Національний технічний університет України «КПІ» імені Ігоря Сікорського
Фізико-технічний інститут

Лабораторна робота 1

Криптографія

Виконали:

студенти ФБ-14

Кот Микита Сергійович

Чавалах Артем Дмитрович

Перевірила:

Селюх П. В.

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

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 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.

Частоти букв та біграм були збережені в файлах .xlsx формату.

Ентропія і надлишковість для букв та непересічних біграм у тексті без пробілів:

```
Letters and not cross bigrams without space:
Letters entropy: 4.458502430665178
Bigrams entropy: 4.128663354359083
Letter redundancy: 0.10005535684616174
Bigrams redundancy: 0.16663307311763065
```

Ентропія і надлишковість для букв та непересічних біграм у тексті з пробілами:

```
Letters and not cross bigrams with space:
Letters entropy: 4.391930259750691
Bigrams entropy: 3.971846934898919
Letter redundancy: 0.12161394804986192
Bigrams redundancy: 0.20563061302021624
```

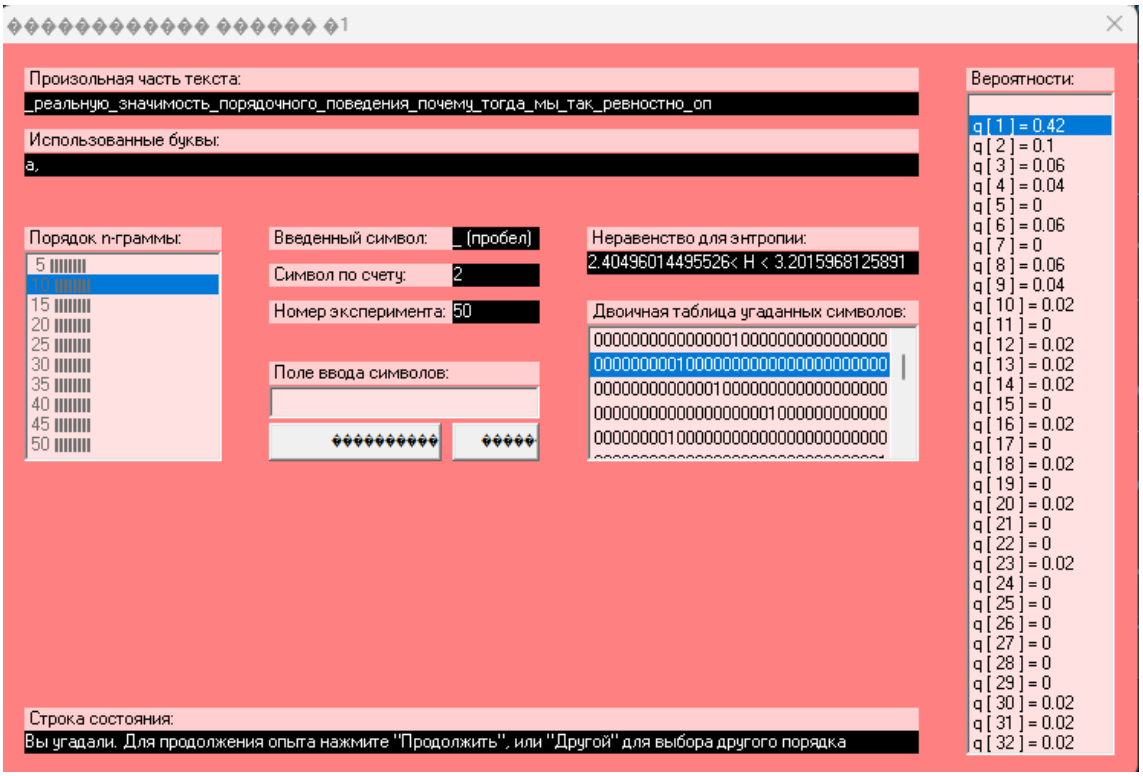
Ентропія і надлишковість для пересічних біграм у тексті без пробілів:

```
Cross bigrams without space:
Bigrams entropy: 4.129496081270868
Bigrams redundancy: 0.16646498795111453
```

Ентропія і надлишковість для пересічних біграм у тексті з пробілами:

```
Cross bigrams with space:
Bigrams entropy: 3.971768447179969
Bigrams redundancy: 0.20564631056400617
```

2. За допомогою програми CoolPinkProgram оцінити значення H^{10}, H^{20}, H^{30}.



$2.40496014495526 < H^{10} < 3.2015968125891$

$2.13655490598199 < H^{20} < 2.9479764393853$

$1.48465845168589 < H^{30} < 2.31092368847664$

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

Надлишковість джерела відкритого тексту (мови): $R = 1 - \frac{H_\infty}{H_0}$; $H_0 = \log_2 32 = 5$

R для H^{10}: $1 - \frac{2.40496014495526}{5} = 0.51900$; $1 - \frac{3.2015968125891}{5} = 0.35968$

$0.35968 < R < 0.51900$

2) R для H^{20}: $1 - \frac{2.13655490598199}{5} = 0.57268$; $1 - \frac{2.9479764393853}{5} = 0.41040$

$0.41040 < R < 0.57268$

3) R для H^{30}: $1 - \frac{1.48465845168589}{5} = 0.70306$; $1 - \frac{2.31092368847664}{5} = 0.53781$

$0.53781 < R < 0.70306$

Труднощі

У ході виконання лабораторної роботи зіткнулися із труднощами з обрахунками ентропії для біграм, вона була у два рази більша або у два рази менша від нормальних значень. Також спочатку було не зовсім зрозуміло поняття ентропії, але після виконання завдання з CoolPinkProgram, усе стало більш зрозуміло.

Висновки

У ході виконання лабораторної роботи, ми набули практичних навичків щодо оцінення ентропії, надлишковості. Покращили свої знання роботи з текстом. На реальному прикладі оцінили значення ентропії за допомогою програми CoolPinkProgram. Виявили, що буква «о» трапляється найбільше всього, приблизно один раз кожні десять символів.