Міністерство освіти і науки України
Національний технічний університет України
“Київський політехнічний інститут імені Ігоря Сікорського”
Фізико-Технічний інститут

КРИПТОГРАФІЯ
КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1
Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконав:
студент групи ФБ-93
Килимчук Денис

Перевірила:
Селюх П.В.

Київ 2021

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

-Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.

-Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1 та H2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H1 та H2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H1 та H2 на тому ж тексті, в якому вилучено всі пробіли.

-За допомогою програми CoolPinkProgram оцінити значення H(10), H(20), H(30).

-Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи

1)

частоти літер

З пробілами	Без пробілів
и: 0.056596685199261976 с: 0.04387783584696652 т: 0.049837876364708836 о: 0.09730226973089408 р: 0.03480172112628088 я: 0.023343345733260537 “ ”: 0.16895617658573067 ж: 0.008352834398650345 з: 0.013296420289066354 н: 0.052120071555596906 б: 0.017297284012169366 а: 0.05950736183502538 е: 0.07128700008953227 м: 0.02972120352441159 в: 0.03468761136673648 п: 0.0240069378733803	и: 0.0681031295879675 с: 0.052798462139696016 т: 0.05997021451883773 о: 0.11708440275885379 р: 0.0418771190468646 я: 0.028089187447849003 ж: 0.010051015558160905 з: 0.0159996620087244 н: 0.06271639363308934 б: 0.02081392524055473 а: 0.07160556418137456 е: 0.0857800733018579 м: 0.03576370184944601 в: 0.04173981009115202 п: 0.02888769183645447 у: 0.027227309695068497

у: 0.022627087550274125 х: 0.007910439638570504 д: 0.027151100632519173 ч: 0.012320342961271148 л: 0.04190637030899167 к: 0.027500452050201272 ы: 0.015490838741843347 й: 0.00853541001392139 ь: 0.017973164895624682 г: 0.015071265933672386 ю: 0.004434480809371747 э: 0.002310283747083618 ф: 0.0003440848133954324 ш: 0.006518300572128779 ц: 0.002373482998523595 щ: 0.002540258800934647	х: 0.009518679299090593 д: 0.032671081677704196 ч: 0.01482514232601371 л: 0.05042618587407713 к: 0.033091458326731944 ы: 0.01864021884935095 й: 0.010270709887301034 ь: 0.021627216747467707 г: 0.018135344381423155 ю: 0.005336037263538135 э: 0.0027799782418116334 ф: 0.00041403931261024324 ш: 0.007843510039397109 ц: 0.0028560262788216778 щ: 0.0030567085987092957

частоти 10 найчастіших біграм

з пробілами (з перетином)	з пробілами (без перетину)	без пробілів (з перетином)	без пробілів (без перетину)
ис: 0.002568351865 ст: 0.011449617819 то: 0.012861070246 ор: 0.005656342933 ри: 0.00432740078 ия: 0.00146411856 я : 0.016545944180 ж: 0.001759048919 жи: 0.00169058294 из: 0.00219968891	ис: 0.002461255415 то: 0.01287858040 ри: 0.0043466964 я : 0.016540619492 жи: 0.00167126615 зн: 0.001369314710 и : 0.019423202511 ро: 0.006534088914 би: 0.000846166269 нз: 3.15995702458	ис: 0.0051649400 ст: 0.014060466766 то: 0.01574408936 ор: 0.00739568722 ри: 0.00528323728 ия: 0.00265957446 яж: 0.00038657833 жи: 0.00204907643 из: 0.003259510249 зн: 0.001939229040	ис: 0.00533180110 то: 0.015674312294 ри: 0.005238853705 яж: 0.000354890089 из: 0.003172886397 ни: 0.01024111401 ро: 0.00797657725 би: 0.0010308712 нз: 0.00013942110 он: 0.00714005061

З повними таблицями біграм та текстом можна ознайомитись у відповідних файлах на git’і.

Текст з пробілами:

Літери

Ентропія: 4.358254800801008

Надлишковість:0.8638045374749685

Біграми

Ентропія біграм з перетином: 3.923394858478649

Ентропія біграм без перетину: 3.9236898792447956

Текст без пробілів:

Літери

Ентропія: 4.455773959254753

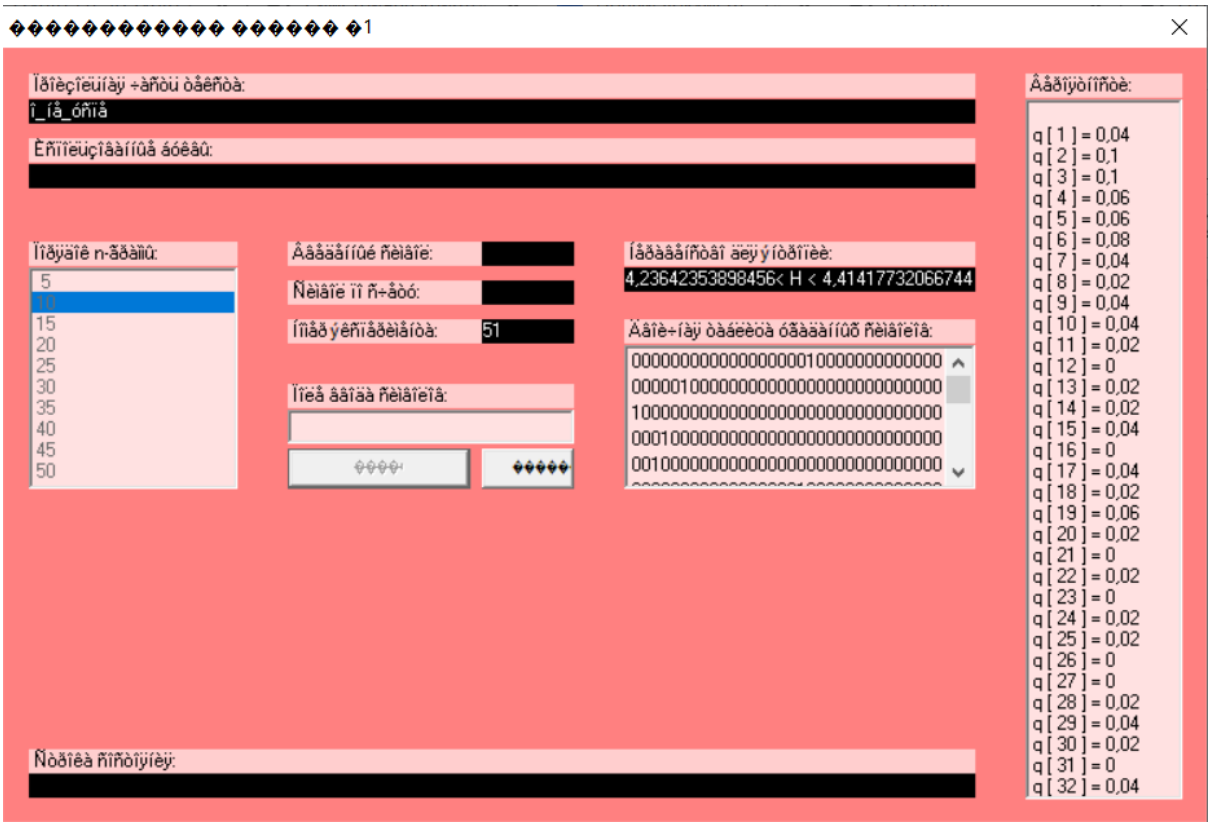
Надлишковість:0.8562653561530725

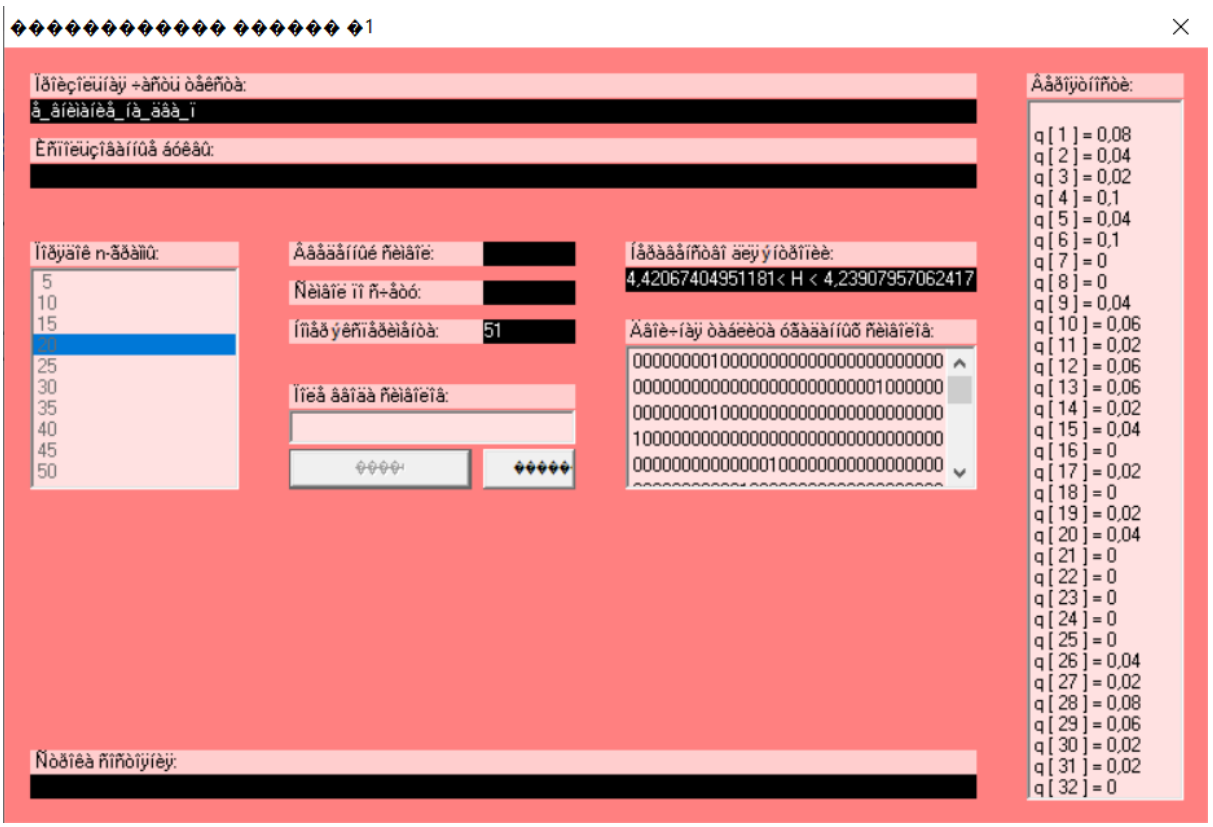
Біграми:

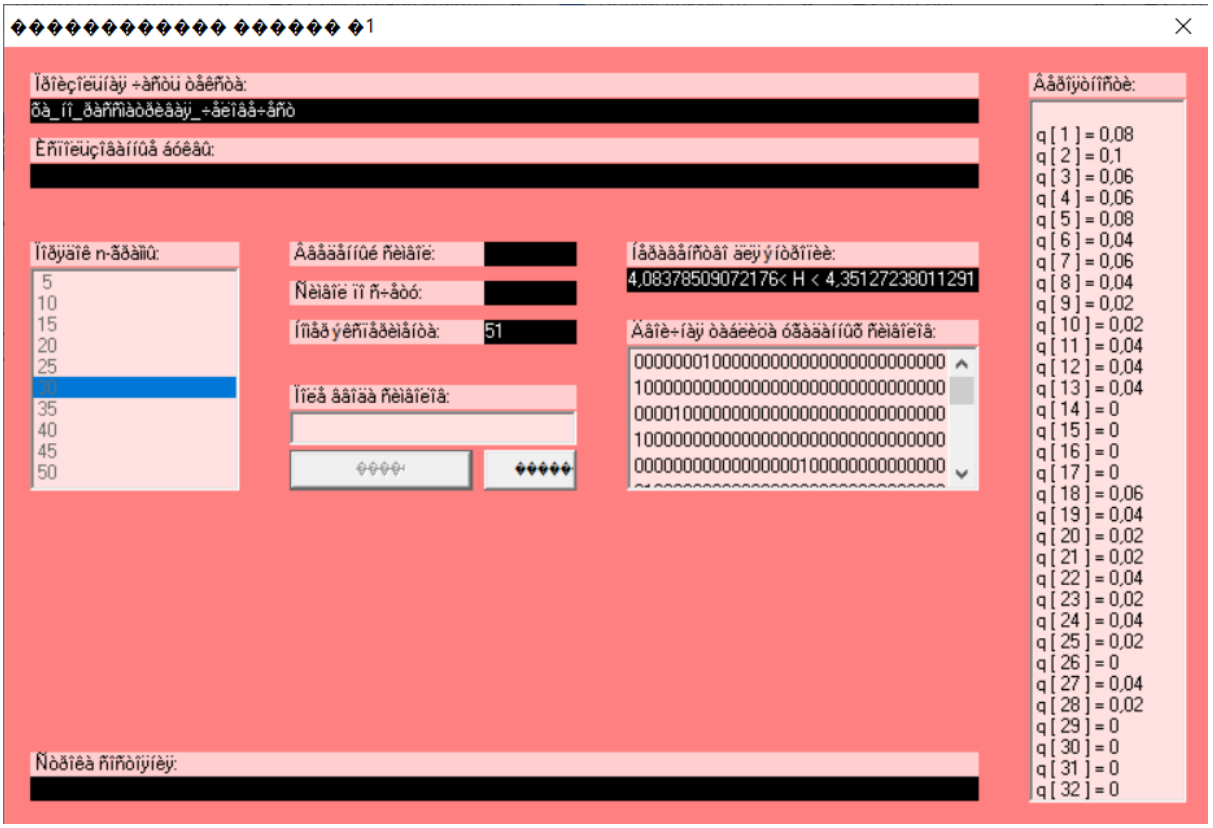
Ентропія біграм з перетином: 4.117710766861664

Ентропія біграм без перетину: 4.116049129335191

2) CoolPinkProgram







11%<R(10)<15%

12%<R(20)<15%

13%<R(30)<18%

Висновки: під час виконання ЛП1 засвоїли поняття ентропії та надлишковості, розробили програму для підрахунку, а також ознайомились та навчились працювати з git.