Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Фізико-технічний інститут

Лабораторна робота №1

З дисципліни «Криптографія»

Виконали:

студенти 3 курсу ФТІ
групи ФБ-83
Волинко Д.В.
Бондаренко.Р.С.

Перевірив:

Чорний О. М.

Вариант - 5

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.

Хід роботи:

Для виконання цієї лабораторної роботи, в якості експериментального тексту, було використано «Волхв». Перш за все, була функція, що очищала текст від непотрібних символів після цього була створена програма для обробки тексту та виконання завдання. Так як по завданню необхідно працювати як з пробілом, так і без нього, то робота програми відбувається в два етапи, які відрізняються один від одного лише алфавітом. Труднощі виникали під час роботи з підрахунком єнтропії біграм, адже її потрібно було порахувати для двух різних випадків випадків.

Результати:

Частоти букв з у тексті пробілами

" "- 0.1601277868756229

о - 0.09062653445798124

е - 0.07050044432333027

а - 0.06745230915306061

н - 0.058653250415830076

и - 0.05515093710661104

т - 0.049749645012652714

с - 0.046812055087523556

л - 0.04333630691473363

в - 0.03704551122034238

р - 0.036641035594712

к - 0.02838270592479667

м - 0.027123861234010574

у - 0.024581320595778885

д - 0.024499054366837112

п - 0.02424797098058774

я - 0.021299240836956047

ь - 0.01631356458713065

ы - 0.015937367977698996

з - 0.014747078477697712

ч - 0.013589352693319212

б - 0.013369119142923006

г - 0.013265429416860978

й - 0.009778541025568516

ж - 0.00817092180166469

х - 0.006858089898135556

ш - 0.006656709025205173

ю - 0.005427857230387431

ц - 0.0032023843495354956

щ - 0.0027490631504709312

э - 0.0020575126634291476

ф - 0.0016470384586050903

Частоти букв у тексті без пробілів:

о - 0.10790514680899475

е - 0.08394187022935813

а - 0.08031258577079696

н - 0.0698359220596623

и - 0.06566586707452328

т - 0.059234779095239894

с - 0.05573711614220428

л - 0.05159869113126133

в - 0.04410850917728396

р - 0.0436269173121052

к - 0.03379407662412265

м - 0.03229522397592857

у - 0.029267929348840864

д - 0.029169978461007894

п - 0.028871024188767688

я - 0.025360097053004696

ь - 0.019423865121627453

ы - 0.018975943874141272

з - 0.017558716965807998

ч - 0.01618026228390861

б - 0.015918039594605762

г - 0.015794580663066292

й - 0.011642891469395959

ж - 0.009728767869660019

х - 0.008165634951325551

ш - 0.00792585934048443

ю - 0.0064627179534794505

ц - 0.00381294237324798

щ - 0.003273192168418389

э - 0.0024497925175724915

ф - 0.0019610584001559053

Энтропия для H(1) - 4.38271934838655, Избыточность - 0.1234561303226901

Энтропия для H(1) тексте без пробелов - 4.462707678657537, Избыточность - 0.09920653138005309

Таблички з частотами біграмм буде надано окремими pdf-файлами, так як

Текст при фільтрації його від небажаних символів було фактично перетворено у рядок тексту, де символи нової стрічки було замінено на пробіл, через що можуть зустрічатись біграми типу "аа", "яя", тощо

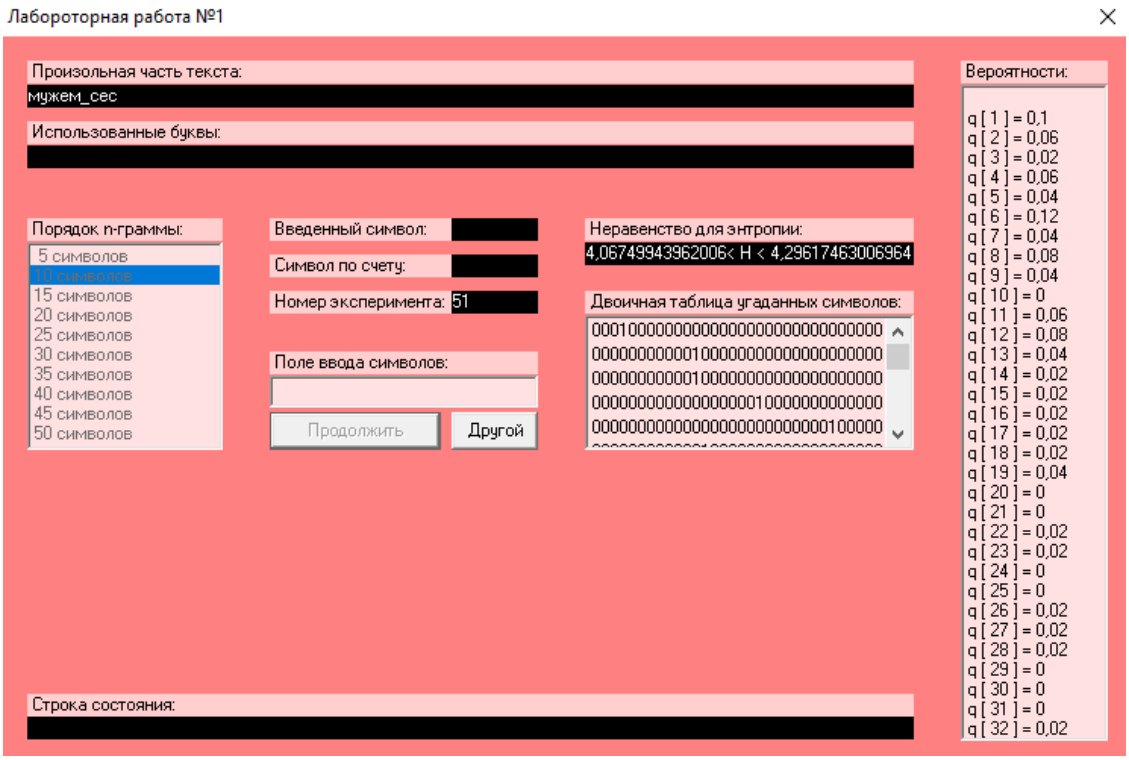
Энтропия пересекающихся биграмм - 3.9943195031842884, Избыточность - 0.20113609936314236

Энтропия непересекающихся биграмм - 3.99469779069679, Избыточность - 0.20106044186064198

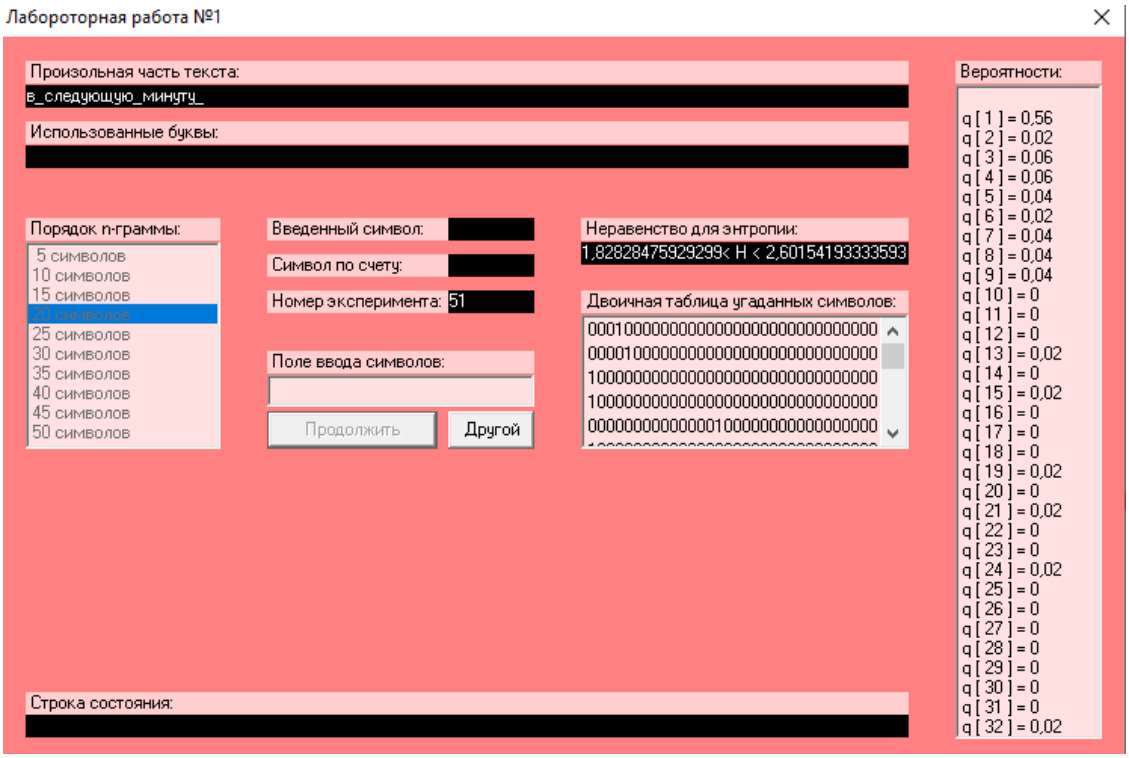
Энтропия пересекающихся биграмм в тексте без пробелов - 4.161299463838363, Избыточность - 0.1600455042296447

Энтропия непересекающихся биграмм в тексте без пробелов - 4.160055110751561, Избыточность - 0.1602966757636014

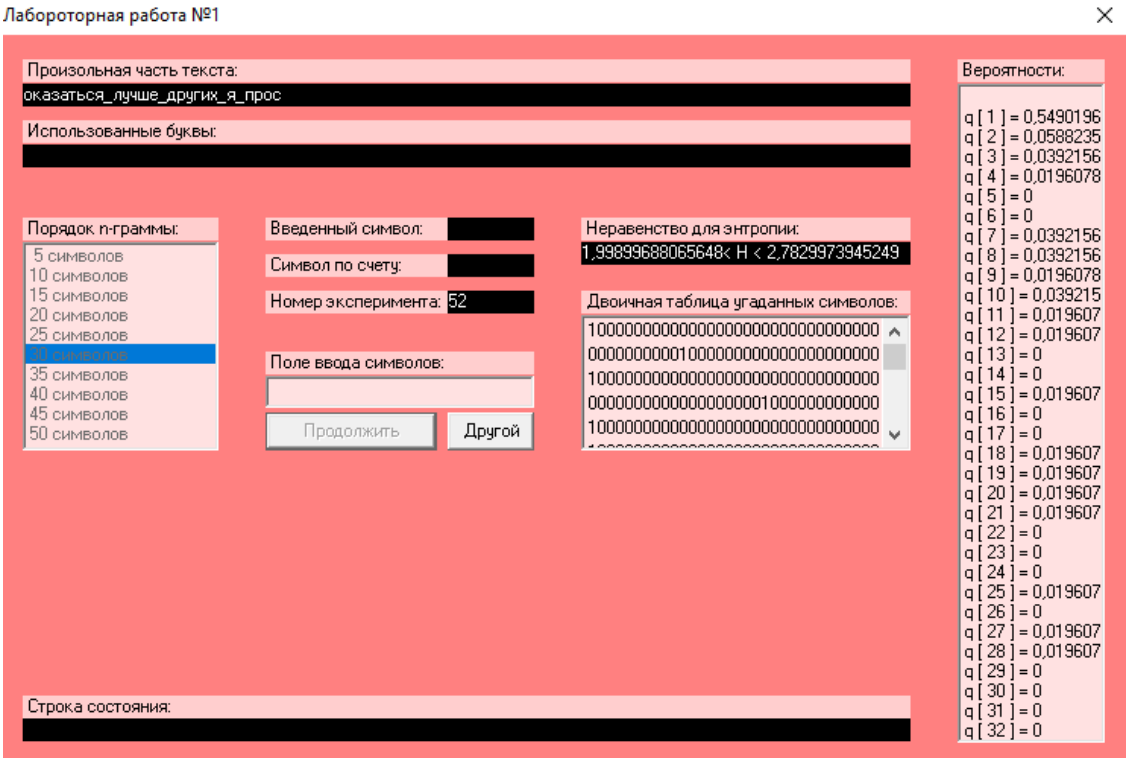
Знайдемо значення для (10)H, (20)H, (30)H



$$0.18650011207598793 < R(10) < 0.14076507398607208$$



$$0.634343048141402 < R(20) < 0.47969161333281396$$



$$0.600200623868704 < R(30) < 0.44340052109502004$$

Висновок: у цій лабораторній роботі ми отримали базові навички роботи в великими текстами на прикладі пошуку ентропій, надлишковості та частот, що знадобляться нам у наступних роботах.