КРИПТОГРАФІЯ
КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

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.

Алгоритм роботи програми:

1. Фільтрування вхідного тексту.
2. Підрахунок кількості букв(для двух випадків) і біграмм (для чотирьох випадків) і вивід частот.
3. Підрахунок ентропій H_1 та H_2.
4. Підрахунок надлишковості.
5. Вивід результату.

Частота букв

```
Частота появи букв без пробілів :
о: 0.1064265
а: 0.0818242
е: 0.0816780
и: 0.0816370
н: 0.0692591
л: 0.0580031
р: 0.0545656
т: 0.0512115
с: 0.0507609
в: 0.0478003
д: 0.0352454
м: 0.0328845
у: 0.0251420
к: 0.0250396
п: 0.0228104
г: 0.0215129
ь: 0.0193743
ы: 0.0178180
з: 0.0173001
я: 0.0172182
б: 0.0166507
х: 0.0111214
ч: 0.0103081
й: 0.0088892
ж: 0.0085704
э: 0.0068340
ш: 0.0050246
ю: 0.0050202
ф: 0.0033307
щ: 0.0032678
ц: 0.0031142
ё: 0.0001902
ъ: 0.0001668
```

```
Частота появи букв з пробілами :
 : 0.1650013
о: 0.0888660
а: 0.0683231
е: 0.0682010
и: 0.0681668
н: 0.0578312
л: 0.0484325
р: 0.0455622
т: 0.0427615
с: 0.0423853
в: 0.0399132
д: 0.0294298
м: 0.0274585
у: 0.0209936
к: 0.0209081
п: 0.0190466
г: 0.0179633
ь: 0.0161776
ы: 0.0148780
з: 0.0144456
я: 0.0143772
б: 0.0139033
х: 0.0092864
ч: 0.0086073
й: 0.0074225
ж: 0.0071562
э: 0.0057064
ш: 0.0041955
ю: 0.0041919
ф: 0.0027811
щ: 0.0027286
ц: 0.0026004
ё: 0.0001588
ъ: 0.0001392
```

Частота біграм

Частота біграм з пробілом

```
Частота появи біграм з пробілом :
('и', ' '): 0.0238590
(' ', 'в'): 0.0173684
('о', ' '): 0.0169678
('а', ' '): 0.0168872
(' ', 'и'): 0.0163070
(' ', 'с'): 0.0162863
('е', ' '): 0.0160994
(' ', 'н'): 0.0150563
(' ', 'п'): 0.0141598
(' ', 'о'): 0.0114287
('н', 'а'): 0.0110281
('л', 'и'): 0.0104577
('о', 'р'): 0.0097114
('с', 'т'): 0.0096174
('ь', ' '): 0.0095453
('н', 'о'): 0.0092656
('я', ' '): 0.0088882
('р', 'а'): 0.0087367
('т', 'о'): 0.0086610
(' ', 'т'): 0.0086378
('а', 'л'): 0.0085682
('г', 'о'): 0.0084912
('в', ' '): 0.0082176
('н', 'и'): 0.0080857
('р', 'о'): 0.0080686
('р', 'е'): 0.0079453
('м', ' '): 0.0075202
('п', 'о'): 0.0074421
('н', 'е'): 0.0074237
(' ', 'к'): 0.0073101
('е', 'н'): 0.0072711
('в', 'о'): 0.0068741
('и', 'л'): 0.0067666
('л', 'а'): 0.0066811
(' ', 'м'): 0.0066213
(' ', 'д'): 0.0065797
('о', 'л'): 0.0065016
```

…….

Частота біграм без пробіла

```
Частота появи біграм без пробіла :
('н', 'а'): 0.0133522
('л', 'и'): 0.0130303
('о', 'р'): 0.0120108
('с', 'т'): 0.0118514
('н', 'о'): 0.0115837
('т', 'о'): 0.0108420
('р', 'а'): 0.0106139
('н', 'и'): 0.0105890
('а', 'л'): 0.0105466
('г', 'о'): 0.0102730
('е', 'н'): 0.0102584
('р', 'о'): 0.0099410
('и', 'н'): 0.0099161
('р', 'е'): 0.0095592
('о', 'н'): 0.0093983
('о', 'с'): 0.0093266
('о', 'в'): 0.0090940
('н', 'е'): 0.0089492
('п', 'о'): 0.0089127
('о', 'т'): 0.0087620
('в', 'о'): 0.0087342
('и', 'л'): 0.0087064
('а', 'н'): 0.0086713
('о', 'л'): 0.0081535
('л', 'а'): 0.0080803
('е', 'р'): 0.0079604
('к', 'о'): 0.0078916
('е', 'л'): 0.0078302
('в', 'а'): 0.0073138
('р', 'и'): 0.0073021
('л', 'е'): 0.0071500
('д', 'а'): 0.0069525
('о', 'д'): 0.0068384
('п', 'р'): 0.0067975
('в', 'е'): 0.0067829
('т', 'а'): 0.0067463
('и', 'с'): 0.0067097
```

…….

Частота біграм з пробілом та з буквами що не перетинаються

```
Частота появи біграм з пробілом с шагом 2 :
('и', ' '): 0.0240886
(' ', 'в'): 0.0173758
('а', ' '): 0.0172048
('о', ' '): 0.0168042
(' ', 'с'): 0.0162008
('е', ' '): 0.0161568
(' ', 'и'): 0.0161544
(' ', 'н'): 0.0151211
(' ', 'п'): 0.0143076
(' ', 'о'): 0.0115765
('н', 'а'): 0.0109047
('л', 'и'): 0.0102305
('о', 'р'): 0.0098079
('ь', ' '): 0.0096662
('с', 'т'): 0.0095759
('н', 'о'): 0.0093242
('т', 'о'): 0.0087893
('я', ' '): 0.0087184
('р', 'а'): 0.0086671
('г', 'о'): 0.0085670
('а', 'л'): 0.0085548
(' ', 'т'): 0.0085059
('в', ' '): 0.0081444
('р', 'о'): 0.0081053
('н', 'и'): 0.0079221
('р', 'е'): 0.0078488
('м', ' '): 0.0076582
('н', 'е'): 0.0073895
('п', 'о'): 0.0073553
('е', 'н'): 0.0072772
(' ', 'к'): 0.0072527
('и', 'л'): 0.0069107
('в', 'о'): 0.0068765
('л', 'а'): 0.0068277
(' ', 'д'): 0.0066665
```

…….

Частота біграм без пробіла та з буквами що не перетинаються

```
Частота появи біграм без пробіла з шагом 2 :
('н', 'а'): 0.0135540
('л', 'и'): 0.0131210
('о', 'р'): 0.0118543
('с', 'т'): 0.0117753
('н', 'о'): 0.0115442
('т', 'о'): 0.0107747
('н', 'и'): 0.0105992
('р', 'а'): 0.0105173
('а', 'л'): 0.0104529
('е', 'н'): 0.0104471
('г', 'о'): 0.0104383
('р', 'о'): 0.0100814
('и', 'н'): 0.0097537
('р', 'е'): 0.0096689
('о', 'н'): 0.0094290
('о', 'с'): 0.0092535
('о', 'в'): 0.0091862
('о', 'т'): 0.0089521
('н', 'е'): 0.0089258
('п', 'о'): 0.0089229
('и', 'л'): 0.0089112
('в', 'о'): 0.0088819
('а', 'н'): 0.0087152
('е', 'р'): 0.0080511
('л', 'а'): 0.0080423
('о', 'л'): 0.0079984
('е', 'л'): 0.0078200
('к', 'о'): 0.0077468
('р', 'и'): 0.0072202
('л', 'е'): 0.0071939
('в', 'а'): 0.0071822
('д', 'а'): 0.0069452
('о', 'д'): 0.0068662
('т', 'а'): 0.0067989
('и', 'с'): 0.0067082
('п', 'р'): 0.0067053
('в', 'е'): 0.0066819
('л', 'ь'): 0.0064157
```

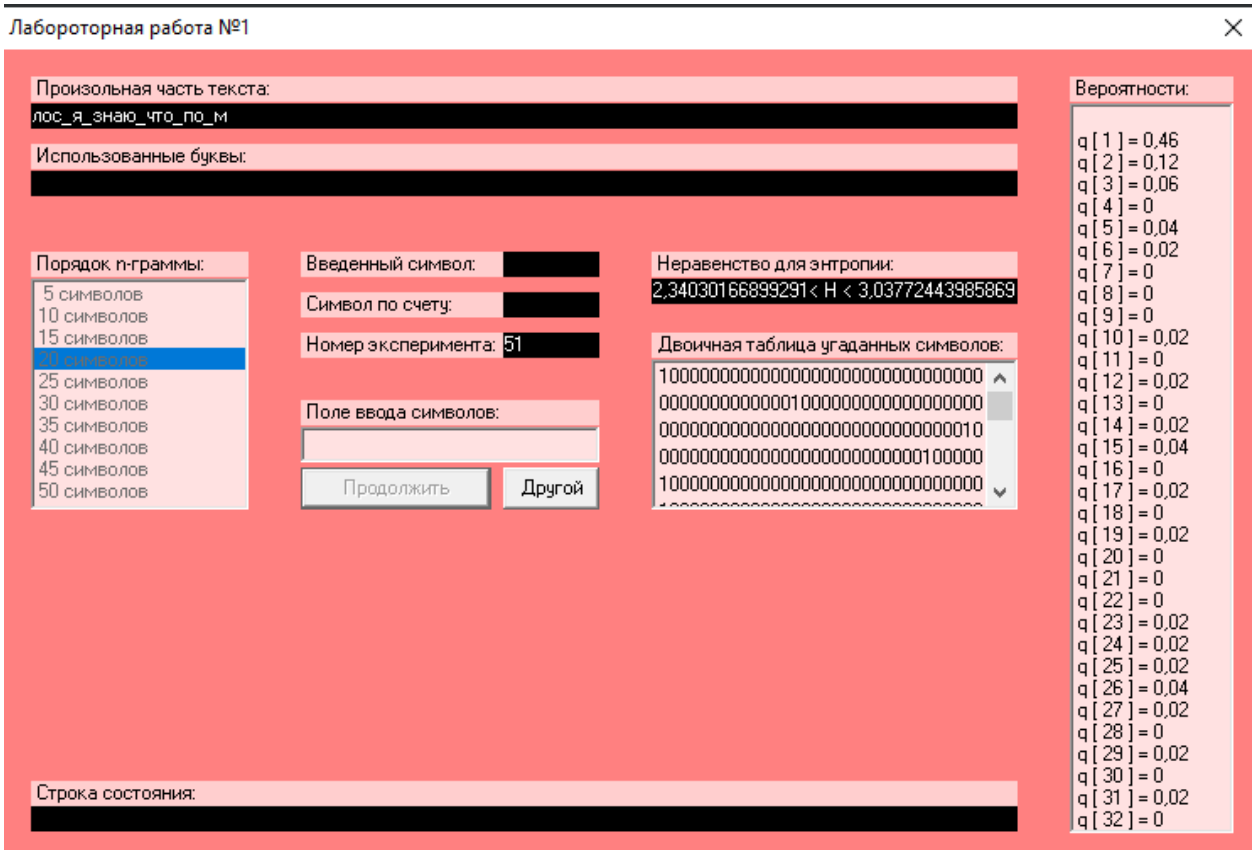
…….

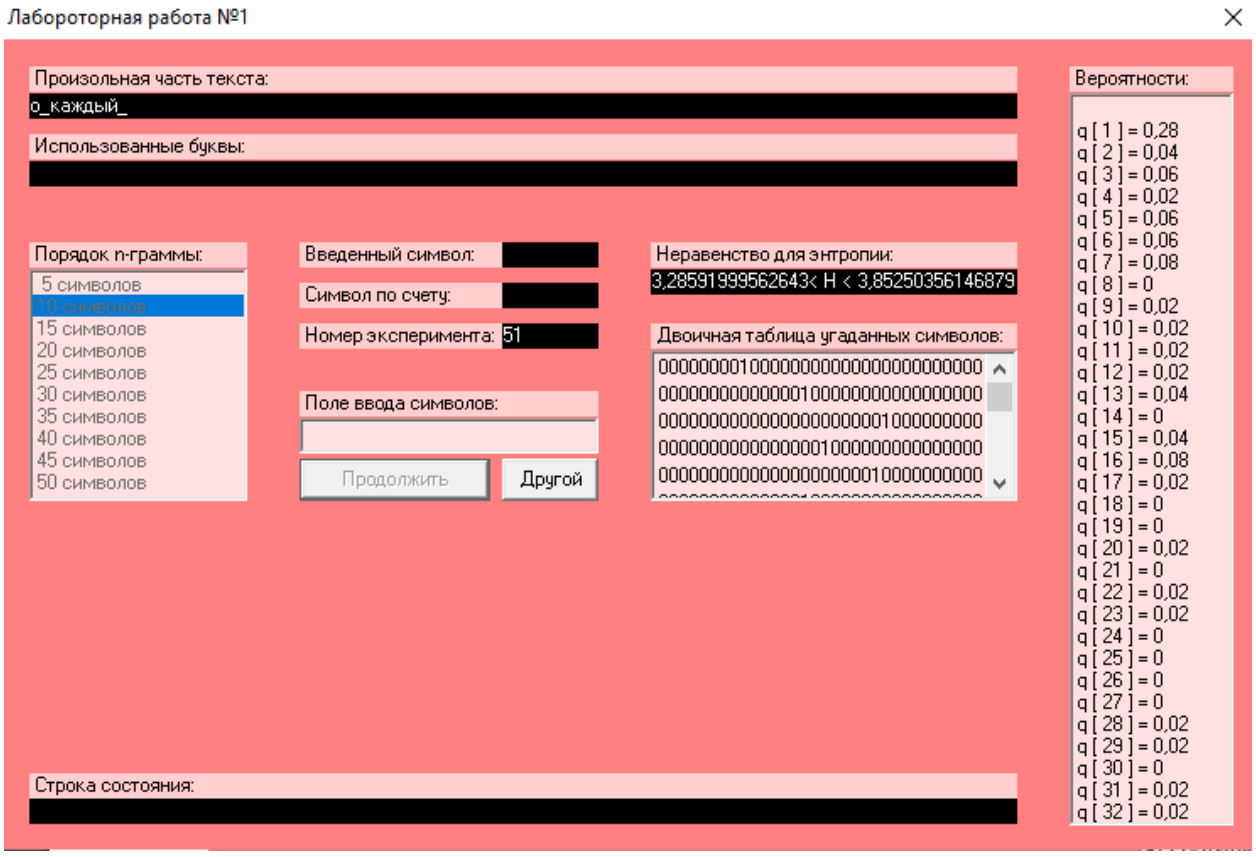
Ентропія та надлишковість для біграм та букв

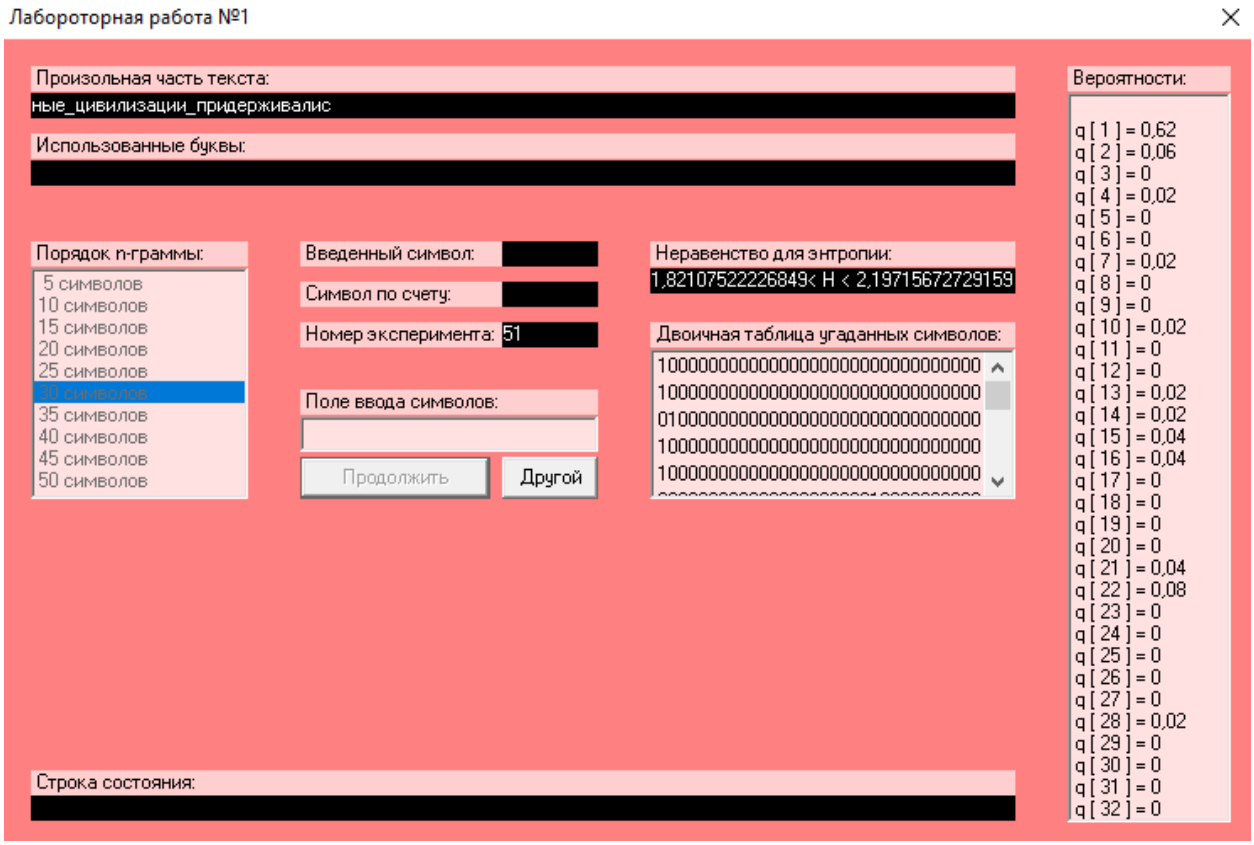
	Ентропія	Надлишковість
Букви з пробілами	4.3521273	0.1445388
Букви без пробілів	4.4383136	0.1201493
Біграми з пробілами	3.9720874	0.2192400
Біграми без пробілів	4.1478485	0.1777311
Біграми з пробілами і не перетинаються	3.9710400	0.2194459
Біграми без пробілів і не перетинаються	4.1469627	0.1779067

H(10),H(20),H(30)

	Ентропія	Надлишковість
H10	3.2859199< H < 3.8525035	0.34281602< R < 0.2294993
H20	2.3403016< H < 3.0377244	0.53193968< R < 0.39245512
H30	1.8210752< H < 2.1971567	0.63578496< R < 0.56056866


Лабороторная работа №1
Произольная часть текста:
лос_я_знаю_что_по_м
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:
2,340301668992911< H < 3,03772443985869
Двоичная таблица угаданных символов:
Строка состояния:
Вероятности:
q [1] = 0,46
q [2] = 0,12
q [3] = 0,06
q [4] = 0
q [5] = 0,04
q [6] = 0,02
q [7] = 0
q [8] = 0
q [9] = 0
q [10] = 0,02
q [11] = 0
q [12] = 0,02
q [13] = 0
q [14] = 0,02
q [15] = 0,04
q [16] = 0
q [17] = 0,02
q [18] = 0
q [19] = 0,02
q [20] = 0
q [21] = 0
q [22] = 0
q [23] = 0,02
q [24] = 0,02
q [25] = 0,02
q [26] = 0,04
q [27] = 0,02
q [28] = 0
q [29] = 0,02
q [30] = 0
q [31] = 0,02
q [32] = 0


Лабороторная работа №1
Произольная часть текста:
о_каждый_
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:
3,28591999562643< H < 3,85250356146879
Двоичная таблица угаданных символов:
Строка состояния:
Вероятности:
q [1] = 0,28
q [2] = 0,04
q [3] = 0,06
q [4] = 0,02
q [5] = 0,06
q [6] = 0,06
q [7] = 0,08
q [8] = 0
q [9] = 0,02
q [10] = 0,02
q [11] = 0,02
q [12] = 0,02
q [13] = 0,04
q [14] = 0
q [15] = 0,04
q [16] = 0,08
q [17] = 0,02
q [18] = 0
q [19] = 0
q [20] = 0,02
q [21] = 0
q [22] = 0,02
q [23] = 0,02
q [24] = 0
q [25] = 0
q [26] = 0
q [27] = 0
q [28] = 0,02
q [29] = 0,02
q [30] = 0
q [31] = 0,02
q [32] = 0,02


Лабороторная работа №1
Произольная часть текста:
ные_цивилизации_придерживалис
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:
1,82107522226849< H < 2,19715672729159
Двоичная таблица угаданных символов:
Вероятности:
q [1] = 0,62
q [2] = 0,06
q [3] = 0
q [4] = 0,02
q [5] = 0
q [6] = 0
q [7] = 0,02
q [8] = 0
q [9] = 0
q [10] = 0,02
q [11] = 0
q [12] = 0
q [13] = 0,02
q [14] = 0,02
q [15] = 0,04
q [16] = 0,04
q [17] = 0
q [18] = 0
q [19] = 0
q [20] = 0
q [21] = 0,04
q [22] = 0,08
q [23] = 0
q [24] = 0
q [25] = 0
q [26] = 0
q [27] = 0
q [28] = 0,02
q [29] = 0
q [30] = 0
q [31] = 0
q [32] = 0
Строка состояния: