

СИМЕТРИЧНА КРИПТОГРАФІЯ

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

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. А також успішне закінчення предмету і успіх у житті.

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

Найбільші труднощі виникли з пошуком файлу через обмеження в його розмірі в 1Мб. Але рішення досить швидко було знайдено, було використано повне зібрання (Старий + Новий заповіт) розмірами 7.5Мб. Наступним випробуванням був запуск програми CoolPinkProgram.exe, через її недоступність на пристроях тас. При написанні програмної частини лабораторної роботи проблем не виникало. Мова виконання: typescript.

Таблиця частот входжень літер руснявого алфавіту в тексті:

Л	Частота	Л	Частота	Л	Частота	Л	Частота
probil	0.18300703901939747	Е	0.07279067875670077	С	0.04623784183418146	Л	0.03666756593674768
О	0.09063020641337455	А	0.06132910401808975	Н	0.04559273568255629	Р	0.03325952110779091
И	0.0741004553327059	Т	0.04697793767852399	В	0.04156506974148388	Д	0.029359537788907677
М	0.026786578497010785	У	0.02170733289536821	П	0.02136495812215932	К	0.021295453468951498
Г	0.0186359994975071	Я	0.016904818783349363	Ы	0.015708566474250327	Б	0.015351518496475342
З	0.013336913252014606	Ь	0.012846261885295699	Х	0.009002139713620235	Ч	0.008397964079995221
Й	0.00793871851954059	Ж	0.0076079278551996685	Ш	0.006406527053270426	Ю	0.00594547952032522
Ц	0.003933705946921098	Щ	0.003018561346351469	Ф	0.0014675778812510425	Э	0.0008253034006824843

Без пробілів:

Л	Частота	Л	Частота	Л	Частота	Л	Частота
-	-	Е	0.0870586453192643	С	0.055301089937970734	Л	0.04385490933911908
О	0.10839496388989721	А	0.07335044549186372	Н	0.05452953417551754	Р	0.039778841206989074
И	0.08862515598141986	Т	0.056186254669436386	В	0.049712390736035256	Д	0.03511440792639617
М	0.03203711349474428	У	0.02596226642811862	П	0.02555278152625799	К	0.02546965301234643
Г	0.02228890976338238	Я	0.0202183940002149	Ы	0.018787659910781553	Б	0.018360625655984036
З	0.01595113028608832	Ь	0.0153643045545497	Х	0.010766681968470894	Ч	0.010044079664210065
Й	0.009494815705624075	Ж	0.009099185556082036	Ш	0.007662293798951534	Ю	0.007110874656671541
Ц	0.004704766004009257	Щ	0.0036102405708404264	Ф	0.001755243177073279	Э	0.0009870741318535226

Таблиця частот біграм з перетином та без:

Багато... Дуже багато...

Найчастіші:

З пробілами				Без пробілів			
З перетином		Без перетину		З перетином		Без перетину	
И_	0.028860913 687812306	И_	0.02884752 0176943402	ГО	0.01473426 415194793	ГО	0.018626528 692380056
_И	0.024053764 694767662	_И	0.02398013 505526392	ТО	0.01237394 2402825288	ВО	0.015202498 974841402
_С	0.017320049 188713164	_С	0.01738337 119191728	ВО	0.01235492 7995608939	НА	0.015093952 770340353
О_	0.017046149 299637468	О_	0.01703018 458376495	НА	0.011980344 173446887	ПО	0.015081892 080951348
Е_	0.016773279 109392886	Е_	0.01673826 5040292105	ПО	0.011893511 713825565	ТО	0.014947616 405753752

Ентропії

З пробілами			Без пробілів		
Літери	Біграми з перетином	Біграми без перетину	Літери	Біграми з перетином	Біграми без перетину
4.307694435 28686	3.91812712 2713653	3.9178001 875692368	4.48810406 6468152	4.1536318 90654613	4.1537137 67057074

Надлишковість

З пробілами			Без пробілів		
Літери	Біграми з перетином	Біграми без перетину	Літери	Біграми з перетином	Біграми без перетину
0.13846111294 262808	0.21637457 545726935	0.21643996 24861526	0.094080293 69799553	0.161593196 87308582	0.161576670 1960366

Оцінки для значень $H^{(10)}$, $H^{(20)}$, $H^{(30)}$

[illegible]

Лабораторная работа №1

Произвольная часть текста:

отношения_итак_воз

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:

2.20803703690249; $H < 3.01648377899751$

Двоичная таблица угаданных символов:

00000000010000000000000000000000

10000000000000000000000000000000

00000100000000000000000000000000

00010000000000000000000000000000

01000000000000000000000000000000

Вероятности:

q[1] = 0.44

q[2] = 0.14

q[3] = 0.1

q[4] = 0.02

q[5] = 0

q[6] = 0.02

q[7] = 0.02

q[8] = 0.02

q[9] = 0

q[10] = 0.02

q[11] = 0.02

q[12] = 0

q[13] = 0

q[14] = 0.02

q[15] = 0.04

q[16] = 0

q[17] = 0.02

q[18] = 0.02

q[19] = 0

q[20] = 0.02

q[21] = 0

q[22] = 0

q[23] = 0

q[24] = 0

q[25] = 0

q[26] = 0.02

q[27] = 0.02

q[28] = 0.02

q[29] = 0

q[30] = 0

q[31] = 0.02

q[32] = 0

Строка состояния:

[illegible]

Вибачте за якість, віндоус-ноутбук 2011 року народження)

Наочно побачили, як зі збільшенням кількості відображених символів у тексті (ми точно мали це отримати), ентропія буде знижуватись

Для знаходження оцінки надлишковості російської мови в різних моделях джерела,

підставимо в формулу: $R = 1 - \frac{H_{\infty}}{H_0}$, на місце H_{∞} отримані вище значення H_1 , H_2 , $H_{(10)}$, $H_{(20)}$, $H_{(30)}$

$$R_{10} = 1 - (2.01916080695114 + 2.63663484475755) / 2 \log_2(32) = 0.5344$$

$$R_{20} = 1 - (2.20803709890249 + 3.01648377899751) / 2 \log_2(32) = 0.4775$$

$$R_{30} = 1 - (1.49590771891671 + 2.34749895285668) / 2 \log_2(32) = 0.6156$$

Висновок:

Під час виконання цього комп'ютерного практикуму, що віндоус-ноутбук є необхідною навичкою на 3 курсі, без якої виконання робіт унеможлиблюється, або значно ускладнюється. Вивчили поняття ентропія, частота, pink, typescript, main, log та багато інших. Перечитали святе письмо. Перелічили символи з тексту, для перевірки показників програми. Визначили значення надлишковості для російської мови, зазначені вище. Втомилися