

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ
ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

ФІЗИКО- ТЕХНІЧНИЙ ІНСТИТУТ

Кафедра інформаційної безпеки

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

з дисципліни

Криптографія

З теми: «Експериментальна оцінка ентропії на символ
джерела відкритого тексту»

Виконали студенти
групи ФБ-92
Рога Владислав та Томич Олег

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 H_1 та H_2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H_1 та H_2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 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.

Код програми знаходиться у файлі Lab1.py
Всі таблиці і дані знаходяться у файлі

H1 з пробілами	4.361476
H2 з пробілами	4.455664
H1 без пробілів	3.960703
H2 без пробілів	4.145339
H2 крок 2 з пробілами	4.144836
H2 крок 2 без пробілів	3.967747

Частота символів з пробілом	
Пробіл	0.1674151172512141
о	0.08937010476913382
е	0.07264587463689826
а	0.07096165160141366
и	0.0556463106519052
н	0.05492868518461177
т	0.05367210834596326
с	0.04409233140822805
л	0.04375255771759116
в	0.034340993863151544
р	0.03408992955972527
к	0.028974494377414926
м	0.026095205257621268
д	0.025931595019888478
у	0.023651931144777906
п	0.021730870783061195
ь	0.01694474827874498
я	0.016527981535057364
ы	0.015958484006785432
б	0.014592694196146498
г	0.014514864262084353
з	0.013303897438558289
ч	0.012215952123711099
ж	0.009746734699513688
й	0.009037478042334462
х	0.00779094377582301
ш	0.007421460809280676
ю	0.004770640205604927
э	0.0038182696146079253
щ	0.0026169269227132017
ц	0.0025453735962367133
ф	0.0008937889201975374

Частота символів без пробілів	
о	0.10734053262422646
е	0.08725341540799696
а	0.08523053093053189
и	0.06683560055545142
н	0.06597367586505325
т	0.06446442814186627
с	0.05295836175004385
л	0.05255026679459003
в	0.041246237560516916
р	0.04094468956387616
к	0.03480064913232077
м	0.03134239619084571
д	0.031145887413034815
у	0.028407831603536757
п	0.026100486849240576
ь	0.020351976873278978
я	0.019851407198855323
ы	0.01916739582647521
б	0.0175269747247495
г	0.01743349484579087
з	0.015979028341993624
ч	0.01467232035655035
ж	0.011706595809588523
й	0.01085472271907839
х	0.00935753691575704
ш	0.008913758780700727
ю	0.005729914516168753
э	0.004586042448911487
щ	0.0031431352849854727
ц	0.003057194105942858
ф	0.0010735108680410888

[illegible]

[illegible][illegible]

[illegible]

Надлишковість мови

H	R
H1 з пробілами	0.140692774193
H2 з пробілами	0.127764612903
H1 без пробілів	0.143731096774
H2 без пробілів	0.133720612903
H2 крок 2 з пробілами	0.127991838709
H2 крок 2 без пробілів	0.133704387096
H10	0.07009101717<R<0.09076061177
H20	0.07357843316<R<0.09469644563
H30	0.07702615552<R<0.10334119329

Висновок:

Під час виконання цієї лабораторної роботи я навчився вимірювати частоти символів та біграм у тексті, визначати ентропію та надлишковість мови.