

**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**

**Виконав:**

**Студент групи ФБ-35**

**Кохта Андрій**

**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.

**Постановка задачі:** 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 (10)  $H_1$ , (20)  $H_2$ , (30)  $H_3$ . Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

## Хід роботи

**1) Знаходимо текст, не менше 1мб.**

**2) Очищуємо його та розділяємо на текст з/без пробілів.**

## З пробілами:

введение моя история в последний день второго года обучения в старших классах средней школы я получил удар бейсбольной битой по лицу когда сзади и ударила прямо между глаз момент удара я не помню удар был такой силы что нос оказался сломан в двух местах и приобрел подковообразную форму несколько трещин в черепе непосредственно после удара у меня начал развиваться отек мозга таким образом эа какуюто долю секунды я получила глаза то увидел что ко мне бегут люди чтобы оказать помощь опустив взгляд я заметил красные пятна на своей одежде один из моих одноклассников скрывающий которая хлестала из сломанного носа потрясенный и растревяный я не осознавал насколько серьезные травмы получил мой учитель обхватив спустившись с холма и наконец вошли в школу чьи руки поддерживали меня не давая упасть мы не торопились и шли медленно никто не осознавал стала задавать мне вопросы какой сейчас год я ответил я на самом деле был й кто президент соединенных штатов был клиентом сказал я правильные десять секунд позже сказал я беззаботно стараясь не обращать внимания на то что мне потребовалась десять секунд чтобы вспомнить имя собственности спрятавшись с быстро распространявшимся отеком мозга и я потерял сознание еще до прибытия скорой помощи спустя несколько минут меня вынесли из организма перестал работать он с трудом справлялся с такими базовыми функциями как глотание и дыхание это был первый приступ в тот день искусственной вентиляции легких они пришли к выводу что местная больница недостаточно хорошо оборудована для лечения такого тяжелого случая катяла меня вытикли из отделения экстренной медицинской помощи и повезли к вертолетной площадке через дорогу коляски с грохотом ехала вручную контролировала каждый мой вдох моя мать которая за несколько минут до этого примчалась в больницу в вертолете села около меня воодушевленно в то время как мама сопровождала меня в полете отец направился домой чтобы присмотреть за моим братом и сестрой и сообщил хорошие приветы и избавиться от плохих слезы когдаobjasnila сестре что в тот вечер не сможет присутствовать на торжестве посвященному окончанию школы в цинциннати чтобы присоединиться к матери когда вертолет с нами приземлился на крышу больницы на площадке уже ожидала команда при подготовке к этому моменту отек мозга стал настолько серьезным что у меня случилось несколько посттравматических приступов сломанные кости существо было противопоказано после еще одного приступа третьего в этот день я был введен в состояние искусственной комы и приprehensione этой больнице десятью годами ранее они были в этом здании на первом этаже тогда мой трехлетней сестре поставили диагноз лейкемия в то же самое время после нескольких сеансов химиотерапии лумбальных пункций и многократной трепанобиопсии моя маленькая сестра покинула жизнь мои родители снова оказались в том же месте но уже с другим ребенком после того как меня ввели в состояние комы больница пригласила

Без пробілів:

введенiemоисториявпоследнийденьвторогогодаобучениявстаршихклассахсреднейшколыяполучилударбесбольнойбитойлицуцокгдамо  
сянепомнодуарбытакойсильносоказалсясломанувхустахиприобрелподковообразнуюформукрометогояолучилчертеномозговутра  
угатакимобразомзакащуюдолескускудышуполчлосмалыйносесколькошательчиннечерепнейпереломьеихлалицизгдагдаоткрыглазатоуви  
ссчиковкорвалсясебяфутболкуипротянулинеявспользоваласячтобыстановитькровькотораяхлясталанаизсломанногоносапотрещенныи  
ккабинитетумедсестрыпредсеклиполупостисьсхолманиаконецощишликовулычаторукиподдержалименянедавнаупастынаторилипсы  
даватьмневопросыакайсейчасгдойтветилянасамомделебыйктопрезидентсоединенныхштатовблклинтоноказалиправильныйответбы  
щательнинаничатончепотребовалосьдесятьсекундчтобыспомнитымасобственноМайамиэтоследнийвопроскоторыйпомномойорганиз  
моимспустянесколькоштатовнунименявынеслииззданншиколыпозвелившмествнубольницувскрепослерибутильникумрганизмпестал  
семполнотыопересталдышатькогдавразчикподключименикапратурускусственноговентиляциигехиконпришивыводуочтеснаболь  
екрупнобольницувцинциннатинакаталкеменявыкатилиизотделенияэкстренноМедицинскойпомощишповезликтвертолетнойплощадкечерез  
уюконтролировалааждыймвойдохможьтакогоразанскошкомунгнудостгопримчаласьвольничузввертолетессаюломеневвремяполе  
квполетеотеествонаправилсядомчтобыприсмотретьзамоимбротомисстрийсообщитьумужаснуовостьонструдомсдерживандклиратомын  
сусствуетвательножречественноспеченошнокончаниенеовьсомогоклассаотштетвездбранитесструктуровщикомасмелминиуваехцинциннат  
симерноиздвадцативрачейимдестеркоторыепозвелимнявтраматологическоеотделениекэтому моментуотекмозгасталнастолькосеръез  
семостояннобиохирургическоешибаштельствомпопротивокапозапоследногонопрладкаттереготовтденешибывелведеносто  
звтойбольницедесятъгодамиранееонибилизметзданинапервэтажетогдамоитехрхлечтнейстремпеставилидиагнозлекреймитвтремя  
цитотерапииймбалыхныхункциймногократноТранебаписсимояменькаястстранаконечнцылаизбольницычастливавдозированилобедивш  
тогокакменяввеливсостояниекомыбольнициапригласиласявященикаисоциальногоработникачтобуспокоитмоихродителейэтобытотже  
решивличноможнъзподдерживалитолькоприбродителитеспокойноспалиабольничнойойкетпроваливаласясъсонтизмежоменитетопр  
влевленесчастъюнаследующеетрмоядыхательнаяфункциявостановиласяиправчилириширеявственимизостоянияискусственной  
хнутычрезносинопнохаталяеткотрогояблочно-госкомаобогнанниневрнулоукомплекснисхокружавшиххрзкийвидуочеснаприв

### 3) Підрахунок кількості та частоти елементів у тексті для монограм:

Монограми з пробілом    Монограми без пробіла

N-грама	Кількість	Частота	N-грама	Кількість	Частота
-	139657	0,1452438029	о	89901	0,1093848479
о	89901	0,09349737659	е	74096	0,09015449982
е	74096	0,07706011742	и	66295	0,08066282344
и	66295	0,06894704821	а	57906	0,07045571241
а	57906	0,0602224568	т	56689	0,06897495735
т	56689	0,05895677224	н	53870	0,0655450079
н	53870	0,05602500169	с	41635	0,05065837022
с	41635	0,04330055588	в	37698	0,04586812155
в	37698	0,03920606114	р	37163	0,04521717335
р	37163	0,03864965914	л	33919	0,04127011552
л	33919	0,03527588699	м	26733	0,03252672538
м	26733	0,02780242009	к	25908	0,03152292676
к	25908	0,026944417	п	23917	0,0291004261
п	23917	0,02487376955	д	23850	0,02901890548
д	23850	0,02480408929	у	19036	0,02316158846
у	19036	0,01979751127	ы	18098	0,02202029985
ы	18098	0,01882198776	я	16887	0,02054684515
я	16887	0,01756254323	ь	16227	0,01996646704
ь	16227	0,01706646144	ч	13843	0,01684313244
ч	13843	0,01439677183	б	13665	0,01662655528
б	13665	0,01421165116	з	13004	0,01582229966
з	13004	0,01352420869	г	11982	0,0145788061
г	11982	0,01246132486	ж	10007	0,01217577305
ж	10007	0,01040731747	й	8305	0,01010490608
й	8305	0,008637231094	х	7622	0,009273882498
х	7622	0,007926908537	ю	5873	0,00714582943
ю	5873	0,006107941989	ш	5019	0,00610674577
ш	5019	0,005219778791	ц	4233	0,005150399451
ц	4233	0,004402335848	э	3472	0,004224471272
э	3472	0,003610892999	щ	2869	0,003490785737
щ	2869	0,002983770742	ф	1973	0,002400599602
ф	1973	0,002051927387	ъ	183	0,0002226607842

### 4) Підрахунок частоти біграм:

Біграми з пробілом та кроком 1(перетинаються)

	-	а	б	в	г	д	е	ж	з	и	й	к	л	м
-	0,0001860	0,004455	0,013403	0,002246	0,006663	0,002895	0,001311	0,002970	0,011038	0,000011	0,007428	0,004073	0,005639	
а	0,011343	0,000001	0,000991	0,003095	0,000575	0,001087	0,002678	0,001635	0,002847	0,000216	0,000489	0,004773	0,004241	0,002479
б	0,000230	0,000544	0,000007	0,000090	0	0,000013	0,001360	0,000005	0,000009	0,000520	0	0,000325	0,001028	0,000901
в	0,005832	0,006770	0,000002	0,000063	0,000004	0,000448	0,004912	0	0,000174	0,003248	0	0,000291	0,001414	0,000079
г	0,000460	0,000911	0,000001	0,000001	0,000014	0,001182	0,000256	0,000001	0,000001	0,001166	0	0,000142	0,000770	0,000010
д	0,000843	0,004082	0,000019	0,000601	0,000042	0,000059	0,005960	0,000245	0,000005	0,002325	0	0,000166	0,000964	0,000044
е	0,018288	0,000214	0,000741	0,001432	0,002101	0,004115	0,001368	0,000638	0,001159	0,000139	0,002175	0,002055	0,004847	0,004196
ж	0,000083	0,000628	0,000027	0,000003	0	0,001288	0,004662	0,000002	0	0,001630	0	0,000029	0,000014	0,000005
з	0,001038	0,003530	0,000406	0,000643	0,000223	0,000526	0,000168	0,000024	0,000094	0,000426	0	0,000056	0,000222	0,000916
и	0,018747	0,000289	0,000657	0,003390	0,000557	0,001159	0,004227	0,000302	0,003148	0,001698	0,001550	0,002121	0,003139	0,003634
й	0,006778	0,000006	0,000005	0,000001	0,000013	0,000075	0,000005	0	0	0	0	0,000069	0,000044	0,000067
к	0,003550	0,005893	0,000001	0,000057	0,000005	0,000012	0,000491	0,000206	0,000004	0,003360	0	0,000036	0,000707	0,000169
л	0,002283	0,003185	0,000012	0,000001	0,000310	0,000026	0,004798	0,001386	0,000002	0,007068	0	0,000187	0,000362	0,000005
м	0,007214	0,003409	0,000154	0,000003	0,000018	0,000001	0,004264	0	0,000001	0,002652	0	0,000029	0,000087	0,000101

**Біграми без пробіла та кроком 1(перетинаються)**

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н
<b>а</b>	0,000193	0,001554	0,004834	0,000866	0,001943	0,003391	0,001994	0,003647	0,001286	0,000574	0,006345	0,005415	0,003395	0,007984
<b>б</b>	0,000642	0,000010	0,000121	0,000006	0,000025	0,001591	0,000006	0,000015	0,000646	0	0,000386	0,001204	0,001058	0,000654
<b>в</b>	0,008017	0,000200	0,000442	0,000225	0,000827	0,005841	0,000037	0,000332	0,004167	0	0,000891	0,001782	0,000284	0,002067
<b>г</b>	0,001070	0,000045	0,000041	0,000034	0,001421	0,000301	0,000001	0,000021	0,001411	0	0,000189	0,000906	0,000027	0,000340
<b>д</b>	0,004782	0,000040	0,000770	0,000064	0,000104	0,006980	0,000288	0,000037	0,002776	0,000001	0,000445	0,001166	0,000079	0,001896
<b>е</b>	0,000425	0,001531	0,003729	0,002794	0,005801	0,002039	0,000959	0,001902	0,001738	0,002546	0,003260	0,006238	0,005787	0,014642
<b>ж</b>	0,000738	0,000037	0,000006	0,000004	0,001509	0,005454	0,000002	0	0,001916	0	0,000041	0,000018	0,000014	0,001819
<b>з</b>	0,004135	0,000501	0,000859	0,000287	0,000680	0,000206	0,000043	0,000135	0,000547	0	0,000161	0,000287	0,001107	0,002892
<b>и</b>	0,000598	0,001491	0,006083	0,001006	0,002325	0,005309	0,000519	0,004061	0,003966	0,001814	0,003603	0,004146	0,005242	0,005690
<b>й</b>	0,000164	0,000231	0,000745	0,000169	0,000479	0,000087	0,000152	0,000164	0,000650	0	0,000573	0,000278	0,000485	0,000837
<b>к</b>	0,006942	0,000097	0,000385	0,000087	0,000189	0,000632	0,000324	0,000062	0,004314	0	0,000236	0,000908	0,000408	0,000605
<b>л</b>	0,003765	0,000128	0,000227	0,000410	0,000122	0,005695	0,001645	0,000059	0,008410	0,000002	0,000335	0,000469	0,000094	0,000605
<b>м</b>	0,004112	0,000425	0,000681	0,000169	0,000445	0,005126	0,000071	0,000150	0,003804	0	0,000492	0,000293	0,000447	0,001951
<b>н</b>	0,011163	0,000169	0,000225	0,000191	0,000574	0,009401	0,000032	0,000074	0,013393	0,000001	0,000429	0,000058	0,000097	0,003633

**Біграми з пробілом та кроком 2(не перетинаються)**

	-	а	б	в	г	д	е	ж	з	и	й	к	л	м
<b>-</b>	0	0,001867	0,004422	0,013409	0,002208	0,006676	0,002930	0,001306	0,002970	0,010994	0,000010	0,007554	0,004218	0,005582
<b>а</b>	0,011446	0	0,000994	0,003101	0,000599	0,001119	0,002666	0,001574	0,002847	0,000216	0,000480	0,004844	0,004322	0,002402
<b>б</b>	0,000214	0,000499	0,000012	0,000101	0	0,000014	0,001343	0	0,000010	0,000517	0	0,000312	0,001006	0,000890
<b>в</b>	0,005788	0,006730	0,000002	0,000081	0,000006	0,000470	0,004923	0	0,000170	0,003280	0	0,000282	0,001418	0,000070
<b>г</b>	0,000472	0,000817	0	0,000002	0,000014	0,001206	0,000270	0	0,000002	0,001133	0	0,000135	0,000777	0,000012
<b>д</b>	0,000879	0,004093	0,000010	0,000057	0,000029	0,000072	0,005830	0,000224	0,000004	0,002350	0	0,000153	0,000967	0,000039
<b>е</b>	0,018071	0,000187	0,000769	0,001439	0,002169	0,004189	0,001360	0,000624	0,001225	0,000153	0,002179	0,002025	0,004944	0,004139
<b>ж</b>	0,000095	0,000682	0,000024	0,000006	0	0,001295	0,004694	0,000002	0	0,001591	0	0,000035	0,000022	0,000006
<b>з</b>	0,001031	0,003546	0,000384	0,000613	0,000241	0,000509	0,000168	0,000016	0,000091	0,000405	0	0,000066	0,000224	0,000888
<b>и</b>	0,018636	0,000289	0,000673	0,003513	0,000580	0,001166	0,004309	0,000284	0,003144	0,001718	0,001535	0,002167	0,003167	0,003588
<b>й</b>	0,006770	0,000004	0,000002	0,000002	0,000012	0,000074	0,000004	0	0	0	0	0,000060	0,000052	0,000064
<b>к</b>	0,003506	0,005915	0	0,000047	0,000006	0,000010	0,000501	0,000222	0,000004	0,003323	0	0,000035	0,000682	0,000176
<b>л</b>	0,002217	0,003147	0,000016	0	0,000341	0,000020	0,004638	0,001395	0	0,007063	0	0,000168	0,000372	0,000004
<b>м</b>	0,007338	0,003419	0,000128	0,000006	0,000016	0	0,004261	0	0,000002	0,002697	0	0,000033	0,000101	0,000091

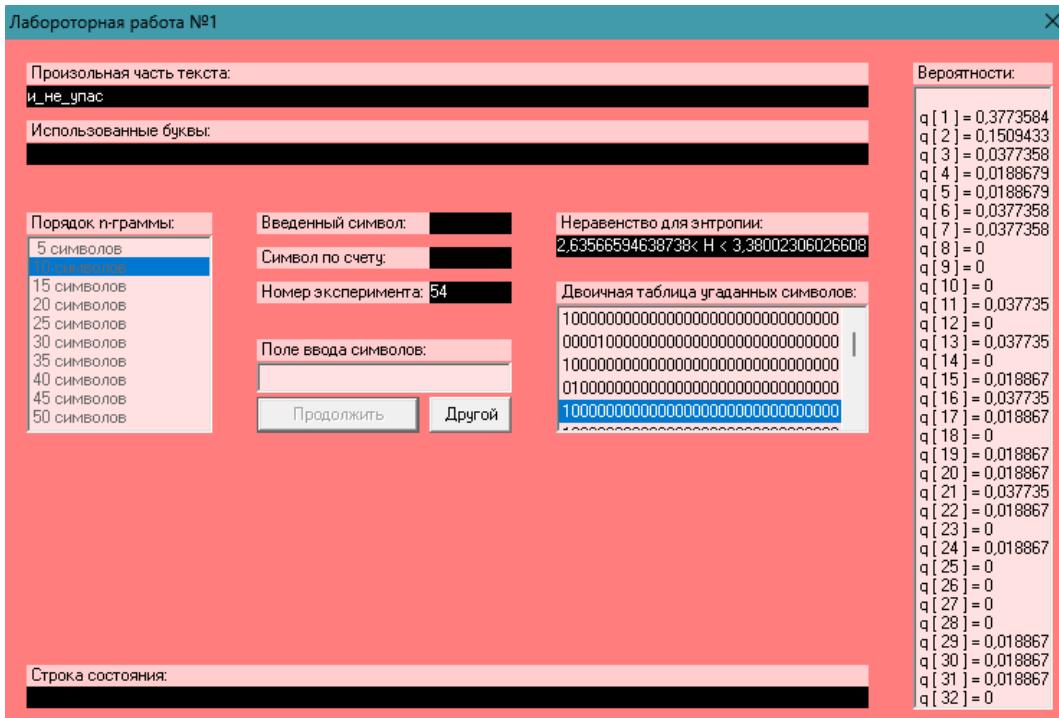
**Біграми без пробіла та кроком 2(не перетинаються)**

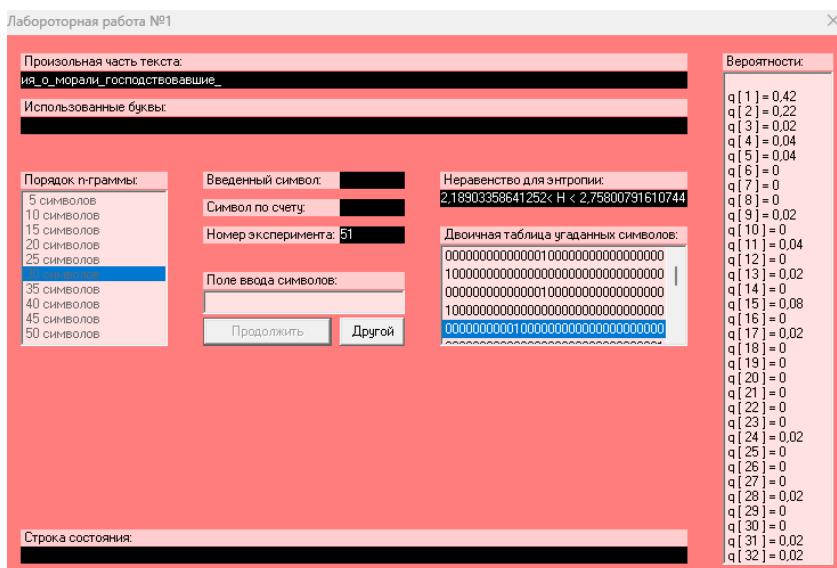
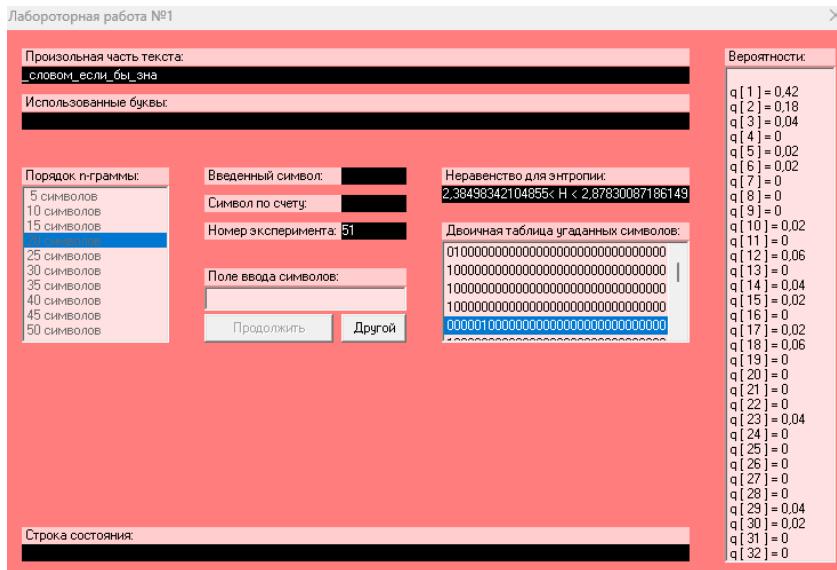
	а	б	в	г	д	е	ж	з	и	й	к	л	м	н
<b>а</b>	0,000184	0,001613	0,004869	0,000895	0,002007	0,003462	0,001961	0,003543	0,001323	0,000584	0,006341	0,005453	0,003387	0,007828
<b>б</b>	0,000584	0,000007	0,000141	0,000009	0,000021	0,001574	0,000004	0,000021	0,000654	0	0,000442	0,001258	0,001043	0,000605
<b>в</b>	0,008149	0,000167	0,000450	0,000216	0,000798	0,005825	0,000046	0,000340	0,004063	0	0,000910	0,001759	0,000262	0,002024
<b>г</b>	0,001095	0,000048	0,000048	0,000036	0,001416	0,000294	0	0,000026	0,001389	0	0,000199	0,000941	0,000036	0,000316
<b>д</b>	0,004752	0,000036	0,000722	0,000073	0,000090	0,007037	0,000313	0,000036	0,002776	0,000002	0,000433	0,001158	0,000090	0,001939
<b>е</b>	0,000403	0,001496	0,003623	0,002886	0,005626	0,002078	0,000978	0,001937	0,001671	0,002530	0,003275	0,006368	0,005684	0,014469
<b>ж</b>	0,000805	0,000041	0,000002	0,000007	0,001445	0,005533	0,000002	0	0,001895	0	0,000041	0,000019	0,000012	0,001805
<b>з</b>	0,004090	0,000540	0,000837	0,000277	0,000681	0,000206	0,000048	0,000138	0,000537	0	0,000141	0,000292	0,001121	0,002937
<b>и</b>	0,000647	0,001440	0,006081	0,001075	0,002404	0,005358	0,000474	0,004144	0,003934	0,001881	0,003628	0,004078	0,005280	0,005742
<b>й</b>	0,000163	0,000214	0,000749	0,000189	0,000450	0,000082	0,000136	0,000146	0,000586	0	0,000591	0,000267	0,000511	0,000868
<b>к</b>	0,006974	0,000087	0,000379	0,000094	0,000175	0,000605	0,000304	0,000063	0,004212	0	0,000206	0,000905	0,000420	0,000610
<b>л</b>	0,003591	0,000128	0,000238	0,000396	0,000111	0,005716	0,001610	0,000070	0,008317	0	0,000326	0,000498	0,000102	0,000603
<b>м</b>	0,004129	0,000450	0,000678	0,000160	0,000455	0,005051	0,000077	0,000131	0,003805	0	0,000489	0,000284	0,000440	0,001997
<b>н</b>	0,011116	0,000114	0,000250	0,000201	0,000596	0,009436	0,000036	0,000085	0,013551	0,000002	0,000440	0,000068	0,000124	0,003562

5) Розраховуємо значення ентропії та надлишковості:

	Ентропія	Надлишковість
<b>H1 (з пробілами)</b>	4.409565017603595	0.12584843426857306
<b>H1 (без пробілів)</b>	4.459466017127637	0.10810679657447264
<b>H2 (з пробілами, крок 1)</b>	3.9825658466810308	0.21049669148620487
<b>H2 (без пробілів, крок 1)</b>	4.115838492726308	0.17683230145473838
<b>H2 (з пробілами, крок 2)</b>	3.982495255190539	0.21051068553362096
<b>H2 (без пробілів, крок 2)</b>	4.114714480773677	0.17705710384526463

## 6) CoolPinkProgram





2.635665949637383<H10<3.380023060260608

0 4728668100725234>R10>0 3239953879478784

2 38493342104855<H20<2 87830087186149

0 52301331579029>R20>0 424339825627702

? 18903568412524<H30<? 75800791610744

0.562192863174952>P30>0.4483984167785

**Висновки:** Виконуючи лабораторну роботу, я навчився аналізувати частоти символів та пар символів, визначати ентропію та надлишковість мови. Видно, що при вгадуванні наступної букви в програмі coolpinkprogram.exe, чим більше символів, ентропія спадає а надлишковість зростає, через, що передбачити наступний символ стає легше.