

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 илья 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 илья 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 илья арнольдович фاینзільберг 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

```
$ 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 илья 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 илья 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 илья арнольдович фاینзільберг 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):
    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)-[~/cr1]
$ python3 crypt1.py
/////////////////////////////////WITHOUT SPACES/////////////////////////////////
quantity  periodicity
o 115667 0.112561
e 91837 0.089371
a 81729 0.079534
н 68872 0.067022
и 65574 0.063813
т 64768 0.063029
с 54652 0.053184
в 48417 0.047117
л 47693 0.046412
р 40107 0.039030

H_1(entropy) without spaces: 4.464430361925528
Excess_1: 0.1149719359173873
periodicity
то 0.017737
не 0.013428
но 0.012729
на 0.011974
ст 0.011971
ов 0.011759
ен 0.010923
по 0.010599
го 0.010138
ко 0.009553

H_2(entropy) cross without spaces: 4.133209748530321
Excess_2: 0.1806330649961222
periodicity
то 0.017600
не 0.013369
но 0.012849
на 0.012116
ст 0.012092
ов 0.011781
ен 0.011002
по 0.010664
го 0.010272
ал 0.009513

H_2(entropy) without spaces: 4.133006232367472
Excess_3: 0.18067341001239845

/////////////////////////////////WITH SPACES/////////////////////////////////
quantity  periodicity
o 205591 0.166715
e 115667 0.093795
e 91837 0.074471
a 81729 0.066275
н 68872 0.055849
и 65574 0.053174
т 64768 0.052521
с 54652 0.044318
в 48417 0.039262
л 47693 0.038675

H_1(entropy) with spaces: 4.370277671141025
Excess_4: 0.14097108764986133
periodicity
о 0.024601
е 0.018590
и 0.017213
в 0.016822
н 0.016355
а 0.015860
с 0.015217
п 0.014849
то 0.014516
ь 0.012363

H_2(entropy) cross with spaces: 3.9525976608708326
Excess_5: 0.2230709522195925
periodicity
о 0.024697
е 0.018618
и 0.017303
в 0.016757
н 0.016119
а 0.015745
с 0.015279
п 0.014877
то 0.014603
ь 0.012486

H_2(entropy) with spaces: 3.9520705649460193
Excess_6: 0.223174559054917
```

Ентропія

	Текст з пробілами	Текст без пробілів
Н1	4.370277671141025	4.464430361925528
Н2	3.9525976608708326	4.133209074853021
Н2(біграми з кроком 2)	3.9520705649460193	4.133006232367472

Надлишковість існує в межах від 0,11-0,22.

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	є	ж	...	ш	щ	ъ	ы	ь	э	ю	я								
а	0.00056	0.001016	0.005894	0.001636	0.003025	0.002198	0.000001	0.002379	...	0.001186	0.000231	0	0	0	0.000345	0.001105	0.002605								
б	0.000642	0.000013	0.000061	0.000005	0.000025	0.002812	0	0.000004	...	0.000007	0.000226	0.000213	0.004812	0.000048	0.000112	0.000002	0.000641								
в	0.007103	0.000179	0.000416	0.000492	0.001235	0.006572	0	0.000005	...	0.000056	0.000018	0.000005	0.003635	0.00015	0.000403	0.000002	0.000224								
г	0.001309	0.000045	0.000122	0.000024	0.00138	0.000956	0	0.000006	...	0.000009	0	0	0	0	0.000006	0	0.000019								
д	0.006381	0.000033	0.000894	0.000026	0.000054	0.005235	0	0.000019	...	0.000092	0	0.000006	0.000576	0.001015	0.000018	0.000015	0.000306								
е	0.000274	0.00266	0.004822	0.004292	0.004481	0.00228	0	0.001145	...	0.001138	0.000991	0	0	0	0.000346	0.000315	0.000668								
є	0.000003	0.000036	0.000061	0.000009	0.00003	0.000052	0	0.000021	...	0.000003	0	0	0	0	0.000157	0	0.000004								
ж	0.001555	0.00006	0.000044	0.000007	0.000992	0.005779	0.000008	0.000009	...	0	0	0	0	0.000025	0.000022	0.000001	0.000016								
з	0.005767	0.000168	0.001296	0.000494	0.000844	0.000311	0	0.000069	...	0.000019	0	0.000039	0.000374	0.001576	0.000037	0.000198	0.000634								
и	0.00029	0.001604	0.005019	0.000888	0.00298	0.002477	0	0.000044	...	0.000764	0.000185	0	0	0	0.000253	0.000314	0.001712								
й	0.00014	0.000234	0.000594	0.000215	0.000659	0.000104	0	0.000157	...	0.000309	0.000005	0	0	0	0.000007	0.000005	0.000079								
к	0.007497	0.000516	0.000697	0.000093	0.000228	0.000654	0	0.000119	...	0.000019	0.000001	0	0	0.000025	0.000111	0	0.000111								
л	0.006919	0.000336	0.000802	0.000477	0.000313	0.004881	0.000016	0.000395	...	0.000022	0.000005	0	0.000779	0.004369	0.000092	0.001109	0.001862								
м	0.002808	0.000339	0.000897	0.000295	0.000505	0.00516	0	0.000164	...	0.000044	0.000017	0	0.001011	0.000085	0.000095	0.000001	0.0006								
н	0.011974	0.000279	0.000446	0.00013	0.000499	0.013428	0.000003	0.000032	...	0.000036	0.000124	0	0.002984	0.001327	0.000039	0.000189	0.004182								
о	0.000269	0.005345	0.011759	0.006251	0.006289	0.003192	0.000002	0.003473	...	0.001107	0.000233	0	0	0	0.00048	0.000704	0.00147								
п	0.001567	0.000003	0.000003	0	0.000001	0.002517	0	0	...	0.000003	0.000001	0	0.000228	0.000069	0	0.000004	0.000584								
р	0.007943	0.000176	0.000524	0.000186	0.000885	0.006091	0	0.000292	...	0.000352	0.000014	0	0.001095	0.001075	0.000006	0.000222	0.001192								
с	0.002298	0.000187	0.001806	0.000159	0.00039	0.004588	0.000783	0.000269	...	0.000155	0	0.000019	0.000239	0.003904	0.000063	0.000275	0.004536								
т	0.007116	0.000567	0.003291	0.000127	0.000392	0.006767	0	0.000086	...	0.000029	0.000025	0.000017	0.001378	0.007792	0.000171	0.000049	0.000689								
у	0.000209	0.00075	0.001549	0.0001832	0.002266	0.000346	0	0.0002041	...	0.000672	0.000297	0	0	0	0.000106	0.000877	0.00015								
ф	0.000255	0.000002	0.000003	0	0.000002	0.000376	0	0	...	0	0	0	0.000004	0.000289	0	0	0.000002								
х	0.000544	0.000101	0.000474	0.000091	0.000199	0.000148	0	0.000042	...	0.000018	0.000005	0	0	0	0.000036	0.000001	0.000044								
ц	0.000714	0.000011	0.00008	0.000008	0.000017	0.000799	0	0.000002	...	0.000002	0	0	0.000233	0	0.000008	0	0.000009								
ч	0.00335	0.000018	0.000068	0.000013	0.000033	0.004765	0.000001	0.000001	...	0.000168	0	0	0	0.000235	0.000007	0	0.000009								
ш	0.001032	0.000005	0.000086	0.000005	0.000003	0.002509	0	0.000002	...	0	0	0	0	0.000528	0.000001	0.000001	0.000003								
щ	0.000347	0.000008	0.000006	0.000002	0.000008	0.001733	0	0	...	0	0	0	0	0.000002	0.000001	0	0.000004								
ъ	0	0	0	0	0	0.000057	0	0	...	0	0	0	0	0	0.000002	0.000002	0.000239								
ы	0.000114	0.000351	0.001488	0.000194	0.000507	0.000818	0	0.000076	...	0.000536	0.000131	0	0	0	0.000087	0.000002	0.000069								
ь	0.000264	0.00052	0.001591	0.000344	0.000677	0.001219	0	0.000089	...	0.000524	0.000064	0	0	0	0.000271	0.000403	0.000847								
э	0	0	0.000004	0.000013	0.000006	0	0	0	...	0.000001	0	0	0	0	0.000009	0	0								
ю	0.000064	0.000616	0.000349	0.000095	0.000517	0.000064	0	0.000114	...	0.000004	0.000215	0	0	0	0.000045	0.000035	0.000084								

Bigram1()

Enter name of excel: bigram1																			
	а	б	в	г	д	е	ё	ж	...	и	ш	щ	ъ	ы	ь	э	ю	я	
а	0.000592	0.001014	0.006016	0.001602	0.003023	0.002158	0.000002	0.002448	...	0.001214	0.000253	0	0	0	0	0.000343	0.001059	0.002612	я
б	0.000627	0.000012	0.000062	0.000004	0.000027	0.00282	0	0.000002	...	0.000012	0.000208	0.000226	0.004751	0.000058	0.000095	0.000002	0.000615	0.000615	я
в	0.0071	0.000162	0.000405	0.000527	0.001255	0.006629	0	0.000049	...	0.000058	0.000021	0.000004	0.003505	0.00015	0.000387	0.000002	0.000232	0.000232	я
г	0.001238	0.000051	0.000125	0.000021	0.001394	0.000934	0	0.000004	...	0.000008	0	0	0	0	0.000008	0	0.000012	0.000012	я
д	0.006226	0.000029	0.000915	0.000025	0.00007	0.005239	0	0.000019	...	0.000097	0	0.000006	0.000609	0.001026	0.000018	0.000016	0.000296	0.000296	я
е	0.000317	0.002768	0.004729	0.004258	0.004496	0.00224	0	0.001211	...	0.001125	0.000944	0	0	0	0.000331	0.000344	0.000687	0.000687	я
ё	0.000004	0.000035	0.000072	0.000008	0.000041	0.00006	0	0.000019	...	0.000002	0	0	0	0	0.000163	0	0.000002	0.000002	я
ж	0.001534	0.000053	0.000054	0.000008	0.000969	0.005718	0.000002	0.000008	...	0	0	0	0	0	0.000021	0.000016	0	0.000016	я
з	0.00586	0.000162	0.00131	0.000487	0.000913	0.000296	0	0.000076	...	0.000019	0	0.000039	0.000368	0.001649	0.000031	0.000218	0.00059	0.00059	я
и	0.000284	0.001549	0.00508	0.000882	0.003009	0.002571	0	0.000473	...	0.00074	0.000202	0	0	0	0.000271	0.000315	0.001752	0.001752	я
й	0.000136	0.000245	0.000566	0.000212	0.000648	0.000121	0	0.000015	...	0.000308	0.000006	0	0	0	0.000058	0.000006	0.000074	0.000074	я
к	0.007466	0.000529	0.000689	0.000099	0.000237	0.000658	0	0.000115	...	0.000023	0.000002	0	0	0.000029	0.000125	0	0.000125	0.000125	я
л	0.007084	0.000372	0.000823	0.000467	0.00035	0.004899	0.000016	0.000413	...	0.000018	0.000006	0	0.000751	0.004185	0.000088	0.001086	0.001903	0.001903	я
м	0.002773	0.000333	0.000913	0.000276	0.0005	0.005127	0	0.000156	...	0.000041	0.000211	0	0.001018	0.000086	0.000113	0.000002	0.000586	0.000586	я
н	0.012116	0.000296	0.000475	0.000121	0.000463	0.013369	0.000006	0.000027	...	0.000041	0.000013	0	0.002993	0.01378	0.000043	0.000177	0.004348	0.004348	я
о	0.000269	0.005345	0.011781	0.006218	0.006185	0.003163	0.000002	0.003501	...	0.001166	0.000234	0	0	0	0.000471	0.000724	0.001467	0.001467	я
п	0.001549	0.000002	0.000002	0	0	0.002485	0	0	...	0.000002	0.000002	0	0.000197	0.000064	0	0.000004	0.000572	0.000572	я
р	0.007873	0.000175	0.000535	0.000189	0.000862	0.006148	0	0.000269	...	0.00036	0.00001	0	0.001133	0.01033	0.000004	0.000193	0.001236	0.001236	я
с	0.002373	0.000191	0.0018	0.000158	0.000395	0.004529	0.000755	0.000267	...	0.000163	0	0.000019	0.000267	0.003924	0.000064	0.000288	0.004611	0.004611	я
т	0.007084	0.000557	0.003229	0.000123	0.000397	0.006569	0	0.000074	...	0.000031	0.000018	0.000014	0.001341	0.007766	0.000154	0.000039	0.000673	0.000673	я
у	0.000199	0.000742	0.00154	0.000185	0.00225	0.000337	0	0.002057	...	0.000695	0.000296	0	0	0	0.000113	0.000827	0.000014	0.000014	я
ф	0.000288	0.000002	0	0	0.000002	0.00037	0	0	...	0	0	0	0.000006	0.00029	0	0	0.000002	0.000002	я
х	0.000512	0.000099	0.000424	0.00008	0.000191	0.000158	0	0.000049	...	0.00001	0.000002	0	0	0	0.000031	0.000002	0.000037	0.000037	я
ц	0.000769	0.000012	0.000076	0.00001	0.000021	0.000815	0	0	...	0.000002	0	0	0.000228	0	0.000006	0	0.00001	0.00001	я
ч	0.003283	0.000018	0.000064	0.000016	0.000023	0.004842	0.000002	0.000002	...	0.000156	0	0	0	0.000239	0.000008	0	0.000012	0.000012	я
ш	0.001072	0.000006	0.000097	0.000008	0.000006	0.002503	0	0	...	0	0	0	0	0.000533	0.000002	0.000002	0.000006	0.000006	я
щ	0.000354	0.000002	0.000012	0	0.000008	0.001763	0	0	...	0	0	0	0	0.000027	0.000002	0	0.000004	0.000004	я
ъ	0	0	0	0	0	0.000058	0	0	...	0	0	0	0	0	0	0.000002	0.000023	0.000023	я
ы	0.000111	0.000385	0.001504	0.000193	0.000537	0.000829	0	0.000088	...	0.000514	0.000127	0	0	0.000095	0.000004	0.000007	0.000007	0.000007	я
ь	0.000234	0.000527	0.001596	0.000339	0.00067	0.001203	0	0.00009	...	0.000525	0.00006	0	0	0.000267	0.000403	0.000886	0.000886	0.000886	я
э	0	0	0.000004	0.000016	0.000006	0	0	0	...	0.00001	0	0	0	0	0.000004	0	0	0	я
ю	0.000066	0.000654	0.000343	0.000093	0.000492	0.000074	0	0.000113	...	0.000054	0.000226	0	0	0	0.000037	0.000037	0.000068	0.000068	я
я	0.0000239	0.000555	0.001944	0.000352	0.001279	0.000455	0	0.000282	...	0.000043	0.000138	0	0	0	0.000165	0.000142	0.000245	0.000245	я

Bigram_cross2()

Enter name of excel: bigram_cross2																
	а	б	в	г	д	е	ё	...	ш	щ	ь	ы	ь	э	ю	я
а	0	0.002484	0.006727	0.016822	0.003636	0.007641	0.00417	0	0.000562	0.000109	0	0	0	0.002921	0.000032	0.002962
б	0.01586	0.000012	0.000356	0.002292	0.000964	0.001782	0.00134	0.000001	0	0.000937	0.000183	0	0	0.000002	0.000092	0.001834
в	0.000353	0.00053	0.000004	0.000046	0.000003	0.000019	0.002313	0	0	0.000006	0.000188	0.000178	0.00401	0.00004	0	0.000002
г	0.00516	0.005869	0.000004	0.000041	0.000193	0.000748	0.005364	0	0	0.000658	0.000015	0.000004	0.003029	0.000125	0	0.000157
д	0.000997	0.001079	0	0.000001	0.000001	0.001098	0.000778	0	0	0.000006	0	0	0	0	0	0
е	0.000788	0.005312	0.000009	0.000665	0.000005	0.000021	0.004338	0	0	0.000075	0	0.000005	0.00048	0.000846	0	0.000012
ё	0.01859	0.000024	0.001441	0.002045	0.003185	0.002901	0.001589	0	0	0.000903	0.000819	0	0	0.000001	0.000255	0.000303
...	0.000657	0	0	0	0	0	0	0	0	0.000002	0	0	0	0	0	0
ш	0.000471	0.001292	0.000041	0	0.000003	0.000809	0.004807	0.000006	0	0	0	0	0.000021	0	0.000001	0
щ	0.001222	0.004792	0.000106	0.000958	0.00038	0.000633	0.000242	0	0.000003	0	0.000032	0.000311	0.001314	0	0.000163	0.000521
ь	0.017213	0.000066	0.000594	0.002236	0.000375	0.001519	0.001711	0	0.000568	0.000147	0	0	0	0	0.000259	0.001143
ы	0.005907	0.000002	0.000031	0.000003	0.000005	0.000234	0.000004	0	0	0.000022	0	0	0	0.000002	0	0
ь	0.004582	0.006188	0	0.000155	0.000003	0.000027	0.00047	0	0	0.000005	0	0	0	0.000021	0	0
э	0.006787	0.005668	0.000032	0.00003	0.000185	0.000022	0.003699	0.000013	0	0.000002	0.000002	0	0.000649	0.003641	0	0.000924
ю	0.007758	0.002243	0.000021	0.000002	0.000072	0.000003	0.004148	0	0	0	0.000009	0	0.000843	0.000071	0.000001	0.000356
я	0.003497	0.009935	0.000003	0.00001	0.000032	0.000285	0.011122	0.000002	0	0.000016	0.000103	0	0.002486	0.001106	0	0.000157
а	0.024601	0.000004	0.00321	0.007125	0.004786	0.004243	0.00195	0.000002	0	0.000843	0.000185	0	0	0.000002	0.000585	0.000577
б	0.000037	0.001304	0	0	0	0	0.002095	0	0	0.000002	0.000001	0	0.00019	0.000058	0	0.000003
в	0.000577	0.006603	0.000123	0.000386	0.000147	0.000718	0.005066	0	0	0.000293	0.000011	0	0.000912	0.000896	0.000002	0.000185
г	0.003595	0.001859	0.000054	0.001246	0.000018	0.000171	0.00375	0.000653	0	0.000115	0	0.000016	0.000199	0.003253	0	0.000228
д	0.005755	0.005826	0.000008	0.002193	0.000012	0.000074	0.005496	0	0	0.00001	0.000021	0.000014	0.001148	0.000493	0.000001	0.00004
е	0.00625	0.000036	0.000485	0.000668	0.00141	0.001633	0.000165	0	0	0.000548	0.000247	0	0	0.000013	0.00073	0.000026
ё	0.000023	0.000211	0	0	0	0	0.000312	0	0	0	0	0.000003	0.000241	0	0	0.000001
ш	0.002608	0.000411	0	0.000195	0	0	0.000068	0	0	0.000004	0	0	0	0.000001	0	0
щ	0.00042	0.000588	0	0.00002	0	0	0.000656	0	0	0	0	0.000194	0	0.000001	0	0
ь	0.000472	0.00278	0	0.000002	0	0	0.003957	0.000001	0	0.00014	0	0	0.000195	0	0	0
ы	0.000087	0.00086	0	0.000059	0	0	0.002091	0	0	0	0	0	0.00044	0	0.000001	0
ь	0.000075	0.000288	0	0	0	0.000001	0.001442	0	0	0	0	0	0.000017	0	0	0
э	0	0	0	0	0	0	0.000048	0	0	0	0	0	0	0	0.000002	0.000199
ю	0.004531	0.000002	0.00014	0.000761	0.000008	0.000171	0.000563	0	0	0.000439	0.000109	0	0	0.000002	0	0.000003
я	0.012363	0.000002	0.000118	0.000033	0.000083	0.000047	0.000614	0	0	0.000411	0.000006	0	0	0.000006	0.000332	0.000405
а	0.000003	0	0	0.000003	0.000011	0.000004	0	0	0	0.000008	0	0	0	0.000006	0	0
б	0.003236	0.000003	0.000431	0.000003	0.000002	0.000284	0.000001	0	0	0.000023	0.00017	0	0	0	0.000028	0
в	0.012151	0.000001	0.000009	0.000259	0.000055	0.000507	0.0001	0	0	0.000008	0.000125	0	0	0.000001	0.000126	0.000075

Bigram2()

Enter name of excel: bigram2																
	а	б	в	г	д	е	ё	...	ш	щ	ь	ы	ь	э	ю	я
а	0	0.002464	0.006758	0.016757	0.00369	0.007503	0.004139	0	0.000576	0.000092	0	0	0	0.002853	0.000023	0.002994
б	0.015745	0.000008	0.000355	0.002283	0.000934	0.001776	0.001296	0.000002	0	0.000973	0.000196	0	0	0	0.000091	0.001863
в	0.000384	0.000547	0	0.000047	0.000003	0.000019	0.002392	0	0	0.000008	0.000182	0.000167	0.003943	0.000036	0	0.000002
г	0.005198	0.005871	0.000002	0.000044	0.000177	0.000735	0.00541	0	0	0.000694	0.000018	0.000003	0.003114	0.00012	0	0.000152
д	0.000947	0.001113	0	0.000002	0.000002	0.001051	0.000788	0	0	0.000005	0	0	0	0	0	0
е	0.000785	0.005381	0.000001	0.000667	0.000003	0.000026	0.004432	0	0	0.000065	0	0.000474	0.000874	0	0.000008	0.000245
ё	0.018618	0.000021	0.001456	0.002043	0.003221	0.002879	0.001537	0	0	0.000892	0.000822	0	0	0.000255	0.000311	0
...	0.000658	0	0	0	0	0	0	0	0	0.000003	0	0	0	0	0	0
ш	0.000495	0.001299	0.000047	0	0.000003	0.0008	0.004805	0.000006	0	0	0	0	0.000023	0	0.000002	0
щ	0.001268	0.004737	0.000102	0.000923	0.000358	0.000658	0.000232	0	0.000003	0	0.000028	0.000295	0.001327	0	0.000157	0.00054
ь	0.017303	0.000073	0.000581	0.002212	0.000383	0.001481	0.001693	0	0	0.000574	0.00013	0	0	0	0.000237	0.001124
ы	0.005877	0.000002	0.000031	0.000002	0.000005	0.000229	0.000003	0	0	0.000201	0	0	0	0.000003	0	0
ь	0.004642	0.00619	0	0.000151	0.000003	0.000031	0.00049	0	0	0.000006	0	0	0.000019	0	0	0
э	0.006875	0.005633	0.000034	0.000029	0.000182	0.000021	0.003631	0.000011	0	0.000005	0.000002	0	0.000645	0.003555	0.000892	0.001439
ю	0.007814	0.002238	0.000026	0	0.000075	0.000005	0.00414	0	0	0.000008	0.000083	0.000083	0	0	0.000358	0.000358
я	0.003386	0.010026	0.000005	0.000011	0.000037	0.000303	0.01122	0.000002	0	0.000024	0.000109	0.002489	0.001087	0	0.000149	0.003432
а	0.024697	0.000006	0.003177	0.007094	0.004869	0.004246	0.00194	0.000003	0	0.000083	0.000206	0	0	0.000018	0.000618	0.000556
б	0.000042	0.001315	0	0	0	0.002068	0	0	0.000003	0.000002	0	0.000191	0.000054	0.000005	0.000496	0.000496
в	0.000587	0.006607	0.000012	0.003063	0.001044	0.000683	0.005063	0	0.0003	0.000015	0	0.000907	0.000908	0.000002	0.000199	0.000991
г	0.003552	0.001901	0.000054	0.001213	0.000021	0.000164	0.003704	0.000649	0	0.000099	0	0.000013	0.000204	0.003211	0.000237	0.003727
д	0.005616	0.005842	0.000008	0.002191	0.000013	0.000076	0.005576	0	0.000008	0.000019	0.000016	0.001119	0.006429	0.000002	0.000041	0.000452
е	0.006246	0.000034	0.000451	0.000675	0.001421	0.001619	0.000159	0	0.00053	0.000247	0	0	0	0.00001	0.000684	0.000024
ё	0.000019	0.000187	0	0	0	0.000294	0	0	0	0	0.000003	0.000245	0	0	0.000002	0
...	0.002645	0.000412	0	0.000201	0	0.000062	0	0	0.000008	0	0	0	0	0	0	0
ш	0.000457	0.000577	0	0.000019	0	0.000641	0	0	0	0	0.000206	0	0	0	0	0
щ	0.00044	0.002781	0	0.000002	0	0.003933	0.000002	0	0.000157	0	0	0.000196	0	0	0	0
ь	0.000071	0.000827	0	0.000052	0	0.002091	0	0	0	0	0	0.000448	0	0.000002	0	0
ы	0.000092	0.000269	0	0	0	0.001448	0	0	0	0	0.000019	0	0	0	0	0
ь	0	0	0	0	0	0.000054	0	0	0	0	0	0	0	0.000003	0.000214	0
э	0.004541	0.000002	0.000123	0.000778	0.000081	0.000191	0.000574	0	0.00043	0.00011	0	0	0.000003	0	0.000003	0
ю	0.012486	0	0.000101	0.000036	0.000088	0.000045	0.000602	0	0.000423	0.000003	0	0	0.00001	0.000334	0.000383	0
я	0.000003	0	0	0.000002	0.000011	0.000002	0	0	0.000011	0	0	0	0.000005	0	0	0
а	0.003274	0.000003	0.000427	0	0.000005	0.000277	0	0	0.000024	0.000177	0	0	0	0	0.000026	0
б	0.012212	0.000002	0.00001	0.00025	0.00006	0.000512	0.000086	0	0.000006	0.000013	0	0	0.000002	0.000135	0.000078	0

Робота з CoolPinkProgram

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

$0,768057 > R > 0,656017$

Лаборотворная 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

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

$0,752332 > R > 0,640292$

Лаборотворная 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

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

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

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	0.1409706240083466	0.11497178841865563
H2	0.22307065957981287	0.18063319649451237
H2(з кроком 2)	0.223174002851259	0.18073335512295907

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.