«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: «Шинель» М.В. Гоголя.

Джерело	Н	R
Н1 текст з пробілами	4.355	0.129
Н1 текст без пробілів	4.433	0.113
H2 перехресна біграма з пробілами	3.8984	0.2203
H2 не перехресна біграма з пробілами	3.9137	0.2172
H2 перехресна біграма без пробілів	4.0693	0.1861
H2 не перехресна біграма без пробілів	4.0613	0.1877

Частоти літер у тексті з пробілами:

{' ': 0.16079456684628385,

'o': 0.09549469119663298,

'e': 0.07374932245002072,

'a': 0.06722890029652775,

'и': 0.059831648758090744,

'h': 0.055319963013742306,

'T': 0.050027101999171,

'c': 0.041163154035009406,

'л': 0.04015878583043714,

'в': 0.03819787647865319,

'ĸ': 0.03791091413448969,

- 'p': 0.03249051430029015,
- 'm': 0.025364282753563117,
- 'д': 0.025093262761853138,
- 'y': 0.024009182795013233,
- 'π': 0.021888849918694,
- '4': 0.017871377100404935,
- 'ь': 0.017568472403787904,
- 'r': 0.014539425437617574,
- 'я': 0.01449159838025699,
- '6': 0.013822019577208813,
- '3': 0.01288142078245066,
- 'ж': 0.009342218537767432,
- 'ш': 0.008608870324905144,
- 'й': 0.008592927972451614,
- 'x': 0.006376940981411217,
- 'ю': 0.004097184580556707,
- 'u': 0.0032841246054267767,
- 'щ': 0.0024072952204827343,
- 'cb': 0.00047827057360584125}
- H1 = 4.252

Частоти літер у тексті без пробілів:

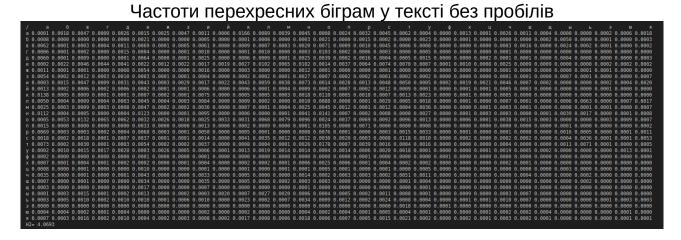
- {'o': 0.1137853085880364,
- 'e': 0.08787493113994263,
- 'a': 0.0801056170810934,
- 'и': 0.07129152973804684,
- 'h': 0.06591569629390422,
- 'T': 0.059609064832931255,
- 'c': 0.04904735672359098,
- 'л': 0.047850616416237675,
- 'B': 0.045514123435214555,
- 'k': 0.045172197633113614,
- 'p': 0.03871359914898467,
- 'm': 0.03022244173014456,
- 'д': 0.02989951180593811,
- 'y': 0.028607792109112323,
- 'π': 0.026081340349144236,
- 'u': 0.02129437911973102,
- 'ь': 0.020933457439735577,
- 'r': 0.017324240639781167.
- 'я': 0.017267253006097678,
- '6': 0.016469426134528807,
- '3': 0.015348669338753491,
- 'ж': 0.01113158444617518,

'ш': 0.010257774063028323, 'й': 0.010238778185133825, 'x': 0.007598351157798757, 'ю': 0.004881940618885702, 'u': 0.003913150846266361, 'щ': 0.002868377562069031, 'ф': 0.0005698763368349068}

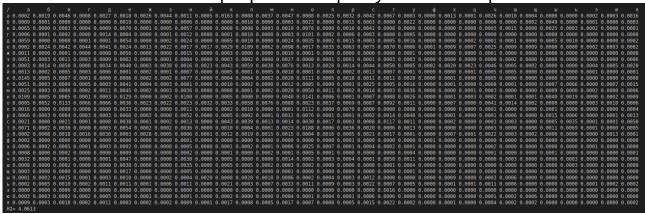
H1 = 4.315

Частоти перехресних біграм у тексті з пробілами

Частоти не перехресних біграм у тексті з пробілами



Частоти не перехресних біграм у тексті без пробілів



Таблиці найбільш частих біграм для різних джерел:

Перехресні з пробілами:

o" " - 0.0245

e" " - 0.0178

и" " - 0.0174

" "н - 0.0165

p,, ,, - 0.0162

c" " - 0.0154

a" " - 0.0132

ь" " - 0.0106

" "и - 0.0104

Не перехресні з пробілами:

o" " - 0.0240

и" " - 0.0179

e" " - 0.0172

" "н - 0.0171

" "п - 0.0154

" "c - 0.0148

a" " - 0.0134

ь" " - 0.0111

" "o - 0.0102

Перехресні без пробілів:

TO - 0.0178

ак — 0.0166

но — 0.0141

не — 0.0123

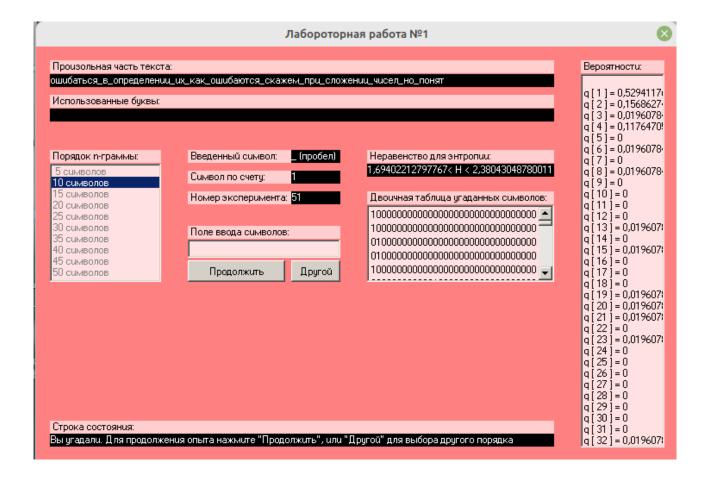
ка — 0.0138

тс — 0.0118

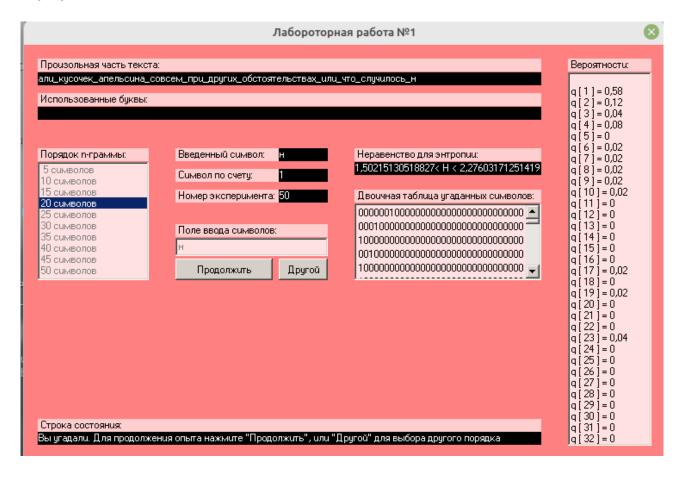
на — 0.0112 ел — 0.0102 нж — 0.0102 Не перехресні без пробілів: ак — 0.0163 ка — 0.0145 но — 0.0141 ов — 0.0133 не — 0.0129 ст — 0.0117 ко — 0.0111 ел — 0.0109 ни — 0.0100

Завдання 2

H(10)



H(20)



H(30)Лабороторная работа №1 Вероятности: Произольная часть текста: й_давай_ть|_же_обещал_каждый_день_люди_произносят_подобное_как_образованнь q[1] = 0.62Использованные буквы: q[2]=0,1 q[3]=0,08 q[4]=0,02 q[5]=0,02 q[6]=0,02 q[7]=0 q[8]=0,02 q[10]=0,02 q[11]=0 q[12]=0 q[14]=0,02 q[15]=0 q[14]=0,02 q[15]=0 q[16]=0 q[17]=0,02 q[18]=0,02 q[19]=0 q[20]=0 q[21]=0 q[21]=0 q[21]=0 q[21]=0 q[22]=0 q[22]=0 q[23]=0 q[24]=0 q[25]=0 q[27]=0,02 Порядок п-граммы: Введенный символ: Неравенство для энтропии: 1,35537332981934< H < 2,18006166784613 5 символов Символ по счету: 10 символов 15 символов Номер эксперимента: 50 Двоичная таблица угаданных символов: 20 символов 30 символов Поле ввода символов:

Другой

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

Продолжить

q[27]=0, q[28]=0 q[29]=0 q[30]=0 q[31]=0

q[32]=0

Висновок:

Строка состояния:

35 символов

40 символов

45 символов

50 символов

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.