

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  $H$  та  $H'$  на тому ж тексті, в якому вилучено всі пробіли.

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

### **With spaces**

Монограми

<b>а</b>	42893
<b>б</b>	8693
<b>в</b>	21483
<b>г</b>	12564
<b>д</b>	17248
<b>е</b>	41887
<b>ж</b>	4755
<b>з</b>	9753
<b>и</b>	36693
<b>й</b>	4651

Біграми, що пересікаються:

<b>ми</b>	<b>0</b>
<b>ис</b>	2255
<b>ст</b>	2076
<b>те</b>	5233
<b>ер</b>	3153
<b>р</b>	4424
<b>и</b>	1129
<b>и</b>	5500
<b>м</b>	12959
<b>сс</b>	3881

Біграми що не пересікаються:

<b>ми</b>	1098
<b>ст</b>	2683
<b>ер</b>	2198
<b>и</b>	2834
<b>м</b>	1967
<b>ис</b>	1074
<b>си</b>	563
<b>с</b>	1130
<b>ду</b>	531
<b>рс</b>	166
<b>зг</b>	96
<b>ла</b>	1717
<b>вг</b>	4

Ентропія та надлишковість:

<b>h1</b>	4,24701271863426
<b>r1</b>	0,165200247911692
<b>h2-crossed</b>	3.9049422773713913
<b>h2-uncrossed</b>	3.9060134979034915
<b>r2-crossed</b>	0.2324381721849238
<b>r2-uncrossed</b>	0.2324381721849238

## Without spaces

### Монограми

<b>а</b>	42893
<b>б</b>	8693
<b>в</b>	21483
<b>г</b>	12564
<b>д</b>	17248
<b>е</b>	41887
<b>ж</b>	4755
<b>з</b>	9753
<b>и</b>	36693
<b>й</b>	4651
<b>к</b>	17490
<b>л</b>	28902
<b>м</b>	17632
<b>н</b>	34613

### Біграми що пересікаються

<b>ми</b>	2536
<b>ис</b>	3189
<b>ст</b>	5432
<b>те</b>	3238
<b>ер</b>	4680
<b>ри</b>	5029
<b>им</b>	1984
<b>сс</b>	828
<b>си</b>	1076
<b>сд</b>	360
<b>ду</b>	1318
<b>ур</b>	617
<b>рс</b>	393
<b>сл</b>	1829
<b>ль</b>	2997

### Біграми що не пересікаються

<b>ми</b>	1258
<b>ст</b>	2737
<b>ер</b>	2348

<b>им</b>	958
<b>ис</b>	1609
<b>си</b>	560
<b>сд</b>	170
<b>ур</b>	304
<b>сл</b>	918
<b>ьп</b>	548
<b>ро</b>	2852
<b>жи</b>	331

Ентропія та надлишковість:

<b>h1</b>	4,41893097852972
<b>r1</b>	0,131407714136014
<b>h2-crossed</b>	4.126965666879817
<b>h2-uncrossed</b>	4.126436885504978
<b>r2-crossed</b>	0.1887968923492837
<b>r2-uncrossed</b>	0.1889008304794165

## Лабораторная работа №1

Лабораторная работа №1

Произвольная часть текста:

равы\_если\_бы\_они\_бы

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:

Символ по счету:

Номер эксперимента:

51

Поле ввода символов:

Продолжить

Другой

Неравенство для энтропии:

1,85446834040817 < H < 2,54444687389046

Двоичная таблица угаданных символов:

00100000000000000000000000000000

00000100000000000000000000000000

00100000000000000000000000000000

10000000000000000000000000000000

10000000000000000000000000000000

Вероятности:

q[1] = 0,56

q[2] = 0,1

q[3] = 0,06

q[4] = 0,04

q[5] = 0

q[6] = 0,02

q[7] = 0

q[8] = 0

q[9] = 0

q[10] = 0

q[11] = 0

q[12] = 0,02

q[13] = 0

q[14] = 0

q[15] = 0,02

q[16] = 0

q[17] = 0,02

q[18] = 0,02

q[19] = 0

q[20] = 0

q[21] = 0,02

q[22] = 0

q[23] = 0,02

q[24] = 0

q[25] = 0

q[26] = 0,02

q[27] = 0,02

q[28] = 0,04

q[29] = 0,02

q[30] = 0

q[31] = 0

q[32] = 0

Строка состояния:

Произвольная часть текста:  
естественным\_потому\_что\_люди\_думают\_что\_каждый\_человек\_знает\_его\_инстинктивно\_  
  
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: а

Символ по счету: 1

Номер эксперимента: 50

Поле ввода символов:  
а

Продолжить Другой

Неравенство для энтропии:  
1,74741456622636 < H < 2,354034403112

Двоичная таблица угаданных символов:  
10000000000000000000000000000000 ^  
10000000000000000000000000000000  
10000000000000000000000000000000  
00001000000000000000000000000000  
10000000000000000000000000000000 v

Вероятности:  
q[1] = 0.58  
q[2] = 0.1  
q[3] = 0  
q[4] = 0.02  
q[5] = 0.08  
q[6] = 0.02  
q[7] = 0.04  
q[8] = 0  
q[9] = 0  
q[10] = 0  
q[11] = 0  
q[12] = 0.02  
q[13] = 0  
q[14] = 0  
q[15] = 0  
q[16] = 0  
q[17] = 0  
q[18] = 0  
q[19] = 0.02  
q[20] = 0  
q[21] = 0  
q[22] = 0  
q[23] = 0.02  
q[24] = 0  
q[25] = 0.02  
q[26] = 0  
q[27] = 0.02  
q[28] = 0.04  
q[29] = 0.02  
q[30] = 0  
q[31] = 0  
q[32] = 0

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