



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 = «start.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.

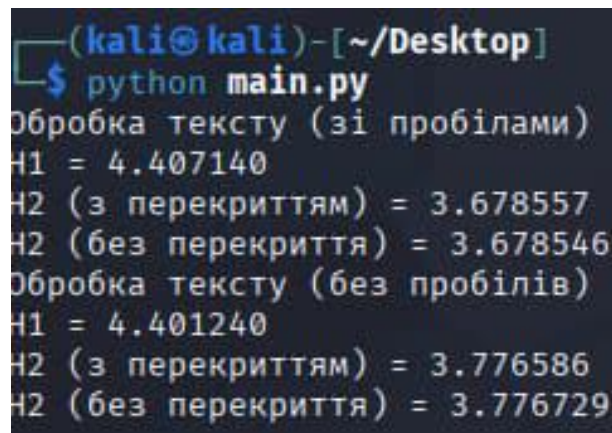
Функція «frequencies» обчислює відносні частоти появи символів або біграм. Для цього вона ділить текст на елементи довжиною n: якщо n=1 – рахує окремі букви, якщо n=2 – біграми. Параметр step визначає, чи біграми беруться з перекриттям (step=1) або без (step=2).

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:



```
(kali@kali)-[~/Desktop]
$ python main.py
Обробка тексту (зі пробілами)
H1 = 4.407140
H2 (з перекриттям) = 3.678557
H2 (без перекриття) = 3.678546
Обробка тексту (без пробілів)
H1 = 4.401240
H2 (з перекриттям) = 3.776586
H2 (без перекриття) = 3.776729
```

Рис.2

<u>Символ</u>	<u>Частота</u>
<u>с</u>	0.04717
<u>о</u>	0.08321
<u>т</u>	0.05617
<u>в</u>	0.04607
<u>е</u>	0.06623
<u>у</u>	0.01237
<u>ю</u>	0.00338
<u>щ</u>	0.00225
<u>и</u>	0.08312
<u>й</u>	0.0213
	0.11689
<u>з</u>	0.02473
<u>н</u>	0.073
<u>а</u>	0.08769
<u>ч</u>	0.0101
<u>п</u>	0.02472
<u>р</u>	0.04609
<u>к</u>	0.02358
<u>л</u>	0.02248
<u>ы</u>	0.02245
<u>х</u>	0.00901
<u>б</u>	0.01125
<u>я</u>	0.02248
<u>м</u>	0.01686
<u>ж</u>	0.00902
<u>ф</u>	0.00563
<u>д</u>	0.01798
<u>ь</u>	0.00674
<u>ц</u>	0.00898
<u>ш</u>	0.00786
<u>г</u>	0.01012
<u>э</u>	0.00111

(1)

<u>Символ</u>	<u>Частота</u>
<u>с</u>	0.05341
<u>о</u>	0.09422
<u>т</u>	0.0636
<u>в</u>	0.05216
<u>е</u>	0.07499
<u>у</u>	0.01401
<u>ю</u>	0.00383
<u>щ</u>	0.00255
<u>и</u>	0.09412
<u>й</u>	0.02412
<u>з</u>	0.02801
<u>н</u>	0.08267
<u>а</u>	0.0993
<u>ч</u>	0.01144
<u>п</u>	0.02799
<u>р</u>	0.05219
<u>к</u>	0.0267
<u>л</u>	0.02545
<u>ы</u>	0.02542
<u>х</u>	0.01021
<u>б</u>	0.01274
<u>я</u>	0.02545
<u>м</u>	0.01909
<u>ж</u>	0.01021
<u>ф</u>	0.00637
<u>д</u>	0.02036
<u>ь</u>	0.00763
<u>ц</u>	0.01017
<u>ш</u>	0.0089
<u>г</u>	0.01145
<u>э</u>	0.00126

(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	0.00112	0.00562	0	0.006741	0.003379	0.004509	0.006755	0	0	0.007874	0.004494	0.003369	0.011266	0
б	0	0	0	0	0	0.001122	0	0	0	0	0	0	0	0	0.002244
в	0.005629	0	0	0	0	0.005607	0	0	0.005607	0	0.00113	0.004505	0	0.006722	0.002252
г	0.003371	0	0	0	0	0	0	0	0	0	0	0	0	0	0.004496
д	0.005626	0	0	0.00113	0	0.004489	0	0	0	0	0.002247	0	0.00112	0.001118	0
е	0.002245	0.001126	0.001118	0.002243	0.002247	0	0	0	0	0.005593	0	0.002248	0.00113	0.016846	0
ж	0	0	0	0	0	0.005635	0	0	0.001128	0	0	0	0	0.002257	0
з	0.013485	0	0.004502	0	0	0	0	0	0.001122	0	0	0	0	0.001118	0.002261
и	0	0	0.006734	0.001125	0.00224	0.001118	0	0.007855	0.006745	0.007861	0.002247	0	0.003366	0.003377	0.001124
й	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00224	0
к	0.006739	0	0	0	0	0.00112	0.002261	0	0.004501	0	0	0	0	0	0.002236
л	0.001133	0	0	0	0	0.003376	0	0	0.005604	0	0	0	0	0.001113	0.003369
м	0.00113	0	0	0	0	0.003355	0	0	0.003368	0	0.001122	0	0	0	0.001129
н	0.011232	0	0	0	0	0.002244	0	0	0.019124	0	0	0	0	0.010095	0.010121
о	0	0.009007	0.011248	0.002247	0.002259	0	0.001128	0.003374	0	0.002248	0.001124	0.006728	0.002261	0.001124	0.003374
п	0	0	0	0	0	0.001112	0	0	0	0	0	0.001133	0	0	0.015716
р	0.020231	0	0	0.003371	0	0.006747	0	0	0.001112	0	0.001124	0	0.002253	0	0.007878
с	0	0	0	0	0	0.001118	0	0	0.001123	0	0	0.003369	0	0.001123	0.007875
т	0.005637	0	0.006727	0	0	0.002242	0	0	0.012344	0	0.00112	0	0	0.002254	0.005612
у	0	0	0	0	0	0.001121	0	0	0	0.001125	0	0	0	0	0
ф	0	0	0	0	0	0.002243	0	0	0.001134	0	0	0	0	0	0.002253
х	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ц	0	0	0	0	0	0.001123	0	0	0.007855	0	0	0	0	0	0
ч	0.002247	0	0	0	0	0.005606	0	0	0	0	0	0	0	0	0
ш	0.001128	0	0	0	0	0.005602	0	0	0.001126	0	0	0	0	0	0
щ	0	0	0	0	0	0.001125	0	0	0.001122	0	0	0	0	0	0
ы	0	0	0.001124	0	0	0.003363	0	0	0	0.005602	0	0	0.001115	0	0
ь	0	0	0	0	0	0	0	0	0	0	0	0	0	0.002245	0
э	0	0	0	0	0	0	0	0	0	0	0.001112	0	0	0	0
ю	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
я	0	0	0	0	0.002247	0.003381	0	0	0	0	0	0	0	0.002243	0
ё	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ё	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.00786	0	0.008992	0	0.002245	0	0.001122	0.006748	0.010101	0	0.003356	0	0.002248	0.007861	0.011266

Рис.4

Біграми з перекриттям з пробілами.

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	0	0.001118	0.005632	0	0.006727	0.003393	0.004529	0.006767	0	0	0.007894	0.004504	0.003357	0.011271	0
б	0	0	0	0	0	0.001128	0	0	0	0	0	0	0	0	0.002232
в	0.005621	0	0	0	0	0.00561	0	0	0.005592	0	0.001133	0.004494	0	0.006712	0.002245
г	0.003375	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00451
д	0.005638	0	0	0.001126	0	0.004476	0	0	0	0	0.00226	0	0.00112	0.00112	0
е	0.002244	0.001123	0.00112	0.002246	0.002255	0	0	0	0	0.00559	0	0.002246	0.001131	0.01682	0
ж	0	0	0	0	0	0.005611	0	0	0.001133	0	0	0	0	0.002251	0
з	0.013495	0	0.004519	0	0	0	0	0	0.001129	0	0	0	0	0.001118	0.002253
и	0	0	0.006731	0.001122	0.002245	0.001118	0	0.007857	0.006727	0.007851	0.002247	0	0.003369	0.003381	0.001123
й	0	0	0	0	0	0	0	0	0	0	0	0	0	0.002236	0
к	0.006718	0	0	0	0	0.001118	0.002264	0	0.004483	0	0	0	0	0	0.002238
л	0.001132	0	0	0	0	0.003377	0	0	0.005594	0	0	0	0	0.001111	0.003362
м	0.001131	0	0	0	0	0.003373	0	0	0.003369	0	0.001128	0	0	0	0.001126
н	0.011244	0	0	0	0	0.002243	0	0	0.019144	0	0	0	0	0.0101	0.010134
о	0	0.008993	0.011247	0.002247	0.002265	0	0.001123	0.00336	0	0.002261	0.001136	0.006735	0.002268	0.001125	0.003388
п	0	0	0	0	0	0.001114	0	0	0	0	0	0.001135	0	0	0.015708
р	0.0202	0	0	0.003366	0	0.00675	0	0	0.001114	0	0.001126	0	0.002248	0	0.007861
с	0	0	0	0	0	0.00112	0	0	0.001111	0	0	0.003376	0	0.001121	0.007855
т	0.005629	0	0.006721	0	0	0.002246	0	0	0.012358	0	0.001121	0	0	0.002256	0.005623
у	0	0	0	0	0	0.001116	0	0	0	0.001121	0	0	0	0	0
ф	0	0	0	0	0	0.002239	0	0	0.001127	0	0	0	0	0	0.002248
х	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ц	0	0	0	0	0	0.001133	0	0	0.00787	0	0	0	0	0	0
ч	0.002248	0	0	0	0	0.005623	0	0	0	0	0	0	0	0	0
ш	0.001123	0	0	0	0	0.005609	0	0	0.001124	0	0	0	0	0	0
щ	0	0	0	0	0	0.001124	0	0	0.001123	0	0	0	0	0	0
ы	0	0	0.001136	0	0	0.003353	0	0	0	0.005596	0	0	0.001114	0	0
ь	0	0	0	0	0	0	0	0	0	0	0	0	0	0.002238	0
э	0	0	0	0	0	0	0	0	0	0	0.00111	0	0	0	0
ю	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
я	0	0	0	0	0.002233	0.003379	0	0	0	0	0	0	0	0.002249	0
ё	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ё	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.00785	0	0.009001	0	0.002249	0	0.001129	0.006736	0.01011	0	0.00335	0	0.002245	0.007831	0.011282

Рис.5

Біграми без перекриття з пробілами.

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
а	0.00086	0.001268	0.007341		0 0.007881	0.003826	0.005228	0.008388	0.001097		0 0.009284	0.005089	0.00406	0.013609	0.001228
б	0	0	0		0 0	0.00127	0	0	0		0 0	0	0	0	0.002541
в	0.006718		0 0.000394		0 0.000096	0.006349	0.000052	0.000293	0.006789		0 0.001424	0.005101	0.000098	0.007954	0.003049
г	0.003817		0		0 0	0	0	0	0		0 0	0	0	0	0.005091
д	0.00637		0 0.001279		0 0.005083	0	0	0	0		0 0.002544	0	0.001269	0.001266	0
е	0.003228	0.001275	0.002043	0.00254	0.002739		0 0.000098	0.000586	0.000877	0.006333	0.000291	0.002546	0.001476	0.019757	0.000978
ж	0	0	0		0 0.006381	0	0	0.001277		0 0	0	0	0	0.002556	0
з	0.01527	0 0.005098			0 0	0	0	0.001271		0 0	0	0	0	0.001266	0.00256
и	0.001629	0 0.009479	0.001274	0.003	0.001266	0.000232	0.010291	0.009721	0.008902	0.00324		0 0.00428	0.00545	0.003595	
й	0.001363	0 0.001565		0 0.000392		0 0.000195	0.001178	0.001761		0 0.000583		0 0.000387	0.003894	0.001948	
к	0.007631	0	0		0 0.001268	0.002561		0 0.005097		0 0	0	0	0	0.002532	
л	0.001283	0	0		0 0.003823	0		0 0.006345		0 0	0	0	0	0.001261	0.003814
м	0.001705	0 0.000497		0 0.000123	0.003799	0.000062	0.000368	0.004365		0 0.001452		0 0.000121	0.000429	0.001895	
н	0.012719	0	0		0 0.002541	0		0 0.021655		0 0	0	0	0	0.011432	0.011461
о	0.000514	0.010199	0.013321	0.002544	0.002706		0 0.001353	0.004257	0.000663	0.002546	0.001491	0.007619	0.002708	0.001786	0.004554
п	0	0	0		0 0.001259	0		0 0	0	0	0	0.001283		0	0.017797
р	0.022909	0	0 0.003817		0 0.007639	0		0 0.001259		0 0.001273		0 0.002551		0	0.00892
с	0.000086	0 0.000099		0 0.000025	0.001266	0.000013	0.000074	0.001384		0 0.000036	0.003814	0.000024	0.001358	0.009038	
т	0.007324	0 0.008697		0 0.000268	0.002539	0.000129	0.000808	0.015192		0 0.00167		0 0.000269	0.003503	0.00771	
у	0	0	0		0 0.00127	0		0 0		0 0.001274		0	0	0	0
ф	0	0	0		0 0.00254	0		0 0.001284		0 0	0	0	0	0.002551	
х	0.000688	0 0.000785		0 0.000196		0 0.000096	0.000589	0.00088		0 0.000296		0 0.000198	0.000687	0.000981	
ц	0	0	0		0 0.001271	0		0 0.008894		0 0	0	0	0	0	0
ч	0.00263	0 0.000101		0 0.000024	0.006348	0.000011	0.000075	0.000108		0 0.000036		0 0.000023	0.000087	0.000123	
щ	0.001277	0	0		0 0.006343	0		0 0.001275		0 0	0	0	0	0	0
ш	0	0	0		0 0.001274	0		0 0.001271		0 0	0	0	0	0	0
ы	0	0 0.001273		0 0 0.003808		0 0		0 0.006343		0 0	0	0.001262		0	0
ь	0.000343	0 0.00039		0 0.000095		0 0.000048	0.00029	0.000437		0 0.000148		0 0.000099	0.002885	0.000496	
э	0	0	0		0 0	0		0 0		0 0.001259		0 0	0	0	0
ю	0.000086	0 0.000097		0 0.000024		0 0.000013	0.000074	0.000111		0 0.000037		0 0.000024	0.000087	0.000126	
я	0.00085	0 0.000983		0 0.002788	0.003829	0.000125	0.000733	0.001106		0 0.000364		0 0.000245	0.003401	0.001232	
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	0.000855	0.001269	0.007346		0 0.007886	0.003811	0.005226	0.008379	0.001098		0 0.00929	0.005096	0.004058	0.013616	0.001234
б	0	0	0		0 0.001275	0		0 0	0		0 0	0	0	0	0.002541
в	0.006735		0 0.000392		0 0.000095	0.006356	0.000051	0.000294	0.006793		0 0.001423	0.005101	0.000096	0.00794	0.003056
г	0.003795		0		0 0	0		0 0	0		0 0	0	0	0	0.005094
д	0.006366		0 0.00128		0 0.005071	0		0 0	0		0 0.002548		0 0.001266	0.001276	0
е	0.003226	0.001276	0.002054	0.002537	0.002722	0	0.0001	0.000584	0.000869	0.006332	0.000291	0.002546	0.001486	0.019734	0.000984
ж	0	0	0		0 0.006388	0		0 0.001281		0 0	0	0	0	0.002559	0
з	0.015252	0 0.005098			0 0	0		0 0.001262		0 0	0	0	0	0.001266	0.002565
и	0.00163	0 0.009467	0.00128	0.002998	0.001266	0.000231	0.010316	0.0097	0.008897	0.003241		0 0.004266	0.005462	0.003607	
й	0.001361	0 0.00156		0 0.000396		0 0.000194	0.00118	0.001764		0 0.000583		0 0.000384	0.003885	0.001946	
к	0.007629	0	0		0 0.001269	0.002554		0 0.005101		0 0	0	0	0	0.002519	
л	0.001281	0	0		0 0.00383	0		0 0.006326		0 0	0	0	0	0.00125	0.003815
м	0.001712	0 0.000497		0 0.000123	0.003801	0.000065	0.000368	0.00438		0 0.001448		0 0.000121	0.000429	0.001885	
н	0.012744	0	0		0 0.00253	0		0 0.02167		0 0	0	0	0	0.011441	0.011456
о	0.000516	0.010185	0.013328	0.002544	0.002712		0 0.001348	0.004263	0.00066	0.002557	0.001491	0.007642	0.002704	0.001774	0.004549
п	0	0	0		0 0.001262	0		0 0	0	0	0	0.001286		0	0.01778
р	0.022895	0	0 0.003838		0 0.007654	0		0 0.001262		0 0.001279		0 0.002538		0	0.008941
с	0.000084	0 0.000098		0 0.000024	0.001276	0.000014	0.000074	0.001383		0 0.000035	0.003814	0.000025	0.001364	0.009057	
т	0.007344	0 0.008718		0 0.000271	0.002552	0.000128	0.000797	0.015196		0 0.001673		0 0.000271	0.003493	0.007705	
у	0	0	0		0 0.001278	0		0 0		0 0.001268		0 0	0	0	0
ф	0	0	0		0 0.002538	0		0 0.001288		0 0	0	0	0	0.002538	
х	0.000691	0 0.000781		0 0.000197	0 0.000096	0.000096	0.000589	0.000871		0 0.000296		0 0.000201	0.000686	0.000989	
ц	0	0	0		0 0.001266	0		0 0.008868		0 0	0	0	0	0	0
ч	0.002639	0 0.000099		0 0.000025	0.006345	0.000011	0.000076	0.000109		0 0.000035		0 0.000023	0.000088	0.000122	
щ	0.001273	0	0		0 0.006351	0		0 0.001276		0 0	0	0	0	0	0
ш	0	0	0		0 0.001277	0		0 0.001281		0 0	0	0	0	0	0
ы	0	0 0.001272		0 0 0.003815		0 0		0 0.006343		0 0	0	0.001258		0	0
ь	0.000345	0 0.000393		0 0.000095		0 0.000047	0.000295	0.000441		0 0.000151		0 0.000099	0.002878	0.000488	
э	0	0	0		0 0	0		0 0.001256		0 0	0	0	0	0	0
ю	0.000085	0 0.000101		0 0.000023		0 0.000013	0.000074	0.000112		0 0.000038		0 0.000023	0.000085	0.000126	
я	0.000839	0 0.000987		0 0.002788	0.003836	0.000122	0.000735	0.001101		0 0.000368		0 0.000248	0.003398	0.001231	
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	4.407140	0.1337
Біграми з перекриттям	3.678557	0.2771
Біграми без перекриття	3.678546	0.2771

	Ентропія	Надлишковість
Монограми	4.401240	0.1276
Біграми з перекриттям	3.776586	0.2661
Біграми без перекриття	3.776729	0.2661

 $H_{10}:$ 

Лабораторная работа №1

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:  
5 символов  
10 символов  
15 символов  
20 символов  
25 символов  
30 символов  
35 символов  
40 символов  
45 символов  
50 символов

Введенный символ: е

Символ по счету: 7

Номер эксперимента: 50

Поле ввода символов:  
е

Продолжить Другой

Неравенство для энтропии:  
3,29318839409308< H < 3,9766912610489

Двоичная таблица угаданных символов:  
000000000000000100000000000000  
010000000000000000000000000000  
001000000000000000000000000000  
000100000000000000000000000000  
000000010000000000000000000000  
.....

Вероятности:  
q[ 1 ] = 0.3  
q[ 2 ] = 0.08  
q[ 3 ] = 0.02  
q[ 4 ] = 0.04  
q[ 5 ] = 0.04  
q[ 6 ] = 0.04  
q[ 7 ] = 0.02  
q[ 8 ] = 0.02  
q[ 9 ] = 0.02  
q[ 10 ] = 0.02  
q[ 11 ] = 0.02  
q[ 12 ] = 0.02  
q[ 13 ] = 0  
q[ 14 ] = 0.02  
q[ 15 ] = 0  
q[ 16 ] = 0.02  
q[ 17 ] = 0.02  
q[ 18 ] = 0.04  
q[ 19 ] = 0.02  
q[ 20 ] = 0.04  
q[ 21 ] = 0  
q[ 22 ] = 0.02  
q[ 23 ] = 0.06  
q[ 24 ] = 0.04  
q[ 25 ] = 0  
q[ 26 ] = 0.02  
q[ 27 ] = 0.02  
q[ 28 ] = 0  
q[ 29 ] = 0.02  
q[ 30 ] = 0.02  
q[ 31 ] = 0  
q[ 32 ] = 0

Строка состояния:  
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.