МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО”

Звіт з лабораторних робіт

курс «Криптографія»

Виконав студент:

Костюковець Остап Юрійович
група: ФБ-96

Київ, 2021 р.

Лабораторна робота 1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

1.1. 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.

1.2. Завдання

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, атакож 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.

1.3. Хід роботи

1.3.1. Аналіз частот символів

1.4. Висновок

Під час виконання даної лабораторної роботи я навчився вимірювати частоти символів та біграм у тексті, визначати ентропію. Також, я навчився визначати надлишковість мови.

Табл. 1: Ентропія, надлишковість

H1 (no spaces)	4.492315117172669
H1 redundancy (no spaces)	0.10153697656546612
H2 (no spaces)	4.1802847959379115
H2 redundancy (no spaces)	0.16394304081241773
H1 (with spaces)	4.406442333151592
H1 redundancy (with spaces)	0.11871153336968165
H2 (with spaces)	2.9080400719491637
H2 redundancy (with spaces)	0.4183919856101672
H10	$3.1173 < R < 3.6639$
H20	$2.375 < R < 2.7555$
H30	$2.56079 < R < 3.0895$

Табл. 2: H1 (no spaces)

о	**0.10584822334367971**
е	0.08399778970225735
а	0.0824350401264014
н	0.06483252245912892
и	0.062443069723669614
т	0.06075944449001265
л	0.052529107290097264
р	0.04690191372068225
с	0.04619608621887991
в	0.04187478145249371
к	0.0377779600505951
д	0.03151616927772478
у	0.03098086279317743
м	0.029899457354636227
п	0.027291996736357238
я	0.022279973925393 82
ь	0.022161256761482107
ы	0.019676830294893435
г	0.019454505424295145
з	0.017880963378992674
б	0.01760683465505109
ч	0.015431072814632862
й	0.011353677857737984
ж	0.010216151578074883
ш	0.009704588526309881
х	0.00845266207051367 8
ю	0.0071014448958095
щ	0.00427813488860013
э	0.004193953626917282
ц	0.0033564579978674083
ф	0.0013792775952651277
ъ	0.0001877889683694 3055

Табл. 3: H2 (no spaces)

то	**0.014304339004416279**
на	0.012314207638478175
не	0.011584636703893491
ст	0.010900394140984187
по	0.010889601671537668
ал	0.010874492214312541
но	0.01060683897203887
ко	0.010501072771462984
ра	0.010036996585262668
ен	0.009972241768583552
он	0.009408874863475261
ов	0.009331169083460324
ер	0.00923187836455235
от	0.008839032476699059
ка	0.008672828447222666
ро	0.008627500075547287
ос	0.008349054363827097
го	0.00828861653492659
ол	0.008228178706026084
ни	0.008143997444343236
ак	0.007850442275397919
ла	0.007833174324283487
ло	0.007546094637006083
во	0.007457596387544627
пр	0.007306501815293361
ор	0.007228796035278424
ре	0.007058275018023424
ет	0.00699352020134431
ве	0.006790621775749753
ел	0.006699965032398994
ли	0.00656182142348355
од	0.00649922510069374

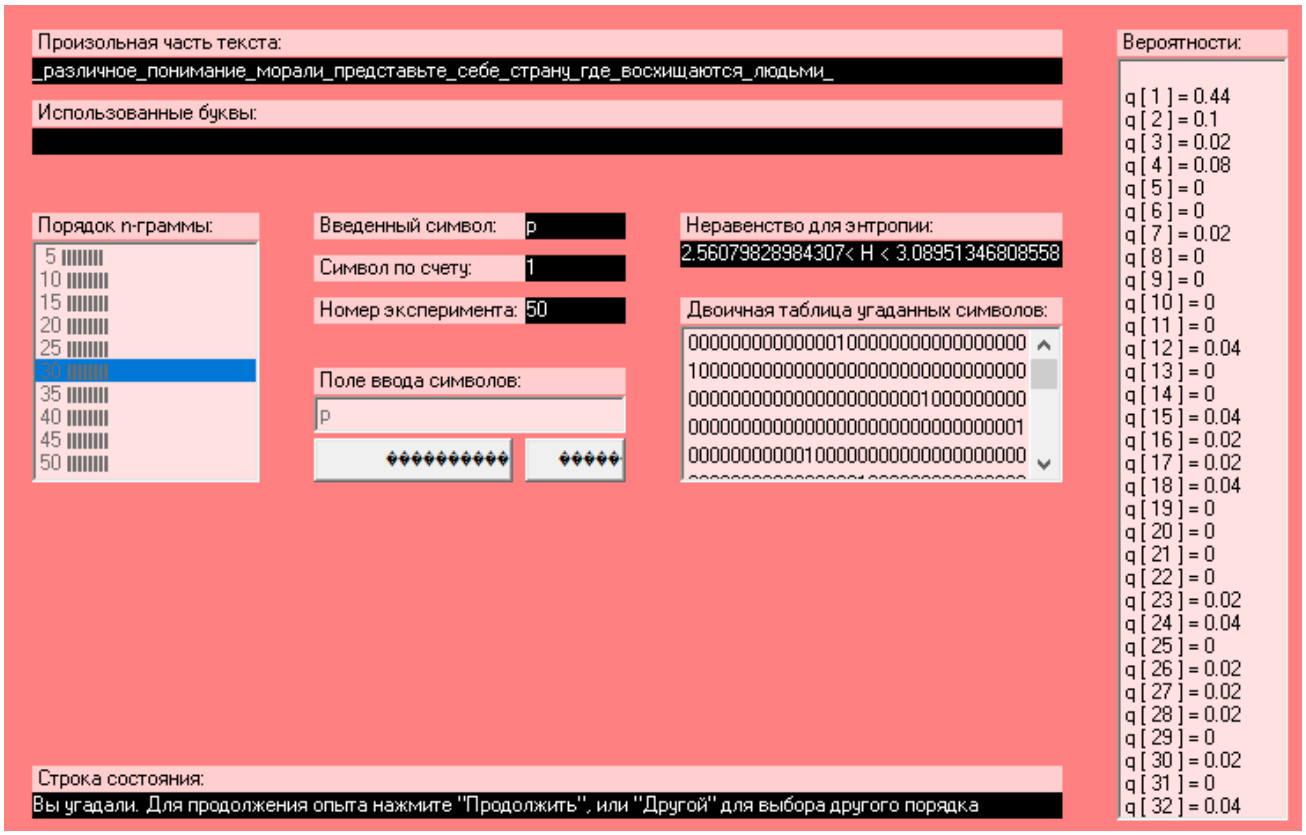
Табл. 4: H1 (with spaces)

	0.16067119346633524
о	0.0888414629727595
е	0.07050176458226141
а	0.06919010384584726
н	0.05441580370018769
и	0.05241026718746603
т	0.05099715202945077
л	0.04408919293007616
р	0.039366127267325156
с	0.03877370591261876
в	0.03514671034037958
к	0.03170813012254245
д	0.026452428746385686
у	0.026003130593580833
м	0.02509547585747103
п	0.022906959048647396
я	0.018700223924401963
ь	0.018600581188900886
ы	0.016515330487778365
г	0.01632872681947635
з	0.015008007652562086
б	0.0147779235178596
ч	0.01295174392903988
й	0.009529468886102918
ж	0.008574710311392607
ш	0.008145340705687969
х	0.0070945627676766215
ю	0.005960447269064373
щ	0.00359076185023878
э	0.0035201060923380173
ц	0.0028171718855304253
ф	0.0011576674179125028
ъ	0.00015761669070170226

Табл. 5: H2 (with spaces)

то	**0.01160566115672534**
на	0.010292188734211156
не	0.009679838832404543
по	0.009132709630198634
ст	0.008886320320595973
ал	0.008802982759995073
но	0.008681599791293761
ко	0.008489561064691687
ра	0.008402600131890748
ер	0.007321023530179067
ка	0.007210510678077873
ро	0.0071779003282775216
го	0.0069152058437746845
ен	0.006859043574674078
ни	0.0066579464175719055
ол	0.006505764785170262
ла	0.00647496612146993
ов	0.006232200184067308
пр	0.006128934076366193
во	0.006027479654765097
от	0.006000304363264803
ре	0.005920590174863943
он	0.005906096686063786
ло	0.005672389179161262
ве	0.005623473654460733
ор	0.005587239932460342
ак	0.005482162138659207
ль	0.005324545447957505
ва	0.005295558470357192
ос	0.005295558470357192
ел	0.0052575130622567814
ли	0.0052575130622567814
за	0.005069097707854746


Произольная часть текста:
онкретном_случае_у_
Использованные буквы:
Порядок n-граммы:
5
10
15
20
25
30
35
40
45
50
Введенный символ:
Символ по счету:
Номер эксперимента: 50
Поле ввода символов:
Неравенство для энтропии:
2.37512225054647< H < 2.75559954320974
Двоичная таблица угаданных символов:
Вероятности:
q[1] = 0.5102040
q[2] = 0.0816326
q[3] = 0
q[4] = 0
q[5] = 0.0408163
q[6] = 0
q[7] = 0.0204081
q[8] = 0
q[9] = 0
q[10] = 0
q[11] = 0
q[12] = 0.040816
q[13] = 0.020408
q[14] = 0
q[15] = 0
q[16] = 0.081632
q[17] = 0
q[18] = 0.020408
q[19] = 0.020408
q[20] = 0.020408
q[21] = 0
q[22] = 0
q[23] = 0
q[24] = 0
q[25] = 0.040816
q[26] = 0.020408
q[27] = 0.020408
q[28] = 0.020408
q[29] = 0
q[30] = 0
q[31] = 0
q[32] = 0.040816
Строка состояния:


Произольная часть текста:
_различное_понимание_морали_представьте_себе_страну_где_восхищаются_людьми_
Использованные буквы:
Порядок n-граммы:
5
10
15
20
25
30
35
40
45
50
Введенный символ: р
Символ по счету: 1
Номер эксперимента: 50
Поле ввода символов: р
Неравенство для энтропии:
2.56079828984307< H < 3.08951346808558
Двоичная таблица угаданных символов:
Вероятности:
q[1] = 0.44
q[2] = 0.1
q[3] = 0.02
q[4] = 0.08
q[5] = 0
q[6] = 0
q[7] = 0.02
q[8] = 0
q[9] = 0
q[10] = 0
q[11] = 0
q[12] = 0.04
q[13] = 0
q[14] = 0
q[15] = 0.04
q[16] = 0.02
q[17] = 0.02
q[18] = 0.04
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0
q[23] = 0.02
q[24] = 0.04
q[25] = 0
q[26] = 0.02
q[27] = 0.02
q[28] = 0.02
q[29] = 0
q[30] = 0.02
q[31] = 0
q[32] = 0.04
Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка