«Криптографія» Комп 'ютерний практикум №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали: студенти групи ФБ-01 Корабельський Тарас та Оліферчук Владислав варіант №10

Мета роботи

Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела івдкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо, оцінки ентропії, на символ джерела.

Постановка задачі:

Визначити надлишковість російської мови в різних моделях джерела, вираховуючи за допомогою самостійно написанної програми частоту входження буквенних символів в текст(розмір від 1 мб, визначення ентропію H1 та H2 для букв та біграм(з букв, які перетинаюся, і не перетинаються) з урахуванням пробілів і без них, за допомогою спеціальної програми CoolPinkProgram оцінити H^{10}, H^{20}, H^{30}

Для лабораторної роботи використовувався текст: «Война и Мир» Льва Толстого

Джерело	Н	R
Н1 текст з пробілами	4.226	0.1548
Н1 текст без пробілів	4.313	0.1374
H2 перехресна біграма з пробілами	3.7547	0.2491
H2 не перехресна біграма з пробілами	3.7532	0.2493
H2 перехресна біграма без пробілів	3.8824	0.2235
H2 не перехресна біграма без пробілів	3.8846	0.2230

Частоти літер у тексті з пробілами:

{' ': 0.17278987407351912,

'o': 0.0931007729298065,

'a': 0.06808690031910179,

'e': 0.06607299088407577,

'и': 0.05498125134123067,

'h': 0.05385874444301944,

'T': 0.04664244726820677,

'c': 0.043163183805425366,

'л': 0.041568309750953304,

'в': 0.037691597578737385,

'p': 0.03735509947011072,

'ĸ': 0.029394697043768882,

'д': 0.024878003563070537,

'м': 0.024526267804241907,

'y': 0.023464711506838532,

'n': 0.021154937697057737,

'я': 0.018856592124928413,

'r': 0.01670554383808473,

'ь': 0.016088419004905253,

'3': 0.014371643748817494,

'6': 0.014303074322908664,

'4': 0.011442205497490232,

'й': 0.009590830997951806,

'ж': 0.008351502114858867,

'ш': 0.007632792946999643,

'x': 0.006989002225966733,

'ю': 0.005316670116301365,

'ц': 0.0033319661774957685,

'щ': 0.002251362817339938,

'ф': 0.0017891540945470798}

H1 = 4.226

Частоти літер у тексті без пробілів:

{'o': 0.11265949811975479,

'a': 0.08238655136749996,

'e': 0.07995844558200005,

'и': 0.06653778240176544,

'н': 0.06517619903090917,

'T': 0.05643348142729591,

'c': 0.0522303995770793,

'л': 0.050303282706815444,

'B': 0.04560535904461721,

'p': 0.04520118700563843,

'ĸ': 0.035567139430132795,

'д': 0.0301023646369444,

'm': 0.02968282483982954,

'y': 0.028387322828958388,

'n': 0.02559961150307432,

'я': 0.022816510504631075,

'r': 0.020217822603820733,

'ь': 0.019466339230966648,

'3': 0.017387081555155653,

'6': 0.017308705988661667,

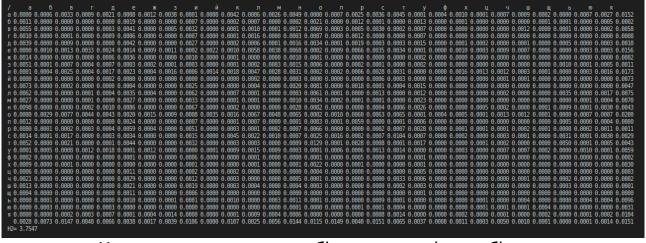
'4': 0.013844813304790284,

'й': 0.011607267720177714,

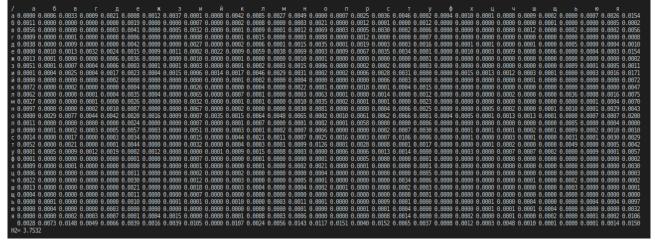
'ж': 0.010107374526096759, 'ш': 0.00923448586396768, 'x': 0.00844765664740062, 'ю': 0.00643448033157475, 'ц': 0.0040324997349061725, 'щ': 0.0027247035175261597, 'ф': 0.002165317121373017}

H1 = 4.313

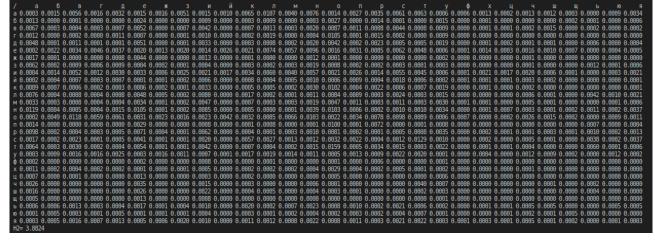
Частоти перехресних біграм у тексті з пробілами



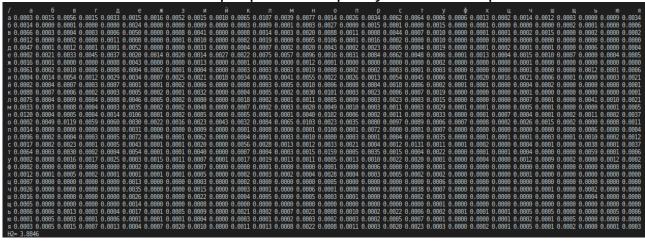
Частоти не перехресних біграм у тексті з пробілами



Частоти перехресних біграм у тексті без пробілів



Частоти не перехресних біграм у тексті без пробілів



Найбільш часті біграми для різних джерел:

Перехресні з пробілами:

o «» - 0.0200

и «» - 0.0173

e «» - 0.0156

a «» - 0.0152

«» «» - 0.0151

«» c - 0.0151

«» п - 0.0149

«» н - 0.0144

«» 0 - 0.0115

«» и - 0.0106

Не перехресні з пробілами:

«» o – 0.0200

«» и -0.0171

 \ll e - 0.0154

 \ll a - 0.0154

 $p \ll -0.0152$

 $\Pi \ll -0.0151$

«» «» – 0.0150

 $B \ll -0.0148$

 $H \ll -0.0143$

Перехресні без пробілів:

OT - 0.0159

TC - 0.0129

ah - 0.0119

BO - 0.0118

ла - 0.0107

eH - 0.0105

OF - 0.0105 OH - 0.0103OK - 0.0102

Не перехресні без пробілів:

ot - 0.0159

TC - 0.0131

aH - 0.0120

BO - 0.0119

ла - 0.0107

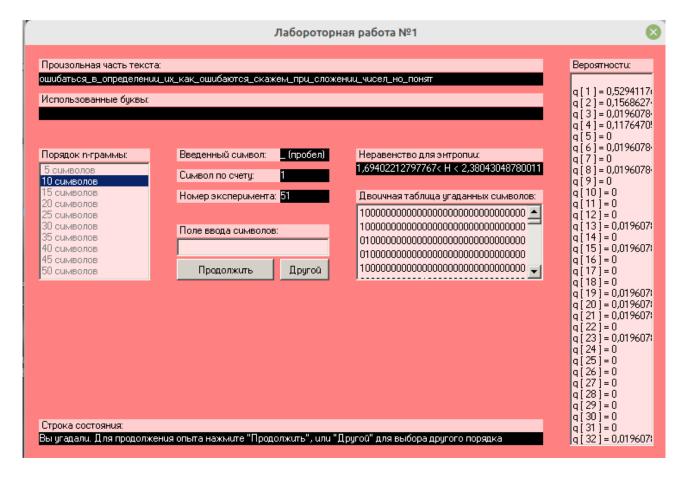
eH - 0.0106

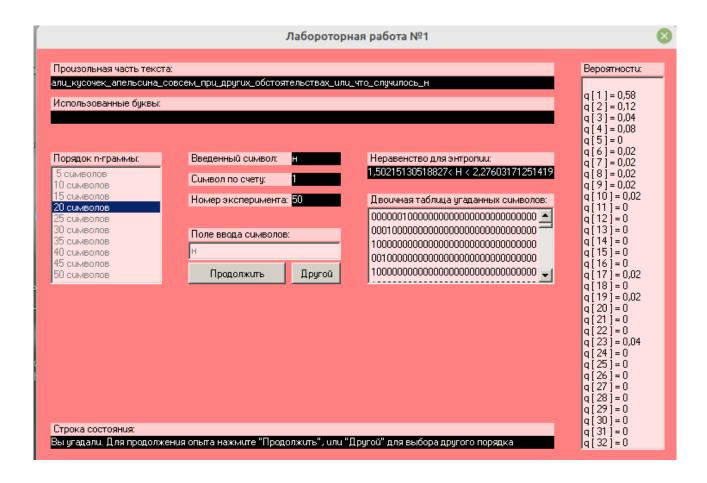
HO - 0.0103

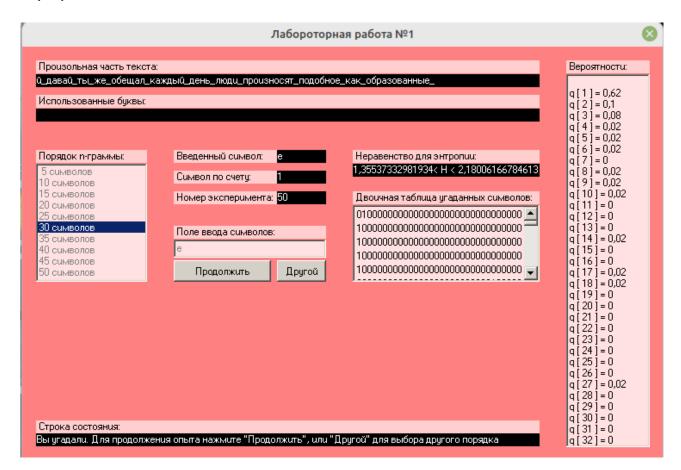
OH - 0.0102

Завдання 2

H(10)







	R(з пробілами)	R(без пробілів)
H1	0.1254	0.1135
H2	0.2026	0.1797
Н2 (крок 2)	0.2032	0.1798

Висновок:

Впродовж виконання лабораторної роботи ознайомились з поняттям ентропії на символ джерела та навчились її визначати, освоїли вимірювання частоти повторювання символів в тексті. Надлишковість може існувати в диапазоні 0.11-0.22. Також помітили за допомогою CoolPinkProgram що H⁽ⁿ⁾ менше, чим більше порядок н-грамми.