

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 замінити 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
```

ильф и 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
ильф и евгений петров легенда о великом комбинаторе часть первая глава первая глава
ава глава седьмая глава глава восьмая часть вторая москва глава девятая глава глава дес
лава четырнадцатая глава глава пятнадцатая глава часть третья глава шестнадцатая
тая глава часть первая глава глава глава глава глава глава глава глава глава
а глава глава глава глава глава глава часть третья глава глава глава глава
овичу катаеву легенда о великом комбинаторе или почему в шанхае ничего не случилось прои
рессказанной литературоведами вымысел практически неотделим от фактов реальность отми
апрот и журнал истинья арнольд ович файнзильберг взял псевдоним ильф ешев одесса в отбы

```
$ python3 crypt1.py
```

и ильф и 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 полн
на обширном историколитературном и реальным комментарием ильф и евгений петров легенда о великом комб
глава третья глава глава четвертая глава глава глава пятая глава глава шестая глава глава седьмая глава глава
есятая глава глава глава одиннадцатая глава глава двенадцатая глава глава тринадцатая глава глава глава четы
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 глава часть вторая глава глава глава
лава глава глава глава часть третья глава глава глава глава глава глава глава глава ильф ил
у петровичу 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 иль арнольд ович файнзильберг взял псе
ого уголовного розыска евгений петрович 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):
    squantity = 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:

```
print('/////////////////WITHOUT SPACES/////////////////')
createDataFrame(quantity1, period1)
print("\nH_1(entropy) without spaces: ",H_1(period1))
print("Excess_1: ",(1-(H_1(period1)/m.log2(len(symbols_2))))))
#createbgDataFrame(bigram_cross1, bg_cross_period1, symbols_2)
print("H_2(entropy) cross without spaces: ",H_2(bigram_cross1, bg_cross_period1))
print("Excess_2: ",(1-(H_2(bigram_cross1, bg_cross_period1)/m.log2(len(symbols_2))))))
#createbgDataFrame(bigram1, bg_period1, symbols_2)
print("H_2(entropy) without spaces: ",H_2(bigram1, bg_period1))
print("Excess_3: ",(1-(H_2(bigram1, bg_period1)/m.log2(len(symbols_2))))), '\n')

print('/////////////////WITH SPACES/////////////////')
createDataFrame(quantity2, period2)
print("\nH_1(entropy) with spaces: ",H_1(period2))
print("Excess_4: ",(1-(H_1(period2)/m.log2(len(symbols))))))
#createbgDataFrame(bigram_cross2, bg_cross_period2, symbols)
print("H_2(entropy) cross with spaces: ",H_2(bigram_cross2, bg_cross_period2))
print("Excess_5: ",(1-(H_2(bigram_cross2, bg_cross_period2)/m.log2(len(symbols))))))
#createbgDataFrame(bigram2, bg_period2, symbols)
print("H_2(entropy) with spaces: ",H_2(bigram2, bg_period2))
print("Excess_6: ",(1-(H_2(bigram2, bg_period2)/m.log2(len(symbols)))))]
```

```
(kali@kali)-[~]
$ python3 crypt1.py
/////////////////WITHOUT SPACES/////////////////
quantity  periodicity
o  77430    0.107115
a  60986    0.084367
e  58369    0.080746
и  52993    0.073309
н  44977    0.062220
т  41673    0.057650
с  39353    0.054440
л  35511    0.049125
р  35343    0.048893
в  34368    0.047544

H_1(entropy) without spaces:  4.465995972453169
Excess_1:  0.11466156950060935
H_2(entropy) cross without spaces:  4.166762456909701
Excess_2:  0.1739815806779923
H_2(entropy) without spaces:  4.166128394668092
Excess_3:  0.1741072770900105

/////////////////WITH SPACES/////////////////
quantity  periodicity
o  125369    0.147799
a  77430    0.091283
e  60986    0.071897
и  58369    0.068812
н  52993    0.062474
т  44977    0.053024
с  41673    0.049129
л  39353    0.046394
р  35511    0.041864
р  35343    0.041666

H_1(entropy) with spaces:  4.410230057145229
Excess_4:  0.1331179814452792
H_2(entropy) cross with spaces:  4.047895630341044
Excess_5:  0.20433902779205404
H_2(entropy) with spaces:  4.047494532339943
Excess_6:  0.20441786826983577
```

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

Для утворення матриці частот попередньо було додано два алфавіти у вигляді списку.
Матриці частот з отриманих даних:

Bigram_cross1()

	а	б	в	г	д	е	є	...	щ	ь	ы	ь	э	ю	я
а	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
б	0	0	0.000436	0.001772	0.006016	0.001584	0.00323	...	0.001471	0.001166	0.000299	0	0	0	0.000297
в	0.000779	0.002763	0	0.001231	0.000022	0.000076	0.00001	...	0.000015	0.000024	0.000039	0.000302	0.000163	0.003157	0.000353
г	0.000026	0.000089	0.000256	0	0.008044	0.000307	0.000575	...	0.000167	0.000137	0.000248	0.001155	0.000021	0.000003	0.003355
д	0.000125	0.000166	0.000014	0.000362	0	0.001881	0.000029	...	0	0.000006	0	0.000026	0.000019	0	0.000001
е	0	0.000001	0.000015	0	0.000004	0	0.005226	...	0.001855	0.000004	0.000059	0.000437	0.000077	0.000098	0
є	0.000037	0.000735	0.000411	0.000011	0.00003	0.000394	0	...	0.006376	0.000773	0.000169	0.000823	0.000686	0.001551	0.000865
ж	0.000735	0	0	0	0.00022	0.000126	0.000314	...	0	0	0	0	0	0	0
з	0	0	0	0	0	0	0	...	0.000011	0.000055	0.000012	0.000303	0.000006	0.000003	0.000004
и	0.000043	0	0	0	0	0.000035	0.000008	...	0.00025	0.000315	0.000232	0.000082	0.000436	0.000011	0.000006
й	0.000028	0.000107	0.000026	0	0.000028	0.000542	0.000112	...	0.002006	0.003373	0.002332	0.006224	0.006138	0.000573	0.000282
к	0.001978	0.001287	0.003284	0.00066	0.000387	0	0.000001	...	0.001058	0.000552	0.001159	0.000379	0.001729	0.000829	0.000191
л	0.000111	0.000093	0.00012	0.000433	0.000241	0.000012	0	...	0.000188	0.000837	0.010615	0.000457	0.002396	0.000876	0.001105
м	0.002256	0.000017	0.00003	0.000266	0.000109	0.000042	0	...	0.00516	0.000188	0.001053	0.007409	0.000791	0.000163	0.00239
н	0.000422	0.001732	0.000033	0.000047	0.000018	0.000375	0.000069	...	0.000645	0.000346	0.000519	0.001779	0.004308	0.001198	0.000346
о	0.00105	0.00032	0.002208	0.000087	0.000086	0.000058	0.000239	...	0	0.000721	0.000039	0.000118	0.003795	0.011723	0.000393
п	0.000142	0.001477	0.001466	0.002859	0.000051	0.000059	0.000699	...	0.001756	0.004605	0.003409	0.007656	0.007314	0.007058	0.001613
р	0.003912	0.008531	0.009626	0.006938	0.000625	0.000342	0.000906	...	0.000028	0.001794	0	0.000127	0.001021	0.000024	0.000253
с	0.010785	0.001122	0.007156	0.000208	0.000154	0.001176	0.000006	...	0.0000393	0.000192	0.006181	0.000001	0.000617	0.000156	0.000638
т	0.001869	0.010093	0.00035	0.000185	0.000982	0.000957	0.00315	...	0	0.00001	0.000093	0.002501	0	0.006229	0.002612
у	0.000952	0.001435	0.003748	0.002453	0.000484	0.001818	0.01429	...	0.006744	0	0.000089	0.000154	0.0049	0.000001	0.000091
ф	0.000302	0.000867	0.002176	0.011467	0.000701	0.004255	0.00184	...	0.002151	0.000422	0	0.001486	0.000697	0.000638	0.000123
х	0.00158	0.002281	0.001407	0.001198	0.000427	0.001657	0.001412	...	0.000003	0.000004	0.000499	0	0	0.000004	0.000504
ц	0	0.000006	0.00009	0.000011	0.000017	0.000477	0.00001	...	0.000534	0.000167	0.000183	0.000148	0	0.000061	0.000097
ч	0.000671	0	0.000293	0.000307	0.000051	0.000499	0.002707	...	0.000019	0.000014	0.000022	0.001161	0	0.000003	0
ш	0.000011	0.001479	0	0.00034	0.000008	0.000022	0.000024	...	0.002525	0.00004	0.000116	0.000022	0.000036	0.003923	0
щ	0.000007	0.000055	0.00197	0	0.000748	0.000097	0.000029	...	0	0.001255	0.000008	0.000047	0.000006	0.000004	0.00233
ъ	0	0	0	0.001902	0	0.000537	0.00057	...	0.000001	0	0.000505	0.00003	0.000017	0.000008	0.000001
ы	0.001462	0	0.000001	0.000001	0.000979	0	0.000004	...	0	0	0	0	0.000001	0	0.000001
э	0	0.000155	0	0	0	0	0	...	0.000001	0.000011	0.000129	0	0.00009	0.000386	0.001616
ю	0.000311	0.000401	0.001732	0	0.000091	0.000243	0.000562	...	0	0.000057	0.000001	0.000042	0	0.000166	0.000422
я	0.001038	0.00039	0.000567	0.000883	0	0.000071	0.000437	...	0	0	0.00013	0.000455	0.000848	0	0
я	0.000001	0.000006	0	0.000004	0	0	0	...	0	0	0	0.000003	0	0	0

Bigram1()

	а	б	в	г	д	е	є	...	щ	ь	ы	ь	э	ю	я
а	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
б	0	0	0.000421	0.001785	0.005929	0.001599	0.003185	...	0.001417	0.00117	0.000335	0	0	0	0.000307
в	0.000769	0.002722	0	0.001245	0.000014	0.000066	0.000006	...	0.000008	0.000039	0.000036	0.000318	0.00018	0.003099	0.000324
г	0.000028	0.000091	0.000266	0	0.007988	0.000307	0.000589	...	0.000018	0.000136	0.000279	0.001151	0.000025	0.000006	0.003345
д	0.000122	0.000163	0.000017	0.00039	0	0.001892	0.000033	...	0	0.000003	0	0.000025	0.000019	0	0
е	0	0.000003	0.000019	0	0.000003	0	0.005279	...	0.001782	0.000006	0.00005	0.000426	0.000083	0.000086	0
є	0.000044	0.000719	0.000443	0.000008	0.000042	0.000398	0	...	0.006388	0.000758	0.000163	0.000802	0.000661	0.001566	0.000866
ж	0.000758	0	0	0.00021	0.000133	0.000329	0	...	0	0	0	0	0	0	0
з	0	0	0	0	0	0	0	...	0.000006	0.000058	0.000017	0.000252	0.000006	0	0.000006
и	0.000033	0	0	0	0	0.000036	0.000008	...	0.000252	0.000324	0.000249	0.0001	0.000432	0.000011	0.000006
й	0.000022	0.000089	0.000036	0	0.000017	0.000553	0.000091	...	0.001964	0.003351	0.002302	0.006109	0.006126	0.000553	0.000299
к	0.002017	0.0013	0.003276	0.000681	0.000401	0	0	...	0.001024	0.000515	0.001176	0.000407	0.001677	0.000772	0.000169
л	0.000108	0.000094	0.0001	0.000412	0.000219	0.000008	0	...	0.000232	0.0008	0.010522	0.000421	0.002476	0.000855	0.001115
м	0.002288	0.000011	0.000042	0.000279	0.000111	0.00003	0	...	0.000515	0.000169	0.001101	0.00744	0.00083	0.000152	0.002363
н	0.000407	0.001785	0.000039	0.00005	0.000019	0.000409	0.000064	...	0.000645	0.00034	0.000523	0.00179	0.004197	0.001287	0.00036
о	0.001035	0.000326	0.002272	0.00008	0.000077	0.00005	0.000216	...	0	0.000719	0.000033	0.00013	0.003663	0.011853	0.00036
п	0.000147	0.001569	0.001513	0.002728	0.000058	0.000066	0.000653	...	0.001804	0.004507	0.003533	0.007609	0.007304	0.007122	0.001607
р	0.003862	0.008591	0.009808	0.006983	0.000528	0.00034	0.000888	...	0.000028	0.001768	0	0.000122	0.000971	0.000028	0.000282
с	0.01066	0.001137	0.007263	0.000224	0.000147	0.001187	0.000006	...	0.000401	0.000191	0.00607	0.000003	0.000589	0.000163	0.0007
т	0.001848	0.009944	0.000329	0.000194	0.000952	0.000924	0.003185	...	0	0.000111	0.000105	0.002388	0	0.006319	0.002565
у	0.00088	0.001458	0.003832	0.00246	0.000498	0.001832	0.014116	...	0.006909	0	0.000077	0.000138	0.004806	0.000003	0.000941
ф	0.000313	0.000877	0.002141	0.011225	0.000694	0.004294	0.001807	...	0.002166	0.00044	0	0.001605	0.000728	0.000642	0.000127
х	0.001574	0.002296	0.00137	0.001179	0.000429	0.001679	0.001464	...	0.000003	0.000003	0.000528	0	0	0.000008	0.000509
ц	0	0.000003	0.000083	0.000006	0.000019	0.000481	0.000017	...	0.000562	0.000163	0.000183	0.00013	0	0.000058	0.000083
ч	0.000678	0	0.000266	0.000296	0.000365	0.000459	0.002786	...	0.000022	0.000224	0.000011	0.000025	0.001226	0	0.000003
ш	0.000017	0.001477	0	0.000371	0.000008	0.000022	0.000019	...	0.002648	0.000039	0.000108	0.000019	0.000036	0.003821	0
щ	0.000008	0.000047	0.002006	0	0.000747	0.000108	0.000033	...	0	0.001281	0.000014	0.000053	0.000011	0.000006	0.002341
ъ	0	0	0	0.001843	0	0.000517	0.000584	...	0.000003	0	0.000512	0.000025	0.000019	0.000011	0
ы	0.001392	0	0.000003	0.000003	0.000932	0	0	...	0	0	0	0	0.000003	0	0
э	0	0.000147	0	0	0	0	0	...	0	0.000017	0.000125	0	0.000091	0.000401	0.001624
ю	0.000285	0.000415	0.001729	0	0.000083	0.000243	0.000581	...	0	0.000053	0	0.000042	0	0.000141	0.000429
я	0.001074	0.000385	0.000534	0.00086	0	0.000086	0.000457	...	0	0	0.000125	0.000451	0.000855	0	0
я	0	0.000003	0.000003	0	0	0	0	...	0	0	0	0.000003	0	0	0

Bigram_cross2()

	а	б	в	г	д	е	ё	...	ш	щ	ъ	ы	ь	э	ю	я
а	0.002519	0.006078	0.013927	0.004565	0.006293	0.002141	0	...	0.001179	0.000103	0	0	0	0.001733	0.00009	0.000984
б	0.0015756	0.000077	0.000826	0.003558	0.000841	0.00208	0.000999	0	...	0.000859	0.000243	0	0	0.000032	0.000648	0.002216
в	0.000235	0.001042	0.000017	0.00005	0.000006	0.000012	0.002193	0	...	0.000032	0.000257	0.000139	0.00269	0.000301	0	0.000075
г	0.006925	0.006753	0.000026	0.000053	0.000042	0.000341	0.005523	0	...	0.000923	0.000011	0.000002	0.002859	0.000106	0.000008	0.000002
д	0.000492	0.001596	0.000001	0.000025	0.000006	0.000726	0.000389	0	...	0.000008	0	0.000001	0	0.000001	0.000007	0
е	0.001078	0.004434	0.000041	0.001157	0.000029	0.000057	0.004874	0	...	0.000072	0	0.000032	0.000626	0.000035	0.000004	0.000026
ё	0.014529	0.00021	0.000892	0.002657	0.002232	0.002974	0.001583	0	...	0.000591	0.000611	0	0	0	0.000009	0.000106
ж	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
з	0.000191	0.001291	0.000035	0.000002	0.000014	0.000805	0.003048	0	...	0	0	0	0	0.000029	0	0
и	0.001326	0.004848	0.000116	0.000917	0.000196	0.000865	0.000568	0	...	0.000009	0	0.000024	0.000462	0.000095	0.000001	0.00001
й	0.015762	0.000487	0.000344	0.002185	0.000671	0.001521	0.002258	0	...	0.000446	0.000312	0	0.000001	0	0.000012	0.000388
к	0.008355	0.000002	0.000027	0.000012	0.000015	0.000127	0.000028	0	...	0.000117	0.000005	0	0	0	0.000002	0.000011
л	0.003662	0.006535	0.000004	0.000383	0.00002	0.000019	0.000752	0	...	0.000015	0	0	0	0	0.000004	0.000002
м	0.006468	0.004981	0.000057	0.000035	0.000162	0.000054	0.004726	0	...	0.000006	0.000005	0.000001	0.000608	0.004467	0.000001	0.000914
н	0.006877	0.00357	0.000288	0.000119	0.000027	0.000009	0.00369	0	...	0.000008	0.000005	0.000002	0.00083	0.000101	0	0.000007
о	0.002686	0.009411	0.000024	0.000041	0.000213	0.000729	0.006878	0	...	0.000018	0.000058	0.000009	0.004502	0.000956	0.000024	0.000098
п	0.01536	0.000048	0.003478	0.008709	0.004244	0.005761	0.001584	0	...	0.000836	0.000223	0	0.000001	0	0.000139	0.000252
р	0.000701	0.001952	0.000013	0.000008	0.000004	0.000005	0.002537	0	...	0.000008	0.000004	0	0.00029	0.000088	0.000001	0.000613
с	0.001636	0.000232	0.000124	0.000397	0.000405	0.000284	0.006078	0	...	0.000222	0.000044	0	0.000001	0.00147	0.000002	0.00015
т	0.002682	0.001825	0.000092	0.001177	0.000035	0.000159	0.002703	0	...	0.000085	0	0.000035	0.000421	0.002699	0.000004	0.000203
у	0.004681	0.006462	0.000026	0.002977	0.000007	0.000226	0.005688	0	...	0.000011	0.000027	0.000014	0.001469	0.003955	0.000004	0.000145
ф	0.005434	0.000121	0.001095	0.000442	0.000789	0.001615	0.000295	0	...	0.000618	0.000222	0	0	0.000028	0.001072	0.00009
х	0.000093	0.000295	0	0	0	0.000421	0	...	0.000001	0	0	0.000017	0.00002	0.000001	0.000005	0
ц	0.003513	0.000712	0.000004	0.000152	0.000009	0.000007	0.000093	0	...	0.000012	0	0.000001	0	0.000008	0	0
ч	0.000448	0.000833	0.000001	0.000153	0.000002	0.000001	0.000981	0	...	0.000001	0	0.000259	0	0	0	0
ш	0.000861	0.002136	0.000004	0.000011	0.000004	0.000001	0.003335	0	...	0.000067	0	0	0.000259	0.000002	0	0
щ	0.000079	0.001068	0	0.000032	0	0	0.001985	0	...	0	0	0	0.000206	0	0.000005	0
ъ	0.000098	0.000426	0	0.000002	0	0.000001	0.001246	0	...	0	0	0	0.000033	0	0	0
ы	0.000006	0	0.000001	0	0	0	0.000132	0	...	0	0	0	0	0	0.000009	0.00011
ь	0.004315	0.000004	0.000151	0.000945	0.000146	0.000125	0.001393	0	...	0.000429	0.000012	0	0	0.000002	0	0.000014
э	0.007409	0.000008	0.000087	0.000061	0.000164	0.000174	0.000021	0	...	0.000485	0.000021	0	0	0.000007	0.000383	0.000657
ю	0.000012	0	0	0.000005	0	0.000004	0	...	0	0	0	0	0	0.000002	0	0
я	0.002193	0.000007	0.000286	0.000017	0.000025	0.000252	0.000001	0	...	0.000024	0.000426	0	0	0.000004	0.000031	0.000001
	0.008545	0.000012	0.000117	0.000309	0.000152	0.000378	0.000093	0	...	0.00002	0.000225	0	0	0.000002	0.000083	0.000072

[34 rows x 34 columns]

Bigram2()

	а	б	в	г	д	е	ё	...	и	щ	ъ	ы	ь	э	ю	я
а	0.002546	0.00626	0.013873	0.004551	0.006189	0.002087	0	...	0.001146	0.000101	0	0	0	0.001683	0.000094	0.000983
б	0.001563	0.000008	0.000823	0.00356	0.000828	0.002035	0.001012	...	0.000875	0.000245	0	0	0	0.000028	0.000604	0.002169
в	0.000243	0.001059	0.000019	0.000057	0.000007	0.000012	0.002136	...	0.000031	0.000271	0.000144	0.002617	0.00033	0	0.000061	0.000193
г	0.006892	0.006701	0.000031	0.000052	0.000047	0.000302	0.005527	...	0.00095	0.000009	0	0.002931	0.000094	0.000009	0.000002	0.000269
д	0.000443	0.001665	0	0.000026	0.000007	0.000712	0.000415	...	0.000005	0	0	0	0.000002	0.000005	0	0
е	0.001049	0.004518	0.000042	0.00112	0.000019	0.000045	0.004951	...	0.000073	0	0.000026	0.000611	0.000335	0.000002	0.000033	0.000333
ё	0.014656	0.000215	0.000905	0.00266	0.002245	0.002985	0.001582	...	0.000578	0.000615	0	0	0	0.000009	0.000106	0.000174
ж	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
з	0.000205	0.001276	0.000052	0.000005	0.000012	0.000785	0.002969	...	0	0	0	0	0.000028	0	0	0
и	0.001311	0.004963	0.000125	0.000931	0.000191	0.000894	0.000592	...	0.000005	0	0.000019	0.000441	0.00009	0	0	0.000205
й	0.015939	0.000495	0.000361	0.00219	0.000655	0.001495	0.002337	...	0.000443	0.000328	0	0	0	0.000014	0.000434	0.001521
к	0.008295	0.000005	0.000026	0.000012	0.000024	0.000127	0.000026	...	0.000118	0.000005	0	0	0	0	0	0.000007
л	0.003763	0.006512	0	0.000384	0.000019	0.000014	0.000726	...	0.000014	0	0	0	0	0	0.000005	0.000002
м	0.006564	0.00519	0.000045	0.000031	0.000172	0.000073	0.004789	...	0.000007	0.000005	0	0.000634	0.004527	0.000955	0.001717	
н	0.006861	0.003622	0.000266	0.000116	0.000026	0.000009	0.003671	...	0.000007	0.000005	0.000002	0.00083	0.000101	0.000007	0.000401	
о	0.002655	0.009521	0.000021	0.000042	0.000226	0.000752	0.006852	...	0.000017	0.000052	0.000009	0.004565	0.000891	0.000026	0.000101	0.001
п	0.015305	0.000047	0.003395	0.008653	0.004277	0.00577	0.001582	...	0.000868	0.000215	0	0	0.000158	0.000269	0.000639	0.000530
р	0.000738	0.002042	0.000012	0.000014	0.000005	0.000002	0.002641	...	0.000009	0.000005	0	0.00029	0.000083	0.000007	0.000323	0.00032
с	0.001622	0.008436	0.000134	0.000396	0.000523	0.000278	0.00613	...	0.0002	0.000033	0.001521	0.000528	0.000002	0.000144	0.000846	0.000846
т	0.002778	0.001844	0.000087	0.001148	0.000035	0.00013	0.002693	...	0.00009	0	0.000047	0.00041	0.002747	0.000179	0.003273	0.003273
у	0.004732	0.006556	0.000014	0.003025	0.000009	0.000219	0.005583	...	0.000005	0.000033	0.000009	0.00154	0.003923	0.000002	0.000156	0.000523
ф	0.000544	0.000146	0.001094	0.000415	0.000698	0.001606	0.002307	...	0.000618	0.000238	0	0	0	0.000031	0.001103	0.000092
х	0.000099	0.000302	0	0	0	0.00042	0	...	0	0	0	0.000007	0.000012	0	0.000005	0
ц	0.003556	0.000693	0.000002	0.00017	0.000009	0.000009	0.000101	...	0.000017	0	0	0	0	0.000007	0.000009	0
ч	0.00046	0.000835	0.000002	0.00016	0.000002	0	0.000905	...	0	0	0	0.000271	0	0	0	0
ш	0.000844	0.002183	0.000002	0.000014	0.000005	0	0.003346	...	0.000054	0	0	0	0.000229	0.000005	0	0
щ	0.00009	0.001103	0	0.000026	0	0	0.001976	...	0	0	0	0	0.000179	0	0.000009	0
ъ	0.00009	0.000436	0	0.000005	0	0	0.001162	...	0	0	0	0	0.000045	0	0	0
ы	0.000005	0	0.000002	0	0	0	0.000127	...	0	0	0	0	0	0	0.000009	0.000116
ь	0.004426	0.000002	0.000134	0.000941	0.000144	0.000108	0.00136	...	0.00045	0.000017	0	0	0	0.000002	0	0.000017
э	0.007415	0.000012	0.000008	0.000061	0.000174	0.00017	0.000589	...	0.000545	0.000019	0	0	0	0.000002	0.000373	0.000703
ю	0.000014	0	0	0.000005	0	0.000005	0	...	0	0	0	0	0	0.000002	0	0
я	0.002087	0.000005	0.000274	0.000019	0.000026	0.00024	0	...	0.000024	0.00049	0	0	0	0.000002	0.000028	0.000002
	0.008512	0.000009	0.000113	0.000307	0.000165	0.000394	0.00009	...	0.000026	0.000233	0	0	0	0	0.000085	0.000073
[34 rows x 34 columns]																

Робота з CoolPinkProgram

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

Лаборотторная работа №1

Произвольная часть текста:
ддельную_доброту_вы_с_таким_же_успехом_можете_представить_себе_страну_где_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: д
Символ по счету: 1
Номер эксперимента: 52
Поле ввода символов:
д
Продолжить Другой

Неравенство для энтропии:
1,17900580392966 < H < 1,75119307737616
Двоичная таблица угаданных символов:
10000000000000000000000000000000
01000000000000000000000000000000
01000000000000000000000000000000
10000000000000000000000000000000
01000000000000000000000000000000

Вероятности:
q[1] = 0,5384615
q[2] = 0,3076923
q[3] = 0,0384615
q[4] = 0,0576923
q[5] = 0
q[6] = 0,0192307
q[7] = 0
q[8] = 0
q[9] = 0
q[10] = 0
q[11] = 0
q[12] = 0,019230
q[13] = 0
q[14] = 0,019230
q[15] = 0
q[16] = 0
q[17] = 0
q[18] = 0
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

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

Лаборотторная 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
00000000000100000000000000000000
10000000000000000000000000000000

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

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

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

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

Поле ввода символов:
Продолжить Другой

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