Міністерство освіти і науки України Національний технічний університет України "Київський політехнічний інститут ім. Ігоря Сікорського" Фізико-технічний інститут

КРИПТОГРАФІЯ КОМП'ЮТЕРНИЙ ПРАКТИКУМ No1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконав: Семенов Олексій ФБ-05

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 і Н та 2 Н за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення і Н та 2 Н на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення і Н та 2 Н на тому ж тексті, в якому вилучено всі пробіли.
- 2. За допомогою програми CoolPinkProgram оцінити значення $_{(10)}H$, $_{(20)}H$, $_{(30)}H$.
- 3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи

До початку роботи над лабораторною, було ознайомлено з теоетичними відомостями з методички. Для більш наглядного сприйняття пробілів у тексті, їх було замінено на "".

В якості тексту для експериментів, під час виконанням роботи, була задіяна книга "Ренегат" на російській мові. Її, зміст знаходиться у файлі "Havelok_Renegat.txt". Копія з пробілами — "2.txt". Копія без пробілів — "3.txt"

Аналіз тексту з пробілами

Частота літер:

0.1725377292
o 0.0911517439
e 0.0730613057
a 0.0643184489
н 0.0572622919
т 0.0546226456
и 0.0501050884
c 0.0437111748
в 0.0367949284
p 0.0350802465
л 0.0320659507
м 0.0278033417
к 0.0261679424
д 0.0253875523
y 0.0245107793
я 0.0234085949
п 0.0227712244

ы --- 0.0170908548 ь --- 0.0158922875 з --- 0.0143812532 г --- 0.0136677093 б --- 0.0133334784 ч --- 0.0121069290 ю --- 0.0107435781 ж --- 0.0093398086 й --- 0.0085578639 x --- 0.0075163074 ш --- 0.0074556794 щ --- 0.0039439238 ц --- 0.0022370147 э --- 0.0020846677 ф --- 0.0006731254 ъ --- 0.0002114204 ë --- 0.0000031091

H1: 4.369264120814455 R1: 0.14117031275640213

Частота біграм без перехресного:

```
\begin{array}{c} \_o & --- & 0.0184673326 \\ o\_ & --- & 0.0166127564 \\ e\_ & --- & 0.0146124126 \\ \underline{e} & --- & 0.0125922634 \\ \underline{H} & --- & 0.0124100538 \\ \underline{H} & --- & 0.0120068162 \\ a\_ & --- & 0.0118966983 \\ \underline{H} & --- & 0.0117643983 \\ \underline{T} & --- & 0.0114332523 \\ \underline{a} & --- & 0.0107376872 \end{array}
```

H2: 4.157817311278825 R2: 0.17575486512389804

Частота перехресних біграм:

о_ --- 0.0205709820
e_ --- 0.0182549074
_H --- 0.0176395024
я_ --- 0.0163011150
_п --- 0.0153629782
и_ --- 0.0150497387
a_ --- 0.0142144331
_c --- 0.0139407439
_B --- 0.0123270854

H2: 3.9917693051821703 R2: 0.21537130987651232

Аналіз тексту без пробілів

Частота літер:

o --- 0.1101581874 e --- 0.0882956339 a --- 0.0777297663 н --- 0.0692022995 т --- 0.0660122492 и --- 0.0605527166 c --- 0.0528255805 в --- 0.0444671977 p --- 0.0423949801 л --- 0.0387521605 м --- 0.0336007365 к --- 0.0316243331 д --- 0.0306812204 y --- 0.0296216277 я --- 0.0282896220 п --- 0.0275193507 ы --- 0.0206545427

H2: 4.478496503560943 R2: 0.11218346592424489 ь --- 0.0192060570
з --- 0.0173799504
г --- 0.0165176223
б --- 0.0161136996
ч --- 0.0146313970
ю --- 0.0129837679
ж --- 0.0112872924
й --- 0.0103423010
х --- 0.0090835650
ш --- 0.0090102953
щ --- 0.0027034643
э --- 0.0025193507
ф --- 0.0008134816
ъ --- 0.0002555046

ë --- 0.000037574

Частота біграм без перехресними:

то --- 0.0106001844

но --- 0.0096068290

не --- 0.0095729325

ен --- 0.0093921513

он --- 0.0092038375

oo --- 0.0086398376

ет --- 0.0083018143

от --- 0.0081323319

ат --- 0.0074252135

по --- 0.0072755040

H2: 4.384278742919081 R2: 0.13086118190212426

Частота біграм перехресних:

то --- 0.0156249706

но --- 0.0145557453

не --- 0.0132892113

ст --- 0.0126972852

на --- 0.0116881921

ен --- 0.0110680790

ет --- 0.0110492877

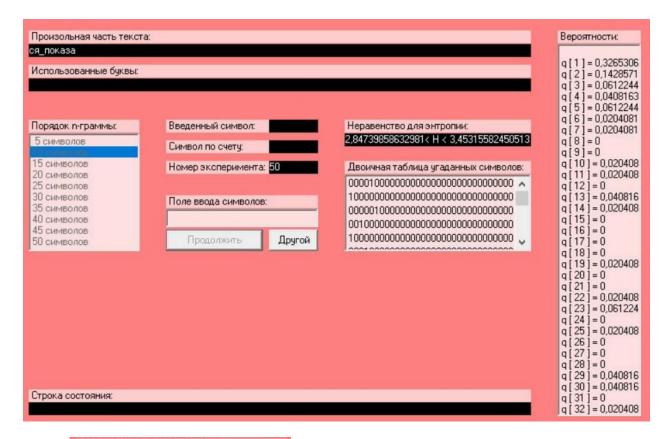
по --- 0.0105381642

от --- 0.0094369937

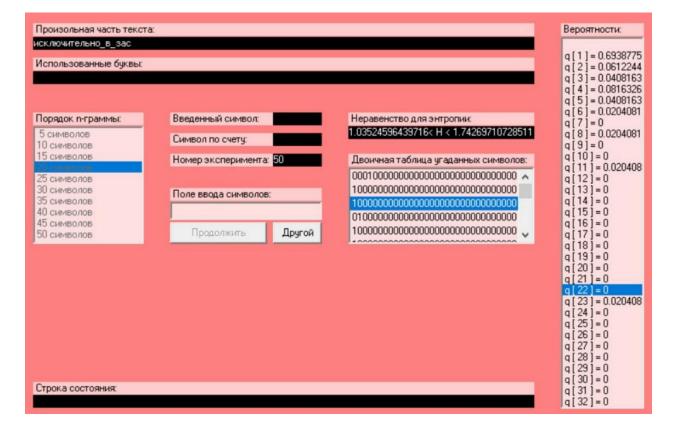
oc --- 0.0089540571

H2: 4.157817311278825 R2: 0.17575486512389793

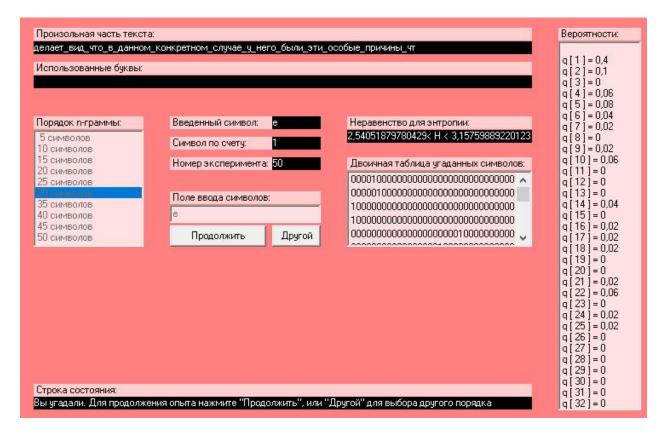
Перейдем до CoolPinkProgram



Неравенство для энтропии: 2,84739058632981 < H < 3,45315582450513



Неравенство для энтропис 1.03524596439716к Н < 1.74269710728511



H30:

Неравенство для энтропии: 2,54051879780429< Н < 3,15759889220123

Висновки

Протягом лабораторної роботи я ознайомився з такими означення як ентропія та надлишковість. Також аналізуючи текст, я навчився їх знаходити та обраховувати. Було проведено експерименти, на тексті великого обсягу в двух видах. З пробілами, та без них. Дослідивши результати, я помітив, що ентропія у тексті без пробілів більша. Здобуті навички знадобляться у наступних лабораторних роботах при аналізі тексту.