

КРИПТОГРАФІЯ

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела
відкритого тексту

Виконали:

Стеденти групи ФБ-11 Тирнавська Єлизавнта та Шестак Максим

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 $H(10)$, $H(20)$, $H(30)$
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела

Хід роботи

Букви:

- З пробілами ($H(1)$): 4.362503181139441, R : 0.1424992540943505

_,0.16827080696469765
о,0.09546058762401124

е,0.07154055090496164
а,0.06624529835221278
н,0.054104963231276365
и,0.05395097371051162
т,0.05387646265207707
с,0.04404100293871614
в,0.03848049101794027
л,0.03825298391952011
р,0.034708244499593664
к,0.027453847850405638
д,0.026633232726846433
м,0.02617722504922697
у,0.02468501691897767
п,0.022837142669800768
ь,0.019129472402097437
я,0.017752508042226906
ч,0.015088986073386438
б,0.01447700191344398
г,0.014016026831928883
ы,0.01373387829065671
з,0.012816895531522151
ж,0.00949072188300373
й,0.00829457102493438
х,0.007086498397515503
ш,0.006838128202733662
ю,0.004695190162155933
э,0.0029436835485543863
щ,0.002488669351714052
ц,0.002286992753551197
ф,0.0009805655289987104
ё,0.0009597024326370358
ъ,0.00020167659816285535

- Без пробілів (H1): 4.459106909759702, R: 0.1160272563463357

о,0.11477364077559733
е,0.08601423576811393
а,0.07964767728118091
н,0.06505117733552958
и,0.06486603351461502
т,0.06477644779481764
с,0.05295113278156493
в,0.04626564913215327
л,0.0459921140677053
р,0.041730222757878466
к,0.03300815707840662
д,0.03202151968437159
м,0.03147325507921169
у,0.029679151730736382
п,0.02745742587976163

ь,0.022999640462644546
я,0.021344096360789166
ч,0.018141705497099214
б,0.017405908118496825
г,0.0168516711320171
ы,0.016512439873051062
з,0.01540993828141144
ж,0.011410831749656887
й,0.009972682327843123
х,0.008520199190861779
ш,0.008221580124870549
ю,0.005645094823498215
э,0.0035392331701280598
щ,0.002992163041232126
ц,0.0027496843596472474
ф,0.0011789480725333766
ё,0.0011538640709901134
ъ,0.0002424786815848789

Біграми, що перетинаються

- З пробілами (H2): 3.947137326489767, R: 0.22414424445807168

о_,0.023776975486855256
е_,0.018094265430246722
и_,0.017515066136015466
а_,0.017018325746451783
_в,0.01679677953270638
_п,0.01597417744758892
_н,0.015951327389668993
_с,0.015743689906831372
то,0.014708482934980657
ь_,0.012016150023545495

- Без пробілів (H2): 4.131419613110301, R: 0.1809879411968437

то,0.01813334416325146
ов,0.01258141848834251
не,0.012264882278391806
на,0.01209885007770068
но,0.011907733875466293
ст,0.011607920333211098
по,0.010910346195055584
ко,0.0106714509422626
он,0.010383582162647054
от,0.009749315266481682

Біграми, що не перетинаються

- З пробілами (H2): 3.9466951243866752, R: 0.22423116442523217

о_,0.023776975486855256

е_,0.018094265430246722
и_,0.017515066136015466
а_,0.017018325746451783
_в,0.01679677953270638
_п,0.01597417744758892
_н,0.015951327389668993
_с,0.015743689906831372
то,0.014708482934980657
б,0.012016150023545495

- Без пробілів (H2): 4.131210488131707, R: 0.18102939810398588

То, 0.01813334416325146
 ов, 0.01258141848834251
 не, 0.012264882278391806
 на, 0.01209885007770068
 но, 0.011907733875466293
 ст, 0.011607920333211098
 по, 0.010910346195055584
 ко, 0.0106714509422626
 он, 0.010383582162647054
 от, 0.009749315266481682

Експерименти в програмі CoolPinkProgram

[illegible]

$$2,313759 < H < 2,999977$$

H(20)

[illegible]

$$1.473898 < H < 2.087860$$

H(30)

[illegible]

$$1,466,716 < H < 2,065,303$$

Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела

$$R = 1 - \frac{H_{\infty}}{H_0} H_0 = \log_2 32 = 5$$

H10: $0.537 < R < 0.4$

H20: $0.705 < R < 0.582$

H30: $0.706 < R < 0.586$

Висновок:

При виконанні даного комп'ютерного практикуму ми отримали важливі навички оцінки джерела символів. Визначили ентропію, надлишковість у різних моделях відкритого тексту та впевнились в коректності отриманих результатів.

Також ми навчилися працювати з програмою CoolPinkProgram, за допомогою якої змогли наближено обчислити значення $H(10)$, $H(20)$ та $H(30)$.

Отримані вміння можуть бути широко застосовані в сфері інформаційної безпеки, а саме в криптографії, тож ці навички є невід'ємною складовою кожного професіоналу в нашій справі.

