

Комп'ютерний практикум №1 з КRYPTOграфії

Грипас Владислав ФБ-12

Мета роботи: Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення на порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навчальних знань щодо оцінки ентропії на символ джерела

Постановка задачі:

- 1) Написати програму для підрахунку частот букв і 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 надовільно обраному тексті російською мовою достатньої довжини, де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, де вилучені 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:

Мною був вибраний текстовий файл зі змістом книги "Брати Карамазови"

Для початку я ініціалізую алфавіт, що буде використовуватись в подальших обрахунках, та створюю його копію, додаючи пробіл, далі, обрахувавши необхідні показники, виводжу головні з них в вивід терміналу, записуючи інші дані в файли з відповідними назвами

При виконанні практику майже не виникало труднощів, за винятком не очевидних позначень, та великої кількості необхідних експериментів в CoolPinkProgram

Приклад виводу після виконання скрипта:

```
[tdoors:...-23-24/cp1/hrypas_fb-12_cp3]$ ./main.py
Total letters: 1407632
Letter entropy is: 4.459193
Bigram entropy without step is: 8.2807
Bigram entropy with step is: 8.7230
Bigram Entropy with space is 7.8698
Bigram Entropy with step and space is 8.3769
*-----*
H1 is 4.4592
H2 is 4.1403
H2Step is 4.3615
H2Space is 3.9349
H2StepSpace is 4.1885
*-----*
Redurancy 1 is: 0.0999
Redurancy 2 without step is: 0.1643
Redurancy 2 with step is: 0.1196
Redurancy 2 with space is: 0.2130
Redurancy 2 with step and space is: 0.1623
*-----*
```

Таким чином, в виводі надані необхідні дані, такі як :
H1 та H2, оцінка надлишковості (Redurancy)
Приклад виконання CoolPinkProgramm для 10 символів:

Лабораторная работа №1

Произвольная часть текста:
ся_как_сл

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:
2.54132912671696 < H < 3.22791993994523

Двоичная таблица угаданных символов:
00000000000000010000000000000000
00000010000000000000000000000000
00000100000000000000000000000000
10000000000000000000000000000000
00000000000000000000000000000100

Вероятности:

q[1] = 0.46
q[2] = 0.06
q[3] = 0.04
q[4] = 0.02
q[5] = 0
q[6] = 0.04
q[7] = 0.04
q[8] = 0
q[9] = 0
q[10] = 0.02
q[11] = 0
q[12] = 0
q[13] = 0.02
q[14] = 0
q[15] = 0.02
q[16] = 0.04
q[17] = 0.04
q[18] = 0.02
q[19] = 0.02
q[20] = 0.04
q[21] = 0
q[22] = 0
q[23] = 0.02
q[24] = 0.02
q[25] = 0.02
q[26] = 0.02
q[27] = 0
q[28] = 0
q[29] = 0.02
q[30] = 0.02
q[31] = 0
q[32] = 0

Строка состояния:

Приклад виконання CoolPinkProgramm для 20 символів:

Приклад виконання CoolPinkProgramm для 30 символів:

Лабораторная работа №1

Произвольная часть текста:

_кто_живет_вокруг_или_вообще_

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:

1.61013135620676 < H < 2.43319752351267

Двоичная таблица угаданных символов:

00000000000000000100000000000000

00000100000000000000000000000000

01000000000000000000000000000000

00000000000000000000000000000001

10000000000000000000000000000000

Вероятности:

q[1] = 0.58

q[2] = 0.14

q[3] = 0.02

q[4] = 0.02

q[5] = 0.02

q[6] = 0.02

q[7] = 0.02

q[8] = 0.02

q[9] = 0

q[10] = 0.02

q[11] = 0

q[12] = 0.02

q[13] = 0

q[14] = 0

q[15] = 0

q[16] = 0.02

q[17] = 0

q[18] = 0.02

q[19] = 0

q[20] = 0.02

q[21] = 0.02

q[22] = 0

q[23] = 0.02

q[24] = 0

q[25] = 0

q[26] = 0

q[27] = 0

q[28] = 0

q[29] = 0

q[30] = 0

q[31] = 0

q[32] = 0.02

Строка состояния:

Таблиця частот букв:

о: 0.1139	к: 0.0329	ж: 0.0111
е: 0.0889	д: 0.0317	й: 0.0101
а: 0.0774	у: 0.0288	ш: 0.0097
и: 0.0672	п: 0.0268	х: 0.0075
т: 0.0656	ь: 0.0210	ю: 0.0066
н: 0.0621	я: 0.0205	э: 0.0036
с: 0.0529	г: 0.0192	ц: 0.0036
в: 0.0467	ч: 0.0179	щ: 0.0029
л: 0.0453	б: 0.0179	ф: 0.0018
р: 0.0410	ы: 0.0170	
м: 0.0330	з: 0.0154	

Таблиця частот для біграм:

('т', 'о'): 0.0180	('к', 'о'): 0.0095	('р', 'а'): 0.0077
('н', 'е'): 0.0126	('о', 'с'): 0.0095	('е', 'с'): 0.0076
('с', 'т'): 0.0122	('а', 'л'): 0.0094	('л', 'и'): 0.0075
('о', 'в'): 0.0121	('о', 'н'): 0.0092	('и', 'т'): 0.0074
('н', 'о'): 0.0119	('п', 'р'): 0.0084	('е', 'т'): 0.0074
('н', 'а'): 0.0110	('р', 'о'): 0.0083	('т', 'ь'): 0.0073
('е', 'н'): 0.0104	('в', 'о'): 0.0082	('о', 'р'): 0.0072
('п', 'о'): 0.0104	('к', 'а'): 0.0082	('о', 'м'): 0.0072
('г', 'о'): 0.0104	('н', 'и'): 0.0079	('е', 'р'): 0.0071
('о', 'т'): 0.0100	('а', 'к'): 0.0077	('т', 'е'): 0.0071

Таблиця частот біграм з кроком:

('o', 'o'): 0.0180	('и', 'o'): 0.0084	('е', 'ь'): 0.0056
('е', 'o'): 0.0129	('a', 'a'): 0.0081	('с', 'a'): 0.0055
('o', 'е'): 0.0114	('и', 'е'): 0.0077	('ч', 'o'): 0.0055
('a', 'o'): 0.0102	('с', 'o'): 0.0076	('е', 'т'): 0.0055
('o', 'и'): 0.0099	('o', 'и'): 0.0075	('т', 'л'): 0.0052
('е', 'е'): 0.0097	('е', 'a'): 0.0073	('a', 'т'): 0.0052
('o', 'a'): 0.0093	('o', 'т'): 0.0073	('в', 'р'): 0.0050
('и', 'a'): 0.0090	('е', 'и'): 0.0072	('и', 'и'): 0.0050
('a', 'е'): 0.0090	('е', 'и'): 0.0069	('л', 'и'): 0.0049
('a', 'и'): 0.0087	('в', 'е'): 0.0067	('o', 'с'): 0.0049

Таблиця частот біграм з пробілом:

(' ', ' '): 0.0260	(' ', 'и'): 0.0119	('г', 'o'): 0.0083
('o', ' '): 0.0235	('ь', ' '): 0.0108	(' ', 'к'): 0.0078
('е', ' '): 0.0187	('я', ' '): 0.0107	('м', ' '): 0.0077
('и', ' '): 0.0173	(' ', 'o'): 0.0105	('a', 'л'): 0.0074
(' ', 'в'): 0.0163	('и', 'е'): 0.0101	('к', 'o'): 0.0074
('a', ' '): 0.0157	('с', 'т'): 0.0096	('o', 'в'): 0.0073
(' ', 'и'): 0.0156	(' ', 'т'): 0.0094	(' ', 'д'): 0.0072
(' ', 'с'): 0.0154	('и', 'o'): 0.0094	('у', ' '): 0.0071
(' ', 'п'): 0.0152	('и', 'a'): 0.0088	('л', ' '): 0.0070
('т', 'o'): 0.0142	('п', 'o'): 0.0084	('е', 'и'): 0.0070

Таблиця біграм з пробілом та кроком:

(' ', ' '): 0.0404	('o', 'o'): 0.0116	(' ', 'ы'): 0.0063
(' ', 'o'): 0.0271	('e', ' '): 0.0116	(' ', 'и'): 0.0063
(' ', 'a'): 0.0221	(' ', 'p'): 0.0095	('e', 'e'): 0.0057
(' ', 'e'): 0.0217	('л', ' '): 0.0084	('a', 'o'): 0.0057
('т', ' '): 0.0176	('o', 'и'): 0.0075	('г', ' '): 0.0057
('н', ' '): 0.0173	('с', ' '): 0.0074	(' ', 'с'): 0.0057
('o', ' '): 0.0143	('e', 'o'): 0.0073	('o', 'н'): 0.0054
('a', ' '): 0.0121	('a', 'и'): 0.0067	('д', ' '): 0.0054
('и', ' '): 0.0116	(' ', 'н'): 0.0066	('с', 'o'): 0.0053
(' ', 'т'): 0.0116	('o', 'e'): 0.0065	('e', 'и'): 0.0053

Висновки:

В даній лабораторній роботі я навчився обраховувати ентропію відкритого тексту та виродженість мови, також експериментальним шляхом я визначив частоти літер та різних біграм в доволі великому тексті, таким чином, можна стверджувати, що такі дані будуть актуальними для усієї мови