорний О. М.

Київ-2020

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 вилучено всі пробіли.

2. За допомогою програми CoolPinkProgram оцінити значення $(10) H$, $(20) H$, $(30) H$. 3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи:

Перед написанням коду я ознайомилась з теоретичними відомостями і методичними вказівками. Створила текстовий файл 1.txt з текстом російською мовою.

Написання коду:

Під час написання коду виникли труднощі з кодуванням файлу, цю проблему вирішила за допомогою: `encoding='utf-8'`.

У кодї використанні формули ентропії: $entr += -(fr) * \text{math.log}((fr), 2)$ та надлишковості: $nadl = 1 - (entr / \text{math.log}(32, 2))$.

Під час розрахування ентропії біграми кінцевий результат треба було поділити на 2. Одержані результати ентропії та надлишковості для тексту з пропусками і тексту без пропусків були занесені у таблиці.

CoolPinkProgram:

Провела по 50 експериментів $H(10)$, $H(20)$, $H(30)$, після чого порахувала за допомогою коду надлишковість на основі результатів експерименту. Усі результати експерименту занесені до таблиці.

Монограми:

| З пропусками: | Без пропусків: |
|--|---|
| ' ': '12.76596', 'а': '7.25338', 'б': '0.77369', 'в': '3.86847', 'г': '2.99807', 'д': '1.93424', 'е': '8.02708', 'ж': '0.87041', 'з': '0.96712', 'и': '7.15667', 'й': '1.16054', 'к': '3.48162', 'л': '3.09478', 'м': '1.64410', 'н': '6.28627', 'о': '9.86460', 'п': '2.70793', 'р': '5.89942', 'с': '5.12573', 'т': '4.93230', 'у': '2.51451', 'ф': '0.29014', 'х': '1.06383', 'ц': '0.48356', 'ч': '0.96712', 'ш': '1.25725', 'щ': '0.19342', 'ы': '0.77369', 'ь': '0.19342', 'э': '0.19342', 'ю': '0.09671', 'я': '1.16054' | 'а': '8.31486', 'б': '0.88692', 'в': '4.43459', 'г': '3.43681', 'д': '2.21729', 'е': '9.20177', 'ж': '0.99778', 'з': '1.10865', 'и': '8.20399', 'й': '1.33038', 'к': '3.99113', 'л': '3.54767', 'м': '1.88470', 'н': '7.20621', 'о': '11.30820', 'п': '3.10421', 'р': '6.76275', 'с': '5.87583', 'т': '5.65410', 'у': '2.88248', 'ф': '0.33259', 'х': '1.21951', 'ц': '0.55432', 'ч': '1.10865', 'ш': '1.44124', 'щ': '0.22173', 'ы': '0.88692', 'ь': '0.22173', 'э': '0.22173', 'ю': '0.11086', 'я': '1.33038' |

Біграми(текст з пропусками):

| Частота біграм з повторенням | | | | Частота біграм без повторення | | | |
|------------------------------|------------------|------------------|------------------|-------------------------------|------------------|------------------|--|
| 'а': '0.58083', | 'еє': '0.29042', | 'нж': '0.09681', | 'ул': '0.09681', | 'а': '0.38685', | 'з': '0.19342', | 'ро': '0.58027', | |
| 'б': '0.09681', | 'ей': '0.29042', | 'ни': '0.77444', | 'ур': '0.48403', | 'б': '0.19342', | 'за': '0.19342', | 'ру': '0.19342', | |
| 'в': '0.96805', | 'ек': '0.29042', | 'нк': '0.09681', | 'ус': '0.38722', | 'в': '1.93424', | 'зв': '0.19342', | 'ры': '0.19342', | |
| 'г': '0.38722', | 'ел': '0.38722', | 'нн': '0.87125', | 'ут': '0.09681', | 'г': '0.58027', | 'зу': '0.19342', | 'ря': '0.38685', | |
| 'д': '0.38722', | 'ем': '0.09681', | 'но': '1.45208', | 'уш': '0.58083', | 'д': '0.77369', | 'н': '0.77369', | 'са': '0.38685', | |
| 'е': '0.09681', | 'ен': '0.96805', | 'нс': '0.09681', | 'фе': '0.09681', | 'е': '0.19342', | 'ив': '0.38685', | 'св': '0.19342', | |
| 'ж': '0.09681', | 'ео': '0.19361', | 'ну': '0.09681', | 'фо': '0.09681', | 'ж': '0.19342', | 'ид': '0.38685', | 'ск': '1.35397', | |
| 'з': '0.09681', | 'еп': '0.09681', | 'нц': '0.09681', | 'фр': '0.09681', | 'з': '0.19342', | 'из': '0.38685', | 'сл': '0.38685', | |
| 'н': '1.25847', | 'ер': '0.96805', | 'ны': '0.38722', | 'х': '0.67764', | 'н': '1.35397', | 'ии': '0.19342', | 'см': '0.19342', | |
| 'к': '0.38722', | 'ес': '0.58083', | 'ня': '0.09681', | 'ха': '0.09681', | 'к': '0.38685', | 'ик': '0.77369', | 'сн': '0.19342', | |
| 'л': '0.38722', | 'ет': '0.96805', | 'о': '2.42014', | 'хо': '0.19361', | 'л': '0.19342', | 'ин': '0.77369', | 'со': '0.19342', | |
| 'м': '0.38722', | 'еп': '0.09681', | 'об': '0.29042', | 'ху': '0.09681', | 'м': '0.58027', | 'ио': '0.19342', | 'ср': '0.19342', | |
| 'н': '0.67764', | 'еч': '0.09681', | 'ов': '0.58083', | 'ц': '0.09681', | 'н': '0.77369', | 'ис': '0.77369', | 'ст': '0.96712', | |
| 'о': '1.06486', | 'еш': '0.09681', | 'ог': '1.16167', | 'ца': '0.09681', | 'о': '1.35397', | 'ит': '0.58027', | 'сх': '0.19342', | |
| 'п': '2.22652', | 'ея': '0.09681', | 'од': '0.58083', | 'ци': '0.29042', | 'п': '2.32108', | 'их': '0.19342', | 'сз': '0.19342', | |
| 'р': '1.25847', | 'жд': '0.09681', | 'ое': '0.29042', | 'ч': '0.09681', | 'р': '1.54739', | 'ич': '0.58027', | 'сд': '0.38685', | |
| 'с': '1.35528', | 'же': '0.19361', | 'ож': '0.48403', | 'ча': '0.29042', | 'с': '1.93424', | 'ию': '0.19342', | 'т': '0.19342', | |
| 'т': '0.29042', | 'жи': '0.19361', | 'оз': '0.19361', | 'че': '0.48403', | 'с': '0.58027', | 'ия': '0.19342', | 'тв': '0.19342', | |
| 'ф': '0.19361', | 'жн': '0.09681', | 'он': '0.19361', | 'чи': '0.09681', | 'ф': '0.19342', | 'й': '1.16054', | 'тв': '0.19342', | |
| 'х': '0.09681', | 'жу': '0.09681', | 'ой': '0.58083', | 'ше': '0.58083', | 'х': '0.19342', | 'к': '0.58027', | 'те': '1.35397', | |
| 'ч': '0.29042', | 'з': '0.19361', | 'ок': '0.19361', | 'ши': '0.09681', | 'ч': '0.38685', | 'ка': '0.19342', | 'ти': '0.96712', | |
| 'я': '0.19361', | 'за': '0.19361', | 'ол': '0.09681', | 'шк': '0.58083', | 'а': '1.54739', | 'кв': '0.19342', | 'тн': '0.19342', | |
| 'а': '1.54889', | 'зв': '0.09681', | 'ом': '0.48403', | 'ша': '0.09681', | 'аб': '0.19342', | 'ки': '0.77369', | 'то': '0.19342', | |
| 'аб': '0.09681', | 'зе': '0.19361', | 'он': '0.48403', | 'ше': '0.09681', | 'ав': '0.38685', | 'ко': '0.19342', | 'ту': '0.77369', | |
| 'ав': '0.48403', | 'зн': '0.09681', | 'оп': '0.09681', | 'ы': '0.19361', | 'ад': '0.19342', | 'кр': '0.58027', | 'у': '0.19342', | |
| 'ар': '0.09681', | 'зу': '0.09681', | 'ор': '0.48403', | 'ык': '0.09681', | 'ае': '0.19342', | 'кт': '0.19342', | 'уб': '0.19342', | |
| 'ар': '0.09681', | 'зй': '0.09681', | 'ос': '0.77444', | 'ым': '0.09681', | 'аз': '0.19342', | 'ла': '0.58027', | 'ур': '0.19342', | |
| 'ае': '0.19361', | 'н': '1.25847', | 'от': '0.29042', | 'ых': '0.38722', | 'ай': '0.19342', | 'ле': '0.77369', | 'ул': '0.19342', | |
| 'аз': '0.38722', | 'иб': '0.09681', | 'оз': '0.19361', | 'ь': '0.09681', | 'ак': '0.19342', | 'ли': '0.77369', | 'уж': '0.38685', | |
| 'ан': '0.09681', | 'ив': '0.38722', | 'пе': '0.38722', | 'ьн': '0.09681', | 'ал': '1.35397', | 'ло': '0.77369', | 'ул': '0.19342', | |
| 'ай': '0.09681', | 'ий': '0.19361', | 'пи': '0.19361', | 'э': '0.19361', | 'ам': '0.19342', | 'ль': '0.19342', | 'ур': '0.19342', | |
| 'ак': '0.09681', | 'не': '0.09681', | 'по': '0.58083', | 'юн': '0.09681', | 'ан': '0.58027', | 'ля': '0.19342', | 'ус': '0.58027', | |
| 'ан': '1.06486', | 'из': '0.19361', | 'пр': '0.77444', | 'я': '0.77444', | 'ар': '0.19342', | 'ма': '0.58027', | 'ут': '0.19342', | |
| 'ам': '0.29042', | 'ии': '0.09681', | 'пу': '0.77444', | 'яз': '0.09681', | 'ас': '0.19342', | 'мо': '0.19342', | 'уш': '0.58027', | |
| 'ан': '0.67764', | 'ий': '0.19361', | 'р': '0.09681', | 'ян': '0.19361', | 'ат': '0.96712', | 'н': '0.58027', | 'фо': '0.19342', | |
| 'ап': '0.09681', | 'ик': '0.67764', | 'ра': '2.03291', | 'ят': '0.09681', | 'аф': '0.19342', | 'на': '0.58027', | 'х': '0.38685', | |
| 'ар': '0.19361', | 'ил': '0.09681', | 'рб': '0.09681', | | 'ах': '0.19342', | 'нд': '0.38685', | 'ха': '0.19342', | |
| 'ас': '0.19361', | 'им': '0.09681', | 'рв': '0.09681', | | 'ац': '0.38685', | 'не': '0.38685', | 'хо': '0.19342', | |
| 'ат': '0.87125', | 'ин': '0.87125', | 'рг': '0.38722', | | 'аш': '0.19342', | 'нж': '0.19342', | 'ци': '0.19342', | |
| 'аф': '0.09681', | 'но': '0.09681', | 'ре': '0.67764', | | 'ба': '0.19342', | 'ни': '1.35397', | 'ча': '0.19342', | |
| 'ах': '0.19361', | 'не': '0.87125', | 'ри': '0.77444', | | 'бр': '0.38685', | 'нн': '0.58027', | 'че': '0.77369', | |
| 'ау': '0.19361', | 'ит': '0.77444', | 'рм': '0.09681', | | 'ва': '0.38685', | 'но': '1.74081', | 'ше': '0.77369', | |
| 'аш': '0.09681', | 'их': '0.19361', | 'рн': '0.19361', | | 'ви': '0.38685', | 'нс': '0.19342', | 'ши': '0.19342', | |
| 'ая': '0.09681', | 'иц': '0.09681', | 'ро': '0.77444', | | 'вл': '0.19342', | 'нр': '0.19342', | 'шк': '0.58027', | |
| 'ба': '0.09681', | 'ич': '0.58083', | 'ру': '0.38722', | | 'вн': '0.38685', | 'ни': '0.19342', | 'ше': '0.19342', | |
| 'бл': '0.19361', | 'ию': '0.09681', | 'ры': '0.09681', | | 'во': '0.77369', | 'ны': '0.19342', | 'ык': '0.19342', | |
| 'бр': '0.38722', | 'ия': '0.19361', | 'ря': '0.19361', | | 'вр': '0.19342', | 'ня': '0.19342', | 'ым': '0.19342', | |
| 'бу': '0.09681', | 'й': '1.06486', | 'са': '0.48403', | | 'вт': '0.19342', | 'о': '2.12766', | 'ых': '0.58027', | |
| 'в': '0.38722', | 'йш': '0.09681', | 'св': '0.29042', | | 'ву': '0.19342', | 'об': '0.19342', | 'эт': '0.38685', | |
| 'ва': '0.38722', | 'к': '0.77444', | 'се': '0.19361', | | 'вш': '0.38685', | 'ов': '0.38685', | 'я': '0.38685', | |
| 'ве': '0.38722', | 'ка': '0.58083', | 'ск': '0.96805', | | 'вы': '0.19342', | 'ог': '0.96712', | 'яз': '0.19342', | |
| 'ви': '0.48403', | 'кв': '0.09681', | 'сл': '0.48403', | | 'г': '0.19342', | 'од': '0.58027', | 'ян': '0.19342', | |
| 'вл': '0.19361', | 'ки': '0.67764', | 'см': '0.09681', | | 'ге': '0.19342', | 'ое': '0.19342', | | |
| 'вн': '0.19361', | 'кл': '0.09681', | 'сн': '0.29042', | | 'ги': '0.19342', | 'ож': '0.77369', | | |
| 'во': '0.87125', | 'ко': '0.67764', | 'со': '0.09681', | | 'го': '2.32108', | 'оз': '0.19342', | | |
| 'вр': '0.19361', | 'кр': '0.29042', | 'сп': '0.09681', | | 'да': '0.19342', | 'он': '0.19342', | | |
| 'вт': '0.09681', | 'кс': '0.19361', | 'ср': '0.48403', | | 'ле': '0.38685', | 'ой': '0.77369', | | |
| 'ву': '0.09681', | 'кт': '0.09681', | 'ст': '1.16167', | | 'ди': '0.38685', | 'ол': '0.19342', | | |
| 'вш': '0.48403', | 'л': '0.19361', | 'сх': '0.19361', | | 'лк': '0.19342', | 'ом': '0.96712', | | |
| 'вы': '0.09681', | 'ла': '0.29042', | 'сь': '0.09681', | | 'е': '0.77369', | 'он': '0.38685', | | |
| 'г': '0.19361', | 'ле': '0.77444', | 'ся': '0.19361', | | 'еа': '0.19342', | 'оп': '0.19342', | | |
| 'га': '0.09681', | 'ли': '0.58083', | 'т': '0.38722', | | 'ев': '0.58027', | 'ор': '0.38685', | | |
| 'ге': '0.48403', | 'ло': '0.77444', | 'та': '0.38722', | | 'ер': '0.96712', | 'ос': '1.16054', | | |
| 'ги': '0.09681', | 'лс': '0.09681', | 'тв': '0.29042', | | 'ед': '0.19342', | 'от': '0.19342', | | |
| 'го': '2.12972', | 'лу': '0.19361', | 'те': '0.96805', | | 'еи': '0.19342', | 'пе': '0.38685', | | |
| 'да': '0.29042', | 'ль': '0.09681', | 'ти': '0.87125', | | 'ей': '0.19342', | 'пи': '0.19342', | | |
| 'дв': '0.09681', | 'ля': '0.09681', | 'тн': '0.38722', | | 'ек': '0.58027', | 'по': '0.77369', | | |
| 'де': '0.58083', | 'м': '0.48403', | 'то': '0.48403', | | 'ел': '0.38685', | 'пр': '0.77369', | | |
| 'ди': '0.19361', | 'ма': '0.48403', | 'тп': '0.09681', | | 'ен': '0.19342', | 'пу': '0.58027', | | |
| 'лк': '0.09681', | 'ме': '0.19361', | 'тр': '0.38722', | | 'еи': '1.35397', | 'р': '0.19342', | | |
| 'дн': '0.09681', | 'мо': '0.19361', | 'тс': '0.09681', | | 'ео': '0.19342', | 'ра': '2.32108', | | |
| 'до': '0.19361', | 'му': '0.09681', | 'ту': '0.48403', | | 'еп': '0.19342', | 'рб': '0.19342', | | |
| 'др': '0.38722', | 'мы': '0.09681', | 'тш': '0.09681', | | 'ер': '0.77369', | 'рв': '0.19342', | | |
| 'е': '0.87125', | 'н': '0.58083', | 'у': '0.38722', | | 'ср': '0.38685', | 'рг': '0.38685', | | |
| 'еа': '0.19361', | 'на': '0.48403', | 'уб': '0.09681', | | 'ет': '1.35397', | 'ре': '0.96712', | | |
| 'ев': '0.38722', | 'нв': '0.09681', | 'уг': '0.09681', | | 'ец': '0.19342', | 'ри': '0.58027', | | |
| 'ет': '0.87125', | 'нд': '0.29042', | 'уд': '0.09681', | | 'ея': '0.19342', | 'рм': '0.19342', | | |
| 'ед': '0.19361', | 'не': '0.77444', | 'уж': '0.19361', | | 'жи': '0.19342', | 'рн': '0.38685', | | |

Біграми(текст без пропусків):

| Частота біграм з повторенням | | | | Частота біграм без повторення | | | |
|------------------------------|------------------|------------------|------------------|-------------------------------|------------------|------------------|--|
| 'aa': '0.22198', | 'за': '0.22198', | 'нн': '1.10988', | 'ур': '0.66593', | 'aa': '0.22173', | 'их': '0.22173', | 'со': '0.22173', | |
| 'аб': '0.11099', | 'зв': '0.11099', | 'но': '1.77580', | 'ус': '0.55494', | 'ав': '0.44346', | 'ич': '0.88692', | 'сп': '0.22173', | |
| 'ав': '0.55494', | 'зе': '0.22198', | 'нр': '0.11099', | 'ут': '0.11099', | 'аг': '0.22173', | 'ио': '0.22173', | 'се': '0.44346', | |
| 'ар': '0.22198', | 'зн': '0.11099', | 'нс': '0.11099', | 'уч': '0.11099', | 'ае': '0.22173', | 'ий': '0.44346', | 'ст': '1.10865', | |
| 'ад': '0.11099', | 'зо': '0.11099', | 'ну': '0.11099', | 'уш': '0.66593', | 'аз': '0.44346', | 'йл': '0.22173', | 'сх': '0.44346', | |
| 'ае': '0.22198', | 'зе': '0.11099', | 'нц': '0.11099', | 'фе': '0.11099', | 'аи': '0.22173', | 'йо': '0.22173', | 'ся': '0.22173', | |
| 'аз': '0.44395', | 'зу': '0.11099', | 'ны': '0.44395', | 'фо': '0.11099', | 'ай': '0.22173', | 'йп': '0.22173', | 'та': '0.66519', | |
| 'аи': '0.22198', | 'зы': '0.11099', | 'ня': '0.11099', | 'фр': '0.11099', | 'ак': '0.22173', | 'йр': '0.22173', | 'тв': '0.22173', | |
| 'ай': '0.11099', | 'на': '0.11099', | 'об': '0.44395', | 'ха': '0.22198', | 'ал': '1.77384', | 'йт': '0.22173', | 'те': '1.55211', | |
| 'ак': '0.11099', | 'нб': '0.11099', | 'ов': '0.66593', | 'хв': '0.11099', | 'ам': '0.44346', | 'ка': '0.88692', | 'ти': '1.10865', | |
| 'ал': '1.22087', | 'нв': '0.66593', | 'ог': '1.44284', | 'хд': '0.11099', | 'ан': '0.66519', | 'кз': '0.22173', | 'тн': '0.22173', | |
| 'ам': '0.33296', | 'иг': '0.22198', | 'од': '0.77691', | 'хи': '0.11099', | 'ап': '0.66519', | 'ки': '0.88692', | 'то': '0.44346', | |
| 'ан': '0.77691', | 'ид': '0.33296', | 'ое': '0.33296', | 'хл': '0.11099', | 'ар': '0.22173', | 'кл': '0.22173', | 'тр': '0.22173', | |
| 'ап': '0.88790', | 'ие': '0.11099', | 'ож': '0.55494', | 'хо': '0.33296', | 'ас': '0.66519', | 'км': '0.22173', | 'тс': '0.22173', | |
| 'ар': '0.22198', | 'иж': '0.11099', | 'оз': '0.22198', | 'хп': '0.11099', | 'ат': '1.10865', | 'ко': '1.33038', | 'ту': '0.66519', | |
| 'ас': '0.55494', | 'из': '0.22198', | 'он': '0.22198', | 'ху': '0.11099', | 'аф': '0.22173', | 'кп': '0.22173', | 'уб': '0.22173', | |
| 'ат': '0.99889', | 'ии': '0.11099', | 'ой': '0.66593', | 'ца': '0.11099', | 'аш': '0.22173', | 'кр': '0.44346', | 'уж': '0.44346', | |
| 'аф': '0.11099', | 'ий': '0.22198', | 'ок': '0.22198', | 'ци': '0.44395', | 'ая': '0.22173', | 'кс': '0.22173', | 'уш': '0.66519', | |
| 'ах': '0.22198', | 'ик': '0.88790', | 'ол': '0.22198', | 'ча': '0.33296', | 'бл': '0.22173', | 'кт': '0.22173', | 'ур': '0.88692', | |
| 'аи': '0.22198', | 'ил': '0.11099', | 'ом': '0.66593', | 'че': '0.55494', | 'бр': '0.88692', | 'ле': '0.88692', | 'ус': '0.88692', | |
| 'ач': '0.11099', | 'им': '0.11099', | 'он': '1.10988', | 'чи': '0.11099', | 'ва': '0.66519', | 'ли': '0.88692', | 'уч': '0.22173', | |
| 'аш': '0.11099', | 'ин': '0.99889', | 'оо': '0.11099', | 'чп': '0.11099', | 'ве': '0.44346', | 'ло': '0.88692', | 'уш': '0.66519', | |
| 'ая': '0.22198', | 'іо': '0.11099', | 'оп': '0.55494', | 'шп': '0.66593', | 'ви': '0.22173', | 'лу': '0.22173', | 'фе': '0.22173', | |
| 'ба': '0.11099', | 'ип': '0.44395', | 'ор': '1.22087', | 'ши': '0.11099', | 'вл': '0.22173', | 'ма': '0.44346', | 'фо': '0.22173', | |
| 'бл': '0.22198', | 'ис': '0.99889', | 'ос': '1.22087', | 'шк': '0.66593', | 'вн': '0.22173', | 'ми': '0.22173', | 'ха': '0.44346', | |
| 'бр': '0.44395', | 'ит': '0.99889', | 'от': '0.33296', | 'ша': '0.11099', | 'во': '1.77384', | 'мо': '0.44346', | 'хд': '0.22173', | |
| 'бу': '0.11099', | 'іх': '0.22198', | 'оз': '0.22198', | 'ше': '0.11099', | 'вр': '0.44346', | 'мр': '0.22173', | 'хи': '0.22173', | |
| 'ва': '0.44395', | 'иц': '0.11099', | 'оя': '0.11099', | 'ьи': '0.11099', | 'вс': '0.44346', | 'мы': '0.22173', | 'хл': '0.22173', | |
| 'ве': '0.44395', | 'ич': '0.66593', | 'пе': '0.44395', | 'ык': '0.11099', | 'вт': '0.44346', | 'на': '0.88692', | 'хо': '0.22173', | |
| 'ви': '0.55494', | 'ию': '0.11099', | 'пи': '0.22198', | 'ым': '0.11099', | 'вш': '0.22173', | 'нв': '0.22173', | 'ху': '0.22173', | |
| 'вл': '0.22198', | 'ія': '0.22198', | 'по': '0.66593', | 'ыр': '0.11099', | 'га': '0.22173', | 'нд': '0.22173', | 'ци': '0.66519', | |
| 'вн': '0.22198', | 'йа': '0.11099', | 'пр': '0.88790', | 'ых': '0.44395', | 'ге': '0.66519', | 'не': '1.10865', | 'ча': '0.22173', | |
| 'во': '0.99889', | 'йв': '0.11099', | 'пу': '0.88790', | 'ье': '0.11099', | 'ги': '0.22173', | 'нж': '0.22173', | 'че': '0.22173', | |
| 'вр': '0.22198', | 'йи': '0.22198', | 'ра': '2.33074', | 'ьн': '0.11099', | 'го': '3.10421', | 'ни': '1.10865', | 'чи': '0.22173', | |
| 'вс': '0.22198', | 'йл': '0.11099', | 'рб': '0.11099', | 'эт': '0.22198', | 'гр': '0.22173', | 'нм': '0.22173', | 'чп': '0.22173', | |
| 'вт': '0.22198', | 'йо': '0.11099', | 'рв': '0.11099', | 'юн': '0.11099', | 'да': '0.66519', | 'нн': '0.88692', | 'ше': '1.10865', | |
| 'ву': '0.11099', | 'йп': '0.22198', | 'рг': '0.44395', | 'яв': '0.22198', | 'де': '0.66519', | 'но': '1.77384', | 'ши': '0.22173', | |
| 'вх': '0.11099', | 'йр': '0.11099', | 'ре': '0.77691', | 'яз': '0.11099', | 'до': '0.22173', | 'нс': '0.22173', | 'шк': '0.66519', | |
| 'вш': '0.55494', | 'йт': '0.11099', | 'ри': '0.88790', | 'ян': '0.11099', | 'др': '0.66519', | 'нц': '0.22173', | 'ьр': '0.22173', | |
| 'вы': '0.11099', | 'йф': '0.11099', | 'рм': '0.11099', | 'як': '0.22198', | 'еа': '0.44346', | 'ны': '0.66519', | 'ых': '0.22173', | |
| 'га': '0.11099', | 'йш': '0.11099', | 'рн': '0.22198', | 'ям': '0.11099', | 'ев': '0.66519', | 'ня': '0.22173', | 'ье': '0.22173', | |
| 'ге': '0.55494', | 'ка': '0.66593', | 'ро': '0.88790', | 'яи': '0.22198', | 'ег': '0.66519', | 'ов': '0.22173', | 'ьн': '0.22173', | |
| 'ги': '0.22198', | 'кв': '0.11099', | 'рс': '0.11099', | 'яс': '0.11099', | 'ед': '0.22173', | 'ог': '1.33038', | 'эт': '0.22173', | |
| 'го': '2.44173', | 'кз': '0.11099', | 'рз': '0.44395', | 'яу': '0.11099', | 'еи': '0.22173', | 'од': '1.10865', | 'яв': '0.22173', | |
| 'гр': '0.11099', | 'ки': '0.88790', | 'ры': '0.11099', | 'яф': '0.11099', | 'еи': '0.22173', | 'ое': '0.22173', | 'як': '0.44346', | |
| 'да': '0.33296', | 'кл': '0.22198', | 'ря': '0.22198', | | 'ек': '0.22173', | 'ож': '0.66519', | 'ян': '0.22173', | |
| 'дв': '0.11099', | 'км': '0.11099', | 'са': '0.55494', | | 'ел': '0.44346', | 'оз': '0.44346', | 'яс': '0.22173', | |
| 'де': '0.66593', | 'ко': '0.99889', | 'св': '0.33296', | | 'ем': '0.22173', | 'ои': '0.22173', | | |
| 'ди': '0.22198', | 'кп': '0.11099', | 'се': '0.22198', | | 'ен': '1.33038', | 'ой': '0.66519', | | |
| 'лк': '0.11099', | 'кр': '0.33296', | 'ск': '1.10988', | | 'ео': '0.22173', | 'ом': '0.88692', | | |
| 'дн': '0.11099', | 'кс': '0.33296', | 'сл': '0.55494', | | 'еп': '0.22173', | 'он': '0.66519', | | |
| 'до': '0.22198', | 'кт': '0.11099', | 'см': '0.11099', | | 'ер': '1.33038', | 'оп': '0.44346', | | |
| 'др': '0.44395', | 'ла': '0.33296', | 'сн': '0.33296', | | 'ес': '1.10865', | 'ор': '0.88692', | | |
| 'еа': '0.33296', | 'лв': '0.11099', | 'со': '0.11099', | | 'ет': '1.33038', | 'ос': '0.88692', | | |
| 'ев': '0.55494', | 'ле': '0.88790', | 'сп': '0.11099', | | 'ец': '0.22173', | 'от': '0.66519', | | |
| 'ег': '0.99889', | 'ли': '0.66593', | 'сз': '0.55494', | | 'еч': '0.22173', | 'оз': '0.22173', | | |
| 'ед': '0.22198', | 'ло': '0.99889', | 'ст': '1.33185', | | 'еи': '0.22173', | 'оя': '0.22173', | | |
| 'ее': '0.33296', | 'лс': '0.11099', | 'сх': '0.22198', | | 'ея': '0.22173', | 'пе': '0.22173', | | |
| 'еи': '0.11099', | 'лу': '0.22198', | 'сь': '0.11099', | | 'же': '0.22173', | 'пи': '0.22173', | | |
| 'ей': '0.33296', | 'ль': '0.11099', | 'ся': '0.22198', | | 'жи': '0.22173', | 'по': '0.88692', | | |
| 'ек': '0.44395', | 'ля': '0.11099', | 'та': '0.44395', | | 'жн': '0.22173', | 'пр': '1.10865', | | |
| 'ел': '0.44395', | 'ма': '0.55494', | 'тв': '0.33296', | | 'зв': '0.22173', | 'пу': '0.88692', | | |
| 'ем': '0.11099', | 'мв': '0.11099', | 'тд': '0.11099', | | 'зн': '0.22173', | 'ра': '1.77384', | | |
| 'ен': '1.10988', | 'ме': '0.22198', | 'те': '1.10988', | | 'зо': '0.22173', | 'рб': '0.22173', | | |
| 'ео': '0.33296', | 'ми': '0.22198', | 'ти': '0.99889', | | 'зы': '0.22173', | 'ре': '0.44346', | | |
| 'еп': '0.33296', | 'мо': '0.22198', | 'тн': '0.44395', | | 'иб': '0.22173', | 'ри': '1.10865', | | |
| 'ер': '1.22087', | 'мп': '0.11099', | 'то': '0.77691', | | 'ив': '0.88692', | 'рм': '0.22173', | | |
| 'ес': '0.77691', | 'му': '0.11099', | 'тп': '0.11099', | | 'иг': '0.22173', | 'рн': '0.22173', | | |
| 'ет': '1.10988', | 'мч': '0.11099', | 'тр': '0.55494', | | 'ид': '0.44346', | 'ро': '0.66519', | | |
| 'еи': '0.11099', | 'мы': '0.11099', | 'тс': '0.11099', | | 'із': '0.22173', | 'рс': '0.22173', | | |
| 'еч': '0.11099', | 'на': '0.55494', | 'ту': '0.55494', | | 'ни': '0.22173', | 'ру': '0.22173', | | |
| 'ещ': '0.11099', | 'нв': '0.22198', | 'тш': '0.11099', | | 'ий': '0.22173', | 'ры': '0.22173', | | |
| 'ея': '0.11099', | 'нд': '0.33296', | 'уб': '0.11099', | | 'нк': '0.44346', | 'ря': '0.44346', | | |
| 'жд': '0.11099', | 'не': '0.88790', | 'уг': '0.11099', | | 'ил': '0.22173', | 'са': '0.66519', | | |
| 'же': '0.22198', | 'нж': '0.11099', | 'уд': '0.11099', | | 'ин': '0.88692', | 'св': '0.22173', | | |
| 'жи': '0.44395', | 'ни': '0.99889', | 'уж': '0.22198', | | 'ип': '0.44346', | 'ск': '1.10865', | | |
| 'жн': '0.11099', | 'нк': '0.11099', | 'ул': '0.11099', | | 'ис': '0.66519', | 'сл': '0.66519', | | |
| 'жу': '0.11099', | 'нм': '0.11099', | 'ун': '0.11099', | | 'ит': '0.66519', | 'сн': '0.22173', | | |

Ентропія:

| | Ентропія | Надлишковість |
|-------------------------------------|-------------------------------------|---|
| Монограми без пропусків | 4.33447938551065 | 0.13310412289787 |
| Монограми з пропуском | 4.332005999067091 | 0.13359880018658177 |
| Біграми з повторенням без пропусків | 3.780136313409568 | 0.24397273731808644 |
| Біграми з повторенням з пропусками | 3.749933547117249 | 0.2500132905765502 |
| Біграми без повторень без пропусків | 3.6392783349042186 | 0.2721443330191563 |
| Біграми без повторень з пропусками | 3.620206555464458 | 0.2759586889071084 |
| H(10) | 2.75209746140447<H<3.27790722401634 | 0.44958050771910596<R<0.34441855519673203 |
| H(20) | 2.84796920139347<H<3.36617789503918 | 0.43040615972132<R<0.32676442099216396 |
| H(30) | 2.12762406575717<H<2.68397148189694 | 0.574475186848566<R<0.34441855519673203 |

CoolPinkProgram:

Лабороторная работа №1



Произвольная часть текста:

или_люди

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

Неравенство для энтропии:

2,75209746140447< H < 3,27790722401634

Двоичная таблица угаданных символов:

00000001000000000000000000000000

10000000000000000000000000000000

00000000000000000000000001000000

00010000000000000000000000000000

00000000000000000001000000000000

Поле ввода символов:

Продолжить

Другой

Вероятности:

q[1] = 0,4285714

q[2] = 0,0612244

q[3] = 0,0204081

q[4] = 0,0408163

q[5] = 0

q[6] = 0,0408163

q[7] = 0,0408163

q[8] = 0,0204081

q[9] = 0

q[10] = 0

q[11] = 0

q[12] = 0

q[13] = 0,020408

q[14] = 0,061224

q[15] = 0

q[16] = 0

q[17] = 0,020408

q[18] = 0

q[19] = 0,020408

q[20] = 0

q[21] = 0,020408

q[22] = 0,020408

q[23] = 0

q[24] = 0,040816

q[25] = 0

q[26] = 0,020408

q[27] = 0,061224

q[28] = 0

q[29] = 0,020408

q[30] = 0

q[31] = 0,020408

q[32] = 0,020408

Строка состояния:

| | |
|----------------------------|--|
| Произвольная часть текста: | |
| ьное_состоит_в_след | |
| Использованные буквы: | |
| | |

| | | |
|-------------------|------------------------|---|
| Порядок n-граммы: | Введенный символ: | Неравенство для энтропии: |
| 5 символов | | $2,84796920139347 < H < 3,36617789503918$ |
| 10 символов | Символ по счету: | |
| 15 символов | | |
| 20 символов | Номер эксперимента: 50 | Двоичная таблица угаданных символов: |
| 25 символов | | 1000000000000000000000000000000000 ^ |
| 30 символов | Поле ввода символов: | 0000000000000000010000000000000000 |
| 35 символов | | 1000000000000000000000000000000000 |
| 40 символов | | 0000000000000000000000000000000000 |
| 45 символов | | 0000000000000000000000000000000000 |
| 50 символов | Продолжить | Другой |

| |
|-------------------|
| Строка состояния: |
| |

| | |
|---|--|
| Произвольная часть текста: | |
| _что_были_усталыми_или_обеспо | |
| Использованные буквы: | |
| | |
| Порядок n-граммы: | |
| <input type="radio"/> 5 символов <input type="radio"/> 10 символов <input type="radio"/> 15 символов <input type="radio"/> 20 символов <input checked="" type="radio"/> 25 символов <input type="radio"/> 35 символов <input type="radio"/> 40 символов <input type="radio"/> 45 символов <input type="radio"/> 50 символов | |
| Введенный символ: | |
| Символ по счету: | |
| Номер эксперимента: 50 | |
| Поле ввода символов: | |
| <input type="button" value="Продолжить"/> <input type="button" value="Другой"/> | |
| Неравенство для энтропии: | |
| 2,12762406575717 < H < 2,68397148189694 | |
| Двоичная таблица угаданных символов: | |
| <pre> 000000000000000000000000000000 ^ 010000000000000000000000000000 000000100000000000000000000000 v 100000000000000000000000000000 000000000000000001000000000000 ~::~:::::::::::::::::::::::::: </pre> | |
| Вероятности: | |
| q[1] = 0,5102040 | |
| q[2] = 0,1224489 | |
| q[3] = 0,0204081 | |
| q[4] = 0,0204081 | |
| q[5] = 0 | |
| q[6] = 0,0408163 | |
| q[7] = 0,0204081 | |
| q[8] = 0 | |
| q[9] = 0 | |
| q[10] = 0,020408 | |
| q[11] = 0 | |
| q[12] = 0 | |
| q[13] = 0 | |
| q[14] = 0 | |
| q[15] = 0 | |
| q[16] = 0 | |
| q[17] = 0,040816 | |
| q[18] = 0,020408 | |
| q[19] = 0,081632 | |
| q[20] = 0 | |
| q[21] = 0,020408 | |
| q[22] = 0 | |
| q[23] = 0 | |
| q[24] = 0 | |
| q[25] = 0,020408 | |
| q[26] = 0,020408 | |
| q[27] = 0,020408 | |
| q[28] = 0 | |
| q[29] = 0 | |
| q[30] = 0,020408 | |
| q[31] = 0 | |
| q[32] = 0 | |
| Строка состояния: | |
| | |

Висновки:

Під час комп'ютерного практикуму №1, я зосвоїла поняття ентропії на символ джерела та надлишковості. Написала програму для підрахунку частоти, ентропії і надлишковості на прикладі російського тексту. Наприклад, за результатами частоти букв буква «о» найбільшу частоту, що означає, що ця буква зустрічається найчастіше у відкритому тексті.