

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

Криптографія
Комп'ютерний практикум №1
Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконав:
Студент гр. ФБ-11
Падик Володимир

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:

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.

Літери з пробілами	
Пробіл	0.1632843675322151
о	0.09488660400873553
а	0.07076458683497411
е	0.06867003395876925
и	0.054741364525091755
н	0.05470634811740358
т	0.04831406716289919
с	0.044129249133879875
л	0.042074715009318656
в	0.038202186167232643
Ентропія	4.381314134135093
Надлишковість	0.13880174246967014

Біграми з пробілами	
‘о ‘	0.02112205513587448
‘и ‘	0.01804918451015941
‘а ‘	0.017491065450000464
‘е ‘	0.01681503391234315
‘с’	0.01603895542408116
‘н’	0.015461541594877025
‘п’	0.015154254532305518
‘в’	0.014888415492127377
‘то’	0.013652121031083873
‘о’	0.012014924146545914
Ентропія	3.969744900372348
Надлишковість	0.21970046283488742

Літери без пробілів	
о	0.11340364674301556
а	0.08457423775657576
е	0.08207093460921226
и	0.06542409679097205
н	0.06538224696012224
т	0.057742517634579237
с	0.052741035808935875
л	0.05028556103968645
в	0.04565731137896901
р	0.04424552014642316
Ентропія	4.468907671211767
Надлишковість	0.11408435473711098

Біграми(step 2)	
‘то’	0.016627208221021955
‘ст’	0.013043493114415827
‘на’	0.012826557071260928
‘ов’	0.01162828436044863
‘ал’	0.010851072552295251
‘не’	0.010839115447554431
‘го’	0.010717836242326102
‘он’	0.010416346387075394
‘ос’	0.010184891002449499
‘ко’	0.010163539029698031
Ентропія	4.1481699411828075
Надлишковість	0.17766735845168813

2. За допомогою програми CoolPinkProgram оцінити значення (10) Н, (20) Н, (30) Н.

[illegible]

$$1,806686 < H(10) < 2,452812$$

[illegible]

$$1,951811 < H(20) < 2,627322$$

$$0.62 < H^{30} < 0.53$$

Висновки: Під час виконання цього комп'ютерного практикуму ми мали можливість дослідити концепції ентропії, біграмів та надлишковості російської мови за допомогою скрипта, написаного власноруч, та програми CoolPinkProgram. Це дало нам можливість оцінити відмінності між початковими даними на практиці порівняно з отриманими результатами через експериментальне "вгадування" наступних літер.