

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ
ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
Фізико-технічний інститут

КРИПТОГРАФІЯ
КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

**Експериментальна оцінка ентропії на символ джерела
відкритого тексту**

Варіант №11

Виконали:

ФБ-32 Пінькас Б. О.

ФБ-32 Драчук О. І.

Київ 2025

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

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

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

Хід роботи

У якості довільного тексту російською мовою достатньої довжини, було обрано “Шерлок Холмс. Великий збірник” Артура Конана Дойля (Sherlock_Holmes.txt, довжини 3 Мб)

Для підрахунку частот монограм і біграм, а також для підрахунку H_1 , H_2 , надлишковості для різних джерел, було написано скрипт lab1.py.

Скрипт для кожної варіації (монограма/біграма, з пробілами/без пробілів, з перетинами/без перетинів) виводить у вигляді датафрейму кількість входжень n-грам, частоту n-грам, у вигляді рядка ентропію та надлишковість

Результати підрахунку кількості n-грам для кожного випадку окремим листом зберігаються в файл counts.xlsx, результати підрахунку частот n-грам - в файл frequencies.xlsx

У підсумку, подаються зібрані ентропії та надлишковості, включно з результатами, отриманими за допомогою програми CoolPinkProgram

Перед безпосереднім запуском скрипта, рекомендується встановити необхідні бібліотеки (pip install -r "requirements.txt")

Перед обробкою, текст пройшов попередню фільтрацію, всі символи, окрім літер російського алфавіту і пробілів було замінено на пробіли, після чого зайві пробіли або фільтруватися, якщо обрано алфавіт з пробілами, або видалялись, якщо алфавіт не містить пробілів. Літеру "ё" можна опціонально замінити на "е", як це і було зроблено в коді (літера "ё" порівняно з іншими, зустрічається дуже рідко, наприклад, "ъ" - 719, "ф" - 3770, тоді як "ё" - 3, тому, в ході експерименту, було вирішено нею знехтувати)

Таблиця частот монограм без пробілів:

Символ	Частота
о	0,1148
е	0,0867
а	0,0757
н	0,0673
и	0,0631
т	0,0621
с	0,057
л	0,0491
в	0,0427
р	0,0424
м	0,036
к	0,034
д	0,0312
у	0,0283
п	0,0264
я	0,0216
ы	0,019
ь	0,0188
г	0,0175
з	0,0166
б	0,0164
ч	0,0156
й	0,0118
ж	0,0102
х	0,0091
ш	0,0082
ю	0,0056
э	0,0046
ц	0,0031
щ	0,0031
ф	0,0016
ъ	0,0003

Таблиця частот монограм з пробілами:

Символ	Частота
—	0,1649
о	0,0959
е	0,0724
а	0,0632
н	0,0562
и	0,0527
т	0,0519
с	0,0476
л	0,041
в	0,0356
р	0,0354
м	0,0301
к	0,0284
д	0,0261
у	0,0236
п	0,0221
я	0,018
ы	0,0159
ь	0,0157
г	0,0146
з	0,0138
б	0,0137
ч	0,013
й	0,0098
ж	0,0085
х	0,0076
ш	0,0069
ю	0,0047
э	0,0038
ц	0,0026
щ	0,0026
ф	0,0013
ъ	0,0002

Таблиця частот біграм без пробілів і з перетинами:

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	0,0002	0,0015	0,0046	0,0013	0,0032	0,0019	0,0014	0,0045	0,0012	0,0011	0,0064	0,0087	0,0052	0,0058	0,0012
б	0,0009	0,0001	0,0001	0	0	0,0021	0	0	0,001	0	0,0002	0,0009	0,0001	0,0004	0,0025
в	0,0066	0,0002	0,0003	0,0003	0,0007	0,0065	0	0,0006	0,0038	0	0,0009	0,0011	0,0004	0,0017	0,0067
г	0,0014	0,0001	0,0001	0	0,0015	0,0005	0	0	0,0008	0	0,0002	0,0016	0,0001	0,0004	0,0087
д	0,005	0,0001	0,0013	0	0,0001	0,006	0,0005	0,0001	0,0027	0	0,0004	0,0008	0,0001	0,0024	0,0053
е	0,0002	0,0019	0,0035	0,0046	0,0048	0,002	0,0012	0,0022	0,0012	0,0026	0,0025	0,0071	0,0058	0,011	0,0018
ж	0,0014	0	0	0	0,001	0,0041	0	0	0,0016	0	0,0002	0	0	0,0012	0,0002
з	0,0062	0,0002	0,0011	0,0005	0,001	0,0004	0,0001	0,0001	0,0004	0	0,0003	0,0004	0,0005	0,0022	0,0007
и	0,0003	0,0015	0,0042	0,0008	0,0028	0,0029	0,0004	0,003	0,0017	0,0013	0,0032	0,0058	0,0041	0,0055	0,0015
й	0,0001	0,0003	0,0007	0,0002	0,0009	0,0001	0,0002	0,0002	0,0005	0	0,0008	0,0002	0,0006	0,0012	0,0005
к	0,0084	0,0002	0,0008	0,0001	0,0002	0,0009	0,0001	0,0002	0,003	0	0,0004	0,001	0,0003	0,0012	0,01
л	0,0053	0,0003	0,0009	0,0003	0,0004	0,0055	0,0004	0,0002	0,0083	0	0,0008	0,0004	0,0018	0,0011	0,008
м	0,0028	0,0004	0,0009	0,0002	0,0006	0,0051	0,0001	0,0002	0,0042	0	0,0006	0,0003	0,0006	0,0036	0,0059
н	0,0126	0,0003	0,0006	0,0004	0,0008	0,0118	0	0,0002	0,0088	0	0,0006	0,0001	0,0002	0,0036	0,0122
о	0,0002	0,0055	0,0097	0,0059	0,0063	0,0034	0,0028	0,0023	0,0023	0,0046	0,0042	0,0083	0,0082	0,0106	0,0023
п	0,0015	0	0	0	0	0,0022	0	0	0,0013	0	0,0001	0,0008	0	0,0003	0,0108
р	0,0082	0,0002	0,0007	0,0003	0,0006	0,0067	0,0004	0,0001	0,0053	0	0,0004	0,0004	0,0004	0,0012	0,0078
с	0,0019	0,0003	0,0023	0,0001	0,0005	0,004	0,0001	0,0001	0,002	0	0,0048	0,004	0,0012	0,0019	0,0041
т	0,0065	0,0003	0,0032	0,0001	0,0004	0,0074	0,0001	0,0002	0,0046	0	0,0011	0,0005	0,0004	0,0022	0,0163
у	0,0002	0,001	0,0017	0,0014	0,0026	0,0004	0,0017	0,0006	0,0006	0,0002	0,0013	0,0017	0,0019	0,0014	0,0009
ф	0,0002	0	0	0	0	0,0003	0	0	0,0002	0	0	0	0	0	0,0002
х	0,0007	0,0001	0,0005	0,0001	0,0002	0,0001	0	0,0001	0,0005	0	0,0002	0,0002	0,0002	0,0005	0,0039
ц	0,0007	0	0,0001	0	0	0,0008	0	0	0,0005	0	0,0001	0	0	0	0,0004
ч	0,0025	0	0	0	0	0,0044	0	0	0,0018	0	0,0003	0	0	0,0011	0,0001
ш	0,001	0	0	0	0	0,0027	0	0	0,0017	0	0,0004	0,0007	0	0,0004	0,0004
щ	0,0004	0	0	0	0	0,0014	0	0	0,0009	0	0	0	0	0,0001	0
ъ	0	0	0	0	0	0,0001	0	0	0	0	0	0	0	0	0
ы	0,0001	0,0004	0,0016	0,0002	0,0004	0,0013	0,0001	0,0003	0,0004	0,0019	0,0004	0,0025	0,0016	0,001	0,0005
ь	0,0001	0,0005	0,0015	0,0002	0,0005	0,0008	0,0001	0,0005	0,0007	0	0,0019	0,0002	0,001	0,0027	0,0008
э	0	0,0001	0	0	0	0	0	0	0	0	0,0002	0,0001	0,0001	0,0002	0
ю	0	0,0004	0,0004	0,0001	0,0005	0,0001	0,0001	0,0001	0,0002	0	0,0003	0,0001	0,0002	0,0003	0,0002
я	0,0001	0,0006	0,0021	0,0003	0,0012	0,0005	0,0003	0,0008	0,0008	0,0001	0,0009	0,0011	0,0009	0,0024	0,0009

п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
0,0026	0,0027	0,0067	0,0059	0,0007	0,0002	0,0013	0,0001	0,0015	0,0016	0,0003	0	0	0	0,0004	0,0011	0,0025
0	0,0014	0,0003	0	0,0013	0	0,0001	0	0	0	0,0002	0,0002	0,0042	0	0,0002	0	0,0003
0,0009	0,001	0,0034	0,0008	0,0008	0	0,0001	0,0001	0,0002	0,0007	0	0	0,0038	0,0002	0,0003	0	0,0004
0,0001	0,001	0,0001	0,0001	0,0007	0	0	0	0,0001	0	0	0	0	0	0	0	0
0,0003	0,0013	0,0006	0,0002	0,0017	0	0,0001	0,0002	0,0001	0,0001	0	0	0,0006	0,0007	0	0	0,0004
0,0032	0,009	0,0079	0,0071	0,0009	0,0001	0,0009	0,0005	0,002	0,0008	0,0008	0	0	0	0,0003	0,0002	0,0006
0	0	0	0	0,0003	0	0	0	0,0001	0	0	0	0	0	0	0	0
0,0002	0,0004	0,0002	0,0001	0,0006	0	0	0	0,0001	0	0	0	0,0004	0,0001	0,0001	0	0,0004
0,0022	0,0011	0,0059	0,0057	0,0006	0,0001	0,0019	0,0011	0,0017	0,0006	0,0002	0	0	0	0,0003	0,0004	0,002
0,0008	0,0003	0,0015	0,0009	0,0002	0,0001	0,0001	0,0001	0,0005	0,0002	0	0	0	0	0,0001	0	0,0001
0,0003	0,0022	0,0007	0,0012	0,0019	0	0,0002	0	0,0002	0	0	0	0	0	0,0002	0	0,0002
0,0006	0,0001	0,0024	0,0005	0,002	0	0,0004	0	0,0006	0,0001	0	0	0,0008	0,0047	0,0001	0,001	0,0019
0,0012	0,0003	0,0026	0,0004	0,0026	0	0,0001	0	0,0005	0,0001	0	0	0,0017	0,0001	0,0002	0	0,0008
0,0006	0,0003	0,0013	0,0012	0,0033	0	0,0001	0,0003	0,0004	0	0,0002	0	0,0038	0,0012	0,0001	0,0002	0,0021
0,004	0,0067	0,0094	0,0089	0,001	0,0003	0,0007	0,0001	0,0031	0,0012	0,0003	0	0	0	0,0005	0,0003	0,0017
0	0,0075	0,0001	0,0001	0,0009	0	0	0	0	0	0	0	0,0004	0	0	0	0,0003
0,0003	0,0002	0,0006	0,001	0,0029	0,0001	0,0004	0,0001	0,0001	0,0003	0	0	0,0016	0,0007	0,0001	0,0002	0,0011
0,0025	0,0004	0,0018	0,0136	0,0011	0	0,0003	0	0,0006	0,0002	0	0	0,0004	0,0035	0,0005	0,0002	0,0045
0,0007	0,0035	0,0025	0,0004	0,002	0	0,0001	0,0001	0,0004	0	0	0	0,0012	0,0068	0,0003	0,0001	0,0007
0,0015	0,0007	0,0021	0,002	0,0002	0	0,0005	0	0,0015	0,0006	0,0003	0	0	0	0,0002	0,0012	0,0002
0	0,0002	0	0	0,0001	0	0	0	0	0	0	0	0	0	0	0	0
0,0004	0,0002	0,0004	0,0001	0,0002	0	0	0	0,0001	0	0	0	0	0	0,0001	0	0,0001
0	0	0	0	0,0002	0	0	0	0	0	0	0	0,0002	0	0	0	0
0	0	0	0,0042	0,0005	0	0	0	0	0,0001	0	0	0	0,0004	0	0	0
0,0001	0	0	0,0001	0,0003	0	0	0	0	0	0	0	0	0,0002	0	0	0
0	0	0	0	0,0001	0	0	0	0	0	0	0	0	0,0001	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0002
0,0009	0,0004	0,0014	0,0011	0,0003	0	0,0012	0	0,0003	0,0006	0,0001	0	0	0	0,0001	0	0,0002
0,0009	0,0002	0,0022	0,0006	0,0003	0,0001	0,0002	0,0001	0,0007	0,0007	0	0	0	0	0,0002	0,0005	0,0005
0,0001	0,0006	0	0,0032	0	0	0	0	0	0	0	0	0	0	0	0	0
0,0003	0,0001	0,0006	0,0005	0,0001	0	0,0001	0	0,0004	0,0001	0,0004	0	0	0	0,0001	0	0,0001
0,0016	0,0005	0,002	0,002	0,0006	0	0,0004	0,0001	0,0005	0,0001	0,0003	0	0	0	0,0002	0,0002	0,0003

Таблиця частот біграм без пробілів і без перетинів:

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	0,0002	0,0015	0,0046	0,0014	0,0031	0,0019	0,0014	0,0045	0,0012	0,0011	0,0064	0,0087	0,0053	0,0058	0,0013
б	0,0009	0	0,0001	0	0	0,0022	0	0	0,001	0	0,0001	0,0009	0,0001	0,0005	0,0024
в	0,0066	0,0002	0,0003	0,0003	0,0007	0,0065	0	0,0006	0,0037	0	0,0009	0,0011	0,0004	0,0017	0,0067
г	0,0014	0,0001	0,0001	0	0,0015	0,0005	0	0	0,0008	0	0,0001	0,0016	0,0001	0,0004	0,0086
д	0,0051	0,0001	0,0013	0,0001	0,0001	0,0059	0,0005	0,0001	0,0027	0	0,0004	0,0008	0,0001	0,0024	0,0053
е	0,0002	0,0019	0,0035	0,0046	0,0048	0,002	0,0011	0,0023	0,0012	0,0026	0,0024	0,0072	0,0058	0,011	0,0018
ж	0,0014	0	0	0	0,001	0,0042	0	0	0,0016	0	0,0002	0	0	0,0011	0,0002
з	0,0062	0,0002	0,0011	0,0005	0,001	0,0004	0,0001	0,0001	0,0004	0	0,0003	0,0004	0,0005	0,0022	0,0007
и	0,0003	0,0015	0,0043	0,0009	0,0028	0,0029	0,0004	0,0028	0,0016	0,0013	0,0031	0,0058	0,0041	0,0056	0,0015
й	0,0001	0,0003	0,0007	0,0002	0,001	0,0001	0,0002	0,0002	0,0005	0	0,0008	0,0002	0,0006	0,0012	0,0006
к	0,0084	0,0003	0,0008	0,0001	0,0003	0,0009	0,0001	0,0001	0,0031	0	0,0004	0,001	0,0003	0,0011	0,0101
л	0,0053	0,0003	0,0009	0,0003	0,0004	0,0055	0,0004	0,0002	0,0084	0	0,0008	0,0004	0,0018	0,0011	0,008
м	0,0028	0,0004	0,0009	0,0002	0,0006	0,0052	0,0001	0,0002	0,0042	0	0,0006	0,0004	0,0006	0,0035	0,0059
н	0,0127	0,0003	0,0006	0,0004	0,0008	0,0119	0	0,0002	0,0089	0	0,0007	0,0001	0,0002	0,0035	0,0123
о	0,0002	0,0054	0,0097	0,0059	0,0064	0,0034	0,0028	0,0022	0,0023	0,0045	0,0043	0,0084	0,0081	0,0106	0,0022
п	0,0015	0	0	0	0	0,0021	0	0	0,0013	0	0,0001	0,0008	0	0,0003	0,0107
р	0,0081	0,0002	0,0007	0,0003	0,0006	0,0066	0,0004	0,0001	0,0053	0	0,0004	0,0004	0,0004	0,0012	0,0077
с	0,002	0,0003	0,0022	0,0001	0,0005	0,0039	0,0001	0,0001	0,002	0	0,0047	0,004	0,0012	0,0018	0,0041
т	0,0065	0,0003	0,0032	0,0001	0,0004	0,0074	0,0001	0,0001	0,0047	0	0,001	0,0005	0,0004	0,0022	0,0164
у	0,0002	0,001	0,0017	0,0014	0,0025	0,0004	0,0017	0,0005	0,0006	0,0002	0,0013	0,0017	0,0019	0,0014	0,0009
ф	0,0003	0	0	0	0	0,0003	0	0	0,0002	0	0	0	0	0	0,0002
х	0,0007	0,0001	0,0005	0,0001	0,0002	0,0001	0	0,0001	0,0005	0	0,0002	0,0002	0,0002	0,0005	0,0039
ц	0,0007	0	0,0001	0	0	0,0009	0	0	0,0005	0	0,0001	0	0	0	0,0004
ч	0,0025	0	0	0	0	0,0044	0	0	0,0018	0	0,0003	0	0	0,0011	0,0001
ш	0,001	0	0	0	0	0,0027	0	0	0,0017	0	0,0004	0,0007	0	0,0004	0,0004
щ	0,0004	0	0	0	0	0,0014	0	0	0,0009	0	0	0	0	0,0001	0
ъ	0	0	0	0	0	0,0001	0	0	0	0	0	0	0	0	0
ы	0,0001	0,0004	0,0016	0,0002	0,0004	0,0013	0,0001	0,0003	0,0004	0,0018	0,0004	0,0025	0,0016	0,001	0,0005
ь	0,0001	0,0005	0,0015	0,0002	0,0006	0,0009	0,0001	0,0006	0,0007	0	0,0019	0,0002	0,0011	0,0027	0,0008
э	0	0,0001	0	0	0	0	0	0	0	0	0,0002	0,0001	0,0001	0,0002	0
ю	0	0,0004	0,0003	0,0001	0,0005	0,0001	0,0001	0,0001	0,0002	0	0,0003	0,0001	0,0002	0,0003	0,0002
я	0,0001	0,0006	0,0021	0,0003	0,0012	0,0005	0,0003	0,0008	0,0008	0,0001	0,0009	0,0011	0,0009	0,0024	0,0009

п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
0,0027	0,0027	0,0067	0,0058	0,0007	0,0002	0,0012	0,0001	0,0015	0,0016	0,0002	0	0	0	0,0004	0,0011	0,0025
0	0,0014	0,0003	0	0,0013	0	0,0001	0	0	0	0,0002	0,0002	0,0042	0	0,0002	0	0,0003
0,0009	0,001	0,0035	0,0008	0,0008	0	0,0001	0	0,0002	0,0007	0	0	0,0037	0,0001	0,0002	0	0,0004
0,0001	0,001	0,0001	0,0001	0,0007	0	0	0	0,0001	0	0	0	0	0	0	0	0
0,0003	0,0013	0,0006	0,0002	0,0017	0	0,0001	0,0002	0,0001	0,0001	0	0	0,0006	0,0007	0	0	0,0004
0,0032	0,0091	0,008	0,0071	0,0009	0,0002	0,0009	0,0005	0,002	0,0008	0,0008	0	0	0	0,0003	0,0002	0,0006
0	0	0	0	0,0003	0	0	0	0	0	0	0	0	0	0	0	0
0,0002	0,0004	0,0002	0,0001	0,0006	0	0	0	0,0001	0	0	0	0,0004	0,0001	0	0	0,0004
0,0022	0,0011	0,0058	0,0058	0,0006	0,0001	0,0019	0,0011	0,0017	0,0006	0,0002	0	0	0	0,0003	0,0004	0,002
0,0008	0,0003	0,0015	0,0009	0,0003	0,0001	0,0001	0,0001	0,0005	0,0002	0	0	0	0	0,0001	0	0,0001
0,0003	0,0022	0,0006	0,0012	0,002	0,0001	0,0002	0	0,0002	0	0	0	0	0	0,0002	0	0,0002
0,0006	0,0002	0,0025	0,0005	0,002	0	0,0004	0	0,0006	0,0001	0	0	0,0008	0,0047	0,0001	0,0009	0,0019
0,0012	0,0003	0,0026	0,0004	0,0025	0	0,0001	0	0,0005	0,0001	0	0	0,0017	0,0001	0,0002	0	0,0008
0,0006	0,0003	0,0013	0,0011	0,0033	0	0,0001	0,0003	0,0004	0	0,0002	0	0,0039	0,0011	0,0001	0,0002	0,002
0,004	0,0068	0,0095	0,0089	0,001	0,0003	0,0007	0,0001	0,0032	0,0012	0,0003	0	0	0	0,0005	0,0003	0,0016
0	0,0076	0,0001	0,0001	0,0009	0	0	0	0	0	0	0	0,0004	0,0001	0	0	0,0003
0,0003	0,0002	0,0006	0,001	0,0029	0,0001	0,0004	0,0001	0,0001	0,0004	0	0	0,0016	0,0007	0,0001	0,0002	0,001
0,0025	0,0004	0,0018	0,0135	0,0012	0	0,0003	0,0001	0,0006	0,0002	0	0	0,0004	0,0035	0,0005	0,0002	0,0044
0,0007	0,0036	0,0025	0,0004	0,002	0	0,0001	0,0001	0,0004	0	0	0	0,0011	0,0068	0,0003	0,0001	0,0007
0,0014	0,0007	0,0021	0,002	0,0002	0	0,0005	0	0,0015	0,0005	0,0003	0	0	0	0,0002	0,0012	0,0002
0	0,0002	0	0,0001	0,0001	0	0	0	0	0	0	0	0	0	0	0	0
0,0004	0,0002	0,0004	0,0001	0,0002	0	0	0	0,0001	0	0	0	0	0	0,0001	0	0,0001
0	0	0	0	0,0002	0	0	0	0	0	0	0	0,0002	0	0	0	0
0	0	0	0,0042	0,0005	0	0	0	0	0,0001	0	0	0	0,0004	0	0	0
0,0001	0	0	0,0001	0,0003	0	0	0	0	0	0	0	0	0,0002	0	0	0
0	0	0	0	0,0001	0	0	0	0	0	0	0	0	0,0001	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0002
0,0009	0,0004	0,0014	0,0011	0,0003	0	0,0012	0	0,0003	0,0006	0,0001	0	0	0	0,0001	0	0,0002
0,0009	0,0002	0,0022	0,0006	0,0003	0	0,0002	0,0001	0,0007	0,0007	0	0	0	0	0,0002	0,0006	0,0005
0,0001	0,0006	0	0,0033	0	0	0	0	0	0	0	0	0	0	0	0	0
0,0003	0,0001	0,0007	0,0005	0,0001	0	0,0001	0	0,0004	0,0001	0,0004	0	0	0	0,0001	0	0,0001
0,0016	0,0005	0,0021	0,002	0,0006	0	0,0004	0,0001	0,0005	0,0001	0,0003	0	0	0	0,0002	0,0001	0,0003

Таблиця частот біграм з пробілами і з перетинами:

	а	б	в	г	д	ж	з	и	и	й	к	л	м	н	о
а	0,0018	0,0064	0,0165	0,0029	0,0076	0,0014	0,0045	0,0012	0,0086	0	0,0088	0,0027	0,0086	0,0161	0,0113
б	0,0153	0	0,0006	0,0025	0,0008	0,002	0	0	0,001	0,0001	0,0009	0,0045	0,007	0,0035	0,0032
в	0,0003	0,0007	0	0,0001	0	0	0,0006	0,0038	0,0008	0	0,0001	0,0007	0	0,0003	0,002
г	0,0057	0,0054	0	0	0	0,0003	0	0,0008	0,003	0	0,0002	0,0007	0,0001	0,0009	0,0053
д	0,0008	0,0011	0	0	0	0,0012	0,0005	0,0001	0,0027	0,0007	0	0,0001	0,0013	0	0,0002
е	0,0012	0,0042	0	0,001	0	0	0,0012	0,0022	0,0012	0,0022	0	0,0003	0,0006	0	0,0019
ж	0,018	0	0,0008	0,0012	0,0035	0,0032	0	0	0,0016	0,0001	0,0022	0,0013	0,0056	0,0038	0,0076
з	0,0003	0,0012	0	0	0	0,0008	0,0001	0,0004	0,0013	0	0,0001	0	0	0,0009	0,0001
и	0,0015	0,0051	0,0001	0,0008	0,0003	0,0007	0,0004	0,003	0,0017	0,0003	0	0,0001	0,0003	0,0003	0,0017
и	0,0157	0,0001	0,0005	0,0017	0,0004	0,0016	0,0002	0,0002	0,0005	0,0006	0,0011	0,0019	0,0046	0,0027	0,0031
й	0,0073	0	0	0	0	0,0003	0,0001	0,0002	0,003	0	0	0,0002	0,0001	0,0001	0,0004
к	0,0045	0,0069	0	0,0002	0	0	0,0004	0,0002	0,0083	0,0023	0	0	0,0008	0	0,0006
л	0,0071	0,0044	0	0	0,0001	0,0001	0,0001	0,0002	0,0042	0,0066	0	0,0004	0,0003	0,0012	0,0003
м	0,0082	0,0022	0,0001	0	0	0	0,0002	0,0088	0,003	0	0,0001	0,0001	0,0001	0,0022	0,0043
н	0,0044	0,0105	0	0	0,0003	0,0005	0,0028	0,0023	0,0023	0,0071	0	0,0003	0	0	0,0026
о	0,0225	0	0,0036	0,0056	0,0046	0,0042	0	0,0013	0,0009	0,0038	0,0026	0,0066	0,0055	0,0067	0,0002
п	0,0001	0,0012	0	0	0	0	0,0004	0,0001	0,0053	0,0011	0	0,0001	0,0007	0	0,0002
р	0,0019	0,0068	0,0001	0,0004	0,0002	0,0004	0,0001	0,0001	0,002	0,0043	0	0,0003	0,0003	0,0002	0,0009
с	0,0045	0,0016	0,0001	0,0014	0	0,0002	0,0001	0,0002	0,0046	0,0015	0	0,0037	0,0033	0,0008	0,0011
т	0,0056	0,0053	0	0,0021	0	0,0001	0,0017	0,0006	0,0006	0,0035	0	0,0006	0,0003	0	0,0013
у	0,0065	0,0001	0,0007	0,0007	0,0011	0,0019	0	0	0,0002	0,0001	0,0001	0,0007	0,0014	0,0011	0,0003
ф	0,0001	0,0002	0	0	0	0	0	0,0001	0,0005	0,0002	0	0	0	0	0,0002
х	0,003	0,0005	0	0,0002	0	0	0	0	0,0005	0,0002	0	0	0,0001	0	0,0001
ц	0,0003	0,0006	0	0	0	0	0	0	0,0018	0,0004	0	0	0	0	0,0003
ч	0,0001	0,0021	0	0	0	0	0	0	0,0017	0,0015	0	0,0003	0	0	0,0009
ш	0,0002	0,0008	0	0	0	0	0	0	0,0009	0,0014	0	0,0003	0,0006	0	0,0003
щ	0	0,0003	0	0	0	0	0	0	0,0008	0	0	0	0	0	0,0001
ъ	0	0	0	0	0	0	0,0001	0,0003	0,0004	0	0	0	0	0	0
ы	0,0054	0	0,0002	0,0008	0,0001	0,0001	0,0001	0,0005	0,0007	0	0,0016	0,0001	0,002	0,0011	0,0002
ь	0,0098	0	0,0001	0,0001	0	0	0	0	0	0	0	0,001	0	0,0004	0,0013
э	0	0	0,0001	0	0	0	0,0001	0,0001	0,0002	0	0	0,0001	0,0001	0	0,0002
ю	0,0027	0	0,0003	0	0	0,0003	0,0003	0,0008	0,0008	0	0	0	0	0	0
я	0,0119	0	0	0,0004	0,0001	0,0005				0,0001	0,0001	0,0002	0,0008	0,0003	0,0006

п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
0,0155	0,0036	0,0157	0,0068	0,0052	0,0007	0,0026	0,0003	0,0056	0,0009	0,0001	0	0	0	0,0028	0,0001	0,004
0,0008	0,002	0,0042	0,0043	0,0001	0,0001	0,0008	0,0001	0,0008	0,0013	0,0002	0	0	0	0	0,0009	0,0016
0	0,0011	0,0002	0	0,0011	0	0,0001	0	0	0	0,0002	0,0002	0,0035	0	0,0001	0	0,0002
0,0003	0,0006	0,0022	0,0003	0,0005	0	0	0	0,0001	0,0005	0	0	0,0032	0,0001	0	0	0,0003
0	0,0008	0	0	0,0006	0	0	0	0	0	0	0	0	0	0	0	0
0,0002	0,0011	0,0004	0,0001	0,0014	0	0	0,0002	0	0,0001	0	0	0,0005	0,0006	0	0	0,0003
0,0008	0,0071	0,0049	0,0052	0,0001	0	0,0005	0,0004	0,001	0,0006	0,0006	0	0	0	0	0,0002	0,0002
0	0	0	0	0,0002	0	0	0	0	0	0	0	0	0	0	0	0
0	0,0003	0	0	0,0004	0	0	0	0	0	0	0	0,0003	0,0001	0	0	0,0003
0,0002	0,0006	0,0035	0,0042	0	0	0,0014	0,0009	0,001	0,0005	0,0002	0	0	0	0	0,0003	0,0012
0	0	0,0005	0,0004	0	0	0	0,0001	0,0002	0,0001	0	0	0	0	0	0	0
0	0,0017	0,0001	0,0008	0,0015	0	0	0	0	0	0	0	0	0	0,0001	0	0
0	0	0,0014	0,0002	0,0015	0	0	0	0,0001	0	0	0	0,0007	0,0039	0	0,0008	0,0013
0,0001	0,0001	0,0014	0	0,0019	0	0	0	0,0001	0	0	0	0,0014	0,0001	0,0001	0	0,0004
0	0,0001	0,0006	0,0008	0,0025	0	0	0,0003	0,0002	0	0,0002	0	0,0032	0,001	0	0,0002	0,0017
0,0012	0,0052	0,0058	0,0064	0,0001	0,0002	0,0004	0,0001	0,0018	0,0009	0,0002	0	0	0	0,0001	0,0003	0,0008
0	0,0063	0	0,0001	0,0007	0	0	0	0	0	0	0	0,0003	0	0	0	0,0002
0,0001	0,0001	0,0003	0,0008	0,0023	0	0,0001	0,0001	0,0001	0,0003	0	0	0,0014	0,0006	0,0001	0,0001	0,0009
0,0016	0,0002	0,0011	0,0111	0,0008	0	0,0002	0	0,0004	0,0001	0	0	0,0003	0,003	0,0004	0,0002	0,0037
0,0002	0,0028	0,0015	0,0001	0,0015	0	0	0,0001	0,0001	0	0	0	0,001	0,0057	0,0001	0,0001	0,0005
0,0007	0,0004	0,0012	0,0015	0	0	0,0003	0	0,0009	0,0004	0,0002	0	0	0	0,0001	0,001	0
0	0,0001	0	0	0,0001	0	0	0	0	0	0	0	0	0	0	0	0
0	0,0001	0	0	0,0001	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0,0001	0	0	0	0	0	0	0	0,0001	0	0	0	0
0	0	0	0,0035	0,0005	0	0	0	0	0,0001	0	0	0	0,0003	0	0	0
0	0	0	0,0001	0,0003	0	0	0	0	0	0	0	0	0,0002	0	0	0
0	0	0	0	0,0001	0	0	0	0	0	0	0	0	0,0001	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0002
0,0001	0,0002	0,0006	0,0007	0	0	0,0009	0	0,0002	0,0005	0	0	0	0	0	0	0
0	0	0,0009	0,0002	0	0	0	0,0001	0,0001	0,0005	0	0	0	0	0	0,0004	0,0002
0,0001	0,0005	0	0,0027	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0,0001	0,0003	0,0003	0	0	0	0	0,0002	0	0,0003	0	0	0	0	0	0
0	0,0001	0,0006	0,0012	0	0	0,0002	0,0001	0,0001	0	0,0002	0	0	0	0	0,0001	0,0001

Таблиця частот біграм з пробілами і без перетинів:

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	0,0018	0,0064	0,0167	0,0029	0,0076	0,0038	0,0017	0,0045	0,0087	0	0,0088	0,0026	0,0086	0,0162	0,0113
б	0,0153	0	0,0006	0,0025	0,0008	0,002	0,0011	0,0011	0,0034	0,0001	0,0009	0,0045	0,007	0,0035	0,0032
в	0,0003	0,0007	0	0	0	0,0018	0	0	0,0008	0	0,0001	0,0007	0	0,0003	0,002
г	0,0056	0,0054	0	0	0	0,0003	0,0053	0	0,0004	0,003	0	0,0002	0,0007	0,0001	0,0009
д	0,0008	0,0011	0	0	0	0,0012	0,0003	0	0	0,0007	0	0,0001	0,0013	0	0,0002
е	0,0012	0,0041	0	0,001	0	0	0,0049	0,0004	0	0,0021	0	0,0003	0,0006	0	0,0019
ж	0,018	0	0,0007	0,0012	0,0035	0,0032	0,0014	0,0008	0,0013	0,0001	0,0023	0,0013	0,0056	0,0038	0,0077
з	0,0003	0,0012	0	0	0	0,0008	0,0035	0	0	0,0013	0	0,0001	0	0	0,0009
и	0,0015	0,0052	0,0001	0,0008	0,0003	0,0007	0,0003	0,0001	0	0,0003	0	0,0001	0,0003	0,0003	0,0017
й	0,0156	0,0001	0,0005	0,0017	0,0004	0,0016	0,002	0,0003	0,002	0,0007	0,0011	0,0019	0,0046	0,0027	0,0031
к	0,0072	0	0	0	0	0,0003	0	0	0	0	0,0002	0,0001	0,0001	0,0005	0,0001
л	0,0045	0,007	0	0,0002	0	0	0,0007	0	0	0,0023	0	0	0,0008	0	0,0006
м	0,0071	0,0044	0	0	0,0001	0,0001	0,0043	0,0003	0	0,0066	0	0,0004	0,0003	0,0012	0,0003
н	0,0082	0,0022	0,0001	0	0	0	0,0041	0	0	0,003	0	0,0001	0,0002	0,0001	0,0022
о	0,0044	0,0105	0	0	0,0003	0,0005	0,0098	0	0	0,007	0	0,0003	0	0	0,0025
п	0,0226	0	0,0036	0,0056	0,0047	0,0042	0,0024	0,002	0,0012	0,0009	0,0038	0,0026	0,0067	0,0055	0,0067
р	0,0001	0,0013	0	0	0	0	0,0018	0	0	0,0011	0	0,0001	0,0007	0	0,0002
с	0,0019	0,0067	0,0001	0,0004	0,0002	0,0004	0,0055	0,0003	0	0,0043	0	0,0003	0,0003	0,0002	0,0009
т	0,0045	0,0016	0,0001	0,0014	0	0,0002	0,0032	0	0	0,0015	0	0,0037	0,0033	0,0008	0,0011
у	0,0056	0,0053	0	0,0021	0	0,0001	0,0062	0	0	0,0036	0	0,0006	0,0003	0	0,0013
ф	0,0065	0,0001	0,0007	0,0007	0,0011	0,0019	0,0002	0,0013	0,0003	0,0001	0,0001	0,0007	0,0014	0,0011	0,0003
х	0,0001	0,0002	0	0	0	0	0,0003	0	0	0,0002	0	0	0	0	0,0002
ц	0,003	0,0005	0	0,0002	0	0	0,0001	0	0	0,0002	0	0	0,0001	0	0,0001
ч	0,0003	0,0006	0	0	0	0	0,0007	0	0	0,0004	0	0	0	0	0,0003
ш	0,0001	0,0021	0	0	0	0	0,0037	0	0	0,0014	0	0,0003	0	0	0,0009
щ	0,0002	0,0008	0	0	0	0	0,0022	0	0	0,0014	0	0,0003	0,0006	0	0,0003
ъ	0	0,0003	0	0	0	0	0,0012	0	0	0,0007	0	0	0	0	0,0001
ы	0	0	0	0	0	0	0,0001	0	0	0	0	0	0	0	0
ь	0,0055	0	0,0001	0,0008	0,0001	0,0001	0,0009	0	0,0001	0	0,0016	0,0001	0,002	0,0011	0,0002
э	0,0097	0	0,0001	0,0001	0	0	0,0003	0	0,0002	0	0	0,0011	0	0,0004	0,0013
ю	0	0	0,0001	0	0	0	0	0	0	0	0	0,0001	0,0001	0,0001	0,0001
я	0,0027	0	0,0003	0	0	0,0003	0	0	0	0	0	0	0	0	0
	0,0119	0	0	0,0003	0,0001	0,0005	0,0001	0,0002	0,0002	0,0001	0,0001	0,0002	0,0008	0,0003	0,0006

п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
0,0155	0,0037	0,0157	0,0068	0,0052	0,0007	0,0026	0,0003	0,0056	0,0009	0,0001	0	0	0	0,0028	0,0001	0,0039
0,0008	0,002	0,0042	0,0043	0,0001	0,0002	0,0008	0,0001	0,0008	0,0013	0,0002	0	0	0	0	0,0009	0,0016
0	0,0011	0,0002	0	0,0011	0	0,0001	0	0	0	0,0002	0,0002	0,0035	0	0,0001	0	0,0002
0,0003	0,0006	0,0022	0,0003	0,0005	0	0	0	0,0001	0,0005	0	0	0,0032	0,0001	0	0	0,0003
0	0,0008	0	0,0001	0,0005	0	0	0	0	0	0	0	0	0	0	0	0
0,0001	0,0011	0,0004	0,0001	0,0014	0	0	0,0002	0	0,0001	0	0	0,0005	0,0006	0	0	0,0003
0,0008	0,0071	0,0049	0,0052	0,0001	0	0,0005	0,0004	0,001	0,0006	0,0006	0	0	0	0	0,0002	0,0002
0	0	0	0	0,0002	0	0	0	0	0	0	0	0	0	0	0	0
0	0,0003	0	0	0,0004	0	0	0	0	0	0	0	0,0003	0,0001	0	0	0,0003
0,0002	0,0006	0,0035	0,0041	0	0	0,0014	0,0009	0,001	0,0005	0,0002	0	0	0	0	0,0004	0,0012
0	0	0,0005	0,0004	0	0	0	0	0,0002	0,0001	0	0	0	0	0	0	0
0	0,0017	0,0001	0,0008	0,0015	0	0	0	0	0	0	0	0	0	0,0001	0	0
0	0	0,0014	0,0002	0,0015	0	0	0	0,0001	0	0	0	0,0006	0,0039	0	0,0008	0,0013
0,0001	0,0001	0,0014	0	0,0019	0	0	0	0,0001	0	0	0	0,0014	0,0001	0,0001	0	0,0004
0	0,0001	0,0006	0,0008	0,0026	0	0	0,0003	0,0002	0	0,0002	0	0,0032	0,001	0	0,0002	0,0017
0,0012	0,0051	0,0058	0,0064	0,0001	0,0002	0,0004	0,0001	0,0018	0,0009	0,0002	0	0	0	0,0001	0,0002	0,0008
0	0,0063	0	0,0001	0,0007	0	0	0	0	0	0	0	0,0003	0	0	0	0,0002
0,0001	0,0001	0,0003	0,0008	0,0024	0	0,0001	0	0,0001	0,0003	0	0	0,0014	0,0006	0,0001	0,0001	0,0009
0,0017	0,0002	0,0011	0,011	0,0008	0	0,0002	0	0,0004	0,0001	0	0	0,0003	0,003	0,0003	0,0002	0,0037
0,0002	0,0028	0,0016	0,0001	0,0015	0	0	0,0001	0,0001	0	0	0	0,001	0,0057	0,0001	0,0001	0,0005
0,0007	0,0004	0,0012	0,0014	0	0	0,0003	0	0,0009	0,0005	0,0002	0	0	0	0,0001	0,001	0
0	0,0001	0	0	0,0001	0	0	0	0	0	0	0	0	0	0	0	0
0	0,0001	0	0	0,0001	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0,0001	0	0	0	0	0	0	0	0,0001	0	0	0	0
0	0	0	0,0036	0,0005	0	0	0	0	0,0001	0	0	0	0,0003	0	0	0
0	0	0	0,0001	0,0003	0	0	0	0	0	0	0	0	0,0002	0	0	0
0	0	0	0	0,0001	0	0	0	0	0	0	0	0	0,0001	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0002
0,0001	0,0002	0,0006	0,0007	0	0	0,0009	0	0,0002	0,0005	0	0	0	0	0	0	0
0	0	0,0009	0,0002	0	0	0	0,0001	0,0001	0,0005	0	0	0	0	0	0,0004	0,0002
0,0001	0,0005	0	0,0027	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0,0001	0,0003	0,0003	0	0	0	0	0,0002	0	0,0003	0	0	0	0	0	0
0	0,0001	0,0006	0,0012	0	0	0,0002	0	0,0001	0	0,0002	0	0	0	0	0,0001	0,0001

Ентропія та надлишковість:

Ентропію H_1, H_2 було обчислено за формулою:

$$H_n = \frac{1}{n} H(x_1, x_2, \dots, x_n), \text{ де}$$

$$H(x_1, x_2, \dots, x_n) = - \sum_{z_1, z_2, \dots, z_n} P(x_1 = z_1, \dots, x_n = z_n) * \log_2 P(x_1 = z_1, \dots, x_n = z_n).$$

$P(x_1, x_2, \dots, x_n)$ - частота(ймовірність) n-грами

Надлишковість R обчислюється за формулою:

$$R = 1 - \frac{H_\infty}{H_0}, \text{ де } H_\infty = \lim_{n \rightarrow \infty} H_n, H_n = \frac{1}{n} H(x_1, x_2, \dots, x_n), H_0 = \log_2 m$$

Оскільки ми обчислюємо R_1 і R_2 , надлишковість для монограм і біграм відповідно,

$$\text{то } R_n = 1 - \frac{H_n}{H_0}$$

Результати:

```
H1 без пробілів = 4.460560004
R1 без пробілів = 0.107887999
H1 з пробілами = 4.370919147
R1 з пробілами = 0.133509586
H2 без пробілів з перетинами = 4.153371277
R2 без пробілів з перетинами = 0.169325745
H2 з пробілами з перетинами = 3.977997489
R2 з пробілами з перетинами = 0.211402322
H2 без пробілів без перетинів = 4.153144601
R2 без пробілів без перетинів = 0.169371080
H2 з пробілами без перетинів = 3.978124583
R2 з пробілами без перетинів = 0.211377127
```

Далі, за допомогою програми CoolPinkProgram порахуємо умовну ентропію джерела для файлу TEXT, було проведено щонайменше 50 експериментів для кожного значення:

$H^{(10)}$:

Произвольная часть текста:
акона_человеческой_природы_то_какая_же_может_быть_разница_между_справедливы

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: o

Символ по счету: 1

Номер эксперимента: 55

Поле ввода символов:
o

Продолжить

Другой

Неравенство для энтропии:
1,72823739142731 < H < 2,51588255884671

Двоичная таблица угаданных символов:
00000100000000000000000000000000
00000000100000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:
q[1] = 0,4909090
q[2] = 0,1818181
q[3] = 0,0727272
q[4] = 0,0363636
q[5] = 0,0363636
q[6] = 0,0363636
q[7] = 0
q[8] = 0,0181818
q[9] = 0,0181818
q[10] = 0,018181
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0
q[17] = 0,018181
q[18] = 0
q[19] = 0
q[20] = 0
q[21] = 0,036363
q[22] = 0
q[23] = 0,036363
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

Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

$H^{(20)}$:

Произвольная часть текста:
_что_человек_делающий_их_не_просто_заявляет_что_ему_не_нравится_поведение_д

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: и

Символ по счету: 1

Номер эксперимента: 53

Поле ввода символов:
и

Продолжить

Другой

Неравенство для энтропии:
1,67597398029812 < H < 2,50127310033339

Двоичная таблица угаданных символов:
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
00001000000000000000000000000000
00100000000000000000000000000000

Вероятности:
q[1] = 0,5283018
q[2] = 0,1320754
q[3] = 0,0754716
q[4] = 0,0188679
q[5] = 0,0566037
q[6] = 0
q[7] = 0,0188679
q[8] = 0
q[9] = 0,0377358
q[10] = 0,037735
q[11] = 0
q[12] = 0
q[13] = 0,018867
q[14] = 0,018867
q[15] = 0,018867
q[16] = 0
q[17] = 0
q[18] = 0,018867
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0,018867
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:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

$H^{(30)}:$

Произвольная часть текста:		
_неорганические_тела_такой_закон_человек_может_нарушить_по_своему_выбору_эт		
Использованные буквы:		
<div>Порядок n-граммы:<ul style="list-style-type: none">5 символов10 символов15 символов20 символов25 символов30 символов35 символов40 символов45 символов50 символов</div>	<div>Введенный символ: к</div> <div>Символ по счету: 1</div> <div>Номер эксперимента: 52</div> <div>Поле ввода символов:<div>к</div></div> <div><div>Продолжить</div><div>Другой</div></div>	<div>Неравенство для энтропии: 1,49649805281308 < H < 2,21602190859436</div> <div>Двоичная таблица угаданных символов:<div>100000000000000000000000000000 100000000000000000000000000000 000000100000000000000000000000 100000000000000000000000000000 100000000000000000000000000000</div></div>

Вероятности:

q [1] = 0,5961538
q [2] = 0,0961538
q [3] = 0,0769230
q [4] = 0
q [5] = 0,0384615
q [6] = 0
q [7] = 0,0384615
q [8] = 0,0192307
q [9] = 0
q [10] = 0,019230
q [11] = 0,038461
q [12] = 0
q [13] = 0
q [14] = 0,019230
q [15] = 0
q [16] = 0,019230
q [17] = 0
q [18] = 0
q [19] = 0
q [20] = 0
q [21] = 0
q [22] = 0
q [23] = 0
q [24] = 0,038461
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:

Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

Аналогічно попередній формулі:

$$R^{(n)} = 1 - \frac{H^{(n)}}{H_0}$$

Результати:

```
1.728237391 < H(10) < 2.515882558
0.496823488 < R(10) < 0.654352522
1.675973980 < H(20) < 2.501273100
0.499745380 < R(20) < 0.664805204
1.496498052 < H(30) < 2.216021909
0.556795618 < R(30) < 0.700700390
```

Висновки:

У ході даного комп'ютерного практикуму, було експериментально досліджено частоти монограм та біграм на тексті довжиною 3 Мб. Отримані результати наближено відповідають [частотам монограм в російській мові](#).

буква ↕	ранг ↕	употреблений ↕		Символ	Частота
о	1	55414481	10,97%	о	0,1148
е	2	42691213	8,45%	е	0,0867
а	3	40487008	8,01%	а	0,0757
и	4	37153142	7,35%	и	0,0673
н	5	33838881	6,70%	и	0,0631
т	6	31620970	6,26%	т	0,0621
с	7	27627040	5,47%	с	0,057
р	8	23916825	4,73%	л	0,0491
в	9	22930719	4,54%	в	0,0427
л	10	22230174	4,40%	р	0,0424
к	11	17653469	3,49%	м	0,036
м	12	16203060	3,21%	к	0,034
д	13	15052118	2,98%	д	0,0312
п	14	14201572	2,81%	у	0,0283
у	15	13245712	2,62%	п	0,0264
я	16	10139085	2,01%	я	0,0216
ы	17	9595941	1,90%	ы	0,019
ь	18	8784613	1,74%	ь	0,0188
г	19	8564640	1,70%	г	0,0175
з	20	8329904	1,65%	з	0,0166
б	21	8051767	1,59%	б	0,0164
ч	22	7300193	1,44%	ч	0,0156
й	23	6106262	1,21%	й	0,0118
х	24	4904176	0,97%	ж	0,0102
ж	25	4746916	0,94%	х	0,0091
ш	26	3678738	0,73%	ш	0,0082
ю	27	3220715	0,64%	ю	0,0056
ц	28	2438807	0,48%	э	0,0046
щ	29	1822476	0,36%	ц	0,0031
э	30	1610107	0,32%	щ	0,0031
ф	31	1335747	0,26%	ф	0,0016
ъ	32	185452	0,04%	ъ	0,0003

Отже, проаналізувавши частоти для тексту розміром 3 Мб, можна отримати наближені частоти для n-грам

Також, для монограм і біграм було обраховано ентропію, і чим більш рівноймовірний розподіл частот, тим, відповідно, більша ентропія, і тим важче, до прикладу, провести частотний аналіз ШТ. У нашому випадку, ентропія тексту без пробілів більша за ентропію тексту з пробілами. Такі результати ми отримали тому, що пробіли мають дуже високу частоту порівняно з іншими символами, і вони значно впливають на частотний розподіл, що робить частоти n-грам менш рівноймовірними.

Для різних ентропій була обрахована надлишковість, яка, власне, характеризує у відсотках відхилення ентропії від максимальної для даного алфавіту. Тобто, чим більше це значення, тим більше, до прикладу, мова вразлива до частотного аналізу. Це допомагає нам оцінити, наскільки наше джерело передбачуване від 0 до 1.