

Міністерство освіти і науки України Національний технічний університет
України “Київський політехнічний інститут ім. Ігоря Сікорського” Фізико-
технічний інститут

Лабораторна робота №1 з предмету «Криптографія»
«Експериментальна оцінка ентропії на символ джерела відкритого тексту»

Виконав студент 3 курсу

Кравець Андрій ФБ-02

Київ - 2022

Мета: 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.

Постановка задачі:

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.

Результати виконання:

H_1 монограм з пробілом 4.379095399823697

надлишковість монограм з пробілом 0.13923786050740905

H_1 монограм без пробілів 4.479179556501572

надлишковість монограм без пробілів 0.11204805760275627

H_2 біграм з пробілами та з перетинами 3.9653930266671646

надлишковість біграм з пробілами та з перетинами 0.22055587423364942

H_2 біграм з пробілами та без перетинів 3.9639366646633825

надлишковість біграм з пробілами та без перетинів 0.22084213912623474

H_2 біграм без пробілів та з перетинами 4.16024639153056

надлишковість біграм без пробілів та з перетинами 0.1752733245871636

H_2 біграм без пробілів та без перетинів 4.159935381649082

надлишковість біграм без пробілів та без перетинів 0.17533497914351248

Оцінки $H^{(10)}$, $H^{(20)}$, $H^{(30)}$ у CoolPinkProgram

Произвольная часть текста:

оне_индии_китае_греции_и_риме_то_его_поразит_факт_насколько_эти_учения_были

Использованные буквы:

Порядок n-граммы:

- 5 символов
- 6 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: 51

Поле ввода символов:

Продолжить Другой

Неравенство для энтропии:
 $1,36338821925068 < H < 2,00731654048116$

Двоичная таблица угаданных символов:

00010000000000000000000000000000	▲
000000000000000000000001000000000000	
1000000000000000000000000000000000	
0100000000000000000000000000000000	
1000000000000000000000000000000000	▼

Вероятности:

q[1] = 0,56862
 q[2] = 0,17647
 q[3] = 0,03921
 q[4] = 0,11764
 q[5] = 0,01960
 q[6] = 0
 q[7] = 0
 q[8] = 0
 q[9] = 0
 q[10] = 0,0196
 q[11] = 0
 q[12] = 0
 q[13] = 0
 q[14] = 0
 q[15] = 0
 q[16] = 0
 q[17] = 0
 q[18] = 0
 q[19] = 0,0196
 q[20] = 0
 q[21] = 0
 q[22] = 0
 q[23] = 0
 q[24] = 0
 q[25] = 0
 q[26] = 0
 q[27] = 0
 q[28] = 0
 q[29] = 0,0196
 q[30] = 0,0196
 q[31] = 0
 q[32] = 0

Строка состояния:

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

$$0,6021325946 < R < 0,7303937861$$

Произвольная часть текста:
азличных_законов_и_

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

Поле ввода символов:

Продолжить

Другой

Неравенство для энтропии:

$$2,10609730976591 < H < 2,95578362989656$$

Двоичная таблица угаданных символов:

10000000000000000000000000000000	▲
10000000000000000000000000000000	
00000000000000000100000000000000	
10000000000000000000000000000000	
00001000000000000000000000000000	
~~~~~	▼

Вероятности:

q[1] = 0,48  
q[2] = 0,1  
q[3] = 0,04  
q[4] = 0,06  
q[5] = 0,02  
q[6] = 0,04  
q[7] = 0,02  
q[8] = 0  
q[9] = 0,04  
q[10] = 0,02  
q[11] = 0  
q[12] = 0  
q[13] = 0,02  
q[14] = 0,02  
q[15] = 0,02  
q[16] = 0  
q[17] = 0,02  
q[18] = 0,04  
q[19] = 0  
q[20] = 0  
q[21] = 0,02  
q[22] = 0  
q[23] = 0  
q[24] = 0  
q[25] = 0  
q[26] = 0  
q[27] = 0  
q[28] = 0  
q[29] = 0  
q[30] = 0,02  
q[31] = 0  
q[32] = 0,02

<b>Произвольная часть текста:</b>		
o_o_чем_я_собираюсь_говорить_		
<b>Использованные буквы:</b>		
<b>Порядок n-граммы:</b>	<b>Введенный символ:</b>	<b>Неравенство для энтропии:</b>
5 символов 10 символов 15 символов 20 символов 25 символов <b>30 символов</b> 35 символов 40 символов 45 символов 50 символов	<b>Символ по счету:</b>	<b>1,46569495757881 &lt; H &lt; 2,10182279672496</b>
	<b>Номер эксперимента:</b> 52	<b>Двоичная таблица угаданных символов:</b>
	<b>Поле ввода символов:</b>	01000000000000000000000000000000 ^ 10000000000000000000000000000000   00001000000000000000000000000000 v 10000000000000000000000000000000 00000000000000000000000010000000000 ~~~~~~
<input type="button" value="Продолжить"/> <input type="button" value="Другой"/>		
<b>Строка состояния:</b>		

0,7092002000 12 0,001027072

**DISCUSSION:**

_____