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.*

ФБ-11 «Проект Нівечення»

Хід роботи:

1) Тут має бути таблиця і вона є. Однак я не придумала кращого способу її запису, а те, як вона виглядає наразі, це просто знущання. З нею, в разі чого, виз можете ознайомитися окремо.

Запишу значення, які ми отримали для H1 I H2X

Для тексту з пробілами:

H1 (Ентропія для одинарних символів): 4.435171349883117

H21 (Ентропія для біграм з зсувом 1): 4.017923178097787

H22 (Ентропія для біграм з зсувом 2): 4.016769024845158

Для тексту без пробілів:

H1 (Ентропія для одинарних символів): 4.473208149868866

H21 (Ентропія для біграм з зсувом 1): 4.1268090159753275

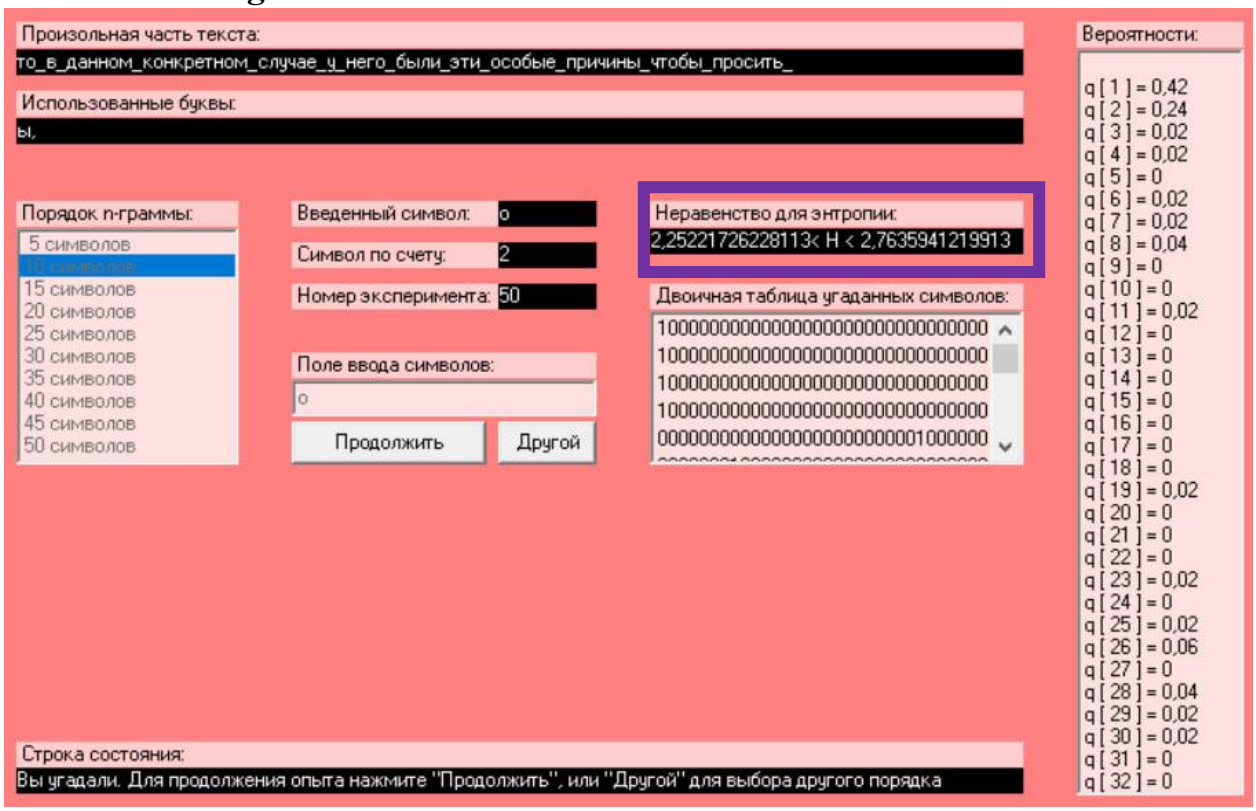
H22 (Ентропія для біграм з зсувом 2): 4.126810472799593

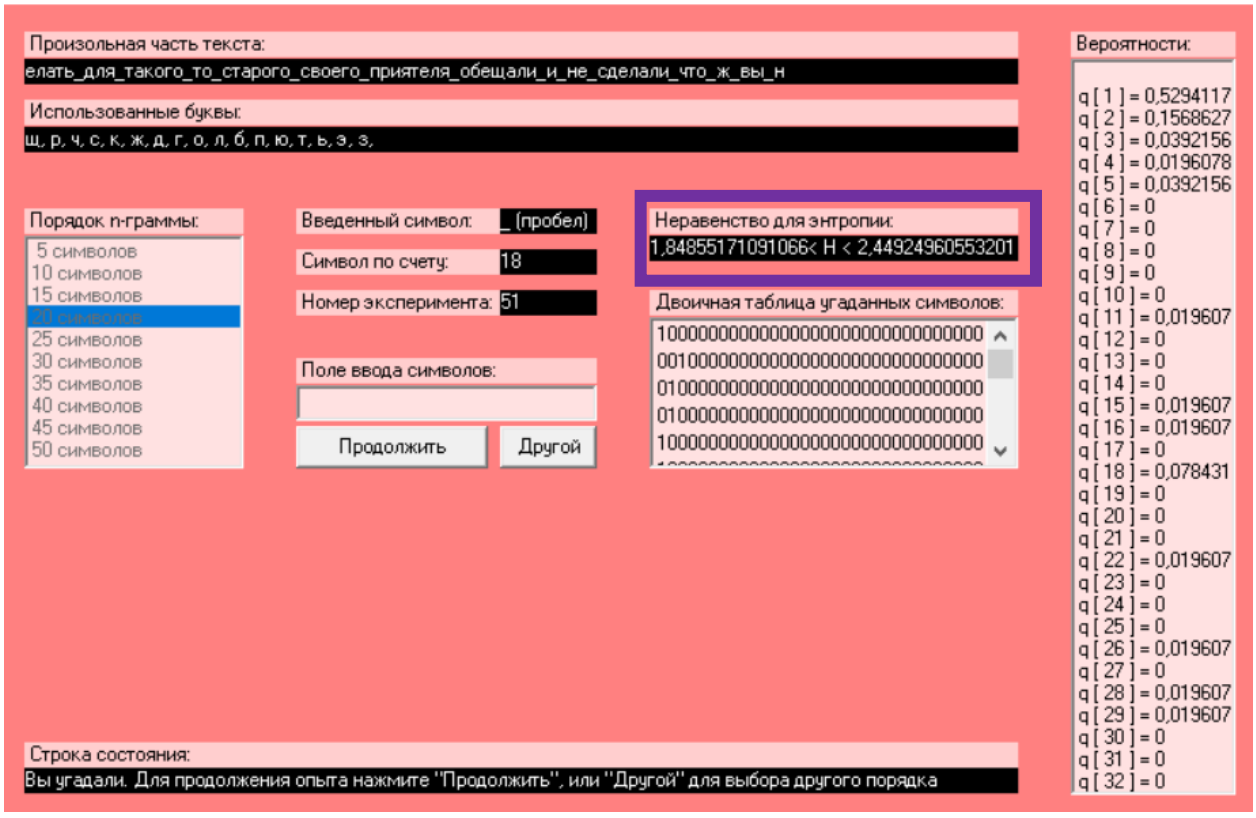
Код також буде на гітхабі, тож не бачу сенсу дублювати його тут

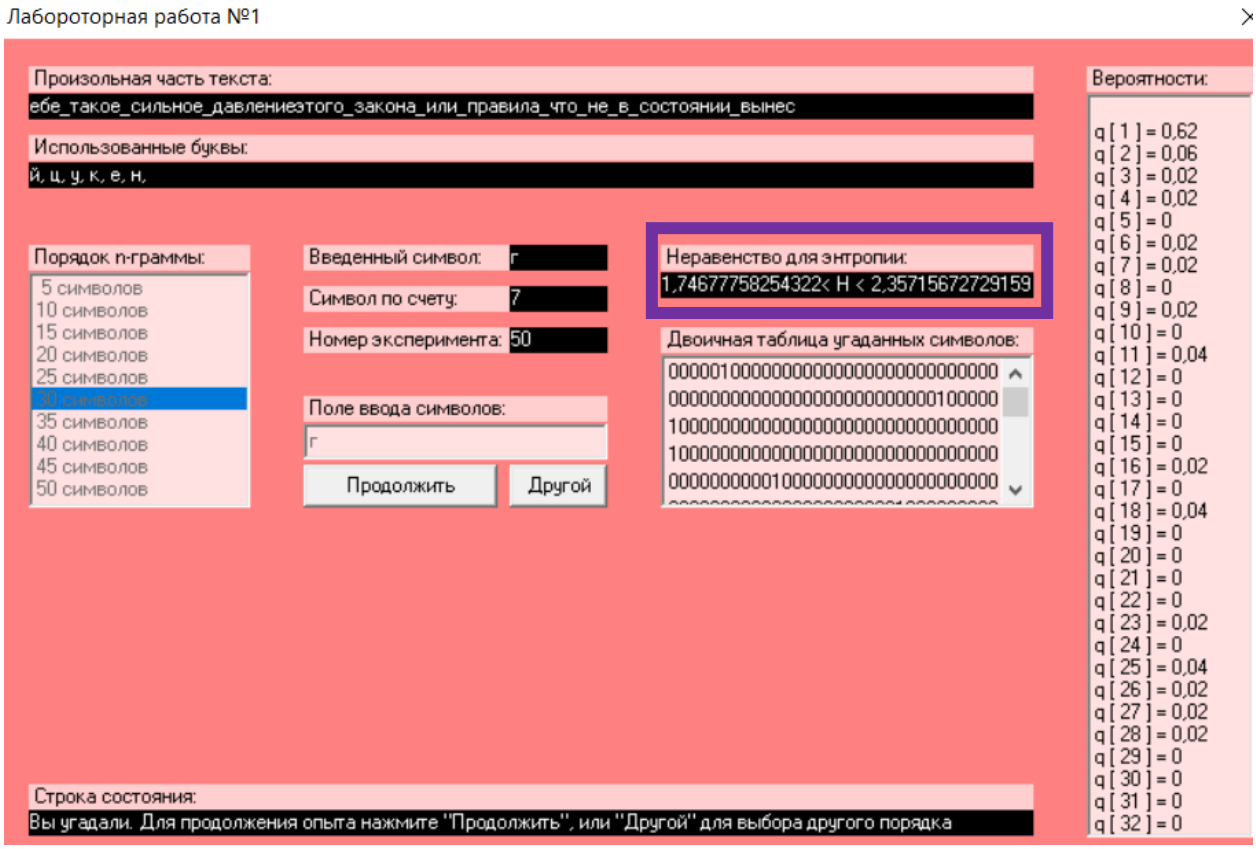
Перейдемо до проблем, які виникли. Найбільшої насправді стала обробка тексту (видалення пробілів). Як я не намагалася зробити якусь функцію, щоб все обраховувалося в одному скрипті, в мене не вийшло, тому я просто створила два різних з різницею в один символ (я впевнена, що вирішення є, просто я не вважаю його пошук доцільним, так як метою роботи не є навчити мене/нас програмувати (маю сумніви, що це хоч в когось вийде(говорю лише за себе))). Ще проблеми виникли під час обрахунку ентропії для біграм, бо спочатку, ми взагалі не подумали, що там біграми з перетином і без, а потім ще виявилося, що та треба було ділити на два. В принципі більше особливих

проблем не було, якщо не враховувати ті, які виникали, бо я не вмію читати завдання.

2. CoolPinkProgram







Проблемою стало те, що я довго не могла знайти екзешник, бо виявилося, що я забула, що вже знайшла його. Також це було просто банально нудно, мені довелося підключати молодшого брата, але і він довго не витримав.

3.

$$R = 1 - \frac{H_\infty}{H_0}$$

Для тексту з першого пункту:

З пробілами:

R1 = 0,112966

R21 = 0,196415364

R22 = 0,196646195

Без пробілів:

R1 = 0,10535837

R21 = 0,174638197

R22 = 0,174637905

Для другого пункту:

R10 = 0,549556548

R20 = 0,630289658

R30 = 0,650644483

Висновки: Під час виконання роботи дізналися про поняття ентропії та надлишковості мови, навчилися їх визначати та оцінювати. Нами було експериментально підтверджено, що найбільш уживаним символом є пробіл (для тексту з пробілами) або літера о (для тексту без пробілів). Також експериментальним шляхом підтверджено, що зі збільшенням значення ентропії (на практиці зі збільшенням символів, які нам достеменно відомі) збільшується значення R (тобто, ми підтвердили достатньо тривіальне твердження, що чим більше у нас початкової інформації, ти легше вгадати наступну літеру). Також ми побачили, що значення надлишковості може суттєво коливатися (у проміжку між 0 та 1) для однієї й тої самої мови.