

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

Виконав:
студент групи ФБ-91
Кузавка Артем

Київ – 2021

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:

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 видалено 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

Вихідний код створеної програми "frequencyCalculator.py" та зразки оброблюваних текстів розміщені на GitHub: https://github.com/Yab0u/fb-labs-2021/tree/lab_1/cp_1

Частоти символів у тексті з пробілами:

```
{
  ' ': 0.16811087538968736, 'о': 0.09594401450184759, 'е': 0.07306202129460862,
  'а': 0.06921992808836565, 'н': 0.05768244340195815, 'и': 0.054918526707901316,
  'т': 0.049859438651009454, 'с': 0.043908302706148344, 'л': 0.04204950647901913,
  'в': 0.039152996484098763, 'р': 0.032772332944890986, 'к': 0.028544909911255863,
  'д': 0.024766312088525016, 'м': 0.023873019193035926, 'у': 0.02269835459806177,
  'п': 0.02008383880638639, 'я': 0.018074396669355883, 'ь': 0.016664425653130945,
  'ы': 0.015425643170884752, 'г': 0.01513119889244131, 'б': 0.014645023455941673,
  'ч': 0.014208648320235854, 'з': 0.013642792402466112, 'ж': 0.00955854523361321,
  'й': 0.008560671706457107, 'ш': 0.007240963735420962, 'х': 0.006372570990328782,
  'ю': 0.005241481658549218, 'э': 0.002966230416031713, 'щ': 0.002365514287706298,
  'ц': 0.002159465543171881, 'ф': 0.0008378900608559676, 'ъ': 0.00024588898516947,
  'ё': 1.1827571438531489e-05
}
```

Частоти біграм у тексті з пробілами та з кроком 1:

```
{
  'а': '0.003136', 'б': '0.006768', 'в': '0.014722', 'г': '0.003367', 'д':
  '0.007384', 'е': '0.005339', 'ж': '0.002541', 'з': '0.004201', 'и':
  '0.011098', 'й': '0.000001', 'к': '0.009047', 'л': '0.003625', 'м':
  '0.005550', 'н': '0.016625', 'о': '0.013655', 'п': '0.014051', 'р':
  '0.003788', 'с': '0.016188', 'т': '0.008300', 'у': '0.004188', 'ф':
  '0.000382', 'х': '0.001529', 'ц': '0.000235', 'ч': '0.006394', 'ш':
  '0.000695', 'щ': '0.000102', 'ы': '0.000001', 'ь': '0.000001', 'э':
  '0.002920', 'ю': '0.000027', 'я': '0.002248', 'ё': '0.000001', 'а ':
  '0.018823', 'аа': '0.000007', 'аб': '0.000581', 'ав': '0.003040', 'аг':
  '0.000581', 'ад': '0.002182', 'ае': '0.000980', 'аж': '0.001116', 'аз':
  '0.004199', 'аи': '0.000120', 'ай': '0.000523', 'ак': '0.004168', 'ал':
  '0.009702', 'ам': '0.002380', 'ан': '0.004335', 'ао': '0.000004', 'ап':
  '0.000566', 'ар': '0.002248', 'ас': '0.003486', 'ат': '0.004133', 'ау':
  '0.000039', 'аф': '0.000237', 'ах': '0.000817', 'ац': '0.000111', 'ач':
  '0.000762', 'аш': '0.000636', 'ащ': '0.000206', 'аз': '0.000001', 'аю':
  '0.000846', 'ая': '0.002390', 'б ': '0.000380', 'ба': '0.000841', 'бб':
  '0.000001', 'бв': '0.000087', 'бг': '0.000004', 'бд': '0.000034', 'бе':
  '0.002013', 'бж': '0.000004', 'бз': '0.000002', 'би': '0.000760', 'бк':
  '0.000248', 'бл': '0.000757', 'бм': '0.000059', 'бн': '0.000299', 'бо':
  '0.001957', 'бп': '0.000001', 'бр': '0.001069', 'бс': '0.000072', 'бт':
  '0.000008', 'бу': '0.000970', 'бх': '0.000055', 'бц': '0.000006', 'бч':
  '0.000006', 'бш': '0.000008', 'бщ': '0.000221', 'бъ': '0.000110', 'бы':
  '0.004143', 'бь': '0.000015', 'бю': '0.000004', 'бя': '0.000511', 'в ':
  '0.005207', 'ва': '0.005417', 'вб': '0.000015', 'вв': '0.000022', 'вг':
  '0.000029', 'вд': '0.000317', 'ве': '0.004179', 'вж': '0.000001', 'вз':
  '0.000543', 'ви': '0.004527', 'вк': '0.000145', 'вл': '0.000753', 'vm':
  '0.000147', 'вн': '0.001243', 'во': '0.006951', 'вп': '0.000204', 'вр':
  '0.001031', 'вс': '0.003462', 'вт': '0.000219', 'ву': '0.000662', 'vx':
  '0.000056', 'вц': '0.000024', 'вч': '0.000079', 'вш': '0.001078', 'вщ':
  '0.000008', 'въ': '0.000004', 'вы': '0.002391', 'вь': '0.000186', 'вю':
  '0.000002', 'вя': '0.000250', 'г ': '0.000656', 'га': '0.000811', 'гв':
  '0.000007', 'гд': '0.001160', 'ге': '0.000321', 'гз': '0.000001', 'ги':
  '0.000740', 'гк': '0.000119', 'гл': '0.001503', 'гм': '0.000001', 'гн':
  '0.000169', 'го': '0.008304', 'гр': '0.000666', 'гс': '0.000029', 'гт':
  '0.000006', 'гу': '0.000591', 'гф': '0.000001', 'гч': '0.000038', 'гш':
  '0.000006', 'гь': '0.000001', 'гю': '0.000002', 'д ': '0.000952', 'да':
  '0.004229', 'дб': '0.000022', 'дв': '0.000756', 'дг': '0.000004', 'дд':
  '0.000037', 'де': '0.004496', 'дж': '0.000009', 'дз': '0.000001', 'ди':
}
```

'0.002026'	'дк':	'0.000205'	'дл':	'0.000786'	'дм':	'0.000070'	'дн':
'0.001501'	'до':	'0.003682'	'дп':	'0.000065'	'др':	'0.001492'	'дс':
'0.000310'	'дт':	'0.000189'	'ду':	'0.001812'	'дх':	'0.000062'	'дц':
'0.000176'	'дч':	'0.000026'	'дш':	'0.000103'	'дщ':	'0.000001'	'дъ':
'0.000059'	'ды':	'0.000505'	'дь':	'0.000784'	'дю':	'0.000006'	'дя':
'0.000400'	'е':	'0.019025'	'еа':	'0.000032'	'еб':	'0.001303'	'ев':
'0.002443'	'ег':	'0.003697'	'ед':	'0.002396'	'ее':	'0.002200'	'еж':
'0.000799'	'ез':	'0.001152'	'ей':	'0.000096'	'ей':	'0.002481'	'ек':
'0.001750'	'ел':	'0.005732'	'ем':	'0.003978'	'ен':	'0.007049'	'ео':
'0.000178'	'еп':	'0.001359'	'ер':	'0.005096'	'ес':	'0.003723'	'ет':
'0.004372'	'еу':	'0.000090'	'еф':	'0.000004'	'ех':	'0.000779'	'ец':
'0.000151'	'еч':	'0.000865'	'еш':	'0.000941'	'ещ':	'0.000609'	'еэ':
'0.000001'	'ею':	'0.000410'	'ея':	'0.000349'	'ж':	'0.000281'	'жа':
'0.001401'	'жб':	'0.000044'	'жг':	'0.000006'	'жд':	'0.000756'	'же':
'0.004084'	'жж':	'0.000011'	'жи':	'0.001434'	'жк':	'0.000090'	'жл':
'0.000004'	'жм':	'0.000002'	'жн':	'0.000985'	'жо':	'0.000040'	'жр':
'0.000001'	'жс':	'0.000096'	'жу':	'0.000241'	'жц':	'0.000001'	'жч':
'0.000041'	'жь':	'0.000042'	'з':	'0.001058'	'за':	'0.005399'	'зб':
'0.000159'	'зв':	'0.000782'	'зг':	'0.000528'	'зд':	'0.000743'	'зе':
'0.000217'	'зж':	'0.000107'	'зз':	'0.000018'	'зи':	'0.000291'	'зк':
'0.000085'	'зл':	'0.000179'	'зм':	'0.000256'	'зн':	'0.001907'	'зо':
'0.000339'	'зр':	'0.000186'	'зс':	'0.000045'	'зт':	'0.000014'	'зу':
'0.000254'	'зц':	'0.000004'	'зч':	'0.000056'	'зш':	'0.000006'	'зь':
'0.000014'	'зы':	'0.000355'	'зь':	'0.000160'	'зю':	'0.000004'	'зя':
'0.000477'	'и':	'0.018205'	'иа':	'0.000078'	'иб':	'0.000376'	'ив':
'0.002660'	'иг':	'0.000401'	'ид':	'0.001539'	'ие':	'0.002346'	'иж':
'0.000309'	'из':	'0.001682'	'ии':	'0.000534'	'ий':	'0.001277'	'ик':
'0.001507'	'ил':	'0.004415'	'им':	'0.002748'	'ин':	'0.004007'	'ио':
'0.000077'	'ип':	'0.000129'	'ир':	'0.000490'	'ис':	'0.002021'	'ит':
'0.003771'	'иу':	'0.000009'	'иф':	'0.000014'	'их':	'0.001271'	'иц':
'0.000791'	'ич':	'0.001849'	'иш':	'0.000389'	'ищ':	'0.000146'	'ию':
'0.000227'	'ия':	'0.001652'	'й':	'0.007014'	'йб':	'0.000001'	'йд':
'0.000209'	'йе':	'0.000001'	'йз':	'0.000001'	'йк':	'0.000065'	'йл':
'0.000085'	'йм':	'0.000054'	'йн':	'0.000250'	'йо':	'0.000002'	'йр':
'0.000001'	'йс':	'0.000475'	'йт':	'0.000235'	'йф':	'0.000004'	'йц':
'0.000042'	'йч':	'0.000075'	'йш':	'0.000048'	'йщ':	'0.000001'	'к':
'0.004721'	'ка':	'0.007690'	'кб':	'0.000001'	'кв':	'0.000161'	'кд':
'0.000004'	'ке':	'0.000487'	'кж':	'0.000008'	'кз':	'0.000002'	'ки':
'0.002774'	'кк':	'0.000013'	'кл':	'0.000312'	'кн':	'0.000608'	'ко':
'0.007684'	'кр':	'0.001366'	'кс':	'0.000941'	'кт':	'0.000365'	'ку':
'0.001393'	'кф':	'0.000001'	'кх':	'0.000001'	'кц':	'0.000005'	'кш':
'0.000011'	'л':	'0.007307'	'ла':	'0.007679'	'лб':	'0.000018'	'лв':
'0.000003'	'лг':	'0.000119'	'лд':	'0.000014'	'ле':	'0.005180'	'лж':
'0.000380'	'лз':	'0.000003'	'ли':	'0.005178'	'лк':	'0.000278'	'лл':
'0.000251'	'лм':	'0.000001'	'лн':	'0.000288'	'ло':	'0.005781'	'лп':
'0.000040'	'лс':	'0.001106'	'лт':	'0.000024'	'лу':	'0.001170'	'лф':
'0.000004'	'лц':	'0.000001'	'лч':	'0.000121'	'лш':	'0.000004'	'лщ':
'0.000001'	'лы':	'0.000947'	'ль':	'0.003447'	'лю':	'0.000995'	'ля':
'0.001708'	'м':	'0.006994'	'ма':	'0.002422'	'мб':	'0.000008'	'мв':
'0.000001'	'мг':	'0.000030'	'ме':	'0.002880'	'мж':	'0.000001'	'мз':
'0.000002'	'ми':	'0.002572'	'мк':	'0.000052'	'мл':	'0.000090'	'мм':
'0.000035'	'мн':	'0.001257'	'мо':	'0.003141'	'мп':	'0.000049'	'мр':
'0.000073'	'мс':	'0.000103'	'мт':	'0.000020'	'му':	'0.002758'	'мф':
'0.000001'	'мх':	'0.000001'	'мц':	'0.000009'	'мч':	'0.000019'	'мш':
'0.000004'	'мщ':	'0.000003'	'мы':	'0.000902'	'мь':	'0.000063'	'мю':
'0.000002'	'мя':	'0.000380'	'н':	'0.005513'	'на':	'0.010870'	'нб':
'0.000011'	'нв':	'0.000004'	'нг':	'0.000074'	'нд':	'0.000635'	'не':
'0.009956'	'нж':	'0.000003'	'нз':	'0.000014'	'ни':	'0.007497'	'нк':
'0.000246'	'нм':	'0.000001'	'нн':	'0.002782'	'но':	'0.009594'	'нр':
'0.000078'	'нс':	'0.001005'	'нт':	'0.000397'	'ну':	'0.002448'	'нф':
'0.000014'	'нц':	'0.000234'	'нч':	'0.000256'	'нш':	'0.000002'	'нщ':
'0.000155'	'ны':	'0.002772'	'нь':	'0.001233'	'ню':	'0.000159'	'ня':
'0.001732'	'о':	'0.024319'	'оа':	'0.000002'	'об':	'0.003351'	'ов':
'0.007215'	'ог':	'0.004515'	'од':	'0.004125'	'ое':	'0.002452'	'ож':

'0.001904'	'оз'	'0.000977'	'ои'	'0.000708'	'ой'	'0.002874'	'ок'
'0.001533'	'ол'	'0.005359'	'ом'	'0.004720'	'он'	'0.008397'	'оо'
'0.000184'	'оп'	'0.001068'	'ор'	'0.005377'	'ос'	'0.005731'	'от'
'0.006282'	'оу'	'0.000046'	'оф'	'0.000120'	'ох'	'0.000324'	'оц'
'0.000079'	'оч'	'0.001582'	'ош'	'0.001247'	'ощ'	'0.000141'	'оз'
'0.000027'	'ою'	'0.000804'	'оя'	'0.000481'	'п'	'0.000041'	'па'
'0.001140'	'пе'	'0.002046'	'пи'	'0.000751'	'пк'	'0.000059'	'пл'
'0.000444'	'пн'	'0.000060'	'по'	'0.008149'	'пп'	'0.000022'	'пр'
'0.006151'	'пс'	'0.000003'	'пт'	'0.000047'	'пу'	'0.000491'	'пф'
'0.000001'	'пц'	'0.000012'	'пч'	'0.000031'	'пш'	'0.000014'	'пщ'
'0.000001'	'пы'	'0.000244'	'пъ'	'0.000036'	'пю'	'0.000003'	'пя'
'0.000337'	'р'	'0.000637'	'ра'	'0.006650'	'рб'	'0.000210'	'рв'
'0.000291'	'рг'	'0.000349'	'рд'	'0.000344'	'ре'	'0.005101'	'рж'
'0.000289'	'рз'	'0.000016'	'ри'	'0.004360'	'рк'	'0.000567'	'рл'
'0.000049'	'рм'	'0.000111'	'рн'	'0.000627'	'ро'	'0.006721'	'рп'
'0.000059'	'рр'	'0.000016'	'рс'	'0.000176'	'рт'	'0.000444'	'ру'
'0.002247'	'рф'	'0.000006'	'рх'	'0.000096'	'рц'	'0.000022'	'рч'
'0.000065'	'рш'	'0.000212'	'рщ'	'0.000030'	'ры'	'0.001213'	'рь'
'0.000877'	'рю'	'0.000111'	'ря'	'0.000876'	'с'	'0.002861'	'са'
'0.001741'	'сб'	'0.000049'	'св'	'0.001804'	'сг'	'0.000011'	'сд'
'0.000324'	'се'	'0.004407'	'сж'	'0.000028'	'сз'	'0.000019'	'си'
'0.001285'	'ск'	'0.004897'	'сл'	'0.002483'	'см'	'0.000816'	'сн'
'0.000931'	'со'	'0.002329'	'сп'	'0.001684'	'ср'	'0.000109'	'сс'
'0.000542'	'ст'	'0.010143'	'су'	'0.000491'	'сф'	'0.000007'	'сх'
'0.000127'	'сц'	'0.000045'	'сч'	'0.000420'	'сш'	'0.000056'	'сщ'
'0.000002'	'съ'	'0.000022'	'сы'	'0.000324'	'сь'	'0.002822'	'сэ'
'0.000002'	'сю'	'0.000121'	'ся'	'0.002991'	'сё'	'0.000011'	'т'
'0.004317'	'та'	'0.004914'	'тб'	'0.000012'	'тв'	'0.002577'	'тг'
'0.000006'	'тд'	'0.000118'	'те'	'0.004941'	'тж'	'0.000004'	'тз'
'0.000002'	'ти'	'0.003729'	'тк'	'0.000468'	'тл'	'0.000256'	'тм'
'0.000013'	'тн'	'0.001036'	'то'	'0.014266'	'тп'	'0.000047'	'тр'
'0.002393'	'тс'	'0.000812'	'тт'	'0.000076'	'ту'	'0.001135'	'тф'
'0.000019'	'тх'	'0.000011'	'тц'	'0.000063'	'тч'	'0.000277'	'тш'
'0.000001'	'тщ'	'0.000013'	'тъ'	'0.000036'	'ты'	'0.001480'	'ть'
'0.006269'	'тэ'	'0.000001'	'тю'	'0.000042'	'тя'	'0.000525'	'у'
'0.006704'	'уа'	'0.000032'	'уб'	'0.000505'	'ув'	'0.001124'	'уг'
'0.000981'	'уд'	'0.001508'	'уе'	'0.000246'	'уж'	'0.001727'	'уз'
'0.000276'	'уй'	'0.000002'	'уй'	'0.000120'	'ук'	'0.000756'	'ул'
'0.001138'	'ум'	'0.001184'	'ун'	'0.000203'	'уо'	'0.000004'	'уп'
'0.000548'	'ур'	'0.000515'	'ус'	'0.000953'	'ут'	'0.001045'	'уу'
'0.000001'	'уф'	'0.000015'	'ух'	'0.000286'	'уц'	'0.000004'	'уч'
'0.000858'	'уш'	'0.000697'	'ущ'	'0.000187'	'уэ'	'0.000013'	'ую'
'0.000950'	'уя'	'0.000116'	'ф'	'0.000030'	'фа'	'0.000062'	'фе'
'0.000106'	'фи'	'0.000285'	'фл'	'0.000019'	'фм'	'0.000002'	'фн'
'0.000001'	'фо'	'0.000078'	'фр'	'0.000138'	'фс'	'0.000009'	'фт'
'0.000019'	'фу'	'0.000024'	'фф'	'0.000004'	'фч'	'0.000001'	'фы'
'0.000006'	'фь'	'0.000050'	'фю'	'0.000001'	'фя'	'0.000001'	'х'
'0.002636'	'ха'	'0.000809'	'хв'	'0.000095'	'хг'	'0.000001'	'хд'
'0.000001'	'хе'	'0.000034'	'хи'	'0.000097'	'хл'	'0.000087'	'хм'
'0.000050'	'хн'	'0.000099'	'хо'	'0.002213'	'хп'	'0.000001'	'хр'
'0.000056'	'хс'	'0.000040'	'хт'	'0.000006'	'ху'	'0.000138'	'хч'
'0.000001'	'хш'	'0.000009'	'хъ'	'0.000001'	'ц'	'0.000171'	'ца'
'0.000463'	'цб'	'0.000001'	'цв'	'0.000037'	'це'	'0.000621'	'ци'
'0.000134'	'цк'	'0.000101'	'цм'	'0.000001'	'цо'	'0.000276'	'цу'
'0.000204'	'цц'	'0.000007'	'цш'	'0.000001'	'цы'	'0.000142'	'ч'
'0.000774'	'ча'	'0.002199'	'чв'	'0.000003'	'че'	'0.003301'	'чи'
'0.001272'	'чк'	'0.000306'	'чл'	'0.000043'	'чм'	'0.000005'	'чн'
'0.000473'	'чо'	'0.000049'	'чр'	'0.000059'	'чт'	'0.004473'	'чу'
'0.000926'	'чш'	'0.000167'	'чь'	'0.000160'	'ш'	'0.000063'	'ша'
'0.000910'	'шв'	'0.000085'	'шг'	'0.000001'	'ше'	'0.002265'	'ши'
'0.001536'	'шк'	'0.000344'	'шл'	'0.000538'	'шм'	'0.000006'	'шн'
'0.000363'	'шо'	'0.000278'	'шп'	'0.000006'	'шр'	'0.000001'	'шт'
'0.000063'	'шу'	'0.000281'	'шц'	'0.000002'	'шш'	'0.000001'	'шь'
'0.000495'	'шю'	'0.000002'	'щ'	'0.000007'	'ща'	'0.000286'	'ще'

```

'0.001276', 'щи': '0.000654', 'щн': '0.000027', 'що': '0.000001', 'щр':
'0.000005', 'щу': '0.000087', 'щъ': '0.000022', 'ъ': '0.000001', 'ье':
'0.000121', 'ью': '0.000007', 'ья': '0.000116', 'ы': '0.004194', 'ыб':
'0.000477', 'ыв': '0.000957', 'ыг': '0.000091', 'ыд': '0.000136', 'ые':
'0.000919', 'ыж': '0.000024', 'ыз': '0.000061', 'ыи': '0.000005', 'ый':
'0.001163', 'ык': '0.000154', 'ыл': '0.002573', 'ым': '0.001150', 'ын':
'0.000327', 'ып': '0.000115', 'ыр': '0.000281', 'ыс': '0.000733', 'ыт':
'0.000702', 'ыу': '0.000006', 'ых': '0.000771', 'ыц': '0.000001', 'ыч':
'0.000168', 'ыш': '0.000419', 'ыщ': '0.000001', 'ья': '0.000001', 'ь':
'0.010594', 'ьб': '0.000082', 'ьв': '0.000025', 'ьг': '0.000047', 'ьд':
'0.000027', 'ье': '0.000483', 'ьж': '0.000001', 'ьз': '0.000166', 'ьи':
'0.000413', 'ьк': '0.001196', 'ьм': '0.000210', 'ьн': '0.001142', 'ьо':
'0.000007', 'ьп': '0.000002', 'ьс': '0.000829', 'ьт': '0.000072', 'ьх':
'0.000001', 'ьц': '0.000097', 'ьч': '0.000061', 'ьш': '0.000411', 'ьщ':
'0.000016', 'ьэ': '0.000002', 'ью': '0.000384', 'ья': '0.000397', 'э':
'0.000001', 'эв': '0.000001', 'эг': '0.000003', 'эд': '0.000001', 'эз':
'0.000003', 'эи': '0.000001', 'эй': '0.000002', 'эк': '0.000037', 'эл':
'0.000035', 'эм': '0.000001', 'эн': '0.000022', 'еп': '0.000003', 'эр':
'0.000004', 'эс': '0.000002', 'эт': '0.002845', 'эф': '0.000002', 'эх':
'0.000001', 'эш': '0.000001', 'ю': '0.003426', 'юа': '0.000001', 'юб':
'0.000627', 'юв': '0.000003', 'юг': '0.000004', 'юд': '0.000233', 'юж':
'0.000002', 'юз': '0.000002', 'юи': '0.000001', 'юк': '0.000002', 'юл':
'0.000012', 'юм': '0.000012', 'юн': '0.000025', 'юп': '0.000009', 'юр':
'0.000037', 'юс': '0.000162', 'ют': '0.000240', 'юф': '0.000002', 'юх':
'0.000010', 'юц': '0.000005', 'юч': '0.000037', 'юш': '0.000026', 'ющ':
'0.000328', 'юю': '0.000034', 'я': '0.011186', 'яа': '0.000001', 'яб':
'0.000023', 'яв': '0.000248', 'яг': '0.000284', 'яд': '0.000540', 'яе':
'0.000065', 'яж': '0.000291', 'яз': '0.000300', 'яи': '0.000039', 'яй':
'0.000119', 'як': '0.000103', 'ял': '0.000817', 'ям': '0.000308', 'ян':
'0.000642', 'яп': '0.000056', 'яр': '0.000045', 'яс': '0.000853', 'ят':
'0.001351', 'яу': '0.000001', 'ях': '0.000176', 'яц': '0.000032', 'яч':
'0.000121', 'яш': '0.000049', 'ящ': '0.000195', 'яю': '0.000096', 'яя':
'0.000133', 'ё': '0.000011', 'ёк': '0.000001'
}

```

За отриманими даними була розрахована питома ентропія на символ та на символ біграми:

$$H_1 = 4.3471$$

$$H_2 = 3.9375$$

Для біграм з кроком 2:

$$H_2 = 3.9373$$

Розрахунок умовної ентропії джерела:

$H^{(50)}$

Лабораторная работа №1

Произвольная часть текста:
_человеческие_существа_во_всех_частях_зеленого_шар

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:
Символ по счету:
Номер эксперимента: 50

Поле ввода символов:

Продолжить

Другой

Неравенство для энтропии:
1,67350951985244 < H < 2,25964913004547

Двоичная таблица угаданных символов:
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
01000000000000000000000000000000
00000000000000000000000000000000

Вероятности:

$q[1]$	= 0,55102041
$q[2]$	= 0,14285711
$q[3]$	= 0,10204081
$q[4]$	= 0,0204081
$q[5]$	= 0
$q[6]$	= 0
$q[7]$	= 0
$q[8]$	= 0
$q[9]$	= 0
$q[10]$	= 0
$q[11]$	= 0,020408
$q[12]$	= 0
$q[13]$	= 0
$q[14]$	= 0,020408
$q[15]$	= 0
$q[16]$	= 0
$q[17]$	= 0,020408
$q[18]$	= 0,020408
$q[19]$	= 0
$q[20]$	= 0
$q[21]$	= 0
$q[22]$	= 0,061224
$q[23]$	= 0
$q[24]$	= 0
$q[25]$	= 0
$q[26]$	= 0,020408
$q[27]$	= 0,020408
$q[28]$	= 0
$q[29]$	= 0
$q[30]$	= 0
$q[31]$	= 0
$q[32]$	= 0

Строка состояния:

$H^{(30)}$

Лабораторная работа №1

Произвольная часть текста:
ия_как_все_организмы_подчиняю

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:
Символ по счету:
Номер эксперимента: 50

Поле ввода символов:

Продолжить

Другой

Неравенство для энтропии:
2,17775849810142 < H < 2,66131527740124

Двоичная таблица угаданных символов:
00010000000000000000000000000000
00010000000000000000000000000000
00000000000000000000000000000001
10000000000000000000000000000000
00000000000000000000000000000010

Вероятности:

$q[1]$	= 0,44
$q[2]$	= 0,12
$q[3]$	= 0,08
$q[4]$	= 0,12
$q[5]$	= 0
$q[6]$	= 0
$q[7]$	= 0,04
$q[8]$	= 0
$q[9]$	= 0,04
$q[10]$	= 0
$q[11]$	= 0
$q[12]$	= 0
$q[13]$	= 0
$q[14]$	= 0
$q[15]$	= 0
$q[16]$	= 0
$q[17]$	= 0
$q[18]$	= 0
$q[19]$	= 0
$q[20]$	= 0
$q[21]$	= 0,04
$q[22]$	= 0
$q[23]$	= 0,04
$q[24]$	= 0
$q[25]$	= 0
$q[26]$	= 0
$q[27]$	= 0
$q[28]$	= 0
$q[29]$	= 0
$q[30]$	= 0
$q[31]$	= 0,04
$q[32]$	= 0,04

Строка состояния:

[illegible][illegible]

Оцінка надлишковості російської мови:

- Для тексту з пробілами та з буквами "ъ" та "ё"
 $R = 0,46$
- Для тексту з пробілами та без букв "ъ" та "ё"
 $R = 0,455$
- Для тексту без пробілів та з буквами "ъ" та "ё"
 $R = 0,38$
- Для тексту без пробілів та без букв "ъ" та "ё"
 $R = 0,37$

Контрольні запитання:

1. Дайте визначення ентропії, сукупної ентропії, умовної ентропії. Який зміст несуть ці поняття?
Ентропія – міра місткості об'єкта щодо інформації.
Сукупна ентропія – міра місткості двох ансамблів щодо інформації.
Умовна ентропія – міра місткості одного ансамблю щодо інформації, якщо поведінка іншого є визначеною.
2. Які визначення ентропії на символ джерела ви знаєте?
$$H_{\infty} = \lim_{n \rightarrow \infty} H_n, \text{ де } H_n = \frac{1}{n} H(x_1, x_2, \dots, x_n),$$
3. Порівняйте одержані значення H_1 , H_2 , $H^{(10)}$, $H^{(20)}$, $H^{(30)}$. Зробіть висновки.
Щодо H_1 , H_2 можна зробити висновок, що питома ентропія на символ та на символ біграми менша, ніж на окремий символ. Щодо умовної ентропії, наглядно, що ми маємо менше інформації зі зменшенням кількості попередніх символів перед тим, який потрібно вгадати, тому це зробити складніше.
4. Що таке надлишковість джерела? Яка надлишковість російського письмового тексту згідно ваших даних?
Надлишковість джерела – кількість інформації, яку можна усунути без втрати сенсу інформації в джерелі. Обчислена надлишковість російської мови приблизно дорівнює 4,2.
5. Які моделі відкритих текстів розглядаються у криптографії?
Моделі, що відрізняються залежно від наявності пробілів, розділових знаків, деяких букв алфавіту, якими можна знехтувати та в залежності від регістру літер.