



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

Навчально-науковий Фізико-технічний інститут
Кафедра інформаційної безпеки

КРИПТОГРАФІЯ

Комп'ютерний практикум №1
Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали:
Студенти ФБ-33
Дохолян Юлія
Терещенко Микола

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:

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 H_{10} , H_{20} , H_{30} .

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

Хід роботи

```
import re, math, os
from collections import Counter
import pandas as pd

def normalize(text, keep_spaces=True):
    text = text.lower().replace('ё', 'е').replace('ь', 'ъ')
    text = re.sub(r'^а-я ', ' ', text)
    text = re.sub(r'\s+', ' ', text).strip()
    return text if keep_spaces else text.replace(' ', '')

def frequencies(seq, n=1, step=1):
    items = [seq[i:i+n] for i in range(0, len(seq)-n+1, step)]
    cnt = Counter(items)
    total = sum(cnt.values())
    return {k: v/total for k,v in cnt.items()}

def bigram_matrix(bigrams):
    abc = list(«абвгдежзийклмнопрстуфхцчшщъыьэюяёь «)
    m = pd.DataFrame(0.0, index=abc, columns=abc)
    for bg, f in bigrams.items():
        if len(bg) == 2: m.at[bg[0], bg[1]] = round(f,6)
    return m

def entropy(freqs): return -sum(p*math.log2(p) for p in freqs.values())

def process(path, keep_spaces=True):
    with open(path, encoding=«utf-8») as f: text = normalize(f.read(),
keep_spaces)
    chars = frequencies(text)
    bigr_1, bigr_2 = frequencies(text,2,1), frequencies(text,2,2)

    out = f»results_{{'spaces' if keep_spaces else 'nospace'}}.xlsx»
    if os.path.exists(out): os.remove(out)
    with pd.ExcelWriter(out) as w:
        pd.DataFrame(chars.items(), columns=[«Символ», «Частота»]).to_excel(w,
«Частоти», index=False)
        bigram_matrix(bigr_1).to_excel(w, «Біграми з перекриттям»)
        bigram_matrix(bigr_2).to_excel(w, «Біграми без перекриття»)

    print(f»{{'3 пробілами' if keep_spaces else 'Без пробілів'}}:»)
    print(f»H1 = {entropy(chars):.6f}»)
    print(f»H2 перекриття = {entropy(bigr_1)/2:.6f}»)
    print(f»H2 без перекриття = {entropy(bigr_2)/2:.6f}\n»)

file = «kafka.txt»
process(file, True)
process(file, False)
```

Рис.1

Скрипт аналізує великий текст російською мовою, підраховує частоти окремих букв, частоти біграм (з перекриттям і без перекриття), ентропію (H_1 – для окремих символів і H_2 – для біграм. Він робить це двічі, для тексту з пробілами і для тексту, де пробіли вилучено.

Результати записуються в два файли Excel.

Функція «normalize» виконує передобробку тексту. Переводить усі літери в нижній регістр, замінює «ё» на «е» та «ъ» на «ь», залишає лише символи російського алфавіту і, за потреби, пробіли. Також вона скорочує послідовності пробілів до одного, забезпечуючи рівномірність обробки тексту.

Функція «frequencies» обчислює відносні частоти появи символів або біграм. Для цього вона ділить текст на елементи довжиною n : якщо $n=1$ – рахує окремі букви, якщо $n=2$ – біграми. Параметр `step` визначає, чи біграми беруться з перекриттям (`step=1`) або без (`step=2`).

Функція «bigram_matrix» створює матрицю частот біграм у вигляді таблиці, яка пізніше зберігається в Excel для зручності перегляду результатів.

Функція «entropy» – ентропія. Для біграм отримане значення ділиться на два, щоб відповідати означенню H_2 .

Основна функція «process» об'єднує всі етапи.

Результат виконання скрипту зображено на рис.2:

```
3 пробілами:
H1 = 4.366882
H2 перекриття = 3.897253
H2 без перекриття = 3.897025

Без пробілів:
H1 = 4.448709
H2 перекриття = 4.050321
H2 без перекриття = 4.050506
```

Рис.2

<u>Символ</u>	<u>Частота</u>
<u>п</u>	0.02751
<u>р</u>	0.03477
<u>о</u>	0.11234
<u>с</u>	0.04154
<u>н</u>	0.06012
<u>у</u>	0.02291
<u>в</u>	0.03651
<u>ш</u>	0.00488
<u>и</u>	0.0454
<u>ь</u>	0.01625
	0.1621
<u>д</u>	0.02686
<u>а</u>	0.06002
<u>ж</u>	0.00952
<u>ы</u>	0.0215
<u>т</u>	0.04642
<u>м</u>	0.02437
<u>л</u>	0.04021
<u>е</u>	0.06792
<u>б</u>	0.01806
<u>к</u>	0.02645
<u>й</u>	0.00927
<u>г</u>	0.01666
<u>з</u>	0.01691
<u>ч</u>	0.01137
<u>я</u>	0.01712
<u>ц</u>	0.00107
<u>х</u>	0.00742
<u>ю</u>	0.00755
<u>щ</u>	0.00339
<u>э</u>	0.00316
<u>ф</u>	0.00043

(1)

<u>Символ</u>	<u>Частота</u>
<u>п</u>	0.03283
<u>р</u>	0.0415
<u>о</u>	0.13407
<u>с</u>	0.04957
<u>н</u>	0.07175
<u>у</u>	0.02734
<u>в</u>	0.04357
<u>ш</u>	0.00583
<u>и</u>	0.05419
<u>ь</u>	0.0194
<u>д</u>	0.03205
<u>а</u>	0.07163
<u>ж</u>	0.01136
<u>ы</u>	0.02566
<u>т</u>	0.0554
<u>м</u>	0.02908
<u>л</u>	0.04799
<u>е</u>	0.08106
<u>б</u>	0.02155
<u>к</u>	0.03156
<u>й</u>	0.01107
<u>г</u>	0.01989
<u>з</u>	0.02018
<u>ч</u>	0.01357
<u>я</u>	0.02044
<u>ц</u>	0.00127
<u>х</u>	0.00886
<u>ю</u>	0.00901
<u>щ</u>	0.00405
<u>э</u>	0.00377
<u>ф</u>	0.00051

(2)

Рис.3

На рис.3(1) зображено частоти символів із пробілами, а на рис.3(2) без пробілів.

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	0	0.000836	0.003818	0	0.001065	0.000623	0.001671	0.004637	0	0.000623	0.00508	0.007406	0.001901	0.002736	0
б	0.000426	0	0	0	0	0.001704	0	0	0	0	0	0	0	0.000426	0.003326
в	0.004244	0	0.000213	0	0.000213	0.003769	0	0.000426	0.002966	0	0	0.000426	0	0.001065	0.007357
г	0.000213	0	0	0	0.001278	0	0	0	0.000852	0	0.000426	0.000852	0	0.000213	0.010094
д	0.005293	0	0.000426	0	0	0.005063	0	0	0.002736	0	0	0.000426	0	0.002327	0.00485
е	0	0.001049	0.001049	0.004227	0.001917	0.001278	0.001278	0.002343	0.000213	0.001032	0.001049	0.006931	0.004031	0.007374	0.000213
ж	0.000852	0.000197	0	0	0.000852	0.004441	0	0	0.001704	0	0.000213	0	0	0.000639	0
з	0.004441	0.000213	0.001688	0.000639	0.001901	0	0	0	0	0	0	0.000213	0.000213	0.001278	0.000639
и	0	0.00041	0.003162	0.000213	0.001065	0.001704	0	0.001704	0.000639	0.000852	0.001917	0.004457	0.001704	0.002753	0
й	0	0	0	0	0.000213	0	0	0	0	0	0.000213	0	0	0.001049	0
к	0.005276	0	0	0	0	0.000639	0	0	0.001688	0	0	0.000213	0	0.000213	0.010143
л	0.002769	0	0	0.000623	0	0.004588	0.000426	0.000213	0.003162	0	0.00041	0.000213	0	0.000213	0.008471
м	0.002769	0	0	0	0	0.003195	0	0.000426	0.003195	0	0.000213	0	0.000426	0.001688	0.002966
н	0.010815	0	0	0	0.000213	0.009881	0	0	0.005686	0	0.000213	0.000197	0	0.003572	0.011552
о	0.000213	0.003834	0.007603	0.005686	0.007161	0.003195	0.002556	0.002294	0.001278	0.003572	0.003605	0.006505	0.005276	0.008045	0.000213
п	0.00213	0	0	0	0	0.001704	0	0	0.001491	0	0.000213	0.001491	0	0	0.013125
р	0.010929	0	0	0.000426	0.000426	0.005047	0.000426	0	0.003359	0	0.000639	0	0	0.001491	0.005489
с	0.001688	0	0.00231	0	0.000213	0.003801	0	0	0.000852	0	0.001049	0.002917	0.000426	0.001049	0.003375
т	0.003752	0	0.001868	0	0.000213	0.004195	0	0	0.003375	0	0.000213	0	0	0.001884	0.013338
у	0	0.000213	0.001049	0.000639	0.001278	0.000213	0.001458	0	0	0.000426	0.000852	0.001065	0.001917	0.000213	0
ф	0	0	0	0	0	0.000213	0	0	0	0	0	0	0	0	0
х	0.000213	0	0	0	0	0	0	0	0.000213	0	0	0.000213	0	0	0.003588
ц	0	0	0	0	0	0.000213	0	0	0.000213	0	0	0	0	0	0.000213
ч	0.001868	0	0	0	0	0.003605	0	0	0.000836	0	0.000213	0	0	0.001032	0
ш	0	0	0	0	0	0.002114	0	0	0.000426	0	0.000426	0.000213	0	0.000639	0.000639
щ	0.000639	0	0	0	0	0.001262	0	0	0.001065	0	0	0	0	0.000213	0
ы	0	0.000213	0.000639	0.000213	0	0.000836	0	0.000213	0	0.002556	0.000213	0.004408	0.002097	0.00041	0
ь	0	0	0	0	0	0.000426	0	0.000213	0	0.001049	0	0.000426	0.000426	0.002278	0
э	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ю	0	0.000393	0	0	0.000623	0	0	0	0	0	0	0	0	0	0
я	0	0	0.000213	0.000426	0.000213	0.000197	0.000639	0.000213	0.001049	0.000213	0	0.000213	0.000426	0.000852	0
е	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ь	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.001491	0.0107	0.012469	0.003572	0.008013	0.004014	0.001065	0.004227	0.008406	0	0.008242	0.001852	0.005522	0.016468	0.012748

Рис.4

Біграми з перекриттям з пробілами.

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	0	0.000852	0.003834	0	0.001114	0.000623	0.001671	0.004785	0	0.000623	0.005112	0.007373	0.001868	0.002818	0
б	0.000459	0	0	0	0	0.001737	0	0	0	0	0	0	0	0.000426	0.003212
в	0.004326	0	0.000197	0	0.000197	0.003834	0	0.000426	0.003015	0	0	0.000459	0	0.001049	0.00721
г	0.000197	0	0	0	0.001344	0	0	0	0.000852	0	0.000426	0.000852	0	0.000197	0.01029
д	0.005309	0	0.000426	0	0	0.004981	0	0	0.002786	0	0	0.000426	0	0.002327	0.004948
е	0	0.001081	0.001081	0.004096	0.001802	0.001245	0.001311	0.002327	0.000229	0.001049	0.001081	0.006947	0.004096	0.007537	0.000229
ж	0.000819	0.000197	0	0	0.000852	0.004457	0	0	0.001704	0	0.000229	0	0	0.000688	0
з	0.004391	0.000229	0.001639	0.000623	0.001933	0	0	0	0	0	0	0.000197	0.000229	0.001311	0.000655
и	0	0.000426	0.003146	0.000229	0.001049	0.00177	0	0.001704	0.000655	0.000852	0.001868	0.004457	0.001704	0.002687	0
й	0	0	0	0	0.000197	0	0	0	0	0	0.000229	0	0	0.001049	0
к	0.005211	0	0	0	0	0.000655	0	0	0.001671	0	0	0.000197	0	0.000229	0.010323
л	0.002818	0	0	0.000623	0	0.004522	0.000426	0.000197	0.003277	0	0.000393	0.000229	0	0.000197	0.00852
м	0.002654	0	0	0	0	0.003212	0	0.000426	0.003212	0	0.000229	0	0.000426	0.001671	0.003048
н	0.010814	0	0	0	0.000197	0.009897	0	0	0.005637	0	0.000229	0.000197	0	0.003506	0.01147
о	0.000229	0.003703	0.00757	0.005604	0.007144	0.003179	0.002556	0.002294	0.001278	0.003572	0.003572	0.006423	0.005211	0.007832	0.000197
п	0.002196	0	0	0	0	0.001737	0	0	0.00154	0	0.000229	0.001442	0	0	0.013043
р	0.010683	0	0	0.000426	0.000426	0.005047	0.000426	0	0.003375	0	0.00059	0	0	0.00154	0.005637
с	0.001671	0	0.002261	0	0.000197	0.003801	0	0	0.000885	0	0.001049	0.002851	0.000426	0.001049	0.00331
т	0.003736	0	0.001901	0	0.000197	0.004129	0	0	0.003441	0	0.000229	0	0	0.00177	0.01337
у	0	0.000229	0.000983	0.000623	0.001245	0.000229	0.001442	0	0	0.000426	0.000852	0.001081	0.001966	0.000229	0
ф	0	0	0	0	0	0.000197	0	0	0	0	0	0	0	0	0
х	0.000197	0	0	0	0	0	0	0	0.000229	0	0	0.000229	0	0	0.003638
ц	0	0	0	0	0	0.000229	0	0	0.000229	0	0	0	0	0	0.000229
ч	0.00177	0	0	0	0	0.003736	0	0	0.000852	0	0.000197	0	0	0.001049	0
ш	0	0	0	0	0	0.002097	0	0	0.000393	0	0.000459	0.000197	0	0.000688	0.000655
щ	0.000623	0	0	0	0	0.001278	0	0	0.001016	0	0	0	0	0.000197	0
ы	0	0.000229	0.00059	0.000229	0	0.000852	0	0.000229	0	0.002556	0.000229	0.004457	0.002097	0.000393	0
ь	0	0	0	0	0	0.000426	0	0.000197	0	0	0.001049	0	0.000426	0.002228	0
э	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ю	0	0.000393	0	0	0.000623	0	0	0	0	0	0	0	0	0	0
я	0	0	0.000197	0.000459	0.000197	0.000197	0.000623	0.000197	0.001016	0.000229	0	0.000197	0.000459	0.000885	0
е	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ь	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.001442	0.010716	0.012682	0.003539	0.008062	0.004129	0.001049	0.004227	0.008553	0	0.008225	0.001868	0.005473	0.016549	0.012781

Рис.5

Біграми без перекриття з пробілами.

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	0.000508	0.001995	0.005319	0.000489	0.00178	0.00176	0.001995	0.006043	0.001271	0.000743	0.007079	0.009093	0.002777	0.004537	0.001525
б	0.000508	0	0	0	0	0.002034	0	0	0	0	0	0	0	0.000508	0.004224
в	0.005065	0.000254	0.000254	0.000763	0.000743	0.004987	0	0.000508	0.00354	0	0.000508	0.000508	0.000508	0.001271	0.009035
г	0.000254	0	0	0	0.001525	0	0	0	0.001017	0	0.000508	0.001017	0.000254	0.000508	0.012301
д	0.006316	0	0.000508	0.000508	0	0.006043	0	0	0.003266	0	0	0.000508	0.000254	0.002777	0.006043
е	0.000254	0.003774	0.003246	0.005045	0.002542	0.001525	0.001525	0.003305	0.001271	0.001232	0.002268	0.00925	0.005319	0.011049	0.001506
ж	0.001017	0.000235	0	0	0.001017	0.0053	0	0	0.002034	0	0.000254	0	0	0.000997	0
з	0.0053	0.000254	0.002014	0.000763	0.002268	0	0	0.000254	0.000763	0	0	0.000254	0.000508	0.00178	0.000763
и	0	0.001995	0.005026	0.000743	0.002777	0.002034	0.000254	0.002288	0.001017	0.001017	0.002542	0.005828	0.002034	0.00528	0.001525
й	0	0.000763	0.000508	0	0.000508	0	0	0.000254	0.000508	0	0.000997	0	0.000254	0.001995	0.000763
к	0.006297	0.000254	0.000254	0	0.000254	0.000763	0	0	0.002523	0	0.000508	0.000254	0.000508	0.001017	0.012868
л	0.003305	0.000978	0.001017	0.000743	0.000235	0.005984	0.000508	0.000508	0.004028	0	0.001252	0.000254	0	0.001017	0.01189
м	0.003305	0.000254	0.000508	0.000254	0.000743	0.004068	0	0.000763	0.004811	0	0.000763	0.000235	0.001017	0.003266	0.004283
н	0.012907	0.000743	0.001017	0	0.000763	0.011792	0	0.000254	0.006786	0	0.000743	0.000235	0.000254	0.006551	0.014295
о	0.000254	0.006571	0.010814	0.007294	0.011069	0.005084	0.003559	0.004478	0.003266	0.004263	0.005573	0.007764	0.007803	0.012125	0.003305
п	0.002542	0	0	0	0	0.002034	0	0	0.00178	0	0.000254	0.00178	0	0	0.015664
р	0.013044	0	0	0.000508	0.000508	0.006023	0.000508	0.000254	0.004244	0	0.000763	0	0	0.002014	0.006805
с	0.002014	0	0.002757	0	0.000254	0.004791	0	0	0.001017	0	0.002014	0.003481	0.000508	0.001252	0.004283
т	0.004732	0.000508	0.003246	0	0.000254	0.00526	0	0	0.004028	0	0.001017	0	0	0.002503	0.016173
у	0.000508	0.000254	0.002014	0.000997	0.001525	0.000489	0.001995	0.000254	0.000508	0.000508	0.001271	0.001271	0.002288	0.001525	0.000254
ф	0	0	0	0	0	0.000254	0	0	0	0	0	0	0	0	0
х	0.000254	0.000254	0.000254	0	0.000254	0	0	0	0.000254	0	0	0.000254	0	0.000254	0.004283
ц	0	0	0	0	0	0.000254	0	0	0.000254	0	0	0	0	0	0.000254
ч	0.002229	0	0	0	0	0.004302	0	0	0.000997	0	0.000254	0	0	0.001232	0
ш	0	0	0	0	0	0.002523	0	0	0.000508	0	0.000508	0.000254	0	0.000763	0.000763
щ	0.000763	0	0	0	0	0.001506	0	0	0.001271	0	0	0	0	0.000254	0
ы	0	0.000508	0.001252	0.000508	0.000763	0.001252	0.000254	0.000254	0.000724	0.003051	0.000489	0.00526	0.002757	0.000743	0
ь	0	0.000489	0.001525	0.000763	0.000508	0.000508	0	0.000254	0.000743	0	0.001252	0	0.000508	0.003735	0.000254
э	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ю	0	0.000724	0	0	0.001252	0	0	0.000254	0.000508	0	0.000489	0	0.000508	0	0.000254
я	0.000254	0.000743	0.002034	0.000508	0.000508	0.000489	0.000763	0.000254	0.001252	0.000254	0.000254	0.000489	0.001017	0.002796	0.000763
е	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ь	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Рис.6

Біграми з перекриттям без пробілів.

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	0.000508	0.002034	0.00528	0.000508	0.001799	0.00176	0.002073	0.006062	0.001252	0.000704	0.007118	0.009074	0.002816	0.00438	0.001486
б	0.000508	0	0	0	0	0.001995	0	0	0	0	0	0	0	0.000508	0.004185
в	0.005241	0.000235	0.000235	0.000743	0.000743	0.004928	0	0.000469	0.00352	0	0.000508	0.000508	0.000508	0.001252	0.009074
г	0.000235	0	0	0	0.001525	0	0	0	0.001017	0	0.000469	0.000978	0.000235	0.000508	0.012477
д	0.006219	0	0.000548	0.000548	0	0.006062	0	0	0.003129	0	0	0.000548	0.000235	0.002777	0.00614
е	0.000274	0.003755	0.003207	0.005045	0.002542	0.001604	0.001486	0.003364	0.001291	0.001252	0.002308	0.009426	0.005241	0.011147	0.001525
ж	0.001017	0.000235	0	0	0.000978	0.005202	0	0	0.002073	0	0.000274	0	0	0.001017	0
з	0.005319	0.000274	0.001995	0.000782	0.002229	0	0	0.000274	0.000704	0	0	0.000274	0.000508	0.001721	0.000743
и	0	0.002073	0.005124	0.000782	0.002777	0.001995	0.000274	0.002268	0.000978	0.001017	0.00262	0.005906	0.002073	0.005358	0.001564
й	0	0.000704	0.000508	0	0.000508	0	0	0.000274	0.000469	0	0.000978	0	0.000235	0.002034	0.000743
к	0.006219	0.000235	0.000274	0	0.000235	0.000743	0	0	0.002542	0	0.000508	0.000274	0.000469	0.001017	0.012868
л	0.003324	0.000978	0.001017	0.000743	0.000235	0.005906	0.000548	0.000508	0.003911	0	0.001252	0.000235	0	0.001056	0.011851
м	0.003364	0.000274	0.000548	0.000274	0.000743	0.00395	0	0.000743	0.004693	0	0.000782	0.000235	0.001056	0.003246	0.004341
н	0.012868	0.000743	0.001017	0	0.000782	0.011733	0	0.000235	0.006805	0	0.000743	0.000235	0.000274	0.006571	0.014354
о	0.000274	0.006688	0.010717	0.007236	0.011069	0.005006	0.003559	0.004498	0.003246	0.004224	0.005593	0.007705	0.007783	0.012164	0.003285
п	0.00266	0	0	0	0	0.002112	0	0	0.001838	0	0.000274	0.001838	0	0	0.015605
р	0.013063	0	0	0.000469	0.000508	0.006101	0.000469	0.000274	0.004263	0	0.000782	0	0	0.001995	0.006962
с	0.002073	0	0.002855	0	0.000274	0.004811	0	0	0.001017	0	0.002034	0.003442	0.000548	0.001252	0.004224
т	0.004654	0.000508	0.003168	0	0.000235	0.005241	0	0	0.004028	0	0.001056	0	0	0.002464	0.01627
у	0.000508	0.000235	0.002073	0.001017	0.001525	0.000469	0.002073	0.000235	0.000548	0.000508	0.001252	0.001291	0.002268	0.001564	0.000235
ф	0	0	0	0	0	0.000274	0	0	0	0	0	0	0	0	0
х	0.000235	0.000274	0.000235	0	0.000274	0	0	0	0.000235	0	0	0.000274	0	0.000235	0.004263
ц	0	0	0	0	0	0.000235	0	0	0.000235	0	0	0	0	0	0.000235
ч	0.002229	0	0	0	0	0.00438	0	0	0.001017	0	0.000274	0	0	0.001252	0
ш	0	0	0	0	0	0.002503	0	0	0.000469	0	0.000508	0.000235	0	0.000743	0.000782
щ	0.000782	0	0	0	0	0.001564	0	0	0.001212	0	0	0	0	0.000274	0
ы	0	0.000508	0.001252	0.000508	0.000782	0.001291	0.000235	0.000274	0.000704	0.003051	0.000508	0.005202	0.002738	0.000743	0
ь	0	0.000508	0.001643	0.000743	0.000508	0.000508	0	0.000235	0.000704	0	0.001252	0	0.000508	0.003755	0.000274
э	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ю	0	0.000743	0	0	0.00133	0	0	0.000274	0.000469	0	0.000469	0	0.000508	0	0.000274
я	0.000235	0.000743	0.001956	0.000508	0.000469	0.000508	0.000704	0.000235	0.001252	0.000274	0.000235	0.000469	0.001017	0.002699	0.000704
е	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ь	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Рис.7

Біграми без перекриття без пробілів.

Надлишковість обраховано за формулою $R = 1 - \frac{H_{\infty}}{H_0}$. Слід зауважити, що значення H_0 буде відрізнятися, оскільки спочатку йде обрахунок враховуючи пробіли, тобто 34 символ, а потім без пробілу – 33 символи відповідно.

	Ентропія	Надлишковість
Монограми	4.366882	0.1416
Біграми з перекриттям	3.897253	0.2334
Біграми без перекриття	3.897025	0.2339

	Ентропія	Надлишковість
Монограми	4.448709	0.1256
Біграми з перекриттям	4.050321	0.2039
Біграми без перекриття	4.050506	0.2038

 $H_{10}:$

The screenshot displays the 'Лабораторная работа №1' application window. The main area shows the following information:

- Произвольная часть текста:** дливы_к_детям_та_не_совсем_чистая_сделка_о_которой_вы_почти_забыли_подверну
- Использованные буквы:** о, р, й, ц, у, к.
- Порядок n-граммы:** A dropdown menu is open, showing options from 5 to 50 symbols. The option '10 символов' is selected.
- Введенный символ:** e
- Символ по счету:** 7
- Номер эксперимента:** 50
- Неравенство для энтропии:** $3,29318839409308 < H < 3,9766912610489$
- Двоичная таблица угаданных символов:** A grid of 0s and 1s representing binary probabilities for each symbol.
- Вероятности:** A list of probabilities $q[1]$ through $q[32]$. The value for $q[10]$ is highlighted as 0.02.
- Поле ввода символов:** A text input field containing the letter 'e'.
- Buttons:** 'Продолжить' (Continue) and 'Другой' (Other).
- Строка состояния:** Displays the message: 'Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка'.

H_{20} :

Лабораторная работа №1

Произвольная часть текста:
ему_закону_тяготения_или_нет_тогда_как_человек_имеет_право_выбора_подчинят

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: 2
Номер эксперимента: 50
Поле ввода символов:
и
Продолжить Другой

Неравенство для энтропии:
2,5100039964676 < H < 2,96888737212111
Двоичная таблица угаданных символов:
00000000000000000000000000000000
00000000000000000000000000000000
00000000000000000000000000000000
10000000000000000000000000000000
00000010000000000000000000000000

Вероятности:
q[1] = 0,42
q[2] = 0,16
q[3] = 0,02
q[4] = 0
q[5] = 0
q[6] = 0,02
q[7] = 0,02
q[8] = 0,04
q[9] = 0,02
q[10] = 0
q[11] = 0,02
q[12] = 0
q[13] = 0
q[14] = 0,06
q[15] = 0
q[16] = 0,04
q[17] = 0
q[18] = 0,04
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] = 0,06
q[28] = 0,02
q[29] = 0
q[30] = 0,04
q[31] = 0
q[32] = 0

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

H_{30} :

Лабораторная работа №1

Произвольная часть текста:
но_когда_мыслители_древности_называли_законы_добра_и_зла_законами_природы_

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: _ (пробел)
Символ по счету: 1
Номер эксперимента: 50
Поле ввода символов:
Продолжить Другой

Неравенство для энтропии:
1,9320894621625 < H < 2,41972183240961
Двоичная таблица угаданных символов:
00010000000000000000000000000000
00001000000000000000000000000000
00000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:
q[1] = 0,6
q[2] = 0,04
q[3] = 0
q[4] = 0,04
q[5] = 0,02
q[6] = 0,02
q[7] = 0
q[8] = 0
q[9] = 0
q[10] = 0
q[11] = 0
q[12] = 0
q[13] = 0
q[14] = 0,04
q[15] = 0,08
q[16] = 0,02
q[17] = 0,02
q[18] = 0,02
q[19] = 0
q[20] = 0,02
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0
q[25] = 0,02
q[26] = 0
q[27] = 0,02
q[28] = 0,02
q[29] = 0
q[30] = 0
q[31] = 0,02
q[32] = 0

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

Зведення результатів експериментів з CoolPinkProgram.exe та відповідний підрахунок надлишковості:

	Ентропія	Надлишковість
H_{10}	$3,29318839409308 < H < 3,9766912610489$	$0,204661755 < R < 0,341362322$
H_{20}	$2,5100039964676 < H < 2,96888737212111$	$0,406222526 < R < 0,497999201$
H_{30}	$1,9320894621625 < H < 2,41972183240961$	$0,516055634 < R < 0,613582108$

У даному випадку $H_0 = \log_2 32 = 5$. 32 – кількість символів, що використовуються в CoolPinkProgram.exe

Висновок: У ході комп'ютерного практикуму було засвоєно теоретичні поняття ентропії та надлишковості джерела відкритого тексту. Було набуто практичних навичок обчислення цих величин і проміжних показників, зокрема частот символів, для різних моделей джерела тексту та здійснено їх порівняння. Експериментальним шляхом за допомогою програми CoolPinkProgram.exe було наближено визначено значення ентропії H_{10} , H_{20} , H_{30} , на основі яких обчислено надлишковість.