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

### **КРИПТОГРАФІЯ** КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали Студенти: Дудченко И.В і Терпило С.Е

#### 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

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

Перед виконанням роботи були розглянуті теоретичні відомості в методичних вказівках. В якості експериментального тексту була взята книга «Good Omens» в перекладі на російську мову. Оригінал тексту можна знайти у файлі badtxt.txt. Відредагований текст з пробілами міститься у файлі spaces.txt, а без пробілів у nospaces.txt. В ході виконання роботи було прийнято рішення пробіли замінити на «\_» для кращого сприйняття. Усі таблиці також наведені у файлах з відповідними назвами. Єдина відмінність - таблиці біграм були виконанні у двох варіантах(таблиці у вигляді матриці і звичайні).

### Аналіз результатів для тексту з пробілами

### Текст без пробілів

### Текст з пробілами

а	0,077776
6	0,016856
В	0,047941
г	0,020603
Д	0,029395
e	0,087385
Э	0,003103
ж	0,00878
3	0,015413
И	0,073477
ы	0,015419
й	0,013549
К	0,037569
л	0,042051
M	0,033064
н	0,067137
0	0,109828
п	0,028976
p	0,050127
С	0,056236
Т	0,053851
у	0,026052
ф	0,002366
X	0,008574
ц	0,004231
ч	0,015297
Ш	0,008547
Щ	0,003281
ь	0,014809
ю	0,006976
Я	0,021133

а	0,065903	
6	0,014283	
В	0,040623	
Γ	0,017458	
Д	0,024908	
е	0,074046	
Э	0,002629	
ж	0,00744	
3	0,01306	
И	0,062261	
Ы	0,013065	
й	0,01148	
К	0,031834	
Л	0,035632	
M	0,028017	
н	0,056889	
0	0,093063	
П	0,024553	
p	0,042475	
С	0,047651	
Т	0,045631	
у	0,022075	
ф	0,002005	
X	0,007265	
ц	0,003585	
ч	0,012962	
Ш	0,007243	
щ	0,00278	
ь	0,012548	
ю	0,005911	
Я	0,017907	

### Біграми

## Текст з пробілами

аб	0,000787
вг	0,000202
де	0,004144
эж	0
зи	0,000551
ый	0,001765
кл	0,000498
мн	0,001576
оп	0,001417
рс	0,000719
ту	0,001207
фх	0
цч	0
шщ	0
ью	0,000362

## Текст без пробілів

аб	0,001658
вг	0,000987
де	0,004973
эж	0
зи	0,000686
ый	0,002083
кл	0,000688
мн	0,002537
оп	0,003907
рс	0,00119
ту	0,001517
фх	1E-05
цч	8,03E-06
шщ	0
ью	0,000435
я	0

#### Перехрестні біграми

#### crossed\_bigrams(text\_with\_space)

#### \б бв 0,000787 6,97E-05 0,000202 ВГ ΓД 0,000655 0,004144 де 1,7E-06 еэ эж 0 0 жз зи 0,000551 иы ый 0,001765 йк 8,67E-05 ΚЛ 0,000498 3,91E-05 ЛΜ МН 0,001576 0,010083 но 0,001417 ОΠ 0,005989 пр 0,000719 рс СТ 0,01211 0,001207 ту уφ 1,36E-05 фх 1,7E-06 ΧЦ 0 ЦЧ 0,000112 ЧШ шщ 2,72E-05 ЩЬ 0,000362 ью

1,7E-06

юя

#### $crossed\_bigrams(text\_without\_space)$

` -	
0,001658	
9,43E-05	
0,000987	
0,000825	
0,004973	
0,000245	
0	
6,02E-06	
0,000686	
0	
0,002083	
0,000777	
0,000688	
0,000544	
0,002537	
0,012102	
0,003907	
0,00707	
0,00119	
0,014432	
0,001517	
5,22E-05	
1E-05	
2,61E-05	
8,03E-06	
0,000151	
0	
3,21E-05	
0,000435	
9,83E-05	

# crossed\_bigrams\_space(text\_with\_space) crossed\_bigrams\_space(text\_without\_space)

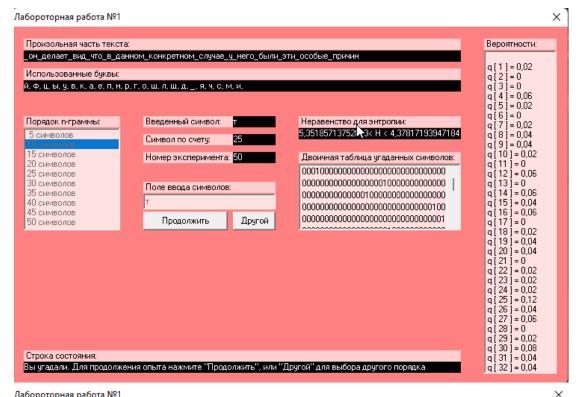
аб	0,000787
бв	6,97E-05
ВГ	0,000202
L <mark>Z</mark>	0,000655
де	0,004144
еэ	1,7E-06
эж	0
жз	0
зи	0,000551
иы	0
ый	0,001765
йк	8,67E-05
кл	0,000498
лм	3,91E-05
мн	0,001576
но	0,010083
оп	0,001417
пр	0,005989
рс	0,000719
ст	0,01211
ту	0,001207
уф	1,36E-05
фх	0
хц	1,7E-06
цч	0
чш	0,000112
шщ	0
щь	2,72E-05
ью	0,000362
юя	1,7E-06
я	0,011411

аб	0,001658	
бв	9,43E-05	
ВГ	0,000987	
гд	0,000825	
де	0,004973	
еэ	0,000245	
эж	0	
жз	6,02E-06	
зи	0,000686	
иы	0	
ый	0,002083	
йк	0,000777	
кл	0,000688	
лм	0,000544	
МН	0,002537	
но	0,012102	
оп	0,003907	
пр	0,00707	
рс	0,00119	
ст	0,014432	
ту	0,001517	
уф	5,22E-05	
фх	1E-05	
хц	2,61E-05	
цч	8,03E-06	
чш	0,000151	
шщ	0	
щь	3,21E-05	
ью	0,000435	
юя	9,83E-05	
Я	0	

### Значення H і R

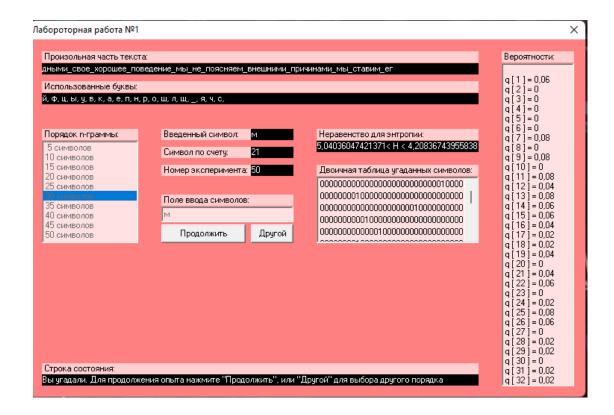
	Н	R
Letters with spaces	4.391661207126057	0.12166775857478862
Letters without spaces	3.977721444187262	0.19710055981277008
Bigrams with spaces	0.09802607833741339	0.9803947843325174
Bigrams without spaces	0.061206655396903696	0.9876454925153897
Crossed bigrams with spaces	0.19804297423171796	0.19804297423171796
Crossed bigrams without spaces	0.16122355129120827	0.16122355129120827

### CoolPinkProgram









#### Отримані результати

5,3518<H(10)<4,3781 4,6870<H(20)<4,2631 5,0403<H(30)<4,2083

#### Висновки

Під час виконання лабораторної роботи, мы отрамали змогу ознайомитись з такими поняттями як ентропія, надлишковість та обрахувати їх на практиці. Успішно були проведені експерименти на різних видах тексту(з пробілом та без), порахована частота монограм та біграм, проведено знайомство з невеличкою програмою. Набуті навички знадобляться у майбутніх лаб.роботах та у професійній діяльності.