

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

Криптографія

Комп'ютерний практикум No1
Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали:
Студенти групи ФБ-32
Коптева Ганна, Чупріна Вікторія

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.

Порядок виконання роботи:

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 ймовірності 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:

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 ймовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилучено всі пробіли.

Текст: **Детская сказка: «Леди и Бродяга»**

Завдання 1

Написати програми для підрахунку частот букв і частот біграм у 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 ймовірність замінити відповідними частотами. Також отримати значення H^1 та H^2 на тому ж тексті, у якому вилучено всі пробіли.

Частоти літер

Без пробілів (unigrams_no_spaces.csv):

char	count	freq
о	65358	0.1133082300
е	48944	0.0848520152
а	46206	0.0801052675
и	38015	0.0659048986
т	37491	0.0649964633
н	36914	0.0639961444
с	30368	0.0526476381
в	26551	0.0460302766
л	26436	0.0458309062
р	25143	0.0435892902
к	19184	0.0332584394
д	18285	0.0316998835
м	18047	0.0312872736
у	17034	0.0295310810
п	15915	0.0275911209
ь	13423	0.0232708524
я	12315	0.0213499626
ч	10074	0.0174648415
б	9931	0.0172169288
г	9807	0.0170019556
ы	9481	0.0164367840
з	9143	0.0158508086
ж	6440	0.0111647388
й	5816	0.0100829381
х	5044	0.0087445563
ш	4882	0.0084637042
ю	3363	0.0058302821
э	1969	0.0034135669
ц	1775	0.0030772378
щ	1736	0.0030096253
ф	904	0.0015672242
ё	695	0.0012048903
ъ	127	0.0002201742

З пробілами (unigrams_with_spaces.csv):

char	count	freq
	117309	0.1690027012
о	65358	0.0941588331
е	48944	0.0705117954
а	46206	0.0665672609
и	38015	0.0547667927
т	37491	0.0540118855
н	36914	0.0531806231
с	30368	0.0437500450
в	26551	0.0382510355
л	26436	0.0380853593
р	25143	0.0362225824
к	19184	0.0276376733
д	18285	0.0263425176
м	18047	0.0259996398
у	17034	0.0245402485
п	15915	0.0229281469
ь	13423	0.0193380155
я	12315	0.0177417612
ч	10074	0.0145132361
б	9931	0.0143072213
г	9807	0.0141285791
ы	9481	0.0136589231
з	9143	0.0131719791
ж	6440	0.0092778678
й	5816	0.0083788943
х	5044	0.0072667027
ш	4882	0.0070333153
ю	3363	0.0048449487
э	1969	0.0028366649
ц	1775	0.0025571763
щ	1736	0.0025009905
ф	904	0.0013023591
ё	695	0.0010012606
ъ	127	0.0001829642

Частоти біграм

Біграми, без пробілів (bigrams_overlap_matrix_no_spaces.csv):

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	п	
а	0.0003571336	0.0013869265	0.0052633860	0.0012534348	0.0032575436	0.0022849614	0.0020283800	0.0048438307	0.0016556435	0.0008824320	0.0075171415	0.0088555256	0.0045803247	0.0056707957	0.0013297158	0.0029038773
б	0.0008234876	0.000052010	0.0000710800	0.0000000000	0.0000156029	0.0024531262	0.000017337	0.0009812505	0.0000000000	0.0001542956	0.0007714779	0.000641453	0.0003536663	0.0025398091	0.000138693	0.000138693
в	0.0067525983	0.0002409785	0.0004802233	0.0002964555	0.0013453187	0.0033598777	0.0000364068	0.0007784125	0.0037549776	0.000104019	0.0006830613	0.0008338895	0.0002617824	0.0023473731	0.0007785586	0.0011563500
г	0.0011494153	0.0000323935	0.0001404263	0.0000242712	0.0012863743	0.0004212789	0.0000334673	0.000537434	0.0007264027	0.0000000000	0.0001248234	0.0015348884	0.0000868829	0.0004108770	0.0091363782	0.0001612302
д	0.0061804911	0.0000693463	0.0012655704	0.0000242712	0.0001352253	0.0049565285	0.0000537451	0.0000667127	0.0029298822	0.0000000000	0.0003467316	0.0006652747	0.0001126878	0.00202017458	0.0043116077	0.0001837678
е	0.003051238	0.0024115184	0.0036025415	0.0043584165	0.0043774867	0.0022537555	0.0011303451	0.0021133292	0.0015186845	0.0023888089	0.0021462687	0.0069433007	0.0058689849	0.0095125820	0.0011793625	0.0035852600
ж	0.0014561469	0.0000693463	0.000537434	0.0000069346	0.0007714779	0.0020121558	0.000069346	0.0000173366	0.0016660454	0.0000000000	0.0002115603	0.0000260049	0.0000277385	0.0009327081	0.0001320897	0.0000502761
з	0.0055623791	0.0001334917	0.001194904	0.0004420828	0.0002704507	0.0000745473	0.0000658790	0.0004784896	0.0000000000	0.0002409785	0.0032393950	0.00022121477	0.0000280652	0.0022121477	0.0007073325	0.0001959034
и	0.0003085911	0.0016808445	0.0052254556	0.0009552456	0.0029073446	0.025534023	0.0004819570	0.0026646325	0.001480795	0.0012308973	0.0039735444	0.0006678034	0.0036393873	0.0007592122	0.0011780552	0.0023716435
й	0.0001109541	0.0002323102	0.000686399	0.0002600487	0.0007385384	0.0002063053	0.001005522	0.0006164312	0.0006466545	0.000069346	0.0005929111	0.0001976370	0.0000420287	0.0009049695	0.0003952740	0.000783613
к	0.0081308565	0.0006327852	0.0007940154	0.0001040195	0.0002721843	0.0005790418	0.0001560292	0.0001740868	0.0028345310	0.0000052010	0.0002495468	0.0007264027	0.0002149736	0.0012274299	0.0108318958	0.0003900731
л	0.0062012950	0.0002929882	0.0007652769	0.0003051238	0.0003397970	0.0048577100	0.0004126106	0.0002167073	0.0064779495	0.0000000000	0.0006795940	0.0001577529	0.0001889867	0.00110357725	0.0007544143	0.0000000000
м	0.0035973406	0.0003016556	0.0008338895	0.0002755616	0.0004646204	0.0004410027	0.0001262939	0.0002357775	0.0039614088	0.0000343673	0.0004715550	0.0002253756	0.0002999929	0.0026126528	0.0051638377	0.0010124563
н	0.0118062117	0.0002791190	0.0006431872	0.0001580283	0.0005426350	0.011690566	0.0000225376	0.0001490496	0.0096287371	0.000104019	0.0005565043	0.0007797483	0.0001438936	0.0034170401	0.0115342074	0.0006310516
п	0.002427121	0.005532130	0.012255942	0.0052200446	0.0071894802	0.027218432	0.003130978	0.001203410	0.0023179009	0.0035071903	0.0030148314	0.0075258098	0.0071236012	0.0080766502	0.0022641575	0.0042405277
р	0.0010592651	0.000069346	0.000034673	0.0000017337	0.0000034673	0.0033424928	0.0000000000	0.0000050210	0.0010124563	0.0000000000	0.0000953512	0.0007194681	0.0000017337	0.0010123087	0.0010769482	0.0000641453
с	0.0094969791	0.0001490946	0.0005409013	0.0002063053	0.0003900731	0.0006909902	0.0002808526	0.0000589444	0.0064977506	0.0000017337	0.0002583151	0.000762810	0.0004334145	0.0009344417	0.0008701744	0.0003137921
т	0.002123995	0.0002895209	0.001746840	0.0001092205	0.0004802233	0.0040444985	0.0002687170	0.000109541	0.0019573000	0.0000017337	0.0045421822	0.0031779953	0.0014545392	0.0013655733	0.0031188509	0.0025238745
у	0.0072224197	0.0004126106	0.003299154	0.0001577629	0.0004576857	0.0068808890	0.001248234	0.0001976370	0.004780389	0.000173366	0.0008304222	0.0003501989	0.0002947219	0.019781039	0.0173625859	0.0007576086
ф	0.0002271092	0.001367275	0.0015169508	0.0017926025	0.0025952862	0.0003588672	0.0021809419	0.0003567004	0.0007055888	0.002877872	0.0011979578	0.0019566893	0.0022357555	0.0015256191	0.0002042448	0.0014680484
х	0.0002687170	0.000017337	0.0000017337	0.0000034673	0.0000017337	0.0000641453	0.0000000000	0.0000000000	0.0004732887	0.0000000000	0.0000017337	0.0000138693	0.0000069346	0.0000000000	0.0000020014	0.0000017337
ц	0.0005270321	0.0001456273	0.0000107272	0.0000066780	0.0000225376	0.0004958262	0.0000329395	0.0000686829	0.0010557799	0.0000000000	0.0001889687	0.0002184409	0.0002305765	0.0004056760	0.0029610880	0.0003954004
ч	0.0006622574	0.0000034673	0.0000739743	0.0000089634	0.0000089634	0.0009483110	0.0000017337	0.0000012156	0.0003848721	0.0000000000	0.0000468088	0.0000017337	0.0000104019	0.0000277385	0.0002288429	0.0000537434
ш	0.003234967	0.0000939134	0.0000089634	0.0000121356	0.0000069346	0.000017337	0.0000069346	0.0018414449	0.0000000000	0.0000665237	0.0000260449	0.0001593910	0.00011035775	0.0001057531	0.0000562761	0.0000000000
щ	0.0011216758	0.0000069346	0.0000104019	0.0000034673	0.0000034673	0.0025831596	0.0000017337	0.0016504425	0.0005898457	0.0005102477	0.0000280652	0.0005984547	0.0000929325	0.0003430438	0.0000000663	0.0000000000
ъ	0.000463540	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0004452028	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.00000485424	0.0000000000	0.0000000000
ы	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
ь	0.0000572107	0.0003883394	0.0015030816	0.0002496468	0.0004299472	0.0009881851	0.0000693463	0.0002115603	0.0004420828	0.0015672269	0.0003952740	0.0022669092	0.0001236034	0.0007073325	0.0002999929	0.0005336967
э	0.0003103248	0.0004958262	0.0013574543	0.0003085911	0.0006657247	0.0009465773	0.001092205	0.0005682739	0.0014129314	0.0000017337	0.0023092326	0.0020211043	0.0009444837	0.0041105034	0.0009240398	0.0012496975
ю	0.0000000000	0.0000000000	0.0000000000	0.0000052010	0.000346732	0.0000000000	0.0000000000	0.0000000000	0.0000017337	0.0000277385	0.000054771	0.0000104019	0.0000346732	0.0000017337	0.0000000000	0.0000000000
я	0.0000936175	0.0004351482	0.0003120585	0.0000641453	0.0005148965	0.0000398741	0.0000936175	0.0000901502	0.0003120585	0.0000034673	0.0002548477	0.0000606780	0.0001959034	0.0003608575	0.000202830	0.0003363297
я	0.0003918067	0.0004732887	0.0018151400	0.0009229922	0.001277714	0.00748496	0.000316556	0.0002963219	0.001581973	0.001581973	0.001282907	0.0003130857	0.0007784125	0.0020873244	0.0008442915	0.0012534348
я	0.0000000000	0.0000710800	0.0001057531	0.0000299126	0.0000381405	0.0000658790	0.0000104019	0.0000329395	0.0000381405	0.0000000000	0.0000450751	0.0000017336	0.0000242712	0.0000745473	0.0000554771	0.0001317580

Біграми, з пробілами (bigrams_overlap_matrix_with_spaces.csv):

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	п	
а	0.0000000000	0.026306539	0.067941751	0.0165013744	0.0031982758	0.0080979767	0.0038264057	0.0021566752	0.045352127	0.0117904006	0.0001033729	0.0082132299	0.024721808	0.0058116417	0.0158055333	0.0113826924
б	0.0169224005	0.0000316946	0.005676219	0.0027646357	0.0007174511	0.018267629	0.0014651561	0.0015271407	0.0035757300	0.001267785	0.0007275357	0.00539818709	0.0070606881	0.0013997165	0.0030902260	0.000115263
в	0.0004048268	0.0006814344	0.000014407	0.0000345760	0.0000000000	0.0000129660	0.0020284650	0.0000014407	0.0007880436	0.0000000000	0.0001181345	0.0000408926	0.0002694043	0.0002399813	0.0020398613	0.0000000000
г	0.0058073197	0.0053939561	0.0000014407	0.0000136946	0.0000187286	0.0008517953	0.0001402919	0.0000000000	0.0005186393	0.0029151835	0.0000000000	0.001152532	0.0009395359	0.001109312	0.0014176141	0.000119944
д	0.0016316150	0.0009391334	0.0000000000	0.0000158473	0.0000014407	0.0011711993	0.0003313529	0.0000000000	0.0009550339	0.0000000000	0.000475419	0.0012796663	0.0000461013	0.0002666230	0.0007474168	0.0000000000
е	0.0008946528	0.0051201225	0.0000288133	0.0000770081	0.0000072833	0.0000916486	0.0000972506	0.0000388979	0.000158473	0.0023727749	0.0000000000	0.0002247547	0.0000795926	0.0000373680	0.0000705926	0.0035388320
ж	0.0174767045	0.0000316946	0.0012634360	0.0011294812	0.0032832750	0.0027516697	0.0015876126	0.0000756349	0.001410065	0.0001267785	0.0019809141	0.0011381252	0.0054601195	0.0041771499	0.0064397176	0.0020430930
з	0.0006857564	0.0012360904	0.0000259320	0.0000000000	0.000014407	0.0006180452	0.004223511	0.0000000000	0.0013095643	0.0000000000	0.0001109312	0.0000312966	0.0000000000	0.000057627	0.0002695958	0.000259320
и	0.0013023610	0.0046331779	0.0009292025	0.0008615175	0.0030414376	0.0006886378	0.001973711	0.0000353046	0.0000129660	0.0003610609	0.0000000000	0.0009796562	0.0002521163	0.001829644	0.0011295054	0.0005348866
й	0.0175761103	0.0000533046	0.0006598245	0.0002439045	0.0004898260	0.0015602400	0.0016913405	0.0004205910	0.0016942218	0.0005546559	0.010027027	0.0023840024	0.0048132610	0.0024808248	0.0039031073	0.001887271
к	0.000386711	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0002522010	0.0000909761	0.0000000000	0.0000000000	0.0000000000	0.000154511	0.0000893212	0.0000475419	0.0000257879	0.0000004320	0.0000000000
л	0.0050351234	0.006791759	0.0000000000	0.0002463537	0.0000000000	0.0000000000	0.00004062675	0.0000216100	0.0000072033	0.0019881174	0.0000000000	0.00000507627	0.0005143173	0.0000004320	0.0005359273	0.0087016135
м	0.0069928860	0.005106321	0.000045913	0.0000158473	0.0001283972	0.00037356438	0.0000000000	0.0000198276	0.0000072033	0.0015674832	0.0000000000	0.00000508263	0.0000069519	0.000144066	0.0002402317	0.0005125913
н	0.0074914569	0.0028879022	0.0000158473	0.00000028813	0.0000763562	0.0000000000	0.00037428471	0.0000000000	0.0000000000	0.0027084498	0.0000000000	0.0000792366	0.0000093212	0.000172880	0.0015266640	0.0037514915
п	0.0044847895	0.0095765226	0.0000000000	0.0000100847	0.0000475419	0.003688616	0.0066719627	0.0000014407	0.0000086440	0.00071704379	0.0000000000	0.00002434733	0.0000000000	0.0000000000	0.0002462960	0.0039081351
а	0.002826329	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
б	0.000547453	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
в	0.000860808	0.007867437	0.0000956245	0.0000394721	0.0001257105	0.0002953363	0.0005061447	0.0000218844	0.0000388979	0.0005290121	0.0001220111	0.0000772018	0.00000590673	0.0000234196	0.0007059257	0.0012777305
г	0.003774231	0.001100691	0.000133279	0.0001774231	0.0000259320	0.0001973711	0.0003334610	0.0000232300	0.0001100847	0.0004565127	0.0000000000	0.00035969673	0.0025470953	0.0010502446	0.0009090595	0.0023943849
д	0.0064570595	0.0000130824	0.0000057627	0.0001251499	0.0001296600	0.0000281179	0.0055400000	0.0000000000	0.0000129660	0.0032141231	0.0000000000	0.0000352969	0.0002160997	0.0000288133	0.001078712	0.0140407023
е	0.0068705303	0.0000002540	0.0000090784	0.0003610312	0.013859195	0.0012388816	0.0020205254	0.0016971031	0.0002650823	0.0000244913	0.0002348824	0.0005605525	0.0015618086	0.0015876126	0.0004999107	0.0000057627
ж	0.000187286	0.0000218624	0.0000000000	0.0000000000	0.0000316946	0.0000000000	0.0000503304	0.0000000000	0.0000000000	0.0003875388	0.0000000000	0.0000014407	0.0000100847	0.0000000000	0.0000000000	0.001829644
з	0.002789577	0.0004033861	0.0000000000	0.0001844051	0.0000000000	0.0000000000	0.0003613602	0.0000000000	0.0000086440	0.0000252485	0.0000000000	0.0000014407	0.000138125	0.0000385586	0.0001455071	0.0023053646
и	0.0032327089	0.0005416900	0.0000000000	0.0000302540	0.0000000000	0.0000000000	0.00007822810	0.0000000000	0.0000000000	0.0000295363	0.0000000000	0.0000230656	0.0000000000	0.00000028813	0.0000000000	0.00000174391
к	0.003688102	0.0025081974	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.00037860670	0.0000000000	0.0000000000	0.0014997320	0.0000000000	0.000813963	0.0000158473	0.0000000000	0.0000810096	0.0000720332
л	0.0005061859	0.0003906694	0.0000000000	0.0000004320	0.0000000000	0.0000000000	0.0002146590	0.0000000000	0.0000000000	0.001368315	0.0000000000	0.0000470293	0.0000571140	0.000144066	0.0000000000	0.0001951678
м	0.0000057627	0.003875388	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0013254516	0.0000000000	0.0000000000	0.0005344866	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
н	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
п	0.0042600458	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
а	0.0181991247	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0013623156	0.0000000000	0.0018416199	0.0000000000	0.0000000000	0.0000000000
б	0.0000259320	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
в	0.0000294978	0.0000000000	0.0002899256	0.0000043020	0.000014407	0.0002388210	0.0000000000	0.0000235056	0.0000096440	0.0000014407	0.0000187286	0.0000043200	0.0000461013	0.0000273726	0.0000000000	0.0000000000
г	0.0140547331	0.0000000000	0.0000172880	0.0002751670	0.0000561859	0.0004912667	0.0001354225	0.001094905	0.0002060151	0.0000230506	0.0001008465	0.0001671171	0.0007959590	0.0000327039	0.0005445713	0.0000000000
д	0.009659234	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000

Ентропія (entropy_summary.csv):

Завдання 2

 $H^{10}:$

$$3.8918 < H^{10} < 4.0279$$

[illegible]

$$2.9540 < H^{20} < 3.4061$$

$$R = 1 - \frac{H_{\infty}}{H_0}$$

H^{10} :

Нижня межа: $R = 1 - \frac{3.8918}{5} \approx 1 - 0,78 \approx 0,22$

Верхня межа: $R = 1 - \frac{4.0279}{5} \approx 1 - 0,81 \approx 0,19$

H^{20} :

Нижня межа: $R = 1 - \frac{2.9540}{5} \approx 1 - 0,59 \approx 0,4$

Верхня межа: $R = 1 - \frac{3.4061}{5} \approx 1 - 0,68 \approx 0,19$

H^{30} :

Нижня межа: $R = 1 - \frac{2.8730}{5} \approx 1 - 0,57 \approx 0,43$

Верхня межа: $R = 1 - \frac{3.3269}{5} \approx 1 - 0,67 \approx 0,33$

Код:

```
#!/usr/bin/env python3
import csv, math, re
from pathlib import Path
from collections import Counter

INPUT_FILE = Path("lab1.txt")
OUT_DIR = Path("out")
LETTERS = "абвгдеёжзийклмнопрстуфхцшщъыьэюяіїєґ"

def read_text(p: Path) -> str:
    return p.read_bytes().decode("utf-8", "ignore")

def normalize(text: str, with_spaces: bool) -> str:
    s = text.lower()
    allowed = LETTERS + (r"\s" if with_spaces else "")
    s = re.sub(f"[^{allowed}]", " " if with_spaces else "", s,
flags=re.IGNORECASE)
    return " ".join(s.split()) if with_spaces else s

def entropy(counts: Counter) -> float:
    n = sum(counts.values())
```



```

        return 0.0 if n == 0 else -sum((c/n)*math.log2(c/n) for c in
counts.values() if c)

def redundancy(h: float, alphabet_size: int) -> float:
    return 0.0 if alphabet_size <= 1 else max(0.0, min(1.0, 1 - h /
math.log2(alphabet_size)))

def save_unigrams(path: Path, counts: Counter):
    total = sum(counts.values())
    rows = [(k, counts[k], (counts[k]/total if total else 0.0)) for k in
counts]
    rows.sort(key=lambda x: (-x[2], x[0]))
    with open(path, "w", newline="", encoding="utf-8") as f:
        w = csv.writer(f); w.writerow(["char", "count", "freq"])
        for ch, cnt, freq in rows:
            w.writerow([repr(ch).strip("'"), cnt, f"{freq:.10f}"])

def save_matrix(path: Path, freqs: dict[str, float], alphabet: list[str]):
    with open(path, "w", newline="", encoding="utf-8") as f:
        w = csv.writer(f)
        w.writerow([""] + [repr(c).strip("'") for c in alphabet])
        for a in alphabet:
            w.writerow([repr(a).strip("'")] + [f"{freqs.get(a+b, 0.0):.10f}"
for b in alphabet])

def process_variant(raw: str, name: str, with_spaces: bool) -> dict:
    text = normalize(raw, with_spaces)
    alphabet = ([" "] + sorted({c for c in text if c != " "})) if (with_spaces
and " " in text) else sorted(set(text))

    uni = Counter(text)
    save_unigrams(OUT_DIR / f"unigrams_{name}.csv", uni)
    h1 = entropy(uni)
    r1 = redundancy(h1, len(alphabet))

    bi_ov = Counter(text[i:i+2] for i in range(len(text)-1))
    total_ov = sum(bi_ov.values())
    freqs_ov = {k: v/total_ov for k, v in bi_ov.items()} if total_ov else {}
    save_matrix(OUT_DIR / f"bigrams_overlap_matrix_{name}.csv", freqs_ov,
alphabet)
    h2_ov_per_char = (entropy(bi_ov) / 2.0)
    r2_ov = redundancy(h2_ov_per_char, len(alphabet))

```

```

bi_no = Counter(text[i:i+2] for i in range(0, len(text)-1, 2))
total_no = sum(bi_no.values())
freqs_no = {k: v/total_no for k, v in bi_no.items()} if total_no else {}
save_matrix(OUT_DIR / f"bigrams_nonoverlap_matrix_{name}.csv", freqs_no,
alphabet)

h2_no_per_char = (entropy(bi_no) / 2.0)
r2_no = redundancy(h2_no_per_char, len(alphabet))

return {
    "variant": name,
    "with_spaces": with_spaces,
    "chars": len(text),
    "alphabet_size": len(alphabet),
    "H1": f"{h1:.10f}",
    "R1": f"{r1:.10f}",
    "H2_overlap_per_char": f"{h2_ov_per_char:.10f}",
    "R2_overlap": f"{r2_ov:.10f}",
    "H2_nonoverlap_per_char": f"{h2_no_per_char:.10f}",
    "R2_nonoverlap": f"{r2_no:.10f}",
}

def main():
    raw = read_text(INPUT_FILE)
    OUT_DIR.mkdir(parents=True, exist_ok=True)
    summaries = []
    for name, s in (("with_spaces", True), ("no_spaces", False)):
        summaries.append(process_variant(raw, name, s))
    if summaries:
        with open(OUT_DIR / "entropy_summary.csv", "w", newline="",
encoding="utf-8") as f:
            keys = list(summaries[0].keys())
            w = csv.DictWriter(f, fieldnames=keys); w.writeheader();
w.writerows(summaries)
if __name__ == "__main__":
    main()

```