



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

ЛАБОРАТОРНА РОБОТА №1

з дисципліни

«Криптографія»

на тему: «Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконали:

студенти 3 курсу ФТІ

групи ФБ-74, ФБ-72

Демиденко Дар'я та Скуратов Ілля

Перевірили:

Чорний О.

Савчук М. М.

Завадська Л. О.

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.

Код програми

Результати виконання програми:

Монограми:

```
| н - 0.057492повторение: 57322 -
| е - 0.0712146повторение: 71004 -
| ђ - 0.0232688повторение: 23200 -
| р - 0.0356484повторение: 35543 -
| и - 0.0558822повторение: 55717 -
| љ - 0.0285524повторение: 28468 -
| ы - 0.0150024повторение: 14958 -
| й - 0.0086856повторение: 8660 -
| - 0.161101повторение: 160625 -
| с - 0.0445919повторение: 44460 -
| ь - 0.0143374повторение: 14295 -
| ѓ - 0.0130827повторение: 13044 -
| а - 0.0712627повторение: 71052 -
| ж - 0.00910793повторение: 9081 -
| т - 0.053416повторение: 53258 -
| о - 0.09199повторение: 91718 -
| к - 0.0253219повторение: 25247 -
| ђ - 0.0409651повторение: 40844 -
| д - 0.0260841повторение: 26007 -
| в - 0.035832повторение: 35726 -
| з - 0.0138149повторение: 13774 -
| у - 0.0228345повторение: 22767 -
| ч - 0.012865повторение: 12827 -
| ь - 0.0181988повторение: 18145 -
| я - 0.0204164повторение: 20356 -
| ш - 0.00674795повторение: 6728 -
| х - 0.0061552повторение: 6137 -
| щ - 0.00301592повторение: 3007 -
| ю - 0.00536085повторение: 5345 -
| ц - 0.00314831повторение: 3139 -
| ф - 0.000854527повторение: 852 -
| а - 0.00262777повторение: 2620 -
| ь - 0.000185549повторение: 185 -
| ё - 0.000934764повторение: 932 -
```

Количество символов: 836418

```
| н - 0.0685327повторение: 57322 -
| е - 0.0848906повторение: 71004 -
| ђ - 0.0277373повторение: 23200 -
| р - 0.0424943повторение: 35543 -
| и - 0.0666138повторение: 55717 -
| љ - 0.0340356повторение: 28468 -
| ы - 0.0178834повторение: 14958 -
| й - 0.0103537повторение: 8660 -
| с - 0.0531552повторение: 44460 -
| ь - 0.0170907повторение: 14295 -
| ѓ - 0.0155951повторение: 13044 -
| а - 0.084948повторение: 71052 -
| ж - 0.010857повторение: 9081 -
| т - 0.0636739повторение: 53258 -
| о - 0.109656повторение: 91718 -
| к - 0.0301847повторение: 25247 -
| ђ - 0.048832повторение: 40844 -
| д - 0.0310933повторение: 26007 -
| в - 0.0427131повторение: 35726 -
| з - 0.0164678повторение: 13774 -
| у - 0.0272196повторение: 22767 -
| ч - 0.0153356повторение: 12827 -
| ь - 0.0216937повторение: 18145 -
| я - 0.0243371повторение: 20356 -
| ш - 0.00804382повторение: 6728 -
| х - 0.00733724повторение: 6137 -
| щ - 0.00359509повторение: 3007 -
```

З пробілом

$H_1 = 4.37681$

Без пробіла

$H_1 = 4.45808$

Біграми:

Перетинаються, з пробілом – $H2 = 3.98984$

Не перетинаються, з пробілом – $H2 = 3.98924$

Перетинаються, без пробіла – $H2 = 4.14112$

Не перетинаються, без пробіла – $H2 = 4.14115$

2. Скріншоти результатів рожевої програми:

10-грамми:

Лабораторная работа №1

Произвольная часть текста:
бы_чтоб_в

Использованные буквы:

Порядок n-граммы:
5 символов
10 символов
15 символов
20 символов
25 символов
30 символов
35 символов
40 символов
45 символов
50 символов

Введенный символ:
Символ по счету:
Номер эксперимента: 51

Неравенство для энтропии:
 $2.68885248779197 < H < 3.40725542509089$

Двоичная таблица угаданных символов:

10000000000000000000000000000000	▲
00100000000000000000000000000000	
00000000000000000000000000000000	
00010000000000000000000000000000	
00000000000000000000000000000000	▼

Поле ввода символов:

Продолжить Другой

Вероятности:

$q(1) = 0.38$
$q(2) = 0.04$
$q(3) = 0.1$
$q(4) = 0.04$
$q(5) = 0.02$
$q(6) = 0.06$
$q(7) = 0.06$
$q(8) = 0.04$
$q(9) = 0.02$
$q(10) = 0.02$
$q(11) = 0$
$q(12) = 0$
$q(13) = 0$
$q(14) = 0$
$q(15) = 0.02$
$q(16) = 0.02$
$q(17) = 0$
$q(18) = 0$
$q(19) = 0.02$
$q(20) = 0$
$q(21) = 0.02$
$q(22) = 0$
$q(23) = 0.04$
$q(24) = 0.02$
$q(25) = 0.02$
$q(26) = 0.04$
$q(27) = 0$
$q(28) = 0.02$
$q(29) = 0$
$q(30) = 0$
$q(31) = 0$
$q(32) = 0$

Строка состояния:

20-грамми:

Лабораторная работа №1

Произвольная часть текста:
говорят_друг_друг_они_говорят_например_такие_вещи_как_бы_вам_понравилось_

Использованные буквы:

Порядок n-граммы:
5 символов
10 символов
15 символов
20 символов
25 символов
30 символов
35 символов
40 символов
45 символов
50 символов

Введенный символ: _ (пробел)
Символ по счету: 1
Номер эксперимента: 50

Неравенство для энтропии:
 $1.90454819185411 < H < 2.67433428621155$

Двоичная таблица угаданных символов:

00100000000000000000000000000000	▲
10000000000000000000000000000000	
10000000000000000000000000000000	
00001000000000000000000000000000	
10000000000000000000000000000000	▼

Поле ввода символов:

Продолжить Другой

Вероятности:

$q(1) = 0.52$
$q(2) = 0.12$
$q(3) = 0.06$
$q(4) = 0$
$q(5) = 0.04$
$q(6) = 0.04$
$q(7) = 0.02$
$q(8) = 0$
$q(9) = 0.02$
$q(10) = 0$
$q(11) = 0$
$q(12) = 0.02$
$q(13) = 0$
$q(14) = 0.02$
$q(15) = 0.02$
$q(16) = 0$
$q(17) = 0$
$q(18) = 0$
$q(19) = 0$
$q(20) = 0.02$
$q(21) = 0.04$
$q(22) = 0$
$q(23) = 0$
$q(24) = 0.02$
$q(25) = 0$
$q(26) = 0.02$
$q(27) = 0$
$q(28) = 0$
$q(29) = 0$
$q(30) = 0$
$q(31) = 0.02$
$q(32) = 0$

Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

30-грамми:

Лабораторная работа №1

Произвольная часть текста:
ся_биологическим_законам_так_

Использованные буквы:

Порядок n-граммы:
5 символов
10 символов
15 символов
20 символов
25 символов
30 символов
35 символов
40 символов
45 символов
50 символов

Введенный символ:
Символ по счету:
Номер эксперимента: 51

Неравенство для энтропии:
 $1.82941849582879 < H < 2.5688984057993$

Двоичная таблица угаданных символов:

10000000000000000000000000000000	▲
00010000000000000000000000000000	
00000000100000000000000000000000	
00010000000000000000000000000000	
10000000000000000000000000000000	▼

Поле ввода символов:

Продолжить Другой

Вероятности:

$q(1) = 0.5$
$q(2) = 0.08$
$q(3) = 0.08$
$q(4) = 0.14$
$q(5) = 0.02$
$q(6) = 0.02$
$q(7) = 0$
$q(8) = 0$
$q(9) = 0.02$
$q(10) = 0.04$
$q(11) = 0$
$q(12) = 0.02$
$q(13) = 0$
$q(14) = 0$
$q(15) = 0.02$
$q(16) = 0$
$q(17) = 0$
$q(18) = 0$
$q(19) = 0$
$q(20) = 0.02$
$q(21) = 0$
$q(22) = 0$
$q(23) = 0$
$q(24) = 0.02$
$q(25) = 0$
$q(26) = 0$
$q(27) = 0$
$q(28) = 0$
$q(29) = 0$
$q(30) = 0.02$
$q(31) = 0$
$q(32) = 0$

Строка состояния:

3. Оцінка надлишковості російської мови в різних моделях джерела.

З пробілом, H1: $R1=1-H1/ H0=0.124638$

Без пробіла, H1: $R1=1-H1/ H0=0.108384$

Перетинаються, з пробілом, H2: $R2=1-H2/ H0= 0.202032$

Не перетинаються, з пробілом, H2: $R2=1-H2/ H0=0.202152$

Перетинаються, без пробіла, H2: $R1=1-H2/ H0=0.171776$

Не перетинаються, без пробіла, H2 $R1=1-H2/ H0=0.17177$

Висновки:

Під час данного комп'ютерного практикуму, ми навчились визначати ентропію на символ джерела та його надлишковості. Порівняли різні моделі джерел відкритого тексту для наближеного визначення ентропії та набули практичних навичок оцінки ентропії на символ джерела.