МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ім. ІГОРЯ СІКОРСЬКОГО”

КРИПТОГРАФІЯ
КОМП’ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали роботу:
студент ФБ-23 Гнидюк Даніїл
студент ФБ-23 Жушман Ілля

Київ 2024

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:
1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1 та H2 за безпосереднім означенням. Підрахувати частоти букв та біграм, атакож значення H1 та H2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H1 та H2 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)
Текст без пробілів

```
агодышлидашлибыстроинеслышнокакподснежныеводыпротекаламолодостьеленывбездействиивнешнемвовнутреннейборьбеитревогеподругунеенебылоизовсехдевицпосещавшихдомстаховых
онанесошласьнисоднойродительскаявластьникогданетяготеланаделенойасшестнадцатилетнеговозрастаонасталапочтисовсемнезависимаоназажиласобственнойсвоеюжизньюножизньюод
инокойеедушаиразгораласьипогасалаодинокоонабиласькакптицавклеткеаклеткинебылоникт онестеснялееиниктонеудерживалаонарвалаcьитомиласьонаиногдасамасебянепонималадажеб
ояласьсамойсебявсечтоокружалоеeказалосьейнетобессмысленнымнетонепонятнымкакжитьбезлюбвиалюбитьнекогодумалаонаистрашностановилосьейотэтихдумотэтихощущенийвосемнадц
атилетоначутьнеумерлаотзлокачественнойлихорадкипотрясенныйдоoснованиявесьееорганизмотприродыздоровыйикрепкийдолгонемогсправитьсяпоследниеследыболезниисчезлинаконе
цноотецеленыниколаевнывсееще небезозлобления толковалобеенервахиногдаейприходиловголовучтоонажелаетчеготочегониктонежелаеточемниктонемыслитвцелойроссиипотомонаутиха
ладажесмеяласьнадсобойбеспечнопроводиладеньзаднемновнезапночтотосильноебезымянноесчемонасовладетьнеумелатакизакипаловнейтакипросилосьвырватьсянаружугрозапроходила
опускалисьусталыеневзлетевшиекрыльянопорывыэтинеобходилисьейдаромкаконанистараласьневыдатьтогочтовнейпроисходилотоскавзволнованнойдушисказываласьвсамомеенаружномс
покойствииродныеeeчастобыливправепожиматьплечамиудивлятьсяинепониматьеестранностейвденьскоторогоначалсянашрассказеленадольшеобыкновенногонеотходилаотокнаонамного
думалаоберсеньевеосвоемразговоресним потребностьвзащитеинформациивозникаетвсвязиснеобходимостьюобеспечитьсекретностьисследованийвстратегическихобластяхправильнорас
пределятьинформациюопромышленныхразработкахирегулироватьинформациюоличностивсовременномобщественачаловосьмидесятыхгодоврассматриваетсякакначальныйпункткогдасоциал
ьныепротестывдемократическихстранахпомоглисплестисьглобальнойсетихакеровполитическийфлиртнапочвенарушенияправчеловекапородилтьмуорганизацийхакероввмассестранмирап
очтиодновременноменеечемзагодэтигруппыузналипрелестьсотрудничестваихчленысвободнообменивалисьидеямичерезнациональныеграницычастопоукраденнымпаролямдающимбесплатны
йдоступктелефоннойсетинесколькопричинобьединившисьвместесделалимеждународныйкомпьютерныйразбойлекгимидейственнымновыетехнологиисоздавшиеболеемощныеидешевыекомпьют
ерыразвитиекоммуникацийдлясвязиимеждународныйхарактерстандартовустановленныхтранснациональнымикорпорациямивпринципеестьлишьдвавидаугрозыраскрытиеивидоизменениедан
ныхраскрытиеданныхпредполагаетчтокомутослучайноилипослецеленаправленныхдействийстализвестенсмыслинформацииэтотвиднарушениявстречаетсянаиболеечастопоследствиямогут
бытьсамыеразныеeслипохищентекстктнигисправочниканакоторуюпотраченымесяцыработыдесятковлюдейтодляколлективаавторовэтокатастрофаипотеримогутвыражатьсявтысячахдолларо
воднакоеслик нигауже изданатодостаточнолишьслегкапожуритьпохитителяирассказатьослучившемсявотделеновостейгазетыилипотелевидениюпохитительможетсделатькнигевеликолепн
уюрекламуоченьважнуюинформациюоберегаемуюотраскрытияпредставляютсведенияолюдяхисторииболезнипись масостояниясчетоввбанкаходнакопомнениюбольшогочисласпециалистовугр
озыличностисвведениемкомпьютеровосталисьнатомжеуровнеивтомжесостояниичтоидообширногоиспользованияэвмвсовременноммиретуризмстановитсявсеболееважнойбыстроразвивающе
йсяотраслью хозяйствадоходыоттуризмастановятсяважнойчастьювалютныхпоступленийвомногихстранахразвитиетуризмаспособствуетростуобщественногопроизводстваулучшениюегост
руктурыроступроизводительноститрудавомногихотрасляхэкономикидаженеимеющихктуризмупрямогоотношениямеждународноетуристскоепотреблениестимулируетмногочисленныеэконом
ическиепроцессыоткрывающиедополнительныерынкидляпродукциинетуристскихотраслейсоздаваятемсамымусловиядляростапроизводствавсеэтифакторыделаютразвитиеиндустриитуризм
аоченьважнымдлястранспереходнымтипомэкономикиэкономическиетрудностикоторыепереживаютэтигосударстванемогутнесказатьсянауровнеразвитиятуризманоприэтомкаждаястранаим
еетвэтомотношениисвоюспецификуцельданнойработырассмотретьипроанализироватьорганизациютуристскойдеятельностивстранеспереходнымтипомэкономикинапримеревенгриивначале
рассматриваютсятеоретикометодическиеположенияисследованиязатемдаетсяоценкаразличныхфакторовразвитияиндустриитуризмавенгрииприродноресурсныйкультурноисторическийи
нфраструктурныйпотенциалкомплексноетуристскоерайонированиедалеепроводитсяанализсовременногосостоянияиндустриитуризмавенгрииееотдельныхкомпонентовнафонеобщегоуровн
яэкономическогоразвитиястраныдаетсяоценкасоциальноэкономическойролииндустриитуризмавэкономикевенгрииивзаключениепроводитсяобщийанализорганизациитуристскойдеятельн
остивстранахспереходнымтипомэкономикивобщемивенгриивчастностивенгрияпринадлежакстранамспереходнымтипомэкономикиимееттемнеменееспецифическиечертыкоторыеотличаютеео
тдругихстранэтоготипавотношенииразвитияиндустриитуризмаосновнойтакойчертойявляетсяточтотуризмвенгрииразвиваетсяужедавноещевначаледвадцатоговекавэтойстранесложилис
ьтрадиционныетуристскиесвязитуризмявляетсяважнойотраслью народногохозяйствасовременной
```

Текст із пробілами

```
а годы шли да шли быстро и неслышно как подснежные воды протекала молодостьелены в бездействии внешнем во внутренней борьбе и тревоге подруг у нее небыло изо всех девиц посещавших
 дом стаховых она не сошлась ни с однойродительская власть никогда не тяготела над еленой а с шестнадцатилетнеговозраста она стала почти совсем независима она зажила собственной с
воеюжизнью но жизнью одинокой ее душа и разгоралась и погасала одиноко она биласькак птица в клетке а клетки не было никто не стеснял ее и никто не удерживала она рвалась и томила
сь она иногда сама себя не понимала даже боялась самойсебя все что окружало ее казалось ей не то бессмысленным не то непонятнымкак жить без любви а любить некого думала она и стра
шно становилось ей от этихдум от этих ощущений восемнадцати лет она чуть не умерла от злокачественнойлихорадки потрясенный до основания весь ее организм от природы здоровый икрепк
ий долго не мог справиться последние следы болезни исчезли наконец но отецелены николаевны все еще не без озлобления толковал об ее нервах иногда ейприходило в голову что она жела
ет чего то чего никто не желает о чем никто немыслит в целой россии потом она утихала даже смеялась над собой беспечнопроводила день за днем но внезапно что то сильное безымянное
 с чем онасовладеть не умела так и закипало в ней так и просилось вырваться наружу грозапроходила опускались усталые не взлетевшие крылья но порывы эти не обходилисьей даром как он
а ни старалась не выдать того что в ней происходило тоскавзволнованной души сказывалась в самом ее наружном спокойствии и родные еечасто были вправе пожимать плечами удивляться и
 не понимать ее странностейв день с которого начался наш рассказ елена дольше обыкновенного не отходилаот окна она много думала о берсеньеве о своем разговоре с нимпотребность в за
щите информации возникает в связи с необходимостью обеспечитьсекретность исследований в стратегических областях правильно распределятьинформацию о промышленных разработках и регул
ировать информацию о личности всовременном обществе начало восьмидесятых годов рассматривается как начальныйпункт когда социальные протесты в демократических странах помогли сплес
тисьглобальной сети хакеров политический флирт на почве нарушения прав человекапородил тьму организаций хакеров в массе стран мира почти одновременно менеечем за год эти группы уз
нали прелесть сотрудничества их члены свободнообменивались идеями через национальные границы часто по украденным паролямдающим бесплатный доступ к телефонной сети несколько причин
 обьединившисьвместе сделали международный компьютерный разбой лекгим и действенным новыетехнологии создавшие более мощные и дешевые компьютеры развитие коммуникацийдля связи и ме
ждународный характер стандартов установленных транснациональнымикорпорациями в принципе есть лишь два вида угрозы раскрытие и видоизменениеданных раскрытие данных предполагает что
 кому то случайно или послецеленаправленных действий стал известен смысл информации этот вид нарушениявстречается наиболее часто последствия могут быть самые разные если похищенте
кст книги справочника на которую потрачены месяцы работы десятков людейто для коллектива авторов это катастрофа и потери могут выражаться в тысячахдолларов однако если книга уже и
здана то достаточно лишь слегка пожуритьпохитителя и рассказать о случившемся в отделе новостей газеты или потелевидению похититель может сделать книге великолепную рекламу очень
важнуюинформацию оберегаемую от раскрытия представляют сведения о людях историиболезни письма состояния счетов в банках однако по мнению большого числаспециалистов угрозы личности
 с введением компьютеров остались на том же уровнеи в том же состоянии что и до обширного использования эвма современном мире туризм становится все более важнойбыстроразвивающейся
 отраслью хозяйства доходы от туризма становятсяважной частью валютных поступлений во многих странах развитиетуризма способствует росту общественного производства улучшениюего стр
уктуры росту производительности труда во многих отрасляхэкономики даже не имеющих к туризму прямого отношениямеждународное туристское потребление стимулирует многочисленныеэкономи
ческие процессы открывающие дополнительные рынки дляпродукции нетуристских отраслей создавая тем самым условия дляроста производства все эти факторы делают развитие индустрии тури
зма оченьважным для стран с переходным типом экономики экономические трудности которыепереживают эти государства не могут не сказаться на уровне развития туризмано при этом каждая
 страна имеет в этом отношении свою специфику цель данной работы рассмотреть и проанализировать организацию туристской деятельности встране с переходным типом экономики на примере
 венгрии в началерассматриваются теоретико методические положения исследования затем даетсяоценка различных факторов развития индустрии туризма венгрииприродноресурсный культурнои
сторический и инфраструктурный потенциалкомплексное туристское районирование далее проводится анализ современногосостояния индустрии туризма венгрии ее отдельных компонентов на фо
не общегоуровня экономического развития страны дается оценка социально экономической роли индустрии туризма в экономике венгрии и в заключение проводится общийанализ организации т
уристской деятельности в странах с переходным типомэкономики в общем и венгрии в частности венгрия принадлежа к странам спереходным типом экономики имеет тем не менее специфически
е черты которыеотличают ее от других стран этого типа в отношении развития индустрии туризмаосновной такой чертой является то что туризм венгрии развивается уже давноеще в начале
двадцатого века в этой стране сложились традиционные туристскиесвязи туризм является важной отраслью народного хозяйства современной
```

Кількість літер у тексті по алфавіту

о	526	**ь**	90
е	444	**я**	88
и	409	**з**	85
а	376	**г**	72
н	367	**ч**	66
т	341	**й**	63
с	306	**б**	56
р	250	**х**	55
в	207	**ю**	37
л	200	**ц**	36
м	153	**ж**	35
д	149	**ш**	28
к	143	**э**	23
п	111	**щ**	17
у	102	**ф**	14
ы	96		

Amount_without_space.xlsx

Кількість літер у тексті по алфавіту (з пробілом)

пробіл	736	**ы**	96
о	526	**ь**	90
е	444	**я**	88
и	409	**з**	85
а	376	**г**	72
н	367	**ч**	66
т	341	**й**	63
с	306	**б**	56
р	250	**х**	55
в	207	**ю**	37
л	200	**ц**	36
м	153	**ж**	35
д	149	**ш**	28
к	143	**э**	23
п	111	**щ**	17
у	102	**ф**	14

Amount_with_space.xlsx

Частота пересічних біграм тексту

	а	б	в	г	д	е	ж	з	и	й
а	0,000202	0,000809	0,004652	0,000607	0,003236	0,002427	0,002427	0,00445	0,001416	0,000405
б	0,000405	0	0,000202	0	0	0,002427	0	0	0,000405	0
в	0,007484	0,000405	0,000607	0,000202	0,000405	0,004652	0	0,000809	0,005057	0
г	0,00182	0	0	0	0,000809	0,000405	0	0	0,001416	0
ё	0	0	0	0	0	0	0	0	0	0
д	0,005259	0	0,000405	0	0	0,005461	0	0	0,003641	0
е	0,000202	0,002225	0,003034	0,002023	0,003641	0,004652	0,001416	0,002023	0,00182	0,003034
ж	0,000607	0	0	0	0,000809	0,002225	0	0	0,001618	0
з	0,003236	0,000202	0,002832	0,000405	0,001011	0,000405	0	0	0,000809	0
и	0,001011	0,000607	0,005259	0,001011	0,002427	0,004045	0	0,005866	0,006068	0,002023
й	0,000405	0,000607	0,000809	0,000202	0,001618	0,000202	0	0	0,000405	0
к	0,005461	0	0	0,000202	0	0,000809	0,000202	0	0,00445	0
л	0,007686	0	0	0,000202	0	0,009506	0	0	0,007282	0
м	0,004854	0,000202	0,000607	0	0,000607	0,004652	0,000405	0,000202	0,004045	0
н	0,011934	0	0	0,001416	0,001214	0,01072	0	0	0,011125	0
о	0,000202	0,004854	0,011529	0,005866	0,009304	0,002225	0,001214	0,002023	0,002225	0,005057
п	0,000607	0	0	0	0	0,002225	0	0	0,000202	0
р	0,013552	0	0,000607	0,000809	0	0,00627	0,000202	0	0,0089	0
с	0,001214	0	0,00182	0	0,000405	0,003236	0	0	0,000809	0
т	0,003843	0,000202	0,004248	0	0,000607	0,006877	0	0,000202	0,0089	0
у	0	0	0	0,001011	0,001214	0,000405	0,001011	0,000202	0	0
ф	0,000607	0	0	0	0	0	0	0	0,000405	0
х	0,000809	0	0	0,000202	0,001011	0	0	0	0,001214	0
ц	0,000809	0	0	0	0	0,001416	0	0	0,004248	0
ч	0,003034	0	0,000202	0	0	0,005461	0	0	0,001011	0
ш	0,000202	0	0	0	0	0,002023	0	0	0,001214	0
щ	0,000202	0	0	0	0	0,00182	0	0	0,001011	0
ы	0	0,000202	0,001214	0	0,000809	0,003034	0	0,000202	0,000202	0,00182
ь	0	0,000405	0,001214	0,000202	0,000405	0,001618	0	0,000405	0,001416	0
э	0	0	0,000202	0	0	0	0	0	0	0
ю	0	0,000607	0,000202	0	0,000405	0,000202	0,000202	0	0,000202	0

Bigram_Crossed_Frequency_without_space.xlsx

Частота пересічних біграм тексту (з пробілом)

	пробіл	**а**	**б**	**в**	**г**	**д**	**е**	**ж**	**з**	**и**
пробіл	0	0,00088	0,003521	0,011972	0,001585	0,00757	0,003345	0,001056	0,001585	0,008803
а	0,014085	0	0,000528	0,002465	0,000176	0,001761	0,001937	0,001937	0,003697	0,000176
б	0,000176	0,000352	0	0,000176	0	0	0,001937	0	0	0,000352
в	0,007394	0,006514	0	0,000176	0	0	0,004049	0	0,000352	0,004401
г	0,000352	0,001585	0	0	0	0,000704	0,000352	0	0	0,001232
ё	0	0	0	0	0	0	0	0	0	0
д	0,000704	0,004577	0	0,000352	0	0	0,004577	0	0	0,003169
е	0,018486	0	0,00088	0,001232	0,001585	0,002113	0,003345	0,001056	0,001761	0,000176
ж	0	0,000528	0	0	0	0,000704	0,001937	0	0	0,001408
з	0,001056	0,002817	0,000176	0,002465	0,000352	0,00088	0,000176	0	0	0,000704
и	0,016901	0,000704	0,000352	0,001937	0,000528	0,001056	0,003345	0	0,00493	0,004049
й	0,00757	0,000176	0,000176	0,000176	0	0,000176	0	0	0	0
к	0,001761	0,004754	0	0	0,000176	0	0,000704	0	0	0,003521
л	0,00088	0,00669	0	0	0,000176	0	0,008099	0	0	0,005986
м	0,00757	0,004225	0	0,000176	0	0,000176	0,003873	0	0	0,003169
н	0,00088	0,010387	0	0	0,001232	0,001056	0,009331	0	0	0,009683
о	0,014261	0,000176	0,003697	0,008627	0,005106	0,007394	0,001585	0,00088	0,001761	0,001056
п	0,000176	0,000528	0	0	0	0	0,001937	0	0	0,000176
р	0,000176	0,011796	0	0,000528	0,000704	0	0,005458	0,000176	0	0,007746
с	0,001761	0,001056	0	0,001408	0	0,000352	0,002817	0	0	0,000704
т	0,005282	0,003345	0	0,002817	0	0,000352	0,00581	0	0	0,007746
у	0,001761	0	0	0	0,000704	0,001056	0,000352	0,00088	0,000176	0
ф	0	0,000528	0	0	0	0	0	0	0	0,000352
х	0,005106	0,000704	0	0	0	0,000352	0	0	0	0,000528
ц	0,000352	0,000704	0	0	0	0	0,001232	0	0	0,003697
ч	0	0,002641	0	0,000176	0	0	0,004754	0	0	0,00088
ш	0,000176	0,000176	0	0	0	0	0,001761	0	0	0,001056
щ	0	0,000176	0	0	0	0	0,001585	0	0	0,00088
ы	0,004577	0	0	0,000528	0	0,000176	0,002641	0	0	0
ь	0,007218	0	0,000176	0,000352	0,000176	0	0,000704	0	0,000176	0,000176
э	0	0	0	0,000176	0	0	0	0	0	0
ю	0,002817	0	0,000352	0	0	0,000352	0,000176	0,000176	0	0,000176

Bigram_Crossed_Frequency_with_space.xlsx

Частота непересічних біграм тексту

	а	б	в	г	д	е	ж	з	и	й
а	0,000405	0,001214	0,00445	0,000405	0,002832	0,002832	0,001618	0,005259	0,001618	0,000405
б	0,000809	0	0,000405	0	0	0,001214	0	0	0,000405	0
в	0,010518	0,000405	0,000809	0,000405	0,000405	0,003236	0	0	0,006068	0
г	0,001618	0	0	0	0,001214	0,000405	0	0	0,002023	0
ё	0	0	0	0	0	0	0	0	0	0
д	0,005663	0	0	0	0	0,003641	0	0	0,004854	0
е	0	0,003641	0,002832	0,001618	0,004045	0,003641	0,001214	0,002832	0,002832	0,002427
ж	0,000405	0	0	0	0,000809	0,002427	0	0	0,002023	0
з	0,004045	0,000405	0,003236	0	0,000809	0	0	0	0,000809	0
и	0,000809	0,000809	0,005663	0,000405	0,001618	0,003641	0	0,002427	0,006877	0,001618
й	0,000405	0,000405	0,001214	0,000405	0,001214	0	0	0	0	0
к	0,004854	0	0	0,000405	0	0,000405	0,000405	0	0,003236	0
л	0,008091	0	0	0,000405	0	0,011327	0	0	0,008091	0
м	0,003641	0,000405	0,000809	0	0,000405	0,00445	0,000405	0	0,003236	0
н	0,010922	0	0	0,001214	0,002023	0,011327	0	0	0,012945	0
о	0,000405	0,004854	0,010518	0,005663	0,0089	0,002427	0,001214	0,002427	0,002427	0,006068
п	0,000809	0	0	0	0	0,001618	0	0	0,000405	0
р	0,01335	0	0,000405	0,000809	0	0,00445	0,000405	0	0,010113	0
с	0,001618	0	0,001618	0	0,000809	0,002832	0	0	0,000809	0
т	0,005259	0	0,003641	0	0,001214	0,007282	0	0,000405	0,008091	0
у	0	0	0	0,000405	0,001618	0	0,001214	0,000405	0	0
ф	0,000405	0	0	0	0	0	0	0	0,000405	0
х	0,000405	0	0	0	0,001214	0	0	0	0,000809	0
ц	0	0	0	0	0	0,002023	0	0	0,00445	0
ч	0,001618	0	0	0	0	0,005663	0	0	0,001618	0
ш	0,000405	0	0	0	0	0,002832	0	0	0,001618	0
щ	0	0	0	0	0	0,002023	0	0	0,001618	0
ы	0	0	0,000809	0	0,001214	0,002023	0	0	0,000405	0,002023
ь	0	0,000809	0,001214	0,000405	0,000405	0,001214	0	0,000405	0,002023	0
э	0	0	0	0	0	0	0	0	0	0
ю	0	0	0,000405	0	0,000809	0,000405	0	0	0	0

Bigram_Uncrossed_Frequency_without_space.xlsx

Частота непересічних біграм тексту (з пробілом)

	пробіл	**а**	**б**	**в**	**г**	**д**	**е**	**ж**	**з**	**и**
пробіл	0	0,001056	0,002465	0,008451	0,001056	0,008803	0,003169	0,000704	0,001408	0,008451
а	0,016197	0	0,001056	0,003521	0	0,001761	0,002817	0,001761	0,003169	0
б	0,000352	0,000352	0	0	0	0	0,002113	0	0	0
в	0,008099	0,007394	0	0	0	0	0,005282	0	0,000352	0,003169
г	0	0,002465	0	0	0	0,000352	0	0	0	0,002113
ё	0	0	0	0	0	0	0	0	0	0
д	0,001056	0,004577	0	0	0	0	0,004225	0	0	0,002817
е	0,01831	0	0,001408	0,001408	0,001761	0,002113	0,002817	0,001056	0,002113	0
ж	0	0,001056	0	0	0	0	0,002817	0	0	0,002465
з	0,000704	0,002817	0,000352	0,001761	0,000352	0,001056	0,000352	0	0	0,000704
и	0,02007	0,001408	0,000352	0,002465	0,000352	0,001761	0,003521	0	0,002817	0,002465
й	0,005986	0,000352	0	0,000352	0	0,000352	0	0	0	0
к	0,002465	0,004577	0	0	0	0	0,000704	0	0	0,003169
л	0,000704	0,005634	0	0	0,000352	0	0,009507	0	0	0,00669
м	0,007394	0,003873	0	0,000352	0	0,000352	0,003521	0	0	0,002465
н	0,000704	0,010563	0	0	0,001761	0,000352	0,010211	0	0	0,01162
о	0,015845	0,000352	0,003873	0,008099	0,00493	0,004225	0,001056	0,000352	0,001408	0,000704
п	0,000352	0,000352	0	0	0	0	0,002113	0	0	0,000352
р	0,000352	0,01338	0	0,000352	0,000352	0	0,00493	0	0	0,009507
с	0,001408	0,001761	0	0,001408	0	0,000352	0,000704	0	0	0,000704
т	0,00493	0,003169	0	0,003521	0	0,000352	0,005634	0	0	0,00669
у	0,001408	0	0	0	0,001056	0,001408	0,000352	0,000352	0	0
ф	0	0,001056	0	0	0	0	0	0	0	0,000352
х	0,005986	0,000704	0	0	0	0	0	0	0	0,000704
ц	0,000352	0,000704	0	0	0	0	0,000704	0	0	0,004225
ч	0	0,002113	0	0,000352	0	0	0,003873	0	0	0
ш	0	0	0	0	0	0	0,002113	0	0	0,001408
щ	0	0	0	0	0	0	0,001408	0	0	0,001408
ы	0,00493	0	0	0,000704	0	0	0,002817	0	0	0
ь	0,008099	0	0,000352	0,000352	0	0	0,001056	0	0	0
э	0	0	0	0,000352	0	0	0	0	0	0
ю	0,002113	0	0,000704	0	0	0,000704	0,000352	0	0	0,000352

Bigram_Uncrossed_Frequency_with_space.xlsx

Значення ентропії

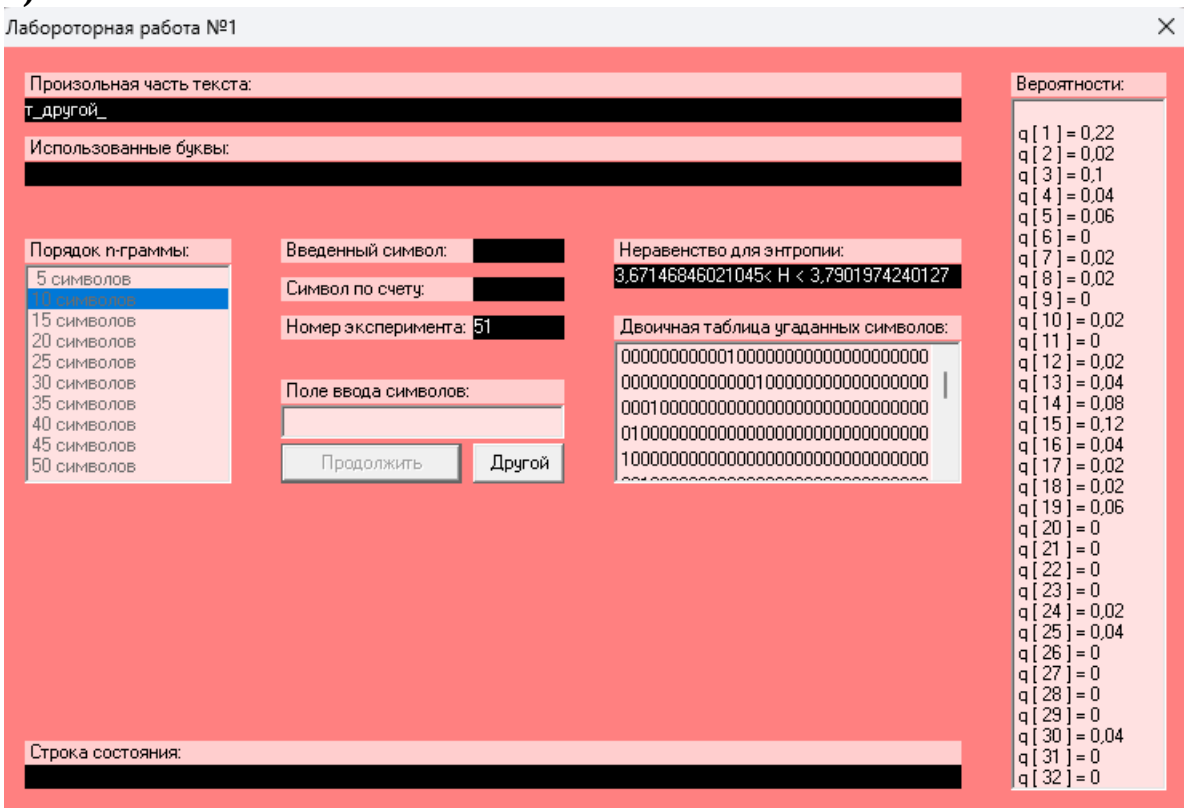
	Текст без пробілів	Текст з пробілами
$H1$	4.42210185111123	4.405413528455854
$H2$ з перетином	4.02077298377599	3.949285426337858
$H2$ без перетину	3.974883235762963	3.911711084412846

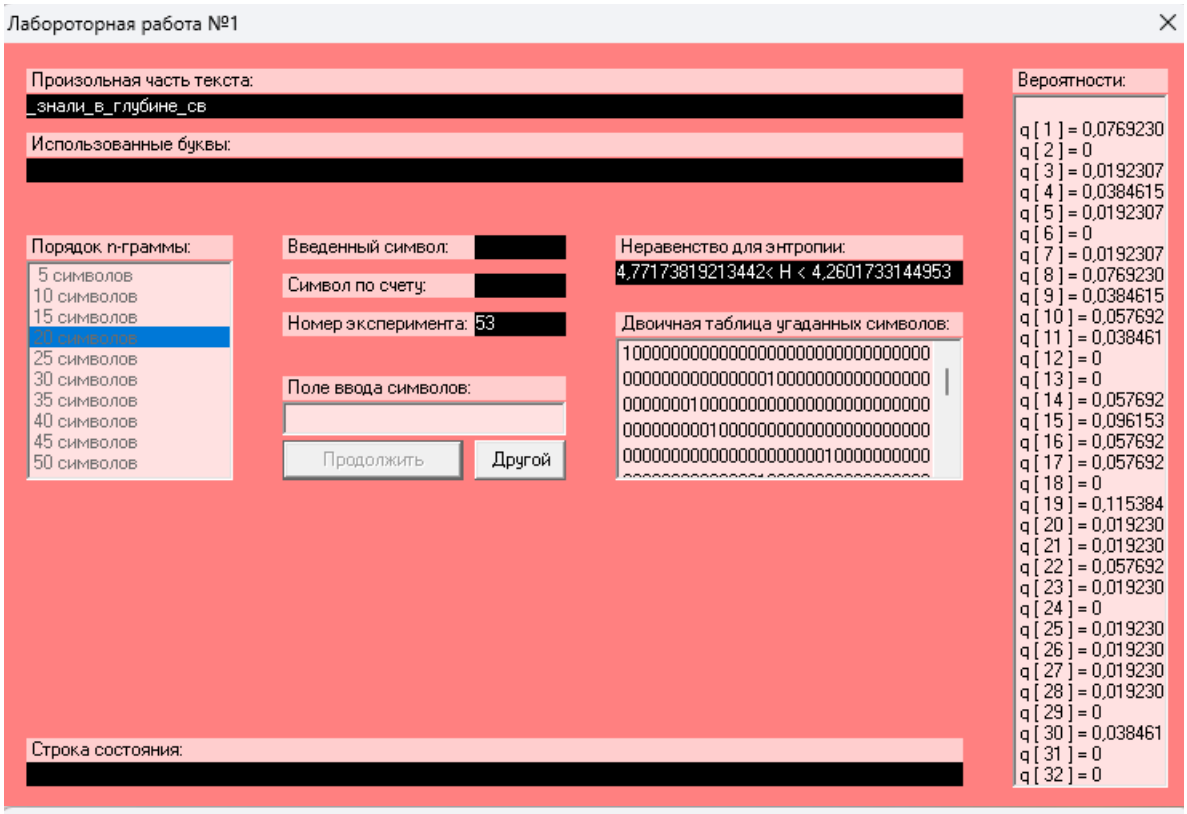
Оцінка надлишковості мови

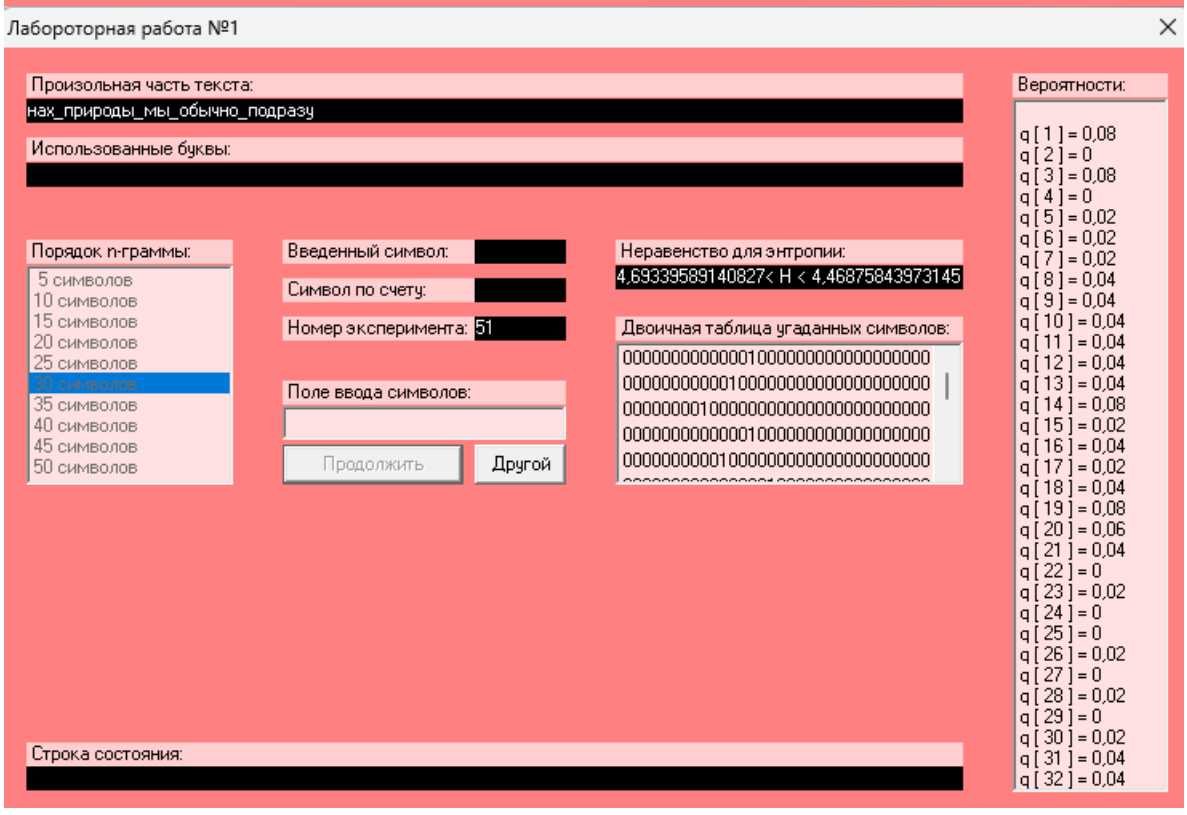
$$R = 1 - \frac{\mathrm{Hn}}{\mathrm{H0}}$$

	Текст без пробілів	Текст з пробілами
$H1$	0.1074	0.1189
$H2$ з перетином	0.1884	0.2101
$H2$ без перетину	0.1977	0.2177

2)




Лабороторная работа №1
Произольная часть текста:
_знали_в_глубине_св
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: 53
Поле ввода символов:
Продолжить
Другой
Неравенство для энтропии:
4,77173819213442< H < 4,2601733144953
Двоичная таблица угаданных символов:
Вероятности:
q [1] = 0,0769230
q [2] = 0
q [3] = 0,0192307
q [4] = 0,0384615
q [5] = 0,0192307
q [6] = 0
q [7] = 0,0192307
q [8] = 0,0769230
q [9] = 0,0384615
q [10] = 0,057692
q [11] = 0,038461
q [12] = 0
q [13] = 0
q [14] = 0,057692
q [15] = 0,096153
q [16] = 0,057692
q [17] = 0,057692
q [18] = 0
q [19] = 0,115384
q [20] = 0,019230
q [21] = 0,019230
q [22] = 0,057692
q [23] = 0,019230
q [24] = 0
q [25] = 0,019230
q [26] = 0,019230
q [27] = 0,019230
q [28] = 0,019230
q [29] = 0
q [30] = 0,038461
q [31] = 0
q [32] = 0
Строка состояния:


Лабороторная работа №1
Произольная часть текста:
нах_природы_мы_обычно_подразу
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: 51
Поле ввода символов:
Продолжить
Другой
Неравенство для энтропии:
4,69339589140827< H < 4,46875843973145
Двоичная таблица угаданных символов:
Вероятности:
q [1] = 0,08
q [2] = 0
q [3] = 0,08
q [4] = 0
q [5] = 0,02
q [6] = 0,02
q [7] = 0,02
q [8] = 0,04
q [9] = 0,04
q [10] = 0,04
q [11] = 0,04
q [12] = 0,04
q [13] = 0,04
q [14] = 0,08
q [15] = 0,02
q [16] = 0,04
q [17] = 0,02
q [18] = 0,04
q [19] = 0,08
q [20] = 0,06
q [21] = 0,04
q [22] = 0
q [23] = 0,02
q [24] = 0
q [25] = 0
q [26] = 0,02
q [27] = 0
q [28] = 0,02
q [29] = 0
q [30] = 0,02
q [31] = 0,04
q [32] = 0,04
Строка состояния:

$3.67146846021045 < H^{(10)} < 3.7901974240127$
$0.26570630795791 > R^{(10)} > 0.24196051519746$

$4.77173819213442 < H^{(20)} < 4.2601733144953$
$0.16565236157312 > R^{(20)} > 0.14796533710094$

$4.69339589140827 < H^{(30)} < 4.46875843973145$
$0.16132082171835 > R^{(30)} > 0.10624831205371$

Висновки:

У процесі виконання комп'ютерного практикуму ми навчилися аналізувати частоти символів і біграм у тексті, а також обчислювати ентропію та надлишковість мови. За результатами дослідження встановлено, що текст без пробілів має вищу ентропію порівняно з текстом, де збережено пробіли, що вказує на меншу передбачуваність символів у ньому.