

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

Криптографія

Комп'ютерний практикум №1
Експериментальна оцінка ентропії на символ
джерела відкритого тексту

Виконали:
Студенти 3 курсу
ФБ-32 Баласанян Юліана та
ФБ-32 Дорогін Артем

скрипт `crypto_lab1.py` робить:

- очищення тексту
- обчислення частот
- обчислення ентропії
- обчислення надлишковості
- обробка двох варіантів тексту
- запис результатів у Excel
- обчислення H_0 (максимальної ентропії)
- обчислення середніх ентропій для моделей довжини $n = 10, 20, 30$

монограми (з пробілами):

[illegible]

монограми (без пробілів):

stats .xlsx

ФайлЗмінитиВиглядВставитиФорматДаніІнструментиДовідка

100%грн.%0.00123За ум...11BIL

A19	A	B	C	D	E	F	G	H	I	J	K
1	Letter	Count									
2	п	23132									
3	р	35262									
4	е	73477									
5	с	44618									
6	т	54582									
7	у	24995									
8	л	38743									
9	н	54864									
10	и	54663									
11	а	67149									
12	к	27840									
13	з	12979									
14	г	14238									
15	ь	19579									
16	ы	13920									
17	й	8441									
18	о	96707									
19	м	26502									
20	в	38996									
21	ж	9617									
22	ю	4735									
23	б	14658									
24	д	26989									
25	ч	15260									
26	э	2973									
27	ц	2337									
28	я	18009									
29	х	7172									
30	ш	6938									
31	ф	1049									
32	щ	2521									
33											
34											

MonogramsBigramsMonograms_no_spaceBigrams_no_space

біграми (із пробілами, перетин):

stats .xlsx

Файл Змінити Вигляд Вставити Формат Дані Інструменти Довідка

100% грн. % .0. .00 123 За ум... 11 B I

A23 fx

	A	B	C	D	E	F	G	H	I	J	K
1	Bigram	Count									
2	пр	6751									
3	ре	5146									
4	ес	4391									
5	ст	9522									
6	ту	1655									
7	уп	659									
8	пл	631									
9	ле	3728									
10	ен	6566									
11	ни	7620									
12	не	1685									
13	е	19180									
14	и	11739									
15	и	17607									
16	н	16033									
17	на	10150									
18	ак	5594									
19	ка	6938									
20	аз	3424									
21	за	4538									
22	ан	3186									
23		3668									
24	г	2975									
25	ге	182									
26	иа	73									
27	ал	7177									
28	ль	4714									
29	ьн	2106									
30	ны	2512									
31	ый	1255									
32	й	6024									
33	р	4892									
34	...	7705									

+ Monograms Bigrams Monograms_no_space Bigrams_no_space

біграми (без пересування , із пробілами):

A1 fx Bigram

	A	B	C	D	E
1	Bigram	Count			
2	пр	3349			
3	ес	2178			
4	ту	851			
5	пл	313			
6	ен	3322			
7	ие	844			
8	и	5891			
9	н	7879			
10	ак	2774			
11	аз	1708			
12	ан	1624			
13		1839			
14	ге	93			
15	ни	3803			
16	ал	3586			
17	ьн	1081			
18	ый	622			
19	р	2445			
20	ом	2641			
21	г	1476			
22	ла	2749			
23	вн	754			
24	ые	337			
25	т	4480			
26	ем	2091			
27	ы	2203			
28	ко	4280			
29	то	7418			
30	ро	3605			
31	го	3873			
32	п	8040			
33	ре	2603			
34	ст	4759			

+ = Monograms Monograms

біграми (перетин, без пробілів):

stats .XLSX

Файл Правка Вид Вставка Формат

100%

р. %

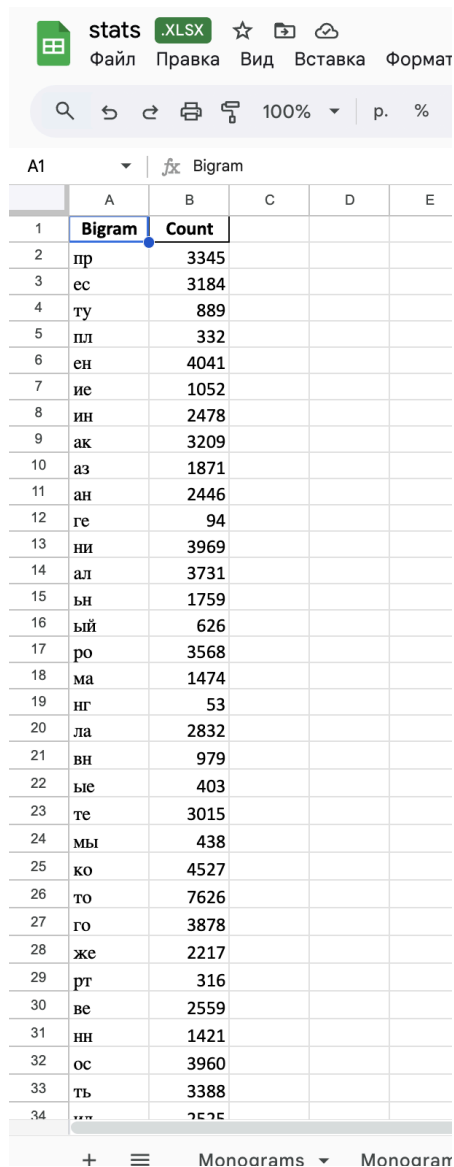
A1

fx Bigram

	A	B	C	D	E
1	Bigram	Count			
2	пр	6751			
3	ре	5154			
4	ес	6335			
5	ст	9803			
6	ту	1818			
7	уп	1262			
8	пл	632			
9	ле	4061			
10	ен	8110			
11	ни	7890			
12	ие	2065			
13	еи	1373			
14	ин	4858			
15	на	10200			
16	ак	6477			
17	ка	6986			
18	аз	3855			
19	за	4552			
20	ан	4853			
21	ег	3683			
22	ге	200			
23	иа	282			
24	ал	7445			
25	ль	4714			
26	ьн	3477			
27	ны	2512			
28	ый	1255			
29	йр	298			
30	ро	7249			
31	ом	6058			
32	ма	3020			
33	нг	92			
34	тв	1208			

+ ≡ Monograms Monogram

біграми (без перетину, без пробілів):



The screenshot shows a Google Sheets spreadsheet with the following data:

	A	B	C	D	E
1	Bigram	Count			
2	пр	3345			
3	ес	3184			
4	ту	889			
5	пл	332			
6	ен	4041			
7	ие	1052			
8	ин	2478			
9	ак	3209			
10	аз	1871			
11	ан	2446			
12	ге	94			
13	ни	3969			
14	ал	3731			
15	ьн	1759			
16	ый	626			
17	ро	3568			
18	ма	1474			
19	нг	53			
20	ла	2832			
21	вн	979			
22	ье	403			
23	те	3015			
24	мы	438			
25	ко	4527			
26	то	7626			
27	го	3878			
28	же	2217			
29	рт	316			
30	ве	2559			
31	нн	1421			
32	ос	3960			
33	ть	3388			
34			

результат коду:

```
/usr/local/bin/python3.7 /Users/uliannabalasan/Downloads/BalasanianYu_FB-32_DorohinAr_FB-32_cp1/crypto_lab1.py
Монограми з пробілами:
H1 = 4.3497, Надлишковість = 0.1301
Монограми без пробілів:
H1 = 4.4487, Надлишковість = 0.1020

Біграми з пересуванням (перетин) з пробілами:
H2 = 3.9502, Надлишковість = 0.2100
Біграми без пересування (кожні дві літери) з пробілами:
H2 = 3.9495, Надлишковість = 0.2101
Біграми з пересуванням (перетин) без пробілів:
H2 = 4.1264, Надлишковість = 0.1671
Біграми без пересування (кожні дві літери) без пробілів:
H2 = 4.1250, Надлишковість = 0.1674

Process finished with exit code 0
```

Для оцінки ентропії тексту при різних довжинах послідовностей ($n = 10, 20, 30$) була використана програма CoolPinkProgram.

Вона проводить статистичне моделювання джерела тексту та обчислює інтервали можливих значень ентропії.

Середні значення $H(n)$, обчислені на основі цих інтервалів, використовувалися далі для оцінки середньої ентропії $H(n)$ та надлишковості $R(n)$ у різних моделях відкритого тексту.

The screenshot displays the CoolPinkProgram interface with the following components:

- Произвольная часть текста:** `носят_подобное_как_образованн`
- Использованные буквы:** (empty field)
- Порядок n-грамм:** A list of values from 5 to 50, with 10 selected.
- Введенный символ:** (empty field)
- Символ по счету:** (empty field)
- Номер эксперимента:** 52
- Неравенство для энтропии:** $1.43418650606007 < H < 2.03380418299189$
- Двоичная таблица угаданных символов:** A grid of 0s and 1s.
- Поле ввода символов:** (empty field)
- Buttons:** `Продолжить` and `Другой`
- Вероятности:** A list of probabilities $q[1]$ through $q[32]$, with $q[17] = 0$ highlighted.
- Строка состояния:** (empty field)

показали, що зі збільшенням довжини блоків ентропія зменшується, а надлишковість зростає, що відповідає властивостям природної мови — високій структурованості та залежності символів між собою.