



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

Симетрична криптографія

Комп'ютерний практикум 1
Експериментальна оцінка ентропії на
символ джерела відкритого тексту

Варіант 4

Виконав:

Волинець Сергій ФІ-03

Київ — 2023

Зміст

1	Вступ	3
1.1	Мета роботи	3
1.2	Порядок виконання роботи	3
2	Основна частина	3
2.1	Хід роботи	3
2.2	Таблиці частот букв:	3
2.3	Таблиці частот біграм:	4
2.4	Обчислення ентропії	9
3	Висновки до роботи	11

1 Вступ

1.1 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.

1.2 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.

2 Основна частина

2.1 Хід роботи

Дану лабораторну роботу я буду виконувати за допомогою мови програмування Python. Вибрав її через те, що в ній багато готових рішень, які легко імплементувати і використовувати на ходу, наприклад пакети «math» чи «collections». Основну частину лабораторної роботи, а саме ітерацію по всіх буквах і біграмах, було виконано за допомоги пакету «itertools». Для обчислень ентропії був створений текст. Я взяв якусь випадкову книгу про російську прозу, видалив усі коми, крапки, тире, пробіли та зменшив великі літери. Таким чином, я отримав одне 1 292 718 буквинне слово. Мій файловий провідник, говорить, що цей файл займає 2.6МВ.

Власне самі обчислення, я виконував прямолінійно. Зробив два словники. Один рахує усі входження літер чи біграм у тексті, для кожної окремо. Це потрібно мені для таблиць, які будуть наведені нижче. Другий словник, за допомоги вбудованої функції, рахує кількість входжень літери чи біграми в тексті, але як ключ, записує кількість входжень, а значення збільшує на один, тобто значення — це кількість літер чи біграм, які згадуються в тексті рівно «ключ» разів. Даний словник використовується для виконання обчислення ентропії.

2.2 Таблиці частот букв:

Усього в моєму російському алфавіті 32 букви.

Для літер, разом з пробілом:

щ:	0.003	ц:	0.002	к:	0.028	я:	0.017	х:	0.007	м:	0.026
б:	0.013	а:	0.069	в:	0.034	и:	0.059	н:	0.054	л:	0.047
й:	0.009	ш:	0.008	д:	0.027	ы:	0.014	ф:	0.003	ъ:	0.000

ж:	0.007	с:	0.044	п:	0.025	у:	0.026	ч:	0.012	э:	0.002
ь:	0.015	т:	0.047	г:	0.021	о:	0.091	р:	0.047	ю:	0.004
з:	0.015	е:	0.065	_:	0.157						

Для літер, без пробілів:

щ:	0.003	ц:	0.003	к:	0.034	я:	0.020	х:	0.008	м:	0.031
б:	0.016	а:	0.082	в:	0.041	и:	0.070	н:	0.064	л:	0.056
й:	0.010	ш:	0.009	д:	0.032	ы:	0.017	ф:	0.004	ъ:	0.000
ж:	0.009	с:	0.053	п:	0.029	у:	0.031	ч:	0.014	э:	0.003
ь:	0.018	т:	0.056	г:	0.025	о:	0.108	р:	0.056	ю:	0.005
з:	0.018	е:	0.077								

2.3 Таблиці частот біграм:

Усього існує 1024 біграми на 32 буквах ($1024 = 32 \times 32$).

Ось частоти біграм з пробілами та з перетинами:

щи:	0.001	ще:	0.001	ца:	0.001	це:	0.001	ка:	0.006	ки:	0.002
кн:	0.001	кл:	0.001	ку:	0.002	кт:	0.001	ко:	0.007	кр:	0.002
ке:	0.001	к_:	0.004	ян:	0.001	ял:	0.001	яд:	0.001	яс:	0.001
ят:	0.001	я_:	0.009	ха:	0.001	хо:	0.002	х_:	0.002	ма:	0.003
ми:	0.004	мн:	0.001	мы:	0.001	му:	0.002	мо:	0.003	ме:	0.003
м_:	0.006	ба:	0.001	би:	0.001	бн:	0.001	бл:	0.001	бы:	0.002
бу:	0.001	бо:	0.002	бр:	0.001	бе:	0.002	ак:	0.004	ая:	0.002
ах:	0.001	ам:	0.004	аб:	0.001	ав:	0.003	ан:	0.003	ал:	0.009
ай:	0.001	аш:	0.001	ад:	0.002	аж:	0.001	ас:	0.004	ап:	0.001
ач:	0.001	ат:	0.004	аг:	0.002	ар:	0.006	аю:	0.001	аз:	0.003
ае:	0.001	а_:	0.016	ва:	0.005	ви:	0.003	вн:	0.001	вш:	0.001
вы:	0.002	вс:	0.002	ву:	0.001	во:	0.005	вр:	0.001	вз:	0.001
ве:	0.005	в_:	0.005	иц:	0.001	ик:	0.002	ия:	0.001	их:	0.001
им:	0.002	иб:	0.001	ив:	0.002	ин:	0.004	ил:	0.006	ий:	0.001
иш:	0.001	ид:	0.002	ис:	0.003	ич:	0.001	ит:	0.003	иг:	0.001
ио:	0.001	ир:	0.001	из:	0.002	ие:	0.001	и_:	0.019	ня:	0.001
на:	0.010	ни:	0.007	нн:	0.002	нд:	0.001	ны:	0.003	нс:	0.001
ну:	0.004	нь:	0.001	нт:	0.001	но:	0.009	не:	0.009	н_:	0.004
ля:	0.002	ла:	0.006	ли:	0.007	лл:	0.001	лы:	0.001	лс:	0.002
лу:	0.001	ль:	0.003	ло:	0.006	лю:	0.001	ле:	0.004	л_:	0.009
й_:	0.006	шк:	0.001	ша:	0.001	ши:	0.001	шл:	0.001	шь:	0.001
ше:	0.002	дя:	0.001	да:	0.004	дв:	0.001	ди:	0.002	дн:	0.002
дл:	0.001	ды:	0.001	дж:	0.001	ду:	0.002	дь:	0.001	до:	0.004
др:	0.001	де:	0.004	д_:	0.002	ых:	0.001	ым:	0.001	ыв:	0.001
ыл:	0.002	ый:	0.001	ыш:	0.001	ыс:	0.001	ыт:	0.001	ые:	0.001
ы_:	0.004	фи:	0.001	фо:	0.001	фе:	0.001	жа:	0.001	жи:	0.001
жн:	0.001	жд:	0.001	же:	0.003	ск:	0.003	ся:	0.004	см:	0.001
са:	0.001	св:	0.001	си:	0.002	сн:	0.001	сл:	0.003	сс:	0.001
сп:	0.002	су:	0.001	сь:	0.003	ст:	0.009	со:	0.003	се:	0.003
с_:	0.004	па:	0.002	пи:	0.001	пл:	0.001	пу:	0.001	по:	0.010
пр:	0.006	пе:	0.002	ук:	0.001	ум:	0.001	уб:	0.001	ув:	0.001
ун:	0.001	ул:	0.003	уш:	0.001	уд:	0.002	уж:	0.001	ус:	0.002
уп:	0.001	уч:	0.001	ут:	0.002	уг:	0.001	ур:	0.001	ую:	0.001

у_:	0.006	чк:	0.001	ча:	0.002	чи:	0.002	чн:	0.001	чу:	0.001
чт:	0.003	че:	0.003	эт:	0.002	ьк:	0.001	ьн:	0.001	ьс:	0.001
ь_:	0.009	тк:	0.001	тя:	0.001	та:	0.005	тв:	0.002	ти:	0.003
тн:	0.001	ты:	0.002	тс:	0.001	ту:	0.002	ть:	0.005	тт:	0.001
то:	0.011	тр:	0.003	те:	0.005	т_:	0.005	га:	0.005	ги:	0.001
гл:	0.002	гд:	0.001	гу:	0.001	го:	0.007	гр:	0.002	ге:	0.001
г_:	0.001	ок:	0.003	оя:	0.001	ох:	0.001	ом:	0.005	об:	0.003
ов:	0.005	ои:	0.001	он:	0.009	ол:	0.006	ой:	0.003	ош:	0.001
од:	0.004	оф:	0.001	ож:	0.002	ос:	0.006	оп:	0.002	оч:	0.002
от:	0.007	ог:	0.004	ор:	0.006	оз:	0.001	ое:	0.002	о_:	0.017
ря:	0.001	рм:	0.001	ра:	0.006	рв:	0.001	ри:	0.009	рн:	0.001
рд:	0.001	ры:	0.001	ру:	0.002	рь:	0.001	рт:	0.001	ро:	0.009
рр:	0.004	ре:	0.005	р_:	0.002	ю_:	0.002	за:	0.006	зв:	0.001
зн:	0.001	зл:	0.001	зд:	0.001	зг:	0.001	зо:	0.001	зе:	0.001
з_:	0.001	ещ:	0.001	ек:	0.001	ем:	0.004	еб:	0.002	ев:	0.002
ен:	0.005	ел:	0.005	ей:	0.002	еш:	0.001	ед:	0.003	еж:	0.001
ес:	0.004	еп:	0.001	еч:	0.001	ет:	0.005	ег:	0.003	ер:	0.008
ез:	0.001	ее:	0.001	е_:	0.014	_к:	0.008	_я:	0.002	_х:	0.002
_м:	0.006	_б:	0.005	_а:	0.001	_в:	0.014	_и:	0.008	_н:	0.014
_л:	0.003	_ш:	0.001	_д:	0.007	_ф:	0.001	_ж:	0.001	_с:	0.014
_п:	0.016	_у:	0.005	_ч:	0.004	_э:	0.002	_т:	0.007	_г:	0.007
_о:	0.010	_р:	0.005	_з:	0.005	_е:	0.003	__:	0.008		

Ось частоти біграм без пробілів але з перетинами:

щи:	0.001	ще:	0.002	ца:	0.001	це:	0.001	ка:	0.008	кв:	0.001
ки:	0.003	кн:	0.001	кл:	0.001	кс:	0.001	ку:	0.002	кт:	0.001
кг:	0.001	ко:	0.009	кр:	0.003	ке:	0.001	як:	0.001	ям:	0.001
яв:	0.002	яи:	0.001	ян:	0.002	ял:	0.001	яд:	0.002	яс:	0.002
яп:	0.001	ят:	0.002	яг:	0.001	яо:	0.001	яз:	0.001	ха:	0.001
хв:	0.001	хн:	0.001	хо:	0.003	мк:	0.001	мя:	0.001	мб:	0.001
ма:	0.004	мв:	0.001	ми:	0.005	мн:	0.002	мы:	0.001	мс:	0.001
мп:	0.001	му:	0.002	мо:	0.004	ме:	0.004	бя:	0.001	ба:	0.001
би:	0.001	бн:	0.001	бл:	0.002	бы:	0.003	бу:	0.001	бо:	0.002
бр:	0.001	бе:	0.002	ак:	0.006	ая:	0.002	ах:	0.001	ам:	0.005
аб:	0.002	ав:	0.005	ай:	0.001	ан:	0.005	ал:	0.011	ай:	0.001
аш:	0.001	ад:	0.003	аж:	0.001	ас:	0.006	ап:	0.003	ау:	0.001
ач:	0.001	ат:	0.006	аг:	0.003	ао:	0.001	ар:	0.007	аю:	0.001
аз:	0.005	ае:	0.002	вк:	0.001	ва:	0.006	ви:	0.004	вн:	0.002
вл:	0.001	вш:	0.001	вд:	0.001	вы:	0.003	вс:	0.004	вп:	0.001
ву:	0.001	вт:	0.001	во:	0.007	вр:	0.001	вз:	0.001	ве:	0.006
иц:	0.001	ик:	0.004	ия:	0.001	их:	0.002	им:	0.003	иб:	0.001
ив:	0.005	ии:	0.002	ин:	0.007	ил:	0.008	ий:	0.001	иш:	0.001
ид:	0.004	иф:	0.001	иж:	0.001	ис:	0.006	ип:	0.003	иу:	0.001
ич:	0.001	ит:	0.005	иг:	0.002	ио:	0.003	ир:	0.002	из:	0.004
ие:	0.002	ня:	0.002	на:	0.012	нв:	0.001	ни:	0.009	нн:	0.003
нд:	0.001	ны:	0.003	нс:	0.001	нп:	0.001	ну:	0.005	нь:	0.001
нт:	0.001	но:	0.011	не:	0.011	лк:	0.001	ля:	0.002	ла:	0.007
лв:	0.001	ли:	0.009	лн:	0.001	лл:	0.001	лд:	0.001	лы:	0.001
лф:	0.001	лс:	0.004	лп:	0.001	лу:	0.002	лч:	0.001	ль:	0.004

лг:	0.001	ло:	0.008	лр:	0.001	лю:	0.001	ле:	0.005	йк:	0.001
йм:	0.001	йв:	0.001	йн:	0.001	йд:	0.001	йс:	0.001	йп:	0.001
йт:	0.001	шк:	0.001	ша:	0.001	ши:	0.002	шл:	0.001	шь:	0.001
ше:	0.003	дя:	0.001	да:	0.005	дв:	0.001	ди:	0.003	дн:	0.002
дл:	0.001	ды:	0.001	дж:	0.001	ду:	0.002	дь:	0.001	до:	0.005
др:	0.002	де:	0.005	ых:	0.001	ым:	0.001	ыв:	0.001	ын:	0.001
ыл:	0.002	ый:	0.002	ыш:	0.001	ыс:	0.001	ып:	0.001	ыт:	0.001
ые:	0.001	фи:	0.001	фо:	0.001	фе:	0.001	жа:	0.002	жи:	0.002
жн:	0.001	жд:	0.001	же:	0.003	ск:	0.004	ся:	0.005	см:	0.001
са:	0.002	св:	0.001	си:	0.002	сн:	0.002	сл:	0.003	сд:	0.001
сс:	0.002	сп:	0.003	су:	0.001	сь:	0.004	ст:	0.011	со:	0.004
се:	0.004	па:	0.002	пи:	0.002	пл:	0.001	пу:	0.001	по:	0.012
пр:	0.008	пе:	0.003	ук:	0.002	ух:	0.001	ум:	0.002	уб:	0.001
ув:	0.002	уи:	0.001	ун:	0.002	ул:	0.004	уш:	0.001	уд:	0.002
уж:	0.001	ус:	0.003	уп:	0.002	уч:	0.002	ут:	0.002	уг:	0.002
ур:	0.001	ую:	0.001	уз:	0.001	чк:	0.001	ча:	0.002	чи:	0.002
чн:	0.001	чу:	0.001	чт:	0.003	че:	0.003	эт:	0.002	ьк:	0.002
ья:	0.001	ьм:	0.001	ьв:	0.001	ьи:	0.001	ьн:	0.002	ьш:	0.001
ьд:	0.001	ьс:	0.002	ьп:	0.001	ьч:	0.001	ьт:	0.001	ьг:	0.001
ьо:	0.001	ьз:	0.001	ье:	0.001	тк:	0.001	тя:	0.001	та:	0.006
тв:	0.003	ти:	0.004	тн:	0.002	тл:	0.001	ты:	0.002	тс:	0.002
тп:	0.001	ту:	0.002	тч:	0.001	ть:	0.006	тт:	0.001	то:	0.013
тр:	0.004	те:	0.006	га:	0.006	ги:	0.001	гн:	0.001	гл:	0.002
гд:	0.001	гу:	0.001	го:	0.008	гр:	0.002	ге:	0.001	ок:	0.004
оя:	0.001	ох:	0.001	ом:	0.006	об:	0.004	ов:	0.009	ои:	0.002
он:	0.012	ол:	0.008	ой:	0.004	ош:	0.001	од:	0.006	оф:	0.001
ож:	0.002	ос:	0.009	оп:	0.004	оу:	0.001	оч:	0.003	от:	0.009
ог:	0.006	оо:	0.002	ор:	0.007	оз:	0.002	ое:	0.002	ря:	0.001
рм:	0.002	ра:	0.008	рв:	0.001	ри:	0.011	рн:	0.002	рд:	0.001
ры:	0.002	рс:	0.001	ру:	0.003	рь:	0.001	рт:	0.001	ро:	0.011
рр:	0.005	ре:	0.006	юп:	0.001	ют:	0.001	за:	0.007	зв:	0.001
зи:	0.001	зн:	0.002	зл:	0.001	зд:	0.001	зу:	0.001	зг:	0.001
зо:	0.001	зе:	0.001	ещ:	0.001	ек:	0.002	ех:	0.001	ем:	0.005
еб:	0.003	ев:	0.004	ей:	0.001	ен:	0.008	ел:	0.007	ей:	0.002
еш:	0.001	ед:	0.004	еж:	0.001	ес:	0.007	еп:	0.003	еу:	0.001
еч:	0.001	ет:	0.007	ег:	0.004	ео:	0.001	ер:	0.009	ез:	0.002
ее:	0.002										

Ось частоти біграм з пробілами але без перетинів:

щи:	0.001	ще:	0.001	ца:	0.001	це:	0.001	ка:	0.006	ки:	0.002
кн:	0.001	кл:	0.001	ку:	0.002	кт:	0.001	ко:	0.007	кр:	0.002
ке:	0.001	к_:	0.004	ян:	0.001	ял:	0.001	яд:	0.001	яс:	0.001
ят:	0.001	я_:	0.009	ха:	0.001	хо:	0.002	х_:	0.002	ма:	0.003
ми:	0.004	мн:	0.001	мы:	0.001	му:	0.002	мо:	0.003	ме:	0.003
м_:	0.006	ба:	0.001	би:	0.001	бн:	0.001	бл:	0.001	бы:	0.002
бу:	0.001	бо:	0.002	бр:	0.001	бе:	0.002	ак:	0.004	ая:	0.002
ах:	0.001	ам:	0.004	аб:	0.001	ав:	0.003	ан:	0.003	ал:	0.009
ай:	0.001	аш:	0.001	ад:	0.002	аж:	0.001	ас:	0.004	ап:	0.001
ач:	0.001	ат:	0.004	аг:	0.002	ар:	0.006	аю:	0.001	аз:	0.003

ае:	0.001	а_:	0.016	ва:	0.005	ви:	0.003	вн:	0.001	вп:	0.001
вы:	0.002	вс:	0.002	ву:	0.001	во:	0.005	вр:	0.001	вз:	0.001
ве:	0.005	в_:	0.005	иц:	0.001	ик:	0.002	ия:	0.001	их:	0.001
им:	0.002	иб:	0.001	ив:	0.002	ин:	0.004	ил:	0.006	ий:	0.001
иш:	0.001	ид:	0.002	ис:	0.003	ич:	0.001	ит:	0.003	иг:	0.001
ио:	0.001	ир:	0.001	из:	0.002	ие:	0.001	и_:	0.019	ня:	0.001
на:	0.010	ни:	0.007	нн:	0.002	нд:	0.001	ны:	0.003	нс:	0.001
ну:	0.004	нь:	0.001	нт:	0.001	но:	0.009	не:	0.009	н_:	0.004
ля:	0.002	ла:	0.006	ли:	0.007	лл:	0.001	лы:	0.001	лс:	0.002
лу:	0.001	ль:	0.003	ло:	0.006	лю:	0.001	ле:	0.004	л_:	0.009
й_:	0.006	шк:	0.001	ша:	0.001	ши:	0.001	шл:	0.001	шь:	0.001
ше:	0.002	дя:	0.001	да:	0.004	дв:	0.001	ди:	0.002	дн:	0.002
дл:	0.001	ды:	0.001	дж:	0.001	ду:	0.002	дь:	0.001	до:	0.004
др:	0.001	де:	0.004	д_:	0.002	ых:	0.001	ым:	0.001	ыв:	0.001
ыл:	0.002	ый:	0.001	ыш:	0.001	ыс:	0.001	ыт:	0.001	ые:	0.001
ы_:	0.004	фи:	0.001	фо:	0.001	фе:	0.001	жа:	0.001	жи:	0.001
жн:	0.001	жд:	0.001	же:	0.003	ск:	0.003	ся:	0.004	см:	0.001
са:	0.001	св:	0.001	си:	0.002	сн:	0.001	сл:	0.003	сс:	0.001
сп:	0.002	су:	0.001	сь:	0.003	ст:	0.009	со:	0.003	се:	0.003
с_:	0.004	па:	0.002	пи:	0.001	пл:	0.001	пу:	0.001	по:	0.010
пр:	0.006	пе:	0.002	ук:	0.001	ум:	0.001	уб:	0.001	ув:	0.001
ун:	0.001	ул:	0.003	уш:	0.001	уд:	0.002	уж:	0.001	ус:	0.002
уп:	0.001	уч:	0.001	ут:	0.002	уг:	0.001	ур:	0.001	ую:	0.001
у_:	0.006	чк:	0.001	ча:	0.002	чи:	0.002	чн:	0.001	чу:	0.001
чт:	0.003	че:	0.003	эт:	0.002	ьк:	0.001	ьн:	0.001	ьс:	0.001
ь_:	0.009	тк:	0.001	тя:	0.001	та:	0.005	тв:	0.002	ти:	0.003
тн:	0.001	ты:	0.002	тс:	0.001	ту:	0.002	ть:	0.005	тт:	0.001
то:	0.011	тр:	0.003	те:	0.005	т_:	0.005	га:	0.005	ги:	0.001
гл:	0.002	гд:	0.001	гу:	0.001	го:	0.007	гр:	0.002	ге:	0.001
г_:	0.001	ок:	0.003	оя:	0.001	ох:	0.001	ом:	0.005	об:	0.003
ов:	0.005	ои:	0.001	он:	0.009	ол:	0.006	ой:	0.003	ош:	0.001
од:	0.004	оф:	0.001	ож:	0.002	ос:	0.006	оп:	0.002	оч:	0.002
от:	0.007	ог:	0.004	ор:	0.006	оз:	0.001	ое:	0.002	о_:	0.017
ря:	0.001	рм:	0.001	ра:	0.006	рв:	0.001	ри:	0.009	рн:	0.001
рд:	0.001	ры:	0.001	ру:	0.002	рь:	0.001	рт:	0.001	ро:	0.009
рр:	0.004	ре:	0.005	р_:	0.002	ю_:	0.002	за:	0.006	зв:	0.001
зн:	0.001	зл:	0.001	зд:	0.001	зг:	0.001	зо:	0.001	зе:	0.001
з_:	0.001	ещ:	0.001	ек:	0.001	ем:	0.004	еб:	0.002	ев:	0.002
ен:	0.005	ел:	0.005	ей:	0.002	еш:	0.001	ед:	0.003	еж:	0.001
ес:	0.004	еп:	0.001	еч:	0.001	ет:	0.005	ег:	0.003	ер:	0.008
ез:	0.001	ее:	0.001	е_:	0.014	_к:	0.008	_я:	0.002	_х:	0.002
_м:	0.006	_б:	0.005	_а:	0.001	_в:	0.014	_и:	0.008	_н:	0.014
_л:	0.003	_ш:	0.001	_д:	0.007	_ф:	0.001	_ж:	0.001	_с:	0.014
_п:	0.016	_у:	0.005	_ч:	0.004	_э:	0.002	_т:	0.007	_г:	0.007
_о:	0.010	_р:	0.005	_з:	0.005	_е:	0.003	_:	0.008		

І нарешті, без пробілів і перетинів:

щи:	0.001	ще:	0.002	ца:	0.001	це:	0.001	ка:	0.008	ки:	0.003
кн:	0.001	кл:	0.001	ку:	0.002	кт:	0.001	ко:	0.009	кр:	0.002

ке:	0.001	ям:	0.001	яв:	0.001	ян:	0.001	ял:	0.001	яд:	0.001
яс:	0.001	ят:	0.002	ха:	0.001	хо:	0.002	ма:	0.004	ми:	0.004
мн:	0.002	мы:	0.001	му:	0.002	мо:	0.004	ме:	0.004	бя:	0.001
ба:	0.001	би:	0.001	бн:	0.001	бл:	0.002	бы:	0.003	бу:	0.001
бо:	0.002	бр:	0.001	бе:	0.002	ак:	0.005	ая:	0.002	ах:	0.001
ам:	0.004	аб:	0.001	ав:	0.003	ан:	0.003	ал:	0.011	ай:	0.001
аш:	0.001	ад:	0.003	аж:	0.001	ас:	0.004	ап:	0.001	ач:	0.001
ат:	0.005	аг:	0.002	ар:	0.007	аю:	0.001	аз:	0.004	ае:	0.001
ва:	0.006	ви:	0.004	вн:	0.001	вл:	0.001	вш:	0.001	вы:	0.003
вс:	0.003	ву:	0.001	во:	0.006	вр:	0.001	вз:	0.001	ве:	0.005
иц:	0.001	ик:	0.003	ия:	0.001	их:	0.001	им:	0.002	иб:	0.001
ив:	0.002	ии:	0.001	ин:	0.005	ил:	0.007	ий:	0.001	иш:	0.001
ид:	0.003	ис:	0.004	ип:	0.001	ич:	0.001	ит:	0.004	иг:	0.001
ио:	0.001	ир:	0.001	из:	0.003	ие:	0.002	ня:	0.002	на:	0.012
ни:	0.008	нн:	0.002	нд:	0.001	ны:	0.003	нс:	0.001	ну:	0.005
нь:	0.001	нт:	0.001	но:	0.010	не:	0.011	лк:	0.001	ля:	0.002
ла:	0.007	ли:	0.009	лл:	0.001	лы:	0.001	лс:	0.003	лу:	0.002
ль:	0.004	ло:	0.007	лю:	0.001	ле:	0.005	шк:	0.001	ша:	0.001
ши:	0.002	шл:	0.001	шь:	0.001	ше:	0.003	дя:	0.001	да:	0.005
дв:	0.001	ди:	0.003	дн:	0.002	дл:	0.001	ды:	0.001	дж:	0.001
ду:	0.002	дь:	0.001	до:	0.005	др:	0.002	де:	0.005	ых:	0.001
ым:	0.001	ыв:	0.001	ыл:	0.002	ый:	0.002	ыш:	0.001	ыс:	0.001
ыт:	0.001	ые:	0.001	фи:	0.001	фо:	0.001	фе:	0.001	жа:	0.002
жи:	0.002	жн:	0.001	жд:	0.001	же:	0.003	ск:	0.004	ся:	0.005
см:	0.001	са:	0.002	св:	0.001	си:	0.002	сн:	0.002	сл:	0.003
сс:	0.002	сп:	0.002	су:	0.001	сь:	0.004	ст:	0.011	со:	0.003
се:	0.004	па:	0.002	пи:	0.002	пл:	0.001	пу:	0.001	по:	0.012
пр:	0.008	пе:	0.003	ук:	0.001	ух:	0.001	ум:	0.001	уб:	0.001
ув:	0.001	ун:	0.001	ул:	0.004	уш:	0.001	уд:	0.002	уж:	0.001
ус:	0.002	уп:	0.001	уч:	0.001	ут:	0.002	уг:	0.001	ур:	0.001
ую:	0.001	уз:	0.001	чк:	0.001	ча:	0.002	чи:	0.002	чн:	0.001
чу:	0.001	чт:	0.003	че:	0.003	эт:	0.002	ьк:	0.001	ьн:	0.001
ьш:	0.001	ьс:	0.001	тк:	0.001	тя:	0.001	та:	0.006	тв:	0.002
ти:	0.004	тн:	0.001	тл:	0.001	ты:	0.002	тс:	0.001	ту:	0.002
ть:	0.006	тт:	0.001	то:	0.012	тр:	0.003	те:	0.006	га:	0.006
ги:	0.001	гл:	0.002	гд:	0.001	гу:	0.001	го:	0.008	гр:	0.002
ге:	0.001	ок:	0.003	оя:	0.001	ох:	0.001	ом:	0.006	об:	0.004
ов:	0.006	ои:	0.001	он:	0.010	ол:	0.007	ой:	0.004	ош:	0.001
од:	0.005	оф:	0.001	ож:	0.002	ос:	0.007	оп:	0.002	оч:	0.003
от:	0.008	ог:	0.005	ор:	0.007	оз:	0.001	ое:	0.002	ря:	0.001
рм:	0.001	ра:	0.008	рв:	0.001	ри:	0.010	рн:	0.001	рд:	0.001
ры:	0.002	рс:	0.001	ру:	0.003	рь:	0.001	рт:	0.001	ро:	0.011
рр:	0.004	ре:	0.006	за:	0.007	зв:	0.001	зн:	0.002	зл:	0.001
зд:	0.001	зг:	0.001	зо:	0.001	зе:	0.001	ещ:	0.001	ек:	0.001
ем:	0.004	еб:	0.002	ев:	0.002	ен:	0.006	ел:	0.006	ей:	0.002
еш:	0.001	ед:	0.003	еж:	0.001	ес:	0.005	еп:	0.001	еч:	0.001
ет:	0.006	ег:	0.004	ер:	0.009	ез:	0.001	ее:	0.001		

2.4 Обчислення ентропії

Для обчислення ентропії, а саме ймовірності, я обчислюю спочатку величину «sum _of _ symbols», яка рівна кількості входжень усіх символів. Її я обчислюю як суму добутків ключа на значення в другому словнику. Для літер, можливо, це зайве, але для біграм іншого способу я не придумав. Ймовірність, власне, я обчислюю як значення ключа (кількість входжень символу), поділене на кількість усіх символів. Далі я множу це значення на логарифм цієї ймовірності і домножую на значення словника по даному ключу, що виключає потребу додавати одне і те саме «значення словника по даному ключу» разів.

При обчисленні ентропії, я отримав наступні результати

Для літер, без пробілів:	$H = 4.447$
Для літер, з пробілами:	$H = 4.390$
Для біграм, без перетинів, без пробілів:	$H = 4.127$
Для біграм, без перетинів, з пробілами:	$H = 3.977$
Для біграм, з перетином, без пробілів:	$H = 4.125$
Для біграм, з перетином, з пробілами:	$H = 3.978$

Для обчислення інших ентропій я скористувався наданою програмою.

Произвольная часть текста:
ния_почему_тогда_мы_так_ревностно_оправдываем_свое_не_совсем_порядочное_пов

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: y

Символ по счету: 1

Номер эксперимента: 81

Поле ввода символов:
y

Продолжить Другой

Неравенство для энтропии:
 $2.25040344889526 < H < 3.08005311522288$

Двоичная таблица угаданных символов

10000000000000000000000000000000
10000000000000000000000000000000
00000010000000000000000000000000
00000000000010000000000000000000
0000000000000000000000001000000000

Вероятности:

q[1]	= 0.4426229508196
q[2]	= 0.1311475409836
q[3]	= 0.0491803279688
q[4]	= 0.0327968852459
q[5]	= 0.0327968852459
q[6]	= 0.0163934426229
q[7]	= 0.0327968852459
q[8]	= 0
q[9]	= 0.0163934426229
q[10]	= 0.0327968852459
q[11]	= 0
q[12]	= 0.0163934426229
q[13]	= 0.0163934426229
q[14]	= 0
q[15]	= 0
q[16]	= 0.0163934426229
q[17]	= 0.0163934426229
q[18]	= 0.0327968852459
q[19]	= 0
q[20]	= 0.0163934426229
q[21]	= 0.0327968852459
q[22]	= 0
q[23]	= 0.0491803279688
q[24]	= 0.0163934426229
q[25]	= 0
q[26]	= 0
q[27]	= 0
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

$$R = 1 - \frac{H_\infty}{H_0} = 1 - \frac{2}{\log_2(32)} = 0.6$$

3 Висновки до роботи

Як ми бачимо, російська мова не надто надлишкова. Ми можемо скоротити повідомлення цією мовою, лише в 0.6 разів без втрати сенсу.