

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

Виконали:
Студенти ФБ-01
Новак О. І.
Тостоган Є. Г.

Київ 2022

Мета роботи: засвоєння поняття 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.

Постановка задачі:

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 заміни відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилучено всі пробіли. За допомогою програми CoolPinkProgram оцінити значення $H(10)$, $H(20)$, $H(30)$. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи

Для початку для визначення частоти входження символів в текст було написано програму на мові python3.

Програма містить в собі такий функціонал:

- ❖ Обробка тексту(видалення зайвих символів та абзаців, заміна регістру літер на малі)
- ❖ Підрахунок кількості входжень кожного символу в текст та відповідно визначення частоти
- ❖ Пошук біграм, що перетинаються та не перетинаються, та визначення частоти їх входжень у текст
- ❖ Обчислення ентропії H_1 та H_2 для тексту з пробілами та без
- ❖ Внесення результатів у датафрейми та таблиці excel.

Код програми міститься у файлі crypto1.py

В ході роботи ми імпортували потрібні для реалізації програмного коду бібліотеки:

```
import collections
import unicodedata
import pandas as pd
import re
import math as m
import numpy as np
```

та провели обробку обраного тексту

```
note = open("/home/kali/lab1.txt").read()
#edit text: delete punctuation marks and spaces, replace capital letters with lowercase letters
note = note.replace("\n", "")
note = note.lower()
new_note1 = re.sub( r'^[а-яё]', '', note )
new_note2 = re.sub( r'^[а-яё ]', '', note )
#print(new_note1)
snote1 = sorted(new_note1)
snote2 = sorted(new_note2)
#print(snote1)
```

```
$ python3 crypt1.py
```

и ильфи е петров завершили роман двенадцать стульев в году но еще до первой публикации цензоры изрядно
к изданию еще десять лет в итоге книга уменьшилась почти на треть публикуемый ныне вариант первый полный
на обширном историколитературном и реальным комментарием ильфи е евгений петров легенда о великом комби
глава третья глава глава четвертая глава глава глава пятая глава глава шестая глава глава седьмая глава глава
десятая глава глава глава одиннадцатая глава глава двенадцатая глава глава тринадцатая глава глава глава четыр
я глава шестнадцатая глава глава семнадцатая глава глава восемнадцатая глава глава девятнадцатая глава глава
а глава глава глава глава глава глава глава глава глава глава часть вторая глава глава глава глава
лава глава глава глава часть третья глава глава глава глава глава глава глава глава ильфи иль
у петрови катаеву легенда о великом комбинаторе или почему в шанхае ничего не случилось происхождение леген
аристами и многократно пересказанной литературоведами вымысел практически неотделим от фактов реальность
емляки одеситы оказались в москве не позже года поэт и журналист ильфи арнольдвич файнзильберг взял псе
ого уголовного розыска евгений петрови катаев свой псевдоним петров выбрал вероятно сменив профессию
вавшейся центральным комитетом профессионального союза рабочих железнодорожного транспорта ссср в гудке ра

```
$ python3 crypt1.py
```

и ильфи е петров завершили роман двенадцать стульев в году но еще до первой публикации цензоры изрядно
к изданию еще десять лет в итоге книга уменьшилась почти на треть публикуемый ныне вариант первый полный
на обширном историколитературном и реальным комментарием ильфи е евгений петров легенда о великом комби
глава третья глава глава четвертая глава глава глава пятая глава глава шестая глава глава седьмая глава глава
десятая глава глава глава одиннадцатая глава глава двенадцатая глава глава тринадцатая глава глава глава четыр
я глава шестнадцатая глава глава семнадцатая глава глава восемнадцатая глава глава девятнадцатая глава глава
а глава глава глава глава глава глава глава глава глава глава часть вторая глава глава глава глава
лава глава глава глава часть третья глава глава глава глава глава глава глава глава ильфи иль
у петрови катаеву легенда о великом комбинаторе или почему в шанхае ничего не случилось происхождение леген
аристами и многократно пересказанной литературоведами вымысел практически неотделим от фактов реальность
емляки одеситы оказались в москве не позже года поэт и журналист ильфи арнольдвич файнзильберг взял псе
ого уголовного розыска евгений петрови катаев свой псевдоним петров выбрал вероятно сменив профессию
вавшейся центральным комитетом профессионального союза рабочих железнодорожного транспорта ссср в гудке ра

Рахуємо кількість та частоту появи літери в тексті:

```
#determine quantity and periodicity of letters in our note
quantity1 = dict(collections.Counter(snote1))
quantity2 = dict(collections.Counter(snote2))
period1 = {l: quantity1[l]/len(snote1) for l in quantity1}
period2 = {l: quantity2[l]/len(snote2) for l in quantity2}
```

Функції для створення біграм з утворених списків літер:

```
#Create bigrams(w/o crossing)
def bigramCross(new_note):
    bigram_cross = []
    for i in range(0, len(new_note)-1):
        bigram_cross.append(new_note[i]+new_note[i+1])
    return bigram_cross

def bigram(new_note):
    bigram=[]
    for i in range(0, len(new_note)-2,2):
        bigram.append(new_note[i]+new_note[i+1])
    return bigram
```

За допомогою отриманих функції створюємо біграми та перетворюємо їх у словники для подальшої роботи з ними.

```
#bigram without spaces
bigram_cross1 = bigramCross(new_note1)
bigram1 = bigram(new_note1)
bg_cross_q1 = dict(collections.Counter(bigram_cross1))
bg_q1 = dict(collections.Counter(bigram1))
bg_cross_period1 = {l: bg_cross_q1[l] / len(bigram_cross1) for l in bg_cross_q1}
bg_period1 = {l: bg_q1[l] / len(bigram1) for l in bg_q1}

#bigram with spaces
bigram_cross2 = bigramCross(new_note2)
bigram2 = bigram(new_note2)
bg_cross_q2 = dict(collections.Counter(bigram_cross2))
bg_q2 = dict(collections.Counter(bigram2))
bg_cross_period2 = {l: bg_cross_q2[l] / len(bigram_cross2) for l in bg_cross_q2}
bg_period2 = {l: bg_q2[l] / len(bigram2) for l in bg_q2}
```

Функції для створення дата фреймів та таблиць для літер та біграм. У ході виконання стикнулися з проблемою перезапису таблиці, тому імя таблиць задаємо окремо з командного рядка.

```
#create dataframe for future analysis
def createDataFrame(quantity, periodicity):
    quantity = sorted(quantity, key=lambda l: quantity[l], reverse=1 )
    speriod = sorted(periodicity, key=lambda l: periodicity[l], reverse=1 )
    temp1 = []
    temp2 = []
    for i in range(0,len(squantity)):
        temp1.append(quantity[squantity[i]])
    for i in range(0,len(speriod)):
        temp2.append(periodicity[speriod[i]])

    df= pd.DataFrame(index = squantity)
    df['quantity'] = temp1
    df['periodicity'] = temp2
    name=input('Enter name of excel: ')
    df.to_excel(f'{name}.xlsx')
    print(df.head(10))

#periodicity matrix for bigram
def createbgDataFrame(bigram, periodicity, symb):
    global bg
    df = pd.DataFrame(index = symb, columns=symb)
    n = 0
    for i in range(0,len(symb)):
        df[symb[i]] = bg[n:len(symb)+n]
        n = len(symb)+n
    df = df.T
    for i in list(periodicity.keys()):
        x,y = np.where(df == i)
        df.iloc[x,y] = periodicity[i]
    for i in bg:
        x,y = np.where(df == i)
        df.iloc[x,y] = 0
    name=input('Enter name of excel: ')
    df.to_excel(f'{name}.xlsx')
    print(df)
```

Функції для підрахунку ентропії з отриманих даних:

```
#H_1
def H_1(periodicity):
    h_1 = []
    for p in periodicity.values():
        h_1.append(p*m.log(p,2))
    H_1 = -sum(h_1)
    return H_1

#H_2
def H_2(bigram, periodicity):
    h_2 = []
    for p in periodicity.values():
        h_2.append(p*m.log(p,2))
    H_2 = -sum(h_2)/2
    return H_2
```

Вивід результатів:

```
(kali@kali)~[~]
$ python3 crypt1.py
/////////////////WITHOUT SPACES/////////////////
quantity  periodicity
о  77418   0.107207
а  60763   0.084143
е  58343   0.080792
и  52986   0.073374
н  44964   0.062265
т  41635   0.057655
с  39339   0.054476
л  35421   0.049050
р  35329   0.048923
в  34264   0.047448

H_1(entropy) without spaces: 4.466033215175493
Excess_1: 0.11465418650843173
periodicity
и и  0.014294
ил  0.012092
ль  0.011735
ьф  0.011473
фи  0.011081
ие  0.010796
еп  0.010623
пе  0.010321
ет  0.010252
тр  0.010102

H_2(entropy) cross without spaces: 4.16681166976208
Excess_2: 0.17397182472887052
periodicity
и и  0.014116
ль  0.012045
фи  0.011865
еп  0.011231
ет  0.011184
ро  0.010671
аз  0.010530
ав  0.010303
ер  0.010175
ши  0.009954

H_2(entropy) without spaces: 4.166160918982793
Excess_3: 0.1741008294743145

/////////////////////////////////WITH SPACES/////////////////////////////////
quantity  periodicity
о  125265  0.147823
а  77418   0.091359
е  60763   0.071705
и  58343   0.068849
н  52986   0.062528
т  44964   0.053061
с  41635   0.049133
л  39339   0.046423
р  35421   0.041800
в  35329   0.041691

H_1(entropy) with spaces: 4.410217046799943
Excess_4: 0.13312053877998453
periodicity
и  0.015778
и  0.015669
ил  0.015604
ль  0.015374
ьф  0.014543
ф  0.014319
е  0.013932
е  0.012017
п  0.011710
пе  0.009843

H_2(entropy) cross with spaces: 4.047886140180645
Excess_5: 0.20434089319347215
periodicity
и  0.015955
ил  0.015674
ьф  0.015492
и  0.015317
е  0.014671
п  0.014423
ет  0.013880
ро  0.011898
в  0.011619
за  0.009752

H_2(entropy) with spaces: 4.047468491754768
Excess_6: 0.20442298684976212
```

Ентропія

	Текст з пробілами	Текст без пробілів
Н1	4.410230057145229	4.465995972453169
Н2	4.047895630341044	4.166762456909701
Н2(біграми з кроком 2)	4.047494532339943	4.166128394668092

Надлишковість існує в межах від 0,11-0,2.

Для утворення матриці частот попередньо було додано два алфавіти у вигляді списку.

Матриці частот з отриманих даних:

Bigram_cross1()

Enter name of excel: bigram_cross1

	а	б	в	г	д	е	...	ь	ы	ь	э	ю	я
а	0.000436	0.001774	0.005899	0.001501	0.003213	0.001461	...	0	0	0	0.000298	0.00078	0.002734
б	0.001232	0.000022	0.000076	0.00001	0.000024	0.002584	...	0.000163	0.00316	0.000353	0.000026	0.000089	0.000256
в	0.007928	0.000307	0.000576	0.000587	0.000781	0.006576	...	0.000003	0.003358	0.000125	0.000166	0.000014	0.00036
г	0.001883	0.000029	0.000093	0.000022	0.000892	0.000467	...	0.000001	0	0.000001	0.000015	0	0.000004
д	0.00523	0.000102	0.001478	0.000069	0.000109	0.005734	...	0.000037	0.000735	0.00041	0.000011	0.00003	0.000395
е	0.000497	0.001962	0.004633	0.003155	0.004308	0.002087	...	0	0	0	0.00022	0.000126	0.000314
ё	0	0	0	0	0	0	...	0	0	0	0	0	0
ж	0.001518	0.000054	0.000033	0.000025	0.000953	0.003585	...	0	0	0.000035	0.000008	0	0.000008
з	0.005711	0.000188	0.001223	0.000309	0.001127	0.000679	...	0.000028	0.000543	0.000112	0.000021	0.000001	0.000281
и	0.000859	0.001203	0.004451	0.001277	0.002621	0.002937	...	0	0.000001	0	0.000217	0.000467	0.001867
й	0.000248	0.000404	0.000769	0.000422	0.000648	0.000123	...	0	0	0	0.000097	0.000017	0.000062
к	0.007749	0.00021	0.000895	0.000152	0.000217	0.000937	...	0	0	0	0.000073	0.000006	0.000044
л	0.005863	0.000307	0.000931	0.000396	0.000309	0.005727	...	0.000001	0.000715	0.005244	0.000069	0.001075	0.001997
м	0.004355	0.00063	0.000886	0.000367	0.000395	0.004413	...	0.000003	0.000975	0.000119	0.000084	0.000014	0.00053
н	0.011081	0.000208	0.0004	0.000319	0.000992	0.008129	...	0.000011	0.005288	0.001123	0.000046	0.000119	0.001167
о	0.000309	0.004971	0.012092	0.005466	0.007573	0.002224	...	0	0.000001	0	0.000363	0.0003	0.000896
п	0.002306	0.000073	0.000104	0.000021	0.000035	0.002987	...	0	0.000341	0.000104	0.000018	0.000015	0.000395
р	0.009702	0.00024	0.000643	0.000627	0.000409	0.007151	...	0.000001	0.001727	0.000555	0.000017	0.000176	0.001005
с	0.002199	0.000231	0.001659	0.000186	0.000321	0.003211	...	0.000042	0.000507	0.003168	0.000069	0.000242	0.00393
т	0.007672	0.000284	0.003988	0.000138	0.000476	0.006751	...	0.000017	0.001724	0.004633	0.000084	0.00017	0.000636
у	0.000028	0.001502	0.00117	0.000982	0.002153	0.000422	...	0	0	0	0.00009	0.001263	0.000151
ф	0.000356	0.000004	0.000004	0.000003	0.000004	0.000499	...	0	0.000019	0.000024	0.000001	0.000007	0
х	0.000932	0.000172	0.000535	0.000168	0.000183	0.000148	...	0.000001	0	0	0.000065	0.000003	0.000019
ц	0.000975	0.000019	0.000228	0.000014	0.000022	0.001162	...	0	0.000305	0	0.000003	0.000001	0.000001
ч	0.002519	0.00004	0.000116	0.000022	0.000036	0.003924	...	0	0	0.000305	0.000008	0	0.000011
ш	0.001256	0.000008	0.000047	0.000006	0.000004	0.002329	...	0	0	0.000242	0	0.000006	0.000001
щ	0.000505	0.00003	0.000017	0.000008	0.000001	0.001464	...	0	0	0.000039	0	0	0
ъ	0	0.000001	0	0.000001	0	0.000155	...	0	0	0	0.000001	0.000011	0.000129
ы	0.00009	0.000386	0.001617	0.000312	0.000402	0.001734	...	0	0	0	0.000057	0.000001	0.000042
ь	0.000166	0.000422	0.001036	0.000391	0.000568	0.000883	...	0	0	0	0.00013	0.000456	0.000843
э	0	0.000001	0.000006	0	0.000004	0	...	0	0	0	0.000003	0	0
ю	0.000064	0.000427	0.000255	0.000115	0.000407	0.000025	...	0	0	0	0.00003	0.000039	0.000036
я	0.000223	0.000525	0.00159	0.000457	0.000865	0.000282	...	0	0	0	0.000105	0.000105	0.000129

[33 rows x 33 columns]

Bigram1()

Enter name of excel: bigram1

	а	б	в	г	д	е	...	ь	ы	ь	э	ю	я
а	0.000421	0.001786	0.005822	0.001515	0.003163	0.001415	...	0	0	0	0.000307	0.00077	0.002698
б	0.001246	0.000014	0.000066	0.000006	0.000019	0.002534	...	0.00018	0.003102	0.000324	0.000028	0.000091	0.000266
в	0.007863	0.000307	0.00059	0.000579	0.000748	0.006567	...	0.000006	0.003348	0.000122	0.000163	0.000017	0.000385
г	0.001894	0.000033	0.000069	0.000017	0.000864	0.000474	...	0	0	0.000003	0.000019	0	0.000003
д	0.005284	0.000108	0.001454	0.000064	0.000105	0.005669	...	0.000044	0.00072	0.000443	0.000008	0.000042	0.000399
е	0.000515	0.00195	0.004758	0.003243	0.004351	0.002063	...	0	0	0	0.00021	0.000133	0.00033
ё	0	0	0	0	0	0	...	0	0	0	0	0	0
ж	0.001496	0.000042	0.000025	0.000019	0.000986	0.003578	...	0	0	0.000036	0.000008	0	0.000008
з	0.005536	0.000172	0.00128	0.000307	0.001155	0.000673	...	0.000017	0.000554	0.000091	0.000025	0.000003	0.000288
и	0.000864	0.001127	0.004478	0.001313	0.002709	0.003005	...	0	0	0	0.000238	0.000479	0.001936
й	0.000244	0.000388	0.000762	0.00044	0.00067	0.00013	...	0	0	0	0.000078	0.000017	0.000072
к	0.007752	0.000222	0.000917	0.000158	0.000213	0.000931	...	0	0	0	0.000069	0.000003	0.000039
л	0.005844	0.000299	0.000939	0.000385	0.000299	0.005705	...	0	0.00069	0.005329	0.000069	0.001136	0.002033
м	0.004246	0.000656	0.000886	0.000363	0.000388	0.004523	...	0	0.000983	0.000125	0.000094	0.000019	0.000515
н	0.011184	0.000219	0.000377	0.000319	0.000953	0.00809	...	0.000006	0.005306	0.001102	0.000039	0.0001	0.001177
о	0.000285	0.005196	0.012045	0.005536	0.007467	0.002216	...	0	0.000003	0	0.000393	0.000299	0.000878
п	0.002282	0.000083	0.000111	0.00003	0.000042	0.003011	...	0	0.000366	0.000114	0.000017	0.000019	0.000388
р	0.009713	0.000219	0.000634	0.000643	0.000399	0.00699	...	0.000003	0.001728	0.000587	0.000017	0.000177	0.001039
с	0.002185	0.000244	0.001681	0.000197	0.00033	0.003196	...	0.000042	0.00051	0.003188	0.000058	0.000255	0.003955
т	0.007846	0.000294	0.003996	0.00013	0.000465	0.006916	...	0.000014	0.001775	0.004642	0.00008	0.000172	0.000573
у	0.000263	0.001498	0.00118	0.000983	0.002169	0.00044	...	0	0	0	0.000075	0.001263	0.000163
ф	0.000341	0.000003	0.000006	0.000003	0.000003	0.000526	...	0	0.000019	0.000019	0.000003	0.000008	0
х	0.000911	0.000174	0.000562	0.000163	0.000183	0.00013	...	0.000003	0	0	0.000075	0.000003	0.000017
ц	0.001011	0.000022	0.000224	0.000011	0.000025	0.001227	...	0	0.000302	0	0.000003	0	0
ч	0.002639	0.000039	0.000108	0.000019	0.000036	0.003819	...	0	0	0.00031	0.000006	0	0.000011
ш	0.001282	0.000014	0.000053	0.000011	0.000006	0.00234	...	0	0	0.000233	0	0.000011	0.000003
щ	0.000512	0.000025	0.000019	0.000011	0	0.001393	...	0	0	0.00003	0	0	0
ъ	0	0.000003	0	0	0	0.000147	...	0	0	0	0	0.000017	0.000125
ы	0.000091	0.000402	0.001626	0.000285	0.000415	0.001731	...	0	0	0	0.000053	0	0.000042
ь	0.000041	0.000429	0.001072	0.000385	0.000535	0.000861	...	0	0	0	0.000125	0.000451	0.000853
э	0	0	0.000003	0	0.000003	0	...	0	0	0	0.000003	0	0
ю	0.000066	0.000388	0.000255	0.000111	0.000391	0.000022	...	0	0	0	0.000033	0.000042	0.000033
я	0.000219	0.000515	0.001523	0.000479	0.000795	0.00031	...	0	0	0	0.000078	0.000127	0.000136

[33 rows x 33 columns]

Bigram_cross2()

Enter name of excel: bigram_cross2

	а	б	в	г	д	...	ь	ы	ь	э	ю	я	
	0	0.002522	0.006084	0.013932	0.004497	0.006293	...	0	0	0	0.001735	0.00009	0.000985
а	0.015669	0.000077	0.000827	0.00346	0.000841	0.00207	...	0	0	0	0.000032	0.000649	0.002191
б	0.000235	0.001043	0.000017	0.00005	0.000006	0.000012	...	0.000139	0.002693	0.000301	0	0.000076	0.000214
в	0.006931	0.006653	0.000026	0.000053	0.000041	0.000341	...	0.000002	0.002862	0.000106	0.000008	0.000002	0.000277
г	0.000492	0.001598	0.000001	0.000025	0.000006	0.000727	...	0.000001	0	0.000001	0.000007	0	0
д	0.001079	0.004437	0.000041	0.001155	0.00003	0.000057	...	0.000032	0.000627	0.000349	0.000004	0.000026	0.000335
е	0.014543	0.00021	0.000893	0.002656	0.002232	0.002976	...	0	0	0	0.000009	0.000106	0.000166
ё	0	0	0	0	0	0	...	0	0	0	0	0	0
ж	0.000191	0.001292	0.000035	0.000002	0.000014	0.000806	...	0	0	0.00003	0	0	0
з	0.001328	0.004852	0.000116	0.000918	0.000196	0.000866	...	0.000024	0.000463	0.000096	0.000001	0.000001	0.000234
и	0.015778	0.000487	0.000345	0.002187	0.000671	0.001522	...	0	0.000001	0	0.000012	0.000388	0.001492
й	0.008362	0.000002	0.000027	0.000012	0.000015	0.000127	...	0	0	0	0	0.000002	0.000011
к	0.003665	0.006541	0.000004	0.000382	0.00002	0.000019	...	0	0	0	0.000004	0.000002	0
л	0.006474	0.004884	0.000057	0.000035	0.000162	0.000054	...	0.000001	0.000609	0.000469	0.000001	0.000915	0.001663
м	0.006881	0.003571	0.000287	0.000119	0.000027	0.000009	...	0.000002	0.000831	0.000101	0	0.000007	0.000397
н	0.002688	0.009409	0.000024	0.000041	0.000214	0.000728	...	0.000009	0.004507	0.000957	0.000024	0.000098	0.000987
о	0.015374	0.000048	0.003481	0.008716	0.004248	0.005766	...	0	0.000001	0	0.000139	0.000253	0.000642
п	0.000702	0.001954	0.000013	0.000008	0.000004	0.000005	...	0	0.00029	0.000089	0.000001	0.000013	0.000328
р	0.001638	0.008237	0.000124	0.000394	0.000485	0.000284	...	0.000001	0.001472	0.000473	0.000002	0.00015	0.000845
с	0.002685	0.001827	0.000092	0.001178	0.000035	0.000159	...	0.000035	0.000432	0.0027	0.000004	0.000203	0.003328
т	0.004686	0.00645	0.000026	0.002979	0.000007	0.000227	...	0.000014	0.001469	0.003949	0.000004	0.000145	0.000515
у	0.005439	0.000122	0.001096	0.000443	0.000709	0.001617	...	0	0	0	0.000028	0.001073	0.00009
ф	0.000092	0.000295	0	0	0	0	...	0	0.000017	0.00002	0.000001	0.000005	0
х	0.003517	0.000713	0.000004	0.000152	0.000009	0.000007	...	0.000001	0	0	0.000008	0	0
ц	0.000448	0.000823	0.000001	0.000153	0.000002	0	...	0	0.00026	0	0	0	0
ч	0.000861	0.002131	0.000004	0.000011	0.000004	0.000001	...	0	0	0.00026	0.000002	0	0
ш	0.000079	0.001069	0	0.000032	0	0	...	0	0	0.000207	0	0.000005	0
щ	0.000098	0.000426	0	0.000002	0	0.000001	...	0	0	0.000033	0	0	0
ъ	0.000006	0	0.000001	0	0	0	...	0	0	0	0	0.000009	0.00011
ы	0.004319	0.000004	0.000151	0.000946	0.000146	0.000125	...	0	0	0	0.000002	0	0.000014
ь	0.00741	0.000008	0.000087	0.000061	0.000164	0.000175	...	0	0	0	0.000007	0.000384	0.000653
э	0.000012	0	0	0.000005	0	0.000004	...	0	0	0	0.000002	0	0
ю	0.002195	0.000007	0.000287	0.000017	0.000025	0.000253	...	0	0	0	0.000004	0.000031	0.000001
я	0.008551	0.000012	0.000117	0.000309	0.000124	0.000379	...	0	0	0	0.000002	0.000083	0.000072

[34 rows x 34 columns]

Bigram2()

Enter name of excel: bigram2

	а	б	в	г	д	...	ь	ы	ь	э	ю	я
а	0	0.002549	0.006266	0.01388	0.004529	0.006191	...	0	0	0.001685	0.000094	0.000984
б	0.015492	0.00008	0.000824	0.003413	0.000828	0.002027	...	0	0	0.000028	0.000604	0.002145
в	0.000243	0.00106	0.000019	0.000057	0.000007	0.000012	...	0.000144	0.00262	0.00033	0	0.000061
г	0.006896	0.006649	0.000031	0.000052	0.000047	0.000302	...	0	0.002934	0.000094	0.000009	0.000002
д	0.000444	0.001666	0	0.000026	0.000007	0.000713	...	0	0	0.000002	0.000005	0
е	0.00105	0.004522	0.000042	0.001119	0.000019	0.000045	...	0.000026	0.000611	0.000335	0.000002	0.000033
ё	0.014671	0.000215	0.000906	0.00266	0.002247	0.002986	...	0	0	0	0.000009	0.000106
ж	0	0	0	0	0	0	...	0	0	0	0	0
з	0.000205	0.001277	0.000052	0.000005	0.000012	0.000786	...	0	0	0.000028	0	0
и	0.001312	0.004968	0.000125	0.000932	0.000191	0.000895	...	0.000019	0.000441	0.00009	0	0.000205
й	0.015955	0.000496	0.000361	0.002193	0.000656	0.001496	...	0	0	0	0.000014	0.000434
к	0.008301	0.000005	0.000026	0.000012	0.000024	0.000127	...	0	0	0	0	0.000007
л	0.003767	0.006519	0	0.000385	0.000019	0.000014	...	0	0	0.000005	0.000002	0
м	0.006571	0.005143	0.000045	0.000031	0.000172	0.000073	...	0	0.000635	0.004529	0	0.000956
н	0.006866	0.003623	0.000264	0.000116	0.000026	0.000009	...	0.000002	0.000831	0.000101	0	0.000007
о	0.002658	0.009519	0.000021	0.000042	0.000227	0.000751	...	0.000009	0.004569	0.000892	0.000026	0.000101
п	0.015317	0.000047	0.003399	0.008659	0.004281	0.005773	...	0	0	0	0.000158	0.000269
р	0.000739	0.002044	0.000012	0.000014	0.000005	0.000002	...	0	0.00029	0.000083	0	0.000007
с	0.001624	0.00844	0.000135	0.000394	0.000524	0.000278	...	0	0.001522	0.000529	0.000002	0.000144
т	0.00278	0.001846	0.000087	0.001149	0.000035	0.00013	...	0.000047	0.000411	0.002747	0	0.000179
у	0.004737	0.006646	0.000014	0.003028	0.000009	0.000219	...	0.000009	0.001539	0.00392	0.000002	0.000156
ф	0.005445	0.000146	0.001095	0.000415	0.000699	0.001607	...	0	0	0	0.000031	0.001105
х	0.000097	0.000302	0	0	0	0	...	0	0.000007	0.000012	0	0.000005
ц	0.003559	0.000694	0.000002	0.00017	0.000009	0.000009	...	0	0	0	0.000009	0
ч	0.00046	0.000826	0.000002	0.00016	0.000002	0	...	0	0.000271	0	0	0
ш	0.000845	0.002176	0.000002	0.000014	0.000005	0	...	0	0	0.000229	0.000005	0
щ	0.00009	0.001105	0	0.000026	0	0	...	0	0	0.000179	0	0.000009
ъ	0.00009	0.000437	0	0.000005	0	0	...	0	0	0.000045	0	0
ы	0.000005	0	0.000002	0	0	0	...	0	0	0	0	0.000009
ь	0.00443	0.000002	0.000135	0.000942	0.000144	0.000109	...	0	0	0	0.000002	0
э	0.007413	0.000012	0.00008	0.000061	0.000175	0.00017	...	0	0	0	0.000002	0.000373
ю	0.000014	0	0	0.000005	0	0.000005	...	0	0	0	0.000002	0
я	0.002089	0.000005	0.000274	0.000019	0.000026	0.000241	...	0	0	0	0.000002	0.000028
	0.008518	0.000009	0.000113	0.000307	0.000139	0.000394	...	0	0	0	0	0.000085

[34 rows x 34 columns]

Для зручності перегляду були створені таблиці

Робота з CoolPinkProgram

$$1,18 < H^{(10)} < 1,75$$

$$0,768057 > R > 0,656017$$

[illegible]

$$1,26 < H^{(20)} < 1,83$$

$$0,752332 > R > 0,640292$$

Лабораторная работа №1

Произвольная часть текста:
есть_законом_природы_сегодня_когда_мы_говорим_о_законах_природы_мы_обычно_п

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: **ы**
Символ по счету: **1**
Номер эксперимента: **53**
Поле ввода символов:
ы
Продолжить Другой

Неравенство для энтропии:
 $1,26051106279087 < H < 1.83392116100822$
Двоичная таблица угаданных символов:
01000000000000000000000000000000 ^
10000000000000000000000000000000
10000000000000000000000000000000
00000000001000000000000000000000
10000000000000000000000000000000 v

Вероятности:
q[1] = 0.6792452
q[2] = 0.0566037
q[3] = 0.0377358
q[4] = 0.0566037
q[5] = 0
q[6] = 0.0188679
q[7] = 0
q[8] = 0
q[9] = 0
q[10] = 0.037735
q[11] = 0
q[12] = 0.018867
q[13] = 0
q[14] = 0
q[15] = 0.037735
q[16] = 0
q[17] = 0
q[18] = 0.056603
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0
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

$$1,56 < H^{(30)} < 2,29$$

$$0,693364 > R > 0,549874$$

Лабораторная работа №1

Произвольная часть текста:
о_что_я_не_пытаюсь_показаться_лучше_других_я_просто_стараюсь_обратить_ваше_

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: 54

Неравенство для энтропии:
 $1,56452114299951 < H < 2,29508131854441$

Двоичная таблица угаданных символов:
01000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Поле ввода символов:
Продолжить Другой

Вероятности:
q[1] = 0,6111111
q[2] = 0,0925925
q[3] = 0
q[4] = 0,0370370
q[5] = 0,0185185
q[6] = 0,0555555
q[7] = 0,0185185
q[8] = 0,0185185
q[9] = 0
q[10] = 0
q[11] = 0,037037
q[12] = 0,018518
q[13] = 0
q[14] = 0
q[15] = 0,018518
q[16] = 0
q[17] = 0
q[18] = 0
q[19] = 0,018518
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0
q[25] = 0,018518
q[26] = 0
q[27] = 0
q[28] = 0,018518
q[29] = 0
q[30] = 0,018518
q[31] = 0
q[32] = 0

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

Оцінка надлишковості мови:

	R(Текст з пробілами)	R(Текст без пробілів)
H1	0.1331179814452792	0.11466156950060935
H2	0.20433902779205404	0.1739815806779923
H2(з кроком 2)	0.20441786826983577	0.1741072770900105

Висновки:

У ході виконання лабораторної роботи ми ознайомились з поняттям ентропії та навчилися її визначати, освоїли вимірювання частоти повторювання символів в тексті. Ми дійшли такого висновку, що коли оброблений текст містить пробіл, ентропія є меншою. За допомогою програми CoolPinkProgram ми впевнилися, що чим більшою є ентропія, тим меншою є надлишковість мови.