



Міністерство освіти і науки України
Національний технічний університет України
“Київський політехнічний інститут імені Ігоря Сікорського”

КРИПТОГРАФІЯ КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Тема: “Експериментальна оцінка ентропії на символ джерела відкритого
тексту”

Виконали: студенти Оласюк Олександр
групи ФБ-32 та Гарбар Дар'я
групи ФБ-33

Київ 2025

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.

Хід роботи:

Був написаний скрипт для розрахунку ентропії та надлишковості російської мови в різних моделях джерела

Значення ентропії (H) та надлишковості (R):

```
=====ENTROPY=====
H1, 4.463556505114775
H1 (with spaces), 4.376529954802351
H2, 3.9624065753489957
H2 (with spaces), 3.968681393177802
H2 (with intersection), 3.9627231619672316
H2 (with all)), 3.9690879769887886
=====REDUNDANCY=====
R1-H1, 0.0990351965349926
R1-H1 (with spaces), 0.12469400903952987
R2-H2, 0.20019185209889878
R2-H2 (with spaces), 0.2062637213644396
R2-H2 (with intersection), 0.20012794937918377
R2-H2 (with all)), 0.20618240460224224
```

Частота символів:

(з пробілом)

Monogram	Frequency
о	0,163429967
а	0,09486411
е	0,070553587
и	0,068691441
н	0,055235783
т	0,054868485
с	0,048452252
л	0,044522705
в	0,041635637
р	0,03830565
д	0,037523792
к	0,02884368
м	0,025454953
у	0,025207837
п	0,023043886
я	0,021339732
ь	0,018817261
г	0,016841685
ы	0,016767415
б	0,015837692
з	0,014739175
ч	0,014481931
й	0,011744752
ж	0,009887333
ш	0,008566006
х	0,007936063
ю	0,007179187
ц	0,005475708
э	0,003055864
щ	0,002594041
ф	0,002418494
	0,001683898

(без пробілу)

Monogram	Frequency
о	0,113396495
а	0,084336737
е	0,082110807
и	0,06602649
н	0,065587438
т	0,057917747
с	0,053220536
л	0,049769457
в	0,045788934
р	0,044854334
к	0,034478501
д	0,030427761
м	0,030132369
у	0,027545675
п	0,025508602
я	0,022493348
ь	0,020131829
г	0,02004305
ы	0,018931699
б	0,017618578
з	0,01731108
ч	0,014039174
й	0,011818894
ж	0,010239437
ш	0,00948643
х	0,008581692
ю	0,006545427
ц	0,003652849
э	0,003100805
щ	0,002890964
ф	0,00201286

Частота біграм з перекриттям і з пробілами(інші результати у файлі calculatings.exe):

		а	б	в	г	д	е	ж	з	и	й	к	л	м	н
	0	0,002928	0,00731	0,014936	0,00489	0,007316	0,004106	0,001921	0,004042	0,010827	6,08E-06	0,010147	0,002604	0,005897	0,015446
а	0,01737	3,38E-06	0,000606	0,003169	0,00087	0,001878	0,000966	0,001079	0,003885	0,00012	0,000739	0,00421	0,008619	0,002707	0,004844
б	0,00029	0,001051	1,69E-05	7,76E-05	7,43E-06	1,96E-05	0,002042	6,75E-06	2,7E-06	0,000743	0	0,00024	0,000824	4,66E-05	0,000261
в	0,006378	0,005541	1,35E-05	3,11E-05	2,9E-05	0,000287	0,004337	6,75E-07	0,000485	0,002947	0	0,000141	0,000913	0,000126	0,001178
г	0,000897	0,000941	0	4,32E-05	0	0,000927	0,000521	0	0	0,000739	0	7,9E-05	0,001527	6,08E-06	0,000255
д	0,001055	0,004157	3,38E-05	0,00083	7,43E-06	3,04E-05	0,004399	1,28E-05	4,73E-06	0,002389	0	0,000185	0,000698	0,000128	0,001588
е	0,016825	3,04E-05	0,001091	0,001456	0,003319	0,002485	0,001557	0,000984	0,001203	0,000167	0,00248	0,001001	0,005915	0,003874	0,007035
ж	0,00023	0,00123	5,81E-05	0	1,82E-05	0,000683	0,003556	1,28E-05	0	0,001238	0	8,24E-05	5,4E-06	6,08E-06	0,001047
з	0,001281	0,005202	0,000139	0,000743	0,000419	0,000735	0,000237	0,000175	8,78E-06	0,000329	0	0,000128	0,000201	0,000248	0,00159
и	0,018008	7,76E-05	0,000395	0,002418	0,000513	0,001599	0,002412	0,00032	0,001762	0,00069	0,001331	0,001736	0,004657	0,002785	0,003134
й	0,008033	0	2,03E-06	0	0	0,000161	6,08E-06	0	0	6,75E-07	0	6,28E-05	0,000135	3,51E-05	0,000381
к	0,004659	0,007107	0	0,000203	2,7E-06	1,42E-05	0,000458	1,49E-05	6,08E-06	0,002432	0	2,3E-05	0,000363	5,4E-06	0,001789
л	0,007486	0,00694	2,03E-05	8,1E-06	0,000134	0,000194	0,003738	0,000346	4,73E-06	0,005857	0	0,000605	9,93E-05	3,38E-06	0,00025
м	0,00747	0,002845	1,55E-05	0	4,05E-05	0	0,002771	0	7,43E-06	0,003368	0	7,09E-05	8,91E-05	2,7E-05	0,001213
н	0,004244	0,010698	1,76E-05	1,62E-05	0,000138	0,001032	0,008883	1,89E-05	2,09E-05	0,007158	0	0,000209	4,73E-06	3,38E-06	0,002672
о	0,021314	6,08E-06	0,003154	0,007636	0,004509	0,004429	0,002235	0,00159	0,000851	0,000818	0,003753	0,00155	0,006568	0,004967	0,006578
п	3,71E-05	0,001055	0	0	1,22E-05	0	0,002439	0	0	0,000798	0	6,14E-05	0,000614	0	5,94E-05
р	0,001369	0,00814	0,000164	0,000295	0,000314	0,000371	0,005892	0,000336	4,19E-05	0,005075	0	0,000222	7,29E-05	0,000202	0,000625
с	0,003048	0,001448	4,86E-05	0,001788	1,96E-05	0,000313	0,003468	2,09E-05	2,97E-05	0,001481	0	0,004649	0,002276	0,000904	0,000761
т	0,004429	0,005595	2,5E-05	0,002383	9,45E-06	0,000129	0,004462	6,75E-07	9,45E-06	0,003253	0	0,000392	0,000274	3,31E-05	0,000932
у	0,006423	2,03E-05	0,000547	0,000849	0,001179	0,001701	0,000188	0,00124	0,000653	1,69E-05	8,78E-05	0,000825	0,001263	0,000995	0,000255
ф	0,000221	0,00016	0	0	0	0	0,000159	0	0	0,000514	0	4,73E-06	6,14E-05	2,03E-06	6,75E-07
х	0,003251	0,000895	0	0,000133	0	0	3,71E-05	0	0	0,000153	0	0	7,22E-05	4,46E-05	0,000113
ц	0,000267	0,000615	0	1,96E-05	1,22E-05	0	0,000909	0	0	0,000221	0	0,000128	2,7E-06	0	1,08E-05
ч	0,000175	0,002038	0	4,05E-06	0	0	0,00308	0	0	0,00133	0	0,000226	4,25E-05	2,03E-06	0,000465
ш	0,000127	0,001469	0	2,03E-05	0	0	0,002175	0	0	0,001764	0	0,000469	0,00044	1,55E-05	0,000365
щ	8,78E-06	0,000363	0	0	0	0	0,001248	0	0	0,000674	0	0	0	0	2,63E-05
ь	0,004196	0	0,000444	0,00099	0,000101	0,000103	0,000953	2,57E-05	5,27E-05	1,15E-05	0,001443	0,000175	0,002289	0,001456	0,0003
ъ	0,010092	0	0,000129	5,4E-06	3,98E-05	6,21E-05	0,001356	0	0,000138	0,000104	0	0,001068	0	0,000305	0,000995
э	1,42E-05	0	0	6,75E-07	6,08E-06	3,38E-06	0	0	4,05E-06	0	1,69E-05	4,46E-05	0,000109	6,08E-06	3,51E-05
ю	0,003214	6,75E-07	0,000472	0	2,7E-06	0,000338	2,7E-06	9,45E-06	3,24E-05	1,35E-06	0	1,55E-05	6,21E-05	3,65E-05	4,93E-05
я	0,011018	0	3,78E-05	0,000248	0,000173	0,000645	9,72E-05	0,000452	0,001238	1,55E-05	3,04E-05	9,45E-05	0,000832	0,000342	0,000615

о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я
0,011889	0,015265	0,004281	0,015984	0,007384	0,003759	0,000683	0,001152	0,00031	0,005226	0,000835	5,47E-05	2,03E-06	1,35E-06	0,002516	6,08E-05	0,001648
4,73E-06	0,000795	0,002801	0,003622	0,004985	0,000115	0,000575	0,000956	7,16E-05	0,000712	0,001358	0,000208	0	0	6,75E-07	0,000772	0,002513
0,002184	0	0,001198	5,54E-05	6,75E-06	0,001098	0	5,81E-05	4,73E-06	1,89E-05	4,73E-06	0,0002	0,003688	0,000133	6,75E-07	6,75E-06	0,000455
0,007062	0,00024	0,000556	0,002996	0,000207	0,00076	0	5E-05	1,08E-05	3,78E-05	0,001055	6,75E-06	0,002552	0,000188	1,35E-06	0	0,000177
0,008827	0	0,001275	5,13E-05	2,03E-06	0,000645	0	0	0	2,5E-05	6,75E-06	0	0	0	0	6,75E-07	0
0,003595	7,97E-05	0,001944	0,000312	0,000257	0,001696	5,4E-06	4,32E-05	0,000174	2,57E-05	0,000112	6,75E-07	0,000593	0,000591	6,75E-07	7,49E-05	0,000431
0,00035	0,000869	0,006576	0,003593	0,003641	6,89E-05	1,01E-05	0,000925	0,000253	0,00093	0,000737	0,000618	0	0	0	0,000321	0,000377
5,47E-05	0	6,75E-07	2,16E-05	0	0,000177	0	0	0	5,87E-05	0	0	0	4,52E-05	6,75E-07	3,98E-05	0
0,000541	0	0,000168	0,000143	1,01E-05	0,000275	0	0	4,05E-06	1,42E-05	8,1E-06	0	0,000492	0,000815	0	9,11E-05	0,000482
0,000213	0,000182	0,000606	0,002794	0,003316	4,05E-06	3,44E-05	0,001591	0,001136	0,001146	0,00036	0,000125	0	0	0	0,000337	0,001552
1,15E-05	0	2,5E-05	0,000585	0,000255	0	0	6,75E-07	4,66E-05	6,41E-05	8,24E-05	6,75E-07	0	0	0	6,75E-07	0
0,008232	0	0,001553	0,000135	0,00041	0,001391	0	1,35E-06	1,08E-05	1,62E-05	4,73E-06	0	0	2,03E-06	0	1,01E-05	0
0,005803	7,43E-05	0	0,001304	5,2E-05	0,00119	5,4E-06	0	0	0,000185	7,43E-06	2,03E-06	0,001088	0,003462	6,75E-07	0,001053	0,001723
0,003431	0,000196	4,59E-05	0,000116	1,15E-05	0,002252	2,03E-06	0	2,3E-05	2,3E-05	4,73E-06	6,08E-06	0,000689	4,19E-05	6,75E-07	3,04E-05	0,000417
0,008264	4,73E-06	5,4E-05	0,000582	0,000466	0,002341	1,55E-05	1,42E-05	0,000503	0,000217	3,11E-05	0,000124	0,003028	0,001031	1,35E-06	0,000148	0,002928
0,000144	0,001041	0,00596	0,006548	0,00636	5,6E-05	0,000266	0,000597	8,98E-05	0,00134	0,001123	0,000145	0	0	3,04E-05	0,000596	0,000612
0,008261	4,93E-05	0,005963	6,75E-06	3,98E-05	0,00063	2,03E-06	0	2,7E-06	2,03E-05	7,43E-06	1,35E-06	0,000201	0,000736	2,03E-06	0	0,000342
0,006652	3,92E-05	2,3E-05	0,00024	0,000579	0,002792	1,89E-05	0,000104	8,03E-05	7,83E-05	0,00024	5,4E-05	0,001506	0,000895	1,35E-06	0,000169	0,000933
0,002723	0,001707	0,000237	0,000783	0,01097	0,000734	1,82E-05	0,000182	2,84E-05	0,000367	7,56E-05	0	0,000379	0,00298	6,75E-07	0,000112	0,002971
0,013645	5,67E-05	0,002737	0,000825	7,49E-05	0,001444	1,76E-05	1,42E-05	0,000137	0,000163	3,38E-06	1,82E-05	0,001468	0,005388	2,7E-06	6,55E-05	0,000465
1,28E-05	0,000547	0,000619	0,00129	0,001242	3,38E-06	1,55E-05	0,000323	5,4E-06	0,000726	0,000724	0,000234	0	0	3,17E-05	0,000964	6,28E-05
6,28E-05	0	0,00039	8,78E-06	8,78E-06	6,95E-05	5,4E-06	0	0	6,75E-07	0	0	1,28E-05	2,03E-06	6,75E-07	6,75E-07	0
0,002222	6,75E-07	8,71E-05	4,25E-05	1,28E-05	0,000111	0	0	0	6,75E-07	2,03E-06	0	0	1,35E-06	0	0	0
0,000339	0	0	0	0	0,000391	0	0	6,75E-07	0	0	0	0,00014	0	0	6,75E-07	0
3,78E-05	0	2,03E-05	0	0,003439	0,000618	0	0	0	0	9,32E-05	0	0	0,000175	6,75E-07	0	0
0,00026	4,66E-05	2,03E-06	2,03E-06	9,99E-05	0,000345	0	0	2,03E-06	0	2,03E-06	0	0	0,000332	0	0	0
6,75E-07	0	1,35E-06	0	0	7,36E-05	0	0	0	0	0	0	0	2,23E-05	6,75E-07	0	0
0	0,000115	0,000279	0,000704	0,000587	4,05E-06	0	0,000978	3,38E-06	0,000112	0,000512	1,35E-06	0	0	0	0	1,35E-06
4,39E-05	0	0	0,000837	8,84E-05	0	4,05E-06	6,75E-07	8,64E-05	3,78E-05	0,000399	9,45E-06	0	0	0	0,000479	0,000561
0	1,49E-05	3,31E-05	4,05E-05	0,002256	0	1,35E-06	6,75E-06	6,75E-07	0	6,75E-07	0	0	0	0	0	0
0	4,05E-06	5,27E-05	0,00014	0,00035	0	0	5,4E-06	2,97E-05	6,08E-05	0,000138	0,000422	0	0	0	3,58E-05	0
0	1,28E-05	3,71E-05	0,00076	0,001341	0	3,38E-06	0,000176	4,12E-05	0,000138	8,1E-06	0,000188	0	0	0	0,000106	0,000165

Аналіз через CoolPinkProgram.exe
H(10):

Лабораторная работа №1

Произвольная часть текста:
человека_без_поддержки_в_воздухе_ч_него_будет_не_больше_свободы_выбора_чем_

Использованные буквы:
к, п, р, с, т, в, л, д, ч, е,

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: б

Символ по счету: 11

Номер эксперимента: 50

Поле ввода символов:
б

Продолжить Другой

Неравенство для энтропии:
2,53007945178473< H < 3,27016036564542

Двоичная таблица угаданных символов:
00000000000100000000000000000000
00000000000000010000000000000000
00000000000000000000000100000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

q[1]=0,38
q[2]=0,1
q[3]=0,1
q[4]=0
q[5]=0,06
q[6]=0,02
q[7]=0,02
q[8]=0,02
q[9]=0,02
q[10]=0,04
q[11]=0,04
q[12]=0,02
q[13]=0,04
q[14]=0
q[15]=0
q[16]=0,04
q[17]=0
q[18]=0
q[19]=0,02
q[20]=0
q[21]=0
q[22]=0
q[23]=0,02
q[24]=0,04
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:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

H(20):

Лабораторная работа №1

Произвольная часть текста:
_слуха_но_рассматривая_человечество_в_целом_люди_полагали_что_человеческая_

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: в

Символ по счету: 1

Номер эксперимента: 50

Поле ввода символов:
в

Продолжить Другой

Неравенство для энтропии:
2,09369203404264< H < 2,74291016606417

Двоичная таблица угаданных символов:
01000000000000000000000000000000
01000000000000000000000000000000
01000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

q[1]=0,42
q[2]=0,22
q[3]=0,04
q[4]=0,06
q[5]=0,02
q[6]=0,02
q[7]=0
q[8]=0,04
q[9]=0
q[10]=0,02
q[11]=0
q[12]=0
q[13]=0
q[14]=0,02
q[15]=0
q[16]=0,02
q[17]=0
q[18]=0
q[19]=0,04
q[20]=0
q[21]=0
q[22]=0
q[23]=0
q[24]=0,04
q[25]=0
q[26]=0
q[27]=0
q[28]=0
q[29]=0
q[30]=0,04
q[31]=0
q[32]=0

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

$H(30)$:

Лабораторная работа №1

Произвольная часть текста:
ются_людьми_которые_убегают_с_поля_битвы_или_где_человек_гордится_тем_что_о

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:
 $1.4170205694229 < H < 2.13972183240961$

Двоичная таблица угаданных символов:
10000000000000000000000000000000
10000000000000000000000000000000
01000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Поле ввода символов:
Продолжить Другой

Вероятности:
 $q[1] = 0.6$
 $q[2] = 0.16$
 $q[3] = 0.02$
 $q[4] = 0.04$
 $q[5] = 0$
 $q[6] = 0$
 $q[7] = 0.02$
 $q[8] = 0$
 $q[9] = 0.02$
 $q[10] = 0.02$
 $q[11] = 0.02$
 $q[12] = 0.02$
 $q[13] = 0$
 $q[14] = 0$
 $q[15] = 0.02$
 $q[16] = 0$
 $q[17] = 0.04$
 $q[18] = 0$
 $q[19] = 0$
 $q[20] = 0$
 $q[21] = 0$
 $q[22] = 0$
 $q[23] = 0$
 $q[24] = 0$
 $q[25] = 0$
 $q[26] = 0$
 $q[27] = 0.02$
 $q[28] = 0$
 $q[29] = 0$
 $q[30] = 0$
 $q[31] = 0$
 $q[32] = 0$

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

Висновок:

Аналізуючи в CoolPinkProgram.exe прийшли до висновку, що короткі відрізки тексту більш випадкові, а при збільшенні довжини ентропія зменшується, що менша ентропія — текст більш передбачуваний.

Аналізуючи результати отриманої скриптом ентропії та надлишковості, для великих текстів, було визначено ентропію, що становить приблизно 4.4 біт/символ та надлишковість – приблизно 10-20%. При переході до біграмної моделі ентропія зменшується, що каже про наявність залежностей між символами, що і є ознакою певної мови.