

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

Фізико-технічний інститут

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

**Експериментальна оцінка ентропії на символ
джерела відкритого тексту**

Виконали:
ФБ-21 Захожий М.
ФБ-21 Хав'юк А.

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.

Хід роботи:

1. Підрахувати частоти, ентропії та надлишковості для букв
2. Підрахувати частоти, ентропії та надлишковості для біграм
3. Оцінити значень $H(10)$, $H(20)$, $H(30)$ з використанням програми CoolPinkProgram
4. Зробити висновки

1.

Частота та надлишковість монограм (текста з пробілами):

```
**** WORKING WITH FILE THAT HAS WHITESPACES ****

Monogram frequencies:
c: 0.044611133043821555
e: 0.07790819716328812
n: 0.050966269631113205
k: 0.023813297641859165
a: 0.056801440497626456
: 0.16845926453737495
n: 0.02154086143298701
p: 0.031766053941629034
m: 0.05657034462172494
v: 0.02155271489509211
t: 0.06536161690081173
y: 0.02466948475248669
l: 0.03058168757763377
ц: 0.001685074095115211
я: 0.01817954223758582
o: 0.09230354443299664
h: 0.008589063389767296
m: 0.027838217480477213
ю: 0.005565559011204811
б: 0.019113554736021107
д: 0.03018689878963535
г: 0.014482008223161584
ж: 0.011766631681318787
з: 0.0129062704814328
х: 0.00743364006163388
ь: 0.022300752024496162
ч: 0.01465533013008772
ш: 0.008020952692748403
щ: 0.002753892521159716
ы: 0.015637487602669156
ъ: 0.00013480592760921686
э: 0.00215089404174747
ф: 0.0009340124984352884

Monogram's entropy: 4.3656249438892925

Monogram's redundancy: 0.13455910846940067
```

Частота та надлишковість монограм (текст без пробілів):

```
**** WORKING WITH FILE THAT HAS >NOT< WHITESPACES ****

Monogram frequencies (without whitespaces):
c: 0.05364876444567961
e: 0.09369137775307441
n: 0.06129136848931194
k: 0.027675490400426134
a: 0.06830866856573797
п: 0.025903795826675003
p: 0.038201440515065195
и: 0.06803075569142407
в: 0.038305657842932904
т: 0.07860302461844879
y: 0.0296671993330091
л: 0.036777137034206445
ц: 0.0020264480418722064
я: 0.021862479446027004
o: 0.11100303388221126
h: 0.010329095162000047
m: 0.032515806294726605
ю: 0.006693068389726487
б: 0.022985710646379026
д: 0.03630236920725353
г: 0.01741587348700456
ж: 0.014150397183816207
з: 0.01558628036777137
х: 0.008939530790430533
ь: 0.0268185923712916
ч: 0.01762430811273999
ш: 0.009645892679311702
щ: 0.003311795085574006
ы: 0.01805437028574075
ъ: 0.0001621158434977652
э: 0.0025938534935964243
ф: 0.0011232312003520229

Monogram's entropy: 4.463349445769817

Monogram's redundancy: 0.11518621658819439
```

Частота та надлишковість біграм, що перетинаються (текста з пробілами):

Частота та надлишковість біграм, що перетинаються (текста без пробілів):

Частота та надлишковість біграм, що не перетинаються (текста з пробілами):

Київ 2024

[illegible]**H(10)**[illegible]

H(20)

Произвольная часть текста:

ласны_со_мной_в_это

Использованные буквы:

т,

Порядок n-граммы:	Введенный символ:	Неравенство для энтропии:
5 ██████████ 10 ██████████ 15 ██████████ 20 ██████████ 25 ██████████ 30 ██████████ 35 ██████████ 40 ██████████ 45 ██████████ 50 ██████████	т	1.9540128635612 < H < 2.5851278491081
	Символ по счету: 1	
	Номер эксперимента: 51	
	Поле ввода символов: 	Двоичная таблица угаданных символов: <pre> 10000000000000000000000000000000 ^ 01000000000000000000000000000000 00100000000000000000000000000000 00100000000000000000000000000000 10000000000000000000000000000000 v ~::~~~~~~</pre>

Строка состояния:

Вы не угадали. Введите другую букву

H(30)

[illegible]

	H	R
Монограми (текст з пробілами)	4.3656249438892925	0.13455910846940067
Монограми (текст без пробілів)	4.463349445769817	0.11518621658819439
Біграми, що не перетинаються (текст з пробілами)	3.9173973525421943	0.22341568484731933
Біграми, що не перетинаються (текст без пробілів)	3.9161950272211934	0.22365403365444103
Біграми, що перетинаються (текст з пробілами)	4.111188048286857	0.18499864383916975
Біграми, що перетинаються (текст без пробілів)	4.11052090047811	0.1851308991294982
	H	R
H(10)	2.7896471294937<H< 3.49608115803978	0.44207057410125994<R<0.30 0783768392044
H(20)	1.9540128635612<H< 2.5851278491081	0.60919742728776<R<0.48297 443017838004
H(30)	1.55975647048236<H< 2.32244472586662	0.688048705903528<R<0.5355 110548266759

Висновки

За результатами виконання даного практикуму було отримано теоретичні знання про ентропію та надлишковість джерела тексту (мови) та практичні навички з їх обрахунку. Також експериментальним шляхом виявлено залежність ентропії та надлишковості від кількості букв за допомогою CoolPinkProgram.