СИМЕТРИЧНА КРИПТОГРАФІЯ
КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

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. А також успішне закриття предмету і успіх у житті.

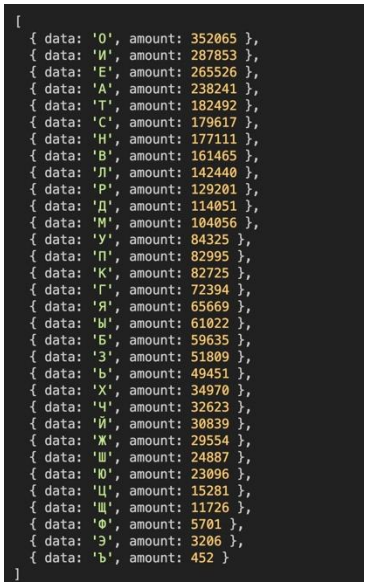
Порядок виконання роботи

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. Підрахуватичастоти букв та біграм, а також 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

Хід роботи

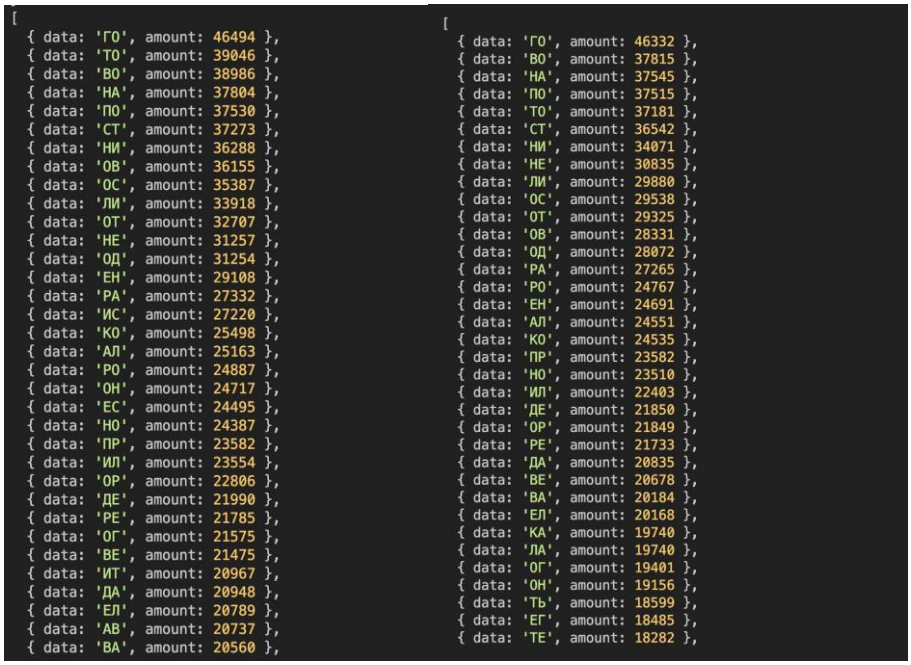
Найбільші труднощі виникли з пошуком файлу через обмеження в його розмірі в 1Мб. Але рішення досить швидко було знайдено, було використано повне зібрання (Старий + Новий заповіт) розмірами 7.5Мб. Наступним випробуванням був запуск програми CoolPinkProgram.exe, через її недоступність на пристроях mac. При написанні програмної частини лабораторної роботи проблем не виникало. Мова виконання: typescript.

Таблиця частот входжень літер руснявого алфавіту в тексті:



```
[
  { data: 'О', amount: 352065 },
  { data: 'И', amount: 287853 },
  { data: 'Е', amount: 265526 },
  { data: 'А', amount: 238241 },
  { data: 'Т', amount: 182492 },
  { data: 'С', amount: 179617 },
  { data: 'Н', amount: 177111 },
  { data: 'В', amount: 161465 },
  { data: 'Л', amount: 142440 },
  { data: 'Р', amount: 129201 },
  { data: 'Д', amount: 114051 },
  { data: 'М', amount: 104056 },
  { data: 'У', amount: 84325 },
  { data: 'П', amount: 82995 },
  { data: 'К', amount: 82725 },
  { data: 'Г', amount: 72394 },
  { data: 'Я', amount: 65669 },
  { data: 'Ы', amount: 61022 },
  { data: 'Б', amount: 59635 },
  { data: 'З', amount: 51809 },
  { data: 'Ь', amount: 49451 },
  { data: 'Х', amount: 34970 },
  { data: 'Ч', amount: 32623 },
  { data: 'Й', amount: 30839 },
  { data: 'Ж', amount: 29554 },
  { data: 'Ш', amount: 24887 },
  { data: 'Ю', amount: 23096 },
  { data: 'Ц', amount: 15281 },
  { data: 'Щ', amount: 11726 },
  { data: 'Ф', amount: 5701 },
  { data: 'Э', amount: 3206 },
  { data: 'ъ', amount: 452 }
]
```

Таблиця частот біграм з перетином та без:



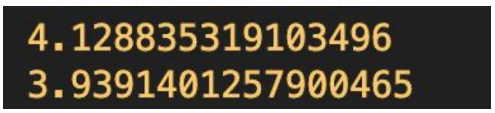
```
[
  { data: 'ГО', amount: 46494 },
  { data: 'ТО', amount: 39046 },
  { data: 'ВО', amount: 38986 },
  { data: 'НА', amount: 37804 },
  { data: 'ПО', amount: 37530 },
  { data: 'СТ', amount: 37273 },
  { data: 'НИ', amount: 36288 },
  { data: 'ОВ', amount: 36155 },
  { data: 'ОС', amount: 35387 },
  { data: 'ЛИ', amount: 33918 },
  { data: 'ОТ', amount: 32707 },
  { data: 'НЕ', amount: 31257 },
  { data: 'ОД', amount: 31254 },
  { data: 'ЕН', amount: 29108 },
  { data: 'РА', amount: 27332 },
  { data: 'ИС', amount: 27220 },
  { data: 'КО', amount: 25498 },
  { data: 'АЛ', amount: 25163 },
  { data: 'РО', amount: 24887 },
  { data: 'ОН', amount: 24717 },
  { data: 'ЕС', amount: 24495 },
  { data: 'НО', amount: 24387 },
  { data: 'ПР', amount: 23582 },
  { data: 'ИЛ', amount: 23554 },
  { data: 'ОР', amount: 22806 },
  { data: 'ДЕ', amount: 21990 },
  { data: 'РЕ', amount: 21785 },
  { data: 'ОГ', amount: 21575 },
  { data: 'ВЕ', amount: 21475 },
  { data: 'ИТ', amount: 20967 },
  { data: 'ДА', amount: 20948 },
  { data: 'ЕЛ', amount: 20789 },
  { data: 'АВ', amount: 20737 },
  { data: 'ВА', amount: 20560 },
```

```
[
  { data: 'ГО', amount: 46332 },
  { data: 'ВО', amount: 37815 },
  { data: 'НА', amount: 37545 },
  { data: 'ПО', amount: 37515 },
  { data: 'ТО', amount: 37181 },
  { data: 'СТ', amount: 36542 },
  { data: 'НИ', amount: 34071 },
  { data: 'НЕ', amount: 30835 },
  { data: 'ЛИ', amount: 29880 },
  { data: 'ОС', amount: 29538 },
  { data: 'ОТ', amount: 29325 },
  { data: 'ОВ', amount: 28331 },
  { data: 'ОД', amount: 28072 },
  { data: 'РА', amount: 27265 },
  { data: 'РО', amount: 24767 },
  { data: 'ЕН', amount: 24691 },
  { data: 'АЛ', amount: 24551 },
  { data: 'КО', amount: 24535 },
  { data: 'ПР', amount: 23582 },
  { data: 'НО', amount: 23510 },
  { data: 'ИЛ', amount: 22403 },
  { data: 'ДЕ', amount: 21850 },
  { data: 'ОР', amount: 21849 },
  { data: 'РЕ', amount: 21733 },
  { data: 'ДА', amount: 20835 },
  { data: 'ВЕ', amount: 20678 },
  { data: 'ВА', amount: 20184 },
  { data: 'ЕЛ', amount: 20168 },
  { data: 'КА', amount: 19740 },
  { data: 'ЛА', amount: 19740 },
  { data: 'ОГ', amount: 19401 },
  { data: 'ОН', amount: 19156 },
  { data: 'ТЬ', amount: 18599 },
  { data: 'ЕГ', amount: 18485 },
  { data: 'ТЕ', amount: 18282 },
```

Ентропія H_1

```
4.438316205882865
```

Ентропія H_2 (з перетином та без)



```
4.128835319103496
3.9391401257900465
```

Ймовірності літер:

```
[
  { data: 'О', frequency: 0.11153728934591021 },
  { data: 'И', frequency: 0.09119436283097807 },
  { data: 'Е', frequency: 0.08412097280576643 },
  { data: 'А', frequency: 0.0754768447617883 },
  { data: 'Т', frequency: 0.057815071101398455 },
  { data: 'С', frequency: 0.05690424580814439 },
  { data: 'Н', frequency: 0.05611032296122451 },
  { data: 'В', frequency: 0.05115353251313648 },
  { data: 'Л', frequency: 0.04512624513777698 },
  { data: 'Р', frequency: 0.04093201346564113 },
  { data: 'Д', frequency: 0.03613236018118929 },
  { data: 'М', frequency: 0.032965856248641685 },
  { data: 'У', frequency: 0.026714901862138752 },
  { data: 'П', frequency: 0.02629354616125948 },
  { data: 'К', frequency: 0.026208007785893012 },
  { data: 'Г', frequency: 0.0229350560973338 },
  { data: 'Я', frequency: 0.020804516933113425 },
  { data: 'Ы', frequency: 0.019332306450417204 },
  { data: 'Б', frequency: 0.01889289264807168 },
  { data: 'З', frequency: 0.016413546997634705 },
  { data: 'Ь', frequency: 0.015666511852767545 },
  { data: 'Х', frequency: 0.01107880365394595 },
  { data: 'Ч', frequency: 0.010335253405852979 },
  { data: 'Й', frequency: 0.00977006651083898 },
  { data: 'Ж', frequency: 0.00936296720585412 },
  { data: 'Ш', frequency: 0.007884420547204828 },
  { data: 'Ю', frequency: 0.007317015990607253 },
  { data: 'Ц', frequency: 0.004841155236944468 },
  { data: 'Щ', frequency: 0.003714899961285965 },
  { data: 'Ф', frequency: 0.0018061269554231012 },
  { data: 'Э', frequency: 0.0010156890052774009 },
  { data: 'Ъ', frequency: 0.0001431975765394214 7 }
]
```

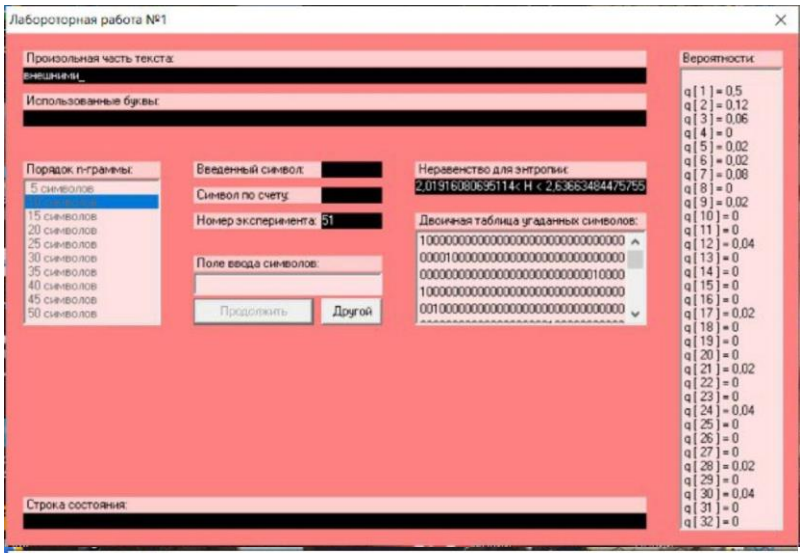
Ймовірності біграм

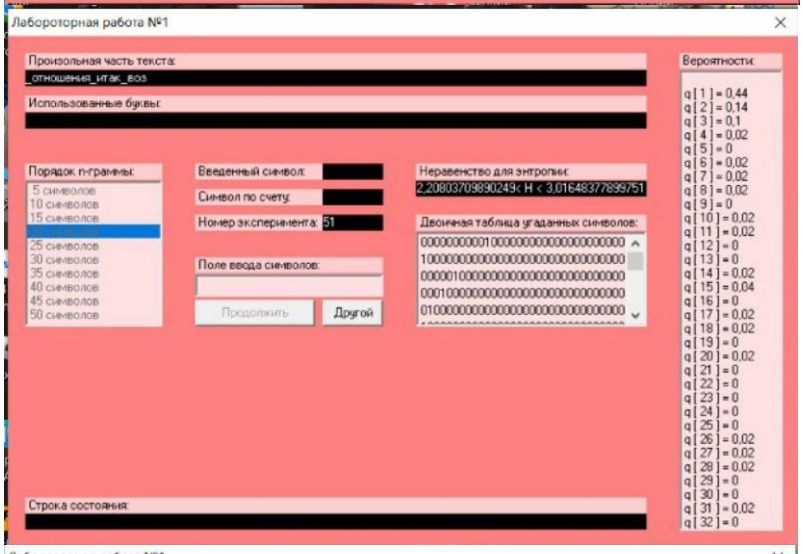
```
[
  { data: 'ГО', frequency: 0.01473426415194793 },
  { data: 'ТО', frequency: 0.012373942402825288 },
  { data: 'ВО', frequency: 0.012354927995608939 },
  { data: 'НА', frequency: 0.011980344173446887 },
  { data: 'ПО', frequency: 0.011893511713825565 },
  { data: 'СТ', frequency: 0.011812066669582209 },
  { data: 'НИ', frequency: 0.011499913484447165 },
  { data: 'ОВ', frequency: 0.01145776488178426 },
  { data: 'ОС', frequency: 0.01121438046941501 },
  { data: 'ЛИ', frequency: 0.010748844399401427 },
  { data: 'ОТ', frequency: 0.010365070280418139 },
  { data: 'НЕ', frequency: 0.0099055554393564 },
  { data: 'ОД', frequency: 0.009904604718995583 },
  { data: 'ЕН', frequency: 0.00922452275422421 },
  { data: 'РА', frequency: 0.008661696300620313 },
  { data: 'ИС', frequency: 0.008626202740483131 },
  { data: 'КО', frequency: 0.008080489253373948 },
  { data: 'АЛ', frequency: 0.00797432547974934 },
  { data: 'РО', frequency: 0.0078685920655414 },
  { data: 'ОН', frequency: 0.007832985052774487 },
  { data: 'ЕС', frequency: 0.007762631746074 },
  { data: 'НО', frequency: 0.007728405813084574 },
  { data: 'ПР', frequency: 0.007473295849598574 },
  { data: 'ИЛ', frequency: 0.007464422459564278 },
  { data: 'ОР', frequency: 0.007227376182933809 },
  { data: 'ДЕ', frequency: 0.006968780244791479 },
  { data: 'РЕ', frequency: 0.006903814353468957 },
  { data: 'ОГ', frequency: 0.00683726392821174 },
  { data: 'ВЕ', frequency: 0.006805573249517826 },
  { data: 'ИТ', frequency: 0.006644584601752748 },
  { data: 'ДА', frequency: 0.006638563372800905 },
  { data: 'ЕЛ', frequency: 0.006588175193677583 },
  { data: 'АВ', frequency: 0.006571696040756748 },
  { data: 'ВА', frequency: 0.006515603539468522 },
  { data: 'МО', frequency: 0.004618282606063948 },
  { data: 'ИИ', frequency: 0.004577718537335739 },
  { data: 'СК', frequency: 0.00458071018810953 },
  { data: 'БО', frequency: 0.004488350823418905 },
  { data: 'АР', frequency: 0.004455392517577235 },
  { data: 'ИЗ', frequency: 0.004445568407182122 },
  { data: 'ДО', frequency: 0.004429406161048226 },
  { data: 'ИХ', frequency: 0.004357785227199983 },
  { data: 'СП', frequency: 0.00435746832041304 4 },
  { data: 'ИП', frequency: 0.004310883022732991 },
  { data: 'ЕД', frequency: 0.004272537301513357 },
  { data: 'ЕВ', frequency: 0.004230705605637391 },
  { data: 'ВС', frequency: 0.00415845085821527 },
  { data: 'СЯ', frequency: 0.004107428865518069 },
  { data: 'СЛ', frequency: 0.004102358356927044 },
  { data: 'АИ', frequency: 0.004096020221188261 },
  { data: 'АЗ', frequency: 0.004045315135278000 5 },
  { data: 'УД', frequency: 0.003974328015003635 },
  { data: 'ИК', frequency: 0.003945806404179113 },
  { data: 'ЕБ', frequency: 0.003793691146448330 5 },
  { data: 'ЖЕ', frequency: 0.003600378006415461 },
  { data: 'ИД', frequency: 0.003507524317842295 6 },
  { data: 'СВ', frequency: 0.003489143724199826 },
  { data: 'МУ', frequency: 0.003473932198426748 },
  { data: 'ДИ', frequency: 0.0034694955034096 },
  { data: 'НЫ', frequency: 0.003324986008565356 6 },
  { data: 'ИО', frequency: 0.003311042109940035 },
  { data: 'БЫ', frequency: 0.003306605414922887 },
  { data: 'ОЕ', frequency: 0.003289492448428174 },
  { data: 'АД', frequency: 0.003207096683824 },
  { data: 'СО', frequency: 0.003150370368961895 8 },
  { data: 'ЧТ', frequency: 0.003109489393446748 },
  { data: 'УС', frequency: 0.003081284689409165 3 },
  { data: 'ОК', frequency: 0.003035016298516052 3 },
  ... 783 more items
]
```

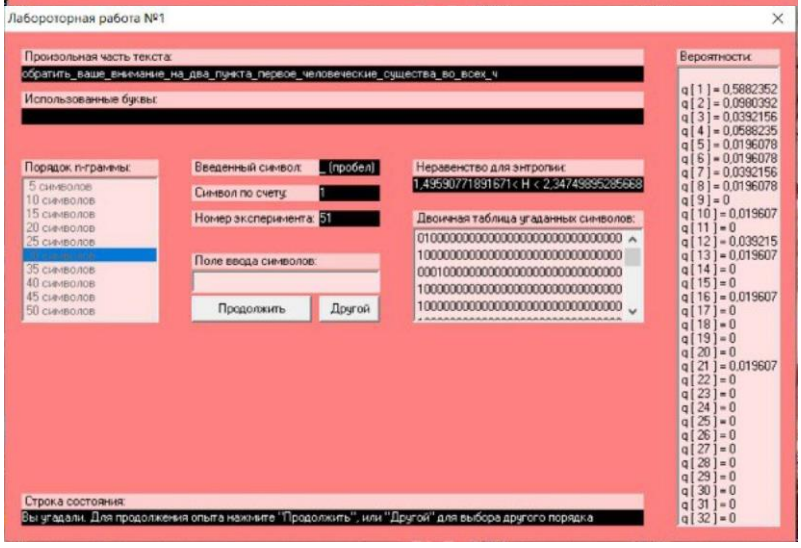
Без перетину:

```
[
  { data: 'ГО', frequency: 0.018626528692380056 },
  { data: 'ВО', frequency: 0.015202498974841402 },
  { data: 'НА', frequency: 0.01509395277034035 3 },
  { data: 'ПО', frequency: 0.015081892080951348 },
  { data: 'ТО', frequency: 0.014947616405753752 },
  { data: 'СТ', frequency: 0.014690723721767936 },
  { data: 'НИ', frequency: 0.013697324939093519 },
  { data: 'НЕ', frequency: 0.012396378576999462 },
  { data: 'ЛИ', frequency: 0.012012446631449453 },
  { data: 'ОС', frequency: 0.011874954772414792 },
  { data: 'ОТ', frequency: 0.011789323877752852 },
  { data: 'ОВ', frequency: 0.011389713035997137 },
  { data: 'ОД', frequency: 0.011285589084272057 },
  { data: 'РА', frequency: 0.01096115653970781 },
  { data: 'РО', frequency: 0.009956903136583288 },
  { data: 'ЕН', frequency: 0.00992634939013114 },
  { data: 'АЛ', frequency: 0.009870066172982448 },
  { data: 'КО', frequency: 0.009863633805308311 },
  { data: 'ПР', frequency: 0.00948050590571757 },
  { data: 'НО', frequency: 0.009451560251183959 },
  { data: 'ИЛ', frequency: 0.009006520812729655 },
  { data: 'ДЕ', frequency: 0.008784202104992322 },
  { data: 'ОР', frequency: 0.008783800082012688 },
  { data: 'РЕ', frequency: 0.0087371654163752 },
  { data: 'ДА', frequency: 0.008376148780664302 },
  { data: 'ВЕ', frequency: 0.008313031172861841 },
  { data: 'ВА', frequency: 0.008114431820922884 },
  { data: 'ЕЛ', frequency: 0.008107999453248747 },
  { data: 'КА', frequency: 0.007935933617965603 },
  { data: 'ЛА', frequency: 0.007935933617965603 },
  { data: 'ОГ', frequency: 0.007799647827869841 },
  { data: 'ОН', frequency: 0.00770115219785963 },
  { data: 'ТЬ', frequency: 0.007477225398203761 },
  { data: 'ЕГ', frequency: 0.007431394778525541 },
  { data: 'ТЕ', frequency: 0.007349784113659937 },
  { data: 'ВИ', frequency: 0.004846789042461667 },
  { data: 'УД', frequency: 0.004680351528893392 },
  { data: 'АТ', frequency: 0.004678341413995224 },
  { data: 'ИЗ', frequency: 0.004664270609708051 },
  { data: 'АН', frequency: 0.004615223806192762 },
  { data: 'МИ', frequency: 0.00460356513978339 },
  { data: 'АЗ', frequency: 0.004571001278433075 },
  { data: 'ЖЕ', frequency: 0.004554116313288468 },
  { data: 'ИЕ', frequency: 0.00446285709691166 },
  { data: 'АС', frequency: 0.004424262890866842 },
  { data: 'НЫ', frequency: 0.004218025102314849 },
  { data: 'МУ', frequency: 0.004200738114190607 },
  { data: 'БЫ', frequency: 0.004194707769496105 },
  { data: 'ЕД', frequency: 0.004175812689453329 },
  { data: 'СВ', frequency: 0.004042341060215002 },
  { data: 'ВС', frequency: 0.004028270255927829 },
  { data: 'ИН', frequency: 0.004018621704416625 },
  { data: 'ДИ', frequency: 0.004004952923109085 },
  { data: 'ЧТ', frequency: 0.003928568556978717 },
  { data: 'СО', frequency: 0.003743235963367666 },
  { data: 'ЕЙ', frequency: 0.003684138585361539 },
  { data: 'ОЕ', frequency: 0.003682530493443005 },
  { data: 'ВЫ', frequency: 0.003644338310377821 },
  { data: 'ЕБ', frequency: 0.003524937485426667 },
  { data: 'ДУ', frequency: 0.003495991830893054 },
  { data: 'ОЙ', frequency: 0.003471468429135409 },
  { data: 'РУ', frequency: 0.003428853993294256 7 },
  { data: 'ЧЕ', frequency: 0.00339870226982174 3 },
  { data: 'ЦА', frequency: 0.003374982914023365 6 },
  { data: 'ОИ', frequency: 0.00337136470720666 4 },
  { data: 'ТР', frequency: 0.003191660435310482 3 },
  { data: 'АД', frequency: 0.003146633861591528 4 },
  { data: 'ТЫ', frequency: 0.00314381970073409 4 },
  { data: 'ЕВ', frequency: 0.003093566828279904 6 },
  ... 657 more items
]
```

Оцінки для значень $H^{(10)}$, $H^{(20)}$, $H^{(30)}$







Вибачте за якість, віндоус-ноутбук 2011 року народження)

Наочно побачили, як зі збільшенням кількості відображених символів у тексті (ми точно мали це отримати), ентропія буде знижуватись

Для знаходження оцінки надлишковості російської мови в різних моделях джерела, підставимо в формулу: $R = 1 - \frac{H_\infty}{H_0}$, на місце H_{oo} отримані вище значення H_1 , H_2 , $H^{(10)}$, $H^{(20)}$, $H^{(30)}$

R1=1-4.44/log(33)=0.12

R2=1-8.26/log(883)=0.57

R2*=1-7.8/log(757)=0.59

R10=1-2.4/log(34)=0.59

R20=1-2.6/log(34)=0.69

R30=1-1.6/log(34)= 0.59

(* - біграми без перетину)

Отримали значення надлишковості 0.12, 0.16, 0.14, 0.59, 0.69, 0.59

Висновок:

Під час виконання цього комп’ютерного практикуму, що віндоус-ноутбук є необхідною навичкою на 3 курсі, без якої виконання робіт унеможливлюється, або значно ускладнюється. Вивчили поняття ентропія, частота, pink, typescript, main, log та багато інших. Перечитали святе письмо. Перелічили символи з тексту, для перевірки показників програми. Визначили значення надлишковості для російської мови, зазначені вище. Втомилися