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

Київ – 2022

**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 1 H та 2 H за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення 1 H та 2 H на довільно обраному тексті російською мовою достатньої  
довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення 1 H та 2 H 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

о --- 0.0911517439

е --- 0.0730613057

а --- 0.0643184489

н --- 0.0572622919

т --- 0.0546226456

и --- 0.0501050884

с --- 0.0437111748

в --- 0.0367949284

р --- 0.0350802465

л --- 0.0320659507

м --- 0.0278033417

к --- 0.0261679424

д --- 0.0253875523

у --- 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

х --- 0.0075163074

ш --- 0.0074556794

щ --- 0.0039439238

ц --- 0.0022370147

э --- 0.0020846677

ф --- 0.0006731254

ъ --- 0.0002114204

ё --- 0.0000031091

H1 : 4.369264120814455

R1 : 0.14117031275640213

Частота біграм без перехресного:

\_о --- 0.0184673326

о\_ --- 0.0166127564

е\_ --- 0.0146124126

\_е --- 0.0125922634

и\_ --- 0.0124100538

\_н --- 0.0120068162

а\_ --- 0.0118966983

н\_ --- 0.0117643983

т\_ --- 0.0114332523

\_а --- 0.0107376872

H2 : 4.157817311278825

R2 : 0.17575486512389804

Частота перехресних біграм:

о\_ --- 0.0205709820

е\_ --- 0.0182549074

\_н --- 0.0176395024

я\_ --- 0.0163011150

\_п --- 0.0153629782

и\_ --- 0.0150497387

а\_ --- 0.0142144331

\_с --- 0.0139407439

\_в --- 0.0138980294

то --- 0.0123270854

H2: 3.9917693051821703

R2: 0.21537130987651232

**Аналіз тексту без пробілів**

Частота літер:

о --- 0.1101581874

е --- 0.0882956339

а --- 0.0777297663

н --- 0.0692022995

т --- 0.0660122492

и --- 0.0605527166

с --- 0.0528255805

в --- 0.0444671977

р --- 0.0423949801

л --- 0.0387521605

м --- 0.0336007365

к --- 0.0316243331

д --- 0.0306812204

у --- 0.0296216277

я --- 0.0282896220

п --- 0.0275193507

ы --- 0.0206545427

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 --- 0.0047662884

ц --- 0.0027034643

э --- 0.0025193507

ф --- 0.0008134816

ъ --- 0.0002555046

ё --- 0.0000037574

H2: 4.478496503560943

R2: 0.11218346592424489

Частота біграм без перехресними:

то --- 0.0106001844

но --- 0.0096068290

не --- 0.0095729325

ен --- 0.0093921513

он --- 0.0092038375

оо --- 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

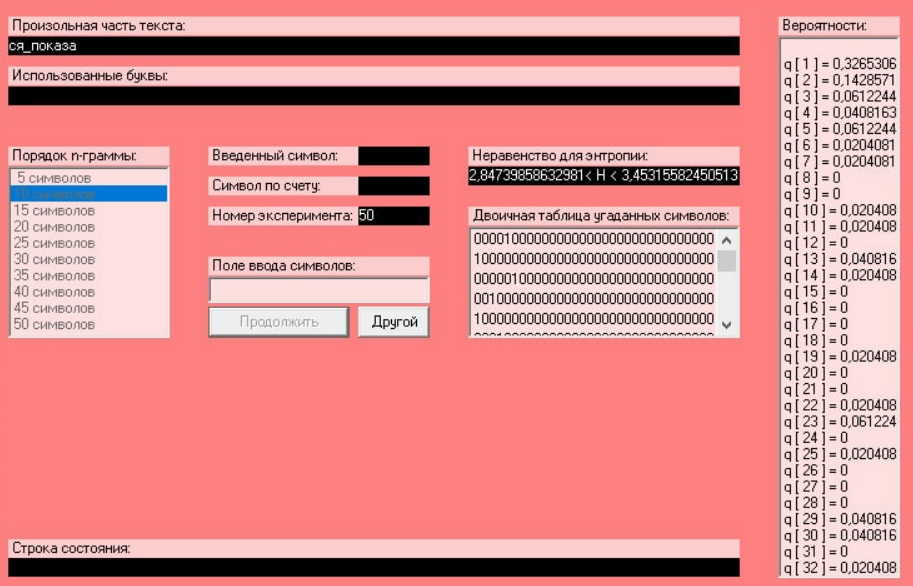
от --- 0.0094369937

ос --- 0.0089540571

H2: 4.157817311278825

R2: 0.17575486512389793

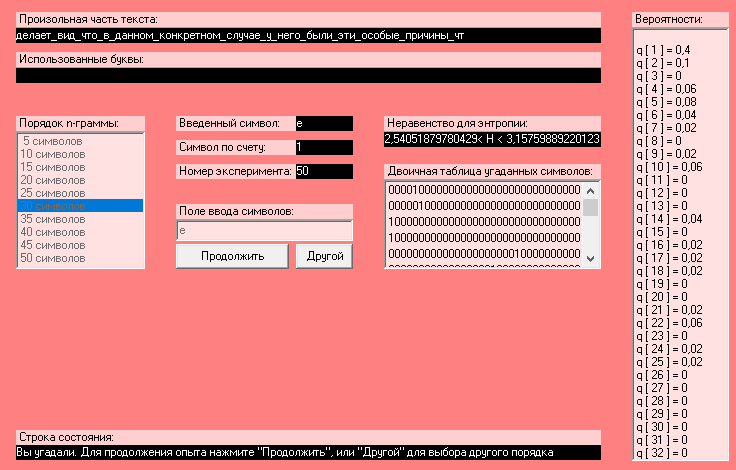
Перейдем до CoolPinkProgram



H10: 



H20: 



H30: 

**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.