КРИПТОГРАФІЯ
КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконала: Левашова Світлана

Група: ФБ-13

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.

Хід роботи:

Відсортована таблиця частот символів:

_: 0.1671

о: 0.0947

а: 0.0726

е: 0.0684

н: 0.0547

т: 0.0533

и: 0.0524

с: 0.0451

л: 0.0410

в: 0.0350

к: 0.0336

р: 0.0328

м: 0.0255

д: 0.0244

у: 0.0243

п: 0.0228

я: 0.0196

ь: 0.0177

з: 0.0143

ы: 0.0141

б: 0.0139

г: 0.0134

ч: 0.0133

й: 0.0084

ж: 0.0084

ш: 0.0080

х: 0.0073

ю: 0.0044

э: 0.0034

ц: 0.0027

щ: 0.0025

ф: 0.0011

Відсортована таблиця частот біграм, що перетинаються (перші 15):

о_: 0.0237

е_: 0.0200

_н: 0.0194

а_: 0.0178

_п: 0.0167

_с: 0.0165

и_: 0.0157

_в: 0.0153

то: 0.0132

я_: 0.0132

ь_: 0.0116

на: 0.0114

_т: 0.0104

но: 0.0101

не: 0.0101

Відсортована таблиця частот біграм, що не перетинаються (перші 15):

о_: 0.0238

е_: 0.0201

_н: 0.0195

а_: 0.0179

_п: 0.0167

_с: 0.0165

и_: 0.0157

_в: 0.0152

то: 0.0133

я_: 0.0132

ь_: 0.0113

на: 0.0112

_т: 0.0104

не: 0.0103

но: 0.0102

Ентропія H1: 4.357789081405551

$R = \frac{4{,}357789081405551}{loll2(32)(=5)} = 1 - 0.8715578162811102 = 0.1284421837188898$

Ентропія H2, що перетинаються: 3.9498866184225405

$R = \frac{3{,}9498866184225405}{loll2(32)(=5)} = 1 - 0.7899773236845081 = 0.2100226763154919$

Ентропія H2, що не перетинаються: 3.9494531633421985

$R = \frac{3{,}9494531633421985}{5} = 1 - 0.7898906326684397 = 0.2101093673315603$

Ентропія H1 для тексту без пробілів: 4.450345774438508

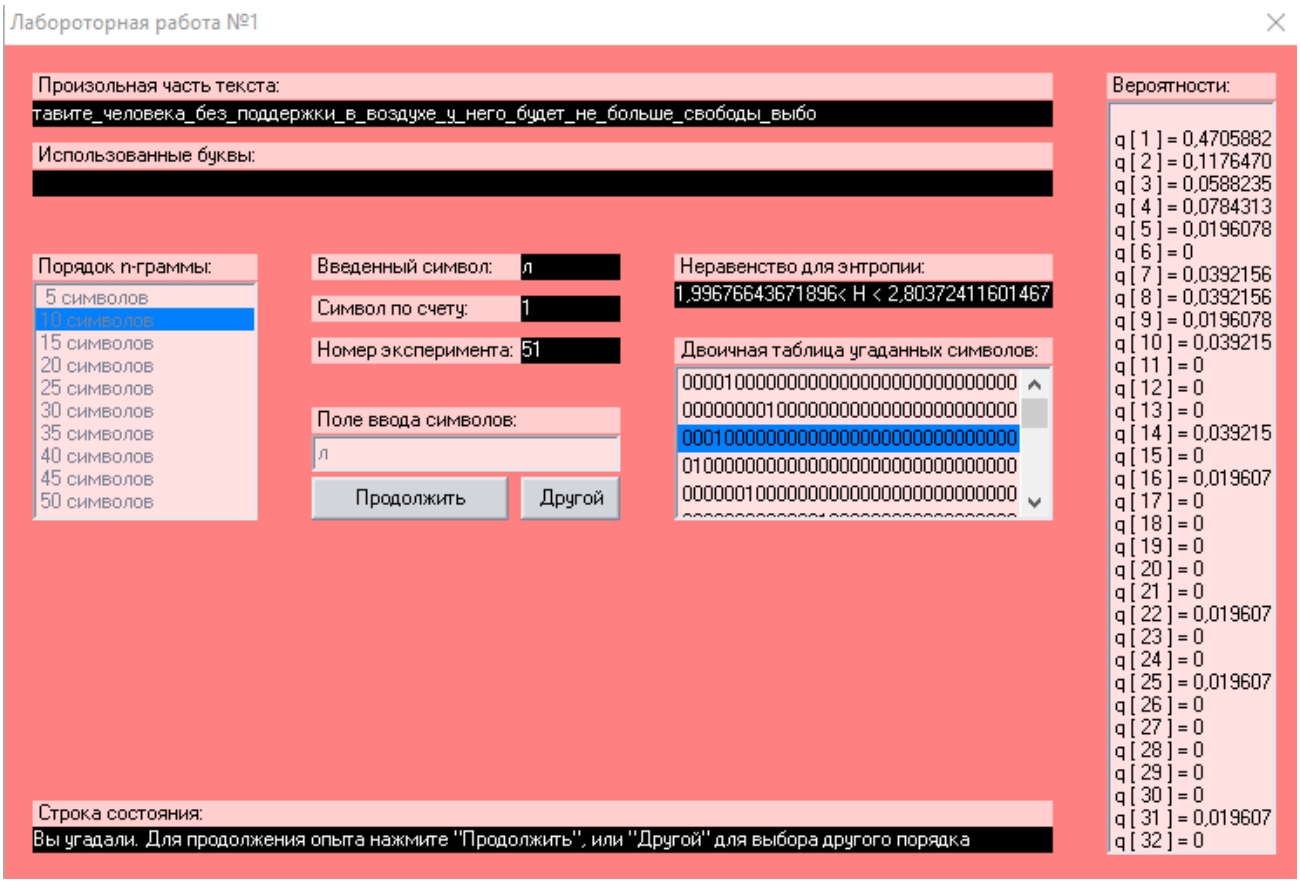
$R = \frac{4{,}450345774438508}{5} = 1 - 0.8900691548877016 = 0.1099308451122984$

Ентропія H2, що перетинаються для тексту без пробілів: 4.132089822886511

$R = \frac{4{,}132089822886511}{5} = 1 - 0.8264179645773022 = 0.1735820354226978$

Ентропія H2, що не перетинаються для тексту без пробілів: 4.131096743499651

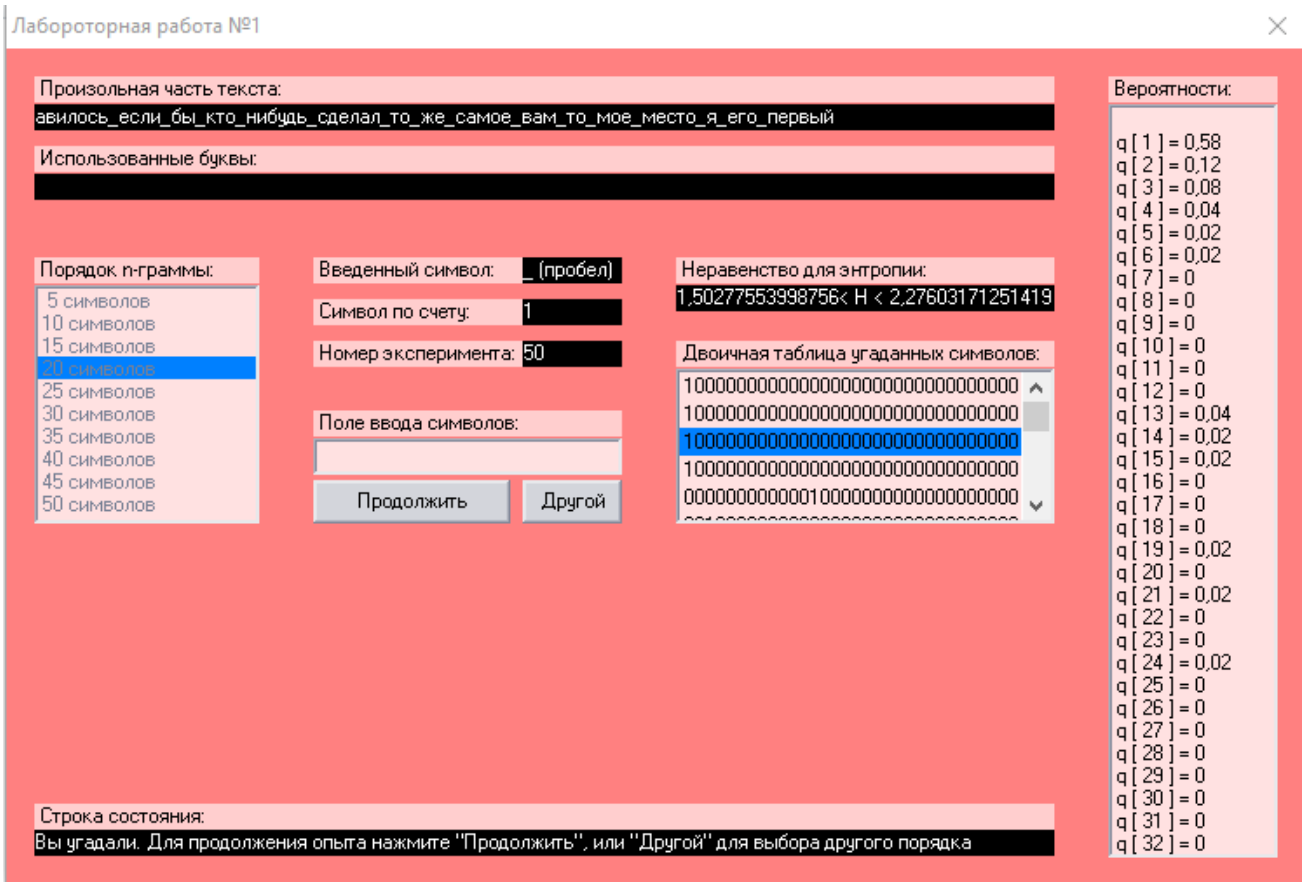
$R = \frac{4{,}131096743499651}{5} = 1 - 0.8262193486999302 = 0.1737806513000698$



$1{,}99676643671896 < H^{(10)} < 2{,}80372411601467$

Оцінка надлишковості російської мови:

$1 - \frac{1{,}99676643671896}{loll2(32)(=5)} < R < 1 - \frac{2{,}80372411601467}{5}$; $0{,}600646713 < R < 0{,}439255177$



$1{,}50277553998756 < H^{(20)} < 2{,}27603171251419$

Оцінка надлишковості російської мови:

$$1 - \frac{1{,}50277553998756}{5} < R < 1 - \frac{2{,}27603171251419}{5};\ 0{,}699444892 < R < 0{,}544793657$$

$1{,}63515048660736 < H^{(30)} < 2{,}25564071234112$

Оцінка надлишковості російської мови:

$1 - \frac{1{,}63515048660736}{5} < R < 1 - \frac{2{,}25564071234112}{5}$; $0{,}672969903 < R < 0{,}548871858$

Висновки:

Реалізовано програму для підрахунку частот букв і біграм в тексті, а також для обчислення значень H1 та H2 за безпосереднім означенням.

Виконано підрахунок частот букв, біграм, H1 та H2 на російському тексті із заміною імовірностей замінити відповідними частотами. Отримано значення H1 та H2 для тексту без пробілів.

Використовуючи програму CoolPinkProgram, оцінено значення ентропії для різних довжин тексту (10, 20, 30).

Робота дозволила освоїти концепції ентропії та надлишковості на символ джерела відкритого тексту.