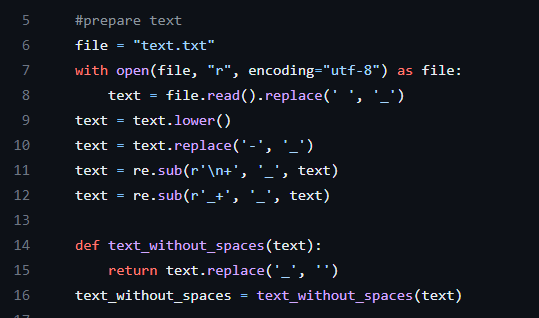
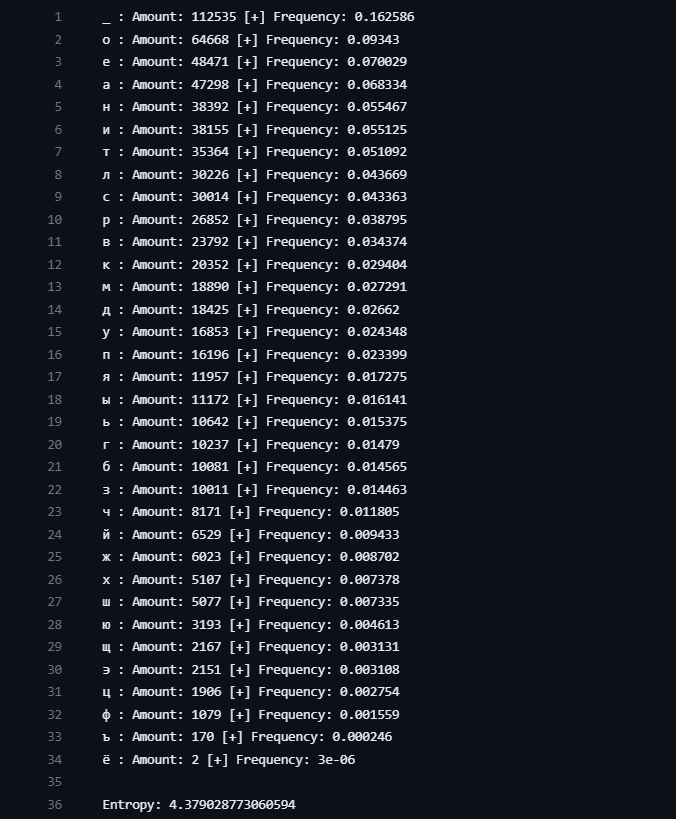
Тарасов Микита

Лабораторна №1

**Мета**: 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.

Код для виконання поставленої задачі міститься у файлі **tarasov\_lab1.py.** Даний код рахує к-сть заданих літер чи біграм у тексті, їх частоту, і на основі частот – ентропію тексту. Для початку підготовлюємо наш текст (залишаємо лише літери а-я та ‘\_’ замість пробілів):



Після цього рахуємо частоту літер, що трапляються в тексті, а далі їх частоти. Так само робимо із біграмами що перетинаються на тексті із пробілами та без, і потім із біграмами, що не перетинаються на тексті із пробілами та без. Результати виконання даного коду наведені у файлі **freqAndEntropy.txt** (нижче наведено лише частину відповідного файлу із результатами)****

Відповідно, частоти наведені у файлі, згаданому вище. Ентропія також обрахована у цьому файлі, але для подальшого обрахування надлишковості запишемо результати нижче:

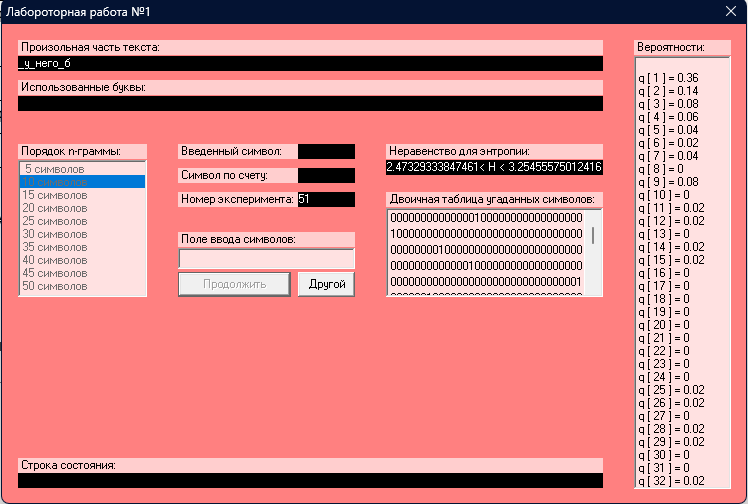
|  |  |  |
| --- | --- | --- |
|  | **Ентропія** | **Надлишковість** |
| **H1 (із пробілом)** | 4.379028773060594 | 0.1392504760606287 |
| **H1 (без пробілу)** | 4.464420005196533 | 0.1149732663024602 |
| **H2 (із пробілом, з перетином)** | 3.9903784772111153 | 0.2156442552450309 |
| **H1 (без пробілів, з перетином)** | 4.162145373626576 | 0.1748961968391469 |
| **H1 (із пробілом, без перетину)** | 3.9903267703185494 | 0.2156544188419076 |
| **H1 (без пробілів, без перетину)** | 4.160968545981837 | 0.1751294911809283 |

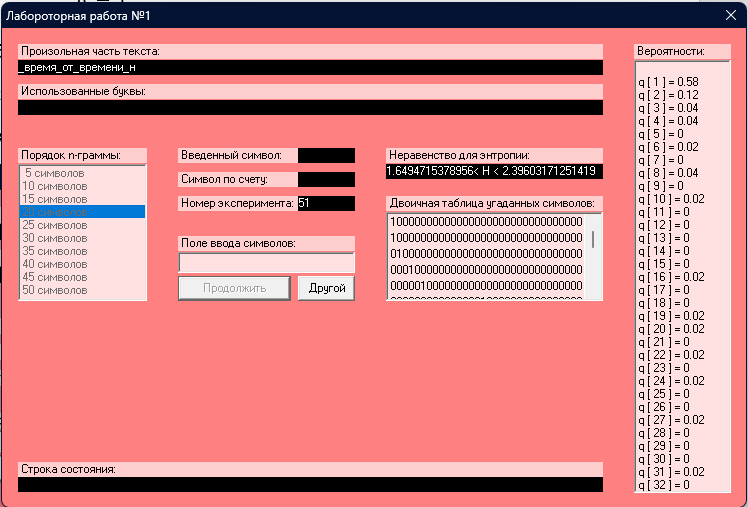
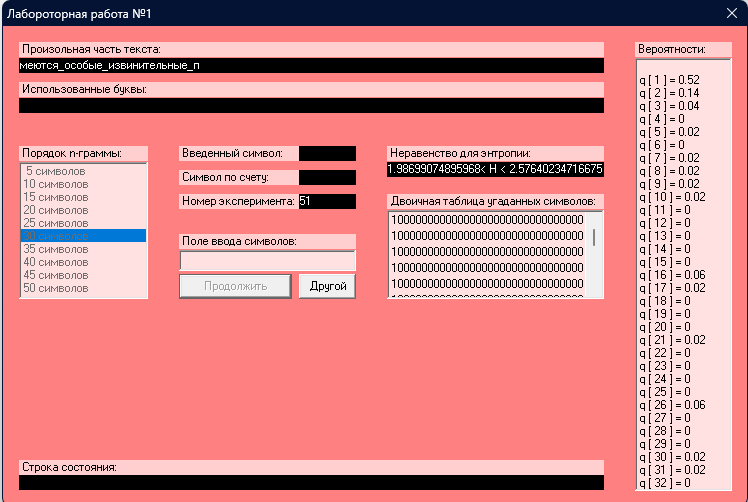
Надлишковість обчислюється за формулою

Де це 5.08746 для тексту із пробілами (log2(34)), 5.04439 для тексті без пробілів (log2(33)).

Переходимо до програми **CoolPinkProgram**

Результати виконання наведені нижче (**H10, H20 та H30 відповідно**):



Отримуємо наступне:

|  |  |  |
| --- | --- | --- |
|  | **Ентропія** | **Надлишковість** |
| **Н(10)** | 2.47329333847461 < H < 3.25455575012416 | 0.349088849975168 < R < 0.505341332305078 |
| **Н(20)** | 1.6494715378956 < H < 2.39603171251419 | 0.520793657497162 < R < 0.67010569242088 |
| **Н(30)** | 1.98699074895968 < H < 2.57640234716675 | 0.48471953056665 < R < 0.602601850208064 |

**Висновки:**Під час виконання даного практикуму, я навчився підготовлювати текст для подальшого його аналізу, а таком ознайомвися із поняттям ентропії та навчився обраховувати частоти літер та біграм, а також, користуючись цими значеннями, обраховувати ентропію та надлишковість у різних текстах.