

Київський національний університет імені Тараса Шевченка
Факультет радіофізики, електроніки та комп'ютерних систем

Лабораторна робота № 4

на тему «Дослідження частотних характеристик української мови»
з предмету “Захист інформації в комп'ютерних системах”

Роботу виконав
студент 4 курсу
напряму підготовки «Комп'ютерна інженерія»
Ліщенко Юрій Вікторович

Київ 2024

Мета: Дослідити вірогіднісні параметри появи літер української мови для різних типів текстів. Аналіз найбільш імовірних літер, біграм та триграм для використання в частотному криптоаналізі.

Хід роботи

Посилання на репозиторій <https://github.com/YuriiLishchenko1/ZICS>

1. Знаходження відносної частоти появи літер українського алфавіту

1. Створіть програму (будь-якою зручною для вас мовою), яка в якості аргументів приймає перелік текстових файлів, та аналізуючи їх вміст обраховує частоти появи літер українського алфавіту. Для спрощення аналізу з тексту виключаються всі знаки пунктуації окрім “пробілу”.
2. Для різних достатньо довгих текстів (оберіть принаймні 2, причому тексти не повинні повторюватись в групі), наприклад, художній твір, вірш, технічна документація, блог, тощо, проаналізуйте отримані результати та в звіті наведіть:
 - а. діаграму відсортовану в алфавітному порядку
 - б. діаграму відсортовану по частотам появи літер
 - с. послідовність літер по мірі спадання частоти появи

Для файлу Dvigun.txt:

```
Аналіз для файлу: Dvigun.txt
Частота літер (алфавітний порядок):
: 11,58%
а: 7,42%
б: 1,26%
в: 5,40%
г: 1,37%
д: 2,90%
е: 4,16%
ж: 0,58%
з: 2,21%
и: 5,61%
й: 0,89%
к: 2,79%
л: 3,45%
м: 2,84%
н: 8,45%
о: 7,19%
п: 3,19%
р: 3,69%
с: 4,40%
т: 3,92%
у: 3,53%
ф: 0,16%
х: 1,21%
ц: 0,82%
ч: 0,87%
ш: 0,53%
щ: 0,18%
ь: 0,84%
ю: 0,53%
я: 2,84%
є: 0,53%
і: 4,16%
```

```
Частота літер (відсортовано по частотам):
: 11,58%
н: 8,45%
а: 7,42%
о: 7,19%
и: 5,61%
в: 5,40%
с: 4,40%
е: 4,16%
і: 4,16%
т: 3,92%
р: 3,69%
у: 3,53%
л: 3,45%
п: 3,19%
д: 2,90%
я: 2,84%
м: 2,84%
к: 2,79%
з: 2,21%
г: 1,37%
б: 1,26%
х: 1,21%
й: 0,89%
ч: 0,87%
ь: 0,84%
ц: 0,82%
ж: 0,58%
ш: 0,53%
ю: 0,53%
є: 0,53%
ї: 0,53%
щ: 0,18%
ф: 0,16%
```



Послідовність літер по мірі спадання частоти:

наоивсеітрулпдямкзгбхйчъцжшюєїщф

Відповідно до даних бачимо що літера «Н» найчастіше з'являється у тексті, а літера «ф» найрідше (9,56% та 0,18% відповідно).

Для файлу Misto.txt

```
Аналіз для файлу: Misto.txt
Частота літер (алфавітний порядок):
: 15,76%
а: 7,16%
б: 1,20%
в: 5,10%
г: 1,29%
д: 2,76%
е: 3,56%
ж: 0,74%
з: 2,19%
и: 5,32%
й: 1,54%
к: 3,03%
л: 2,57%
м: 2,93%
н: 5,39%
о: 8,17%
п: 2,77%
р: 3,70%
с: 4,21%
т: 4,19%
у: 3,24%
ф: 0,09%
х: 0,70%
ц: 0,63%
ч: 1,16%
ш: 0,65%
щ: 0,43%
ь: 1,34%
ю: 0,83%
я: 1,92%
є: 0,39%
і: 4,48%
```

```
Частота літер (відсортовано по частотам):
: 15,76%
о: 8,17%
а: 7,16%
н: 5,39%
и: 5,32%
в: 5,10%
і: 4,48%
с: 4,21%
т: 4,19%
р: 3,70%
е: 3,56%
у: 3,24%
к: 3,03%
м: 2,93%
п: 2,77%
д: 2,76%
л: 2,57%
з: 2,19%
я: 1,92%
й: 1,54%
ь: 1,34%
г: 1,29%
б: 1,20%
ч: 1,16%
ю: 0,83%
ж: 0,74%
х: 0,70%
ш: 0,65%
ц: 0,63%
ї: 0,57%
щ: 0,43%
є: 0,39%
ф: 0,09%
```



Послідовність літер по мірі спадання частоти:

оанивістрекмплзйяьгбчюжхшщїщєф

2. Знаходження відносної частоти появи біграм українського алфавіту

1. Аналогічно до першого завдання, створіть програму яка підрахує відносні частоти появи біграм української мови аналізуючи вміст текстів, що задаються.
2. Для текстів, що були використані в першому завданні проаналізуйте отримані результати та в звіті наведіть:
 - a.
 - b. таблицю з відносними частотами біграм, відсортовану за спаданням частоти (використовуйте декілька колонок, щоб результат був максимально компактним)
 - c. послідовність з 30-ти найбільш імовірних біграм
 - d. діаграму відсортовану по частотам появи 30-ти найбільш імовірних біграм
 - e. матрицю частот появи біграм (імовірність позначена кольором)

Dvigun.txt:

Аналіз біграм для файлу: Dvigun.txt

Таблиця з біграмами, відсортована за спаданням частоти:

нн: 2,295%	ня: 2,089%	на: 1,952%	ст: 1,849%	ан: 1,610%
ни: 1,507%	ва: 1,507%	ис: 1,507%	те: 1,438%	ем: 1,438%
ал: 1,438%	си: 1,404%	ви: 1,370%	ма: 1,267%	ен: 1,267%
ов: 1,233%	по: 1,199%	тр: 1,027%	но: 1,027%	ро: 0,959%
па: 0,959%	ун: 0,925%	ло: 0,925%	ор: 0,890%	ві: 0,890%
дв: 0,856%	ив: 0,822%	ск: 0,822%	од: 0,788%	ів: 0,788%
иг: 0,753%	гу: 0,753%	во: 0,753%	пр: 0,753%	ід: 0,719%
ий: 0,685%	та: 0,685%	ко: 0,685%	их: 0,685%	за: 0,651%
ол: 0,651%	ув: 0,651%	ус: 0,616%	ли: 0,616%	ер: 0,616%
ля: 0,616%	пу: 0,582%	ра: 0,582%	ле: 0,582%	вн: 0,548%
рі: 0,548%	мі: 0,548%	об: 0,548%	ос: 0,548%	он: 0,548%
ом: 0,514%	ре: 0,514%	то: 0,514%	ри: 0,479%	ль: 0,479%
же: 0,479%	лі: 0,479%	ін: 0,479%	го: 0,445%	ря: 0,445%
га: 0,445%	ку: 0,445%	лю: 0,445%	ох: 0,445%	іт: 0,445%
кт: 0,411%	хо: 0,411%	дж: 0,411%	ил: 0,411%	бе: 0,411%
ас: 0,411%	ої: 0,411%	ти: 0,377%	ну: 0,377%	ні: 0,377%
кр: 0,377%	ип: 0,377%	іл: 0,377%	ру: 0,377%	бл: 0,377%
ок: 0,377%	оп: 0,377%	ці: 0,377%	ьо: 0,342%	ян: 0,342%
ми: 0,342%	аз: 0,342%	кн: 0,342%	ци: 0,342%	пе: 0,342%
ач: 0,342%	че: 0,342%	ат: 0,308%	оз: 0,308%	из: 0,308%

30 найбільш імовірних біграм:

нн: 2,295%
ня: 2,089%
на: 1,952%
ст: 1,849%
ан: 1,610%
ни: 1,507%
ва: 1,507%
ис: 1,507%
те: 1,438%
ем: 1,438%
ал: 1,438%
си: 1,404%
ви: 1,370%
ма: 1,267%
ен: 1,267%
ов: 1,233%
по: 1,199%
тр: 1,027%
но: 1,027%
ро: 0,959%
па: 0,959%
ун: 0,925%
ло: 0,925%
ор: 0,890%
ві: 0,890%
дв: 0,856%
ив: 0,822%
ск: 0,822%
од: 0,788%
ів: 0,788%

Таблиця

	Біграма	Відносна частота (%)																		
1	нн	2,295	36	ий	0,685	71	кт	0,411	106	ш	0,274	141	кі	0,205	176	ит	0,137	211	дд	0,103
2	ня	2,089	37	та	0,685	72	хо	0,411	107	ог	0,274	142	ет	0,205	177	ош	0,137	212	фі	0,103
3	на	1,952	38	ко	0,685	73	дж	0,411	108	ть	0,274	143	уп	0,205	178	ащ	0,137	213	ам	0,103
4	ст	1,849	39	их	0,685	74	ил	0,411	109	ап	0,274	144	ча	0,205	179	ке	0,137	214	ій	0,103
5	ан	1,610	40	за	0,651	75	бе	0,411	110	вн	0,274	145	ші	0,205	180	ве	0,137	215	ім	0,103
6	ни	1,507	41	ол	0,651	76	ас	0,411	111	вп	0,274	146	ті	0,205	181	тв	0,137	216	ет	0,103
7	ва	1,507	42	ув	0,651	77	ої	0,411	112	ап	0,274	147	ся	0,205	182	юч	0,137	217	че	0,103
8	ис	1,507	43	ус	0,616	78	ти	0,377	113	дн	0,274	148	ав	0,205	183	ук	0,137	218	ют	0,103
9	те	1,438	44	ли	0,616	79	ну	0,377	114	др	0,274	149	сл	0,205	184	рс	0,137	219	рн	0,103
10	ем	1,438	45	ер	0,616	80	ні	0,377	115	ав	0,274	150	шн	0,171	185	ду	0,137	220	щу	0,103
11	ал	1,438	46	ля	0,616	81	кр	0,377	116	бо	0,274	151	от	0,171	186	вл	0,137	221	лу	0,103
12	си	1,404	47	пу	0,582	82	ип	0,377	117	су	0,274	152	тн	0,171	187	вт	0,137	222	уч	0,103
13	ви	1,370	48	ра	0,582	83	іл	0,377	118	ум	0,274	153	ут	0,171	188	мо	0,137	223	ім	0,103
14	ма	1,267	49	ле	0,582	84	ру	0,377	119	зі	0,274	154	ме	0,171	189	ьт	0,137	224	фу	0,103
15	ен	1,267	50	вн	0,548	85	бл	0,377	120	со	0,274	155	із	0,171	190	сп	0,137	225	нк	0,103
16	ов	1,233	51	рі	0,548	86	ок	0,377	121	ел	0,274	156	ту	0,171	191	ій	0,137	226	ар	0,103
17	по	1,199	52	мі	0,548	87	оп	0,377	122	ак	0,240	157	ді	0,171	192	нз	0,137	227	вц	0,103
18	тр	1,027	53	об	0,548	88	ці	0,377	123	зм	0,240	158	ки	0,171	193	ьс	0,137	228	юо	0,103
19	но	1,027	54	ос	0,548	89	юо	0,342	124	дл	0,240	159	им	0,171	194	де	0,137	229	ич	0,103
20	ро	0,959	55	он	0,548	90	ян	0,342	125	на	0,240	160	вк	0,171	195	ик	0,137	230	рш	0,068
21	па	0,959	56	ом	0,514	91	ми	0,342	126	ла	0,240	161	ак	0,171	196	ул	0,137	231	чо	0,068
22	ун	0,925	57	ре	0,514	92	аз	0,342	127	сн	0,240	162	еч	0,171	197	чн	0,137	232	зг	0,068
23	ло	0,925	58	то	0,514	93	кн	0,342	128	да	0,240	163	чу	0,171	198	ир	0,103	233	рп	0,068
24	ор	0,890	59	ри	0,479	94	ци	0,342	129	дп	0,240	164	пі	0,171	199	нь	0,103	234	ша	0,068
25	ві	0,890	60	ль	0,479	95	пе	0,342	130	ац	0,240	165	кл	0,171	200	ех	0,103	235	зо	0,068
26	дв	0,856	61	же	0,479	96	ач	0,342	131	ць	0,240	166	ин	0,171	201	ха	0,103	236	ще	0,068
27	ив	0,822	62	лі	0,479	97	че	0,342	132	ек	0,240	167	до	0,171	202	ши	0,103	237	чи	0,068
28	ск	0,822	63	ін	0,479	98	ат	0,308	133	ії	0,240	168	вц	0,171	203	пн	0,103	238	шу	0,068
29	од	0,788	64	го	0,445	99	оз	0,308	134	ад	0,240	169	бу	0,171	204	зк	0,103	239	ює	0,068
30	ів	0,788	65	ря	0,445	100	из	0,308	135	ає	0,240	170	бі	0,171	205	ед	0,103	240	зв	0,068
31	иг	0,753	66	га	0,445	101	юв	0,308	136	як	0,240	171	ди	0,171	206	рт	0,103	241	ух	0,068
32	гу	0,753	67	ку	0,445	102	му	0,308	137	іс	0,205	172	нт	0,171	207	оч	0,103	242	ож	0,068
33	во	0,753	68	лю	0,445	103	нд	0,308	138	ує	0,205	173	ей	0,171	208	нс	0,103	243	жн	0,068
34	пр	0,753	69	ох	0,445	104	ез	0,308	139	зи	0,205	174	аг	0,171	209	ед	0,103	244	чі	0,068
35	ід	0,719	70	іт	0,445	105	зн	0,308	140	ям	0,205	175	не	0,137	210	ес	0,103	245	яв	0,068

Загальна кількість біграм 343.

30 найбільш імовірних біграм:

нн ня на ст ан ни ва ис те ем ал си ви ма ен ов по тр но ро па ун ло ор ві дв ив ск од ів

Діаграма:



Матриця частот появи біграм: (якщо клітинка пуста – частота = 0)

Column1	а	б	в	г	ґ	д	е	є	ж	з	и	і	ї	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ь	ю	я
а		0,274	0,250	0,171		0,240		0,240	0,340	0,342				0,130	0,240	1,438	0,130	1,610		0,274	0,130	0,411	0,380			0,171	0,240	0,342	0,340	0,137		0,680	
б		0,680					0,411	0,340				0,340	0,171		0,377		0,340	0,274		0,377	0,340	0,274			0,171		0,340	0,340					
в		1,570	0,340				0,137					1,370	0,890		0,171	0,137	0,548	0,753	0,274		0,340	0,137	0,340										
г		0,445										0,340				0,340					0,445	0,680		0,753									
ґ																																	
д		0,240	0,680	0,856			0,130	0,137		0,411		0,171	0,171			0,340	0,240		0,274	0,171	0,240	0,274	0,340		0,137								
е			0,340	0,340			0,130			0,680	0,380			0,171	0,240	0,274	1,438	1,267	0,340		0,616	0,130	0,250			0,130	0,680	0,171					
є							0,130																0,130						0,340				
ж		0,340						0,479				0,340						0,680	0,340														
з		0,651	0,340	0,680	0,680		0,340	0,680	0,680		0,250	0,274			0,130		0,240	0,380	0,680	0,274	0,340		0,340	0,340								0,340	
и			0,340	0,822	0,753		0,340			0,680	0,380			0,685	0,137	0,411	0,171	0,171		0,377	0,130	1,570	0,137			0,685	0,680	0,130		0,340			
і		0,340	0,340	0,788	0,340		0,719			0,680	0,171			0,240	0,137	0,340	0,377	0,130	0,479	0,340		0,340	0,250	0,445					0,274			0,340	
ї																																	
й																	0,130	0,340	0,340					0,340		0,340							
к		0,240					0,137				0,171	0,250				0,171	0,342	0,685		0,377	0,340	0,411	0,445			0,171							
л		0,240					0,582				0,616	0,479					0,925						0,130							0,479	0,445	0,616	
м		1,267					0,171				0,342	0,548			0,340			0,340	0,137	0,340				0,380									
н		1,952					0,380	0,137			0,137	1,570	0,377		0,130			2,295	1,270			0,130	0,171	0,377				0,340	0,680	0,130	0,340	2,890	
о			0,548	1,233	0,274		0,788		0,340	0,680	0,380		0,411	0,377	0,651	0,514	0,548	0,340	0,377	0,890	0,548	0,171				0,445	0,680	0,130	0,137	0,340		0,680	0,340
п		0,959					0,342					0,340	0,171				0,130	1,199		0,753			0,582										
р		0,582	0,340		0,340		0,514				0,479	0,548			0,680		0,340	0,130	0,959	0,680		0,137	0,130	0,377				0,680		0,340	0,445		
с		0,680		0,340			0,680				1,440				0,822	0,250		0,240	0,274	0,137		1,849	0,274									0,250	
т		0,685		0,137			1,438				0,377	0,250			0,340			0,171	0,514		1,270		0,171						0,274			0,340	
у			0,340	0,651				0,250						0,137	0,137	0,274	0,925			0,250	0,340	0,616	0,171	0,340		0,680	0,130	0,340					
ф												0,130												0,130									
х		0,130										0,340	0,340			0,340			0,411														
ц							0,130					0,342	0,377																		0,240		
ч		0,250		0,340							0,342	0,680						0,137	0,680					0,171									
ш		0,680					0,680				0,130	0,250						0,171					0,680										
щ							0,680											0,680					0,130										
ь															0,340		0,340	0,274	0,342		0,137	0,137						0,340					
ю			0,380					0,680															0,130						0,137				
я			0,680							0,340				0,240		0,250	0,342	0,130	0,340		0,340					0,340	0,130					0,340	

Для файлу Misto.txt:

Аналіз біграм для файлу: Misto.txt

Таблиця з біграмами, відсортована за спаданням частоти:

на: 1,973%	ст: 1,860%	ти: 1,465%	по: 1,371%	ом: 1,315%
ав: 1,221%	ві: 1,146%	го: 1,146%	ро: 1,108%	ко: 1,052%
ог: 1,052%	ан: 1,033%	ва: 0,996%	ов: 0,977%	ся: 0,939%
до: 0,939%	пр: 0,921%	ра: 0,902%	ри: 0,864%	ит: 0,864%
ка: 0,845%	те: 0,808%	ла: 0,789%	ін: 0,770%	то: 0,770%
ер: 0,770%	ни: 0,751%	за: 0,751%	йо: 0,733%	ло: 0,714%
но: 0,714%	сь: 0,714%	ал: 0,714%	од: 0,695%	та: 0,695%
ат: 0,695%	ку: 0,695%	не: 0,676%	ос: 0,658%	ні: 0,658%
ви: 0,639%	ин: 0,639%	ли: 0,620%	ив: 0,601%	им: 0,601%
па: 0,582%	ад: 0,582%	ма: 0,582%	ід: 0,582%	ми: 0,564%
ор: 0,564%	ис: 0,564%	ол: 0,545%	іс: 0,545%	во: 0,545%
ув: 0,545%	ді: 0,526%	ть: 0,526%	му: 0,526%	чи: 0,526%
ен: 0,507%	ий: 0,507%	ій: 0,507%	ар: 0,507%	ки: 0,507%
бу: 0,507%	еп: 0,488%	ве: 0,488%	рі: 0,488%	як: 0,488%
що: 0,470%	вс: 0,470%	ів: 0,470%	оп: 0,451%	ас: 0,451%
нн: 0,451%	ою: 0,432%	мі: 0,432%	ру: 0,432%	ус: 0,432%
тр: 0,432%	ди: 0,413%	ам: 0,413%	ьк: 0,413%	мо: 0,395%
ле: 0,395%	ту: 0,395%	об: 0,395%	зн: 0,395%	ре: 0,395%
да: 0,395%	ці: 0,376%	ля: 0,376%	ну: 0,376%	де: 0,357%
ак: 0,357%	ня: 0,357%	пе: 0,338%	ел: 0,338%	ча: 0,338%
се: 0,319%	ай: 0,319%	он: 0,319%	ає: 0,301%	сп: 0,301%
рн: 0,301%	са: 0,301%	ик: 0,301%	ті: 0,301%	оз: 0,301%
че: 0,282%	аз: 0,282%	ль: 0,282%	уд: 0,282%	кі: 0,282%
лі: 0,282%	оч: 0,282%	пи: 0,282%	нк: 0,263%	зо: 0,263%
ьс: 0,263%	ир: 0,263%	си: 0,263%	кр: 0,263%	от: 0,263%
нь: 0,263%	ок: 0,263%	ск: 0,263%	пі: 0,263%	хо: 0,263%
ме: 0,244%	ут: 0,244%	ши: 0,244%	ьо: 0,244%	же: 0,244%
іт: 0,244%	чі: 0,244%	іл: 0,244%	зу: 0,244%	св: 0,225%
ил: 0,225%	бо: 0,225%	сі: 0,225%	бі: 0,207%	йк: 0,207%

30 найбільш імовірних біграм:

на: 1,973%
ст: 1,860%
ти: 1,465%
по: 1,371%
ом: 1,315%
ав: 1,221%
ві: 1,146%
го: 1,146%
ро: 1,108%
ко: 1,052%
ог: 1,052%
ан: 1,033%
ва: 0,996%
ов: 0,977%
ся: 0,939%
до: 0,939%
пр: 0,921%
ра: 0,902%
ри: 0,864%
ит: 0,864%
ка: 0,845%
те: 0,808%
ла: 0,789%
ін: 0,770%
то: 0,770%
ер: 0,770%
ни: 0,751%
за: 0,751%
йо: 0,733%
ло: 0,714%

Таблиця:

[illegible]

Загальна кількість біграм 453.

Діаграма:



Матриця частот появи біграм: (якщо клітинка пуста – частота = 0)

Column1	а	б	в	г	д	е	ж	з	и	і	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ь	ю	я		
а		0,190	1,221	0,113		0,582		0,310	0,132	0,282		0,750	0,319	0,357	0,714	0,413	1,330	0,190	0,188	0,570	0,451	0,695	0,190		0,270	0,380	0,188	0,190		0,940		
б	0,188					0,190	0,132			0,150	0,270				0,750	0,190	0,190	0,225		0,150		0,570										
в	0,996		0,190	0,190		0,940	0,488		0,150	0,380	0,639	1,146			0,113	0,132	0,190	0,113	0,545	0,560	0,380	0,470	0,190	0,132	0,190		0,380	0,270	0,380		0,560	
г	0,150								0,113	0,190					0,190	0,113		0,940	1,146		0,750		0,380	0,560								
д		0,395	0,190	0,750	0,190			0,357		0,560		0,413	0,526			0,940	0,750		0,132	0,939	0,190	0,270					0,150		0,560	0,113		
е		0,188	0,270	0,190		0,113			0,560	0,190			0,750	0,560	0,150	0,338	0,150	0,570		0,488	0,770	0,132	0,750		0,190	0,190	0,560	0,150	0,750		0,132	
є			0,190									0,560				0,380					0,190	0,113								0,190		
ж	0,132	0,190				0,380	0,244			0,270	0,750			0,560		0,380		0,380			0,190		0,113									
з	0,751	0,750	0,132			0,560	0,190			0,560	0,380			0,190	0,380	0,750	0,395	0,263	0,380	0,113		0,244				0,560	0,190				0,190	
и		0,750	0,610	0,940		0,150		0,380	0,190	0,940			0,380	0,570	0,310	0,225	0,610	0,639		0,560	0,263	0,564	0,864		0,190	0,188	0,940	0,380	0,150		0,380	
і	0,190	0,940	0,470	0,380		0,582		0,560	0,560	0,188		0,190	0,560	0,570	0,169	0,244	0,169	0,770		0,270	0,545	0,244		0,190	0,380	0,190	0,150	0,169	0,190		0,560	
ї			0,750						0,190					0,132	0,190		0,380	0,380	0,560					0,380						0,560	0,750	
й		0,190						0,190					0,270	0,190	0,380	0,750	0,733			0,750	0,380					0,940						
к	0,845	0,190	0,940		0,190	0,750				0,570	0,282			0,380		0,560	1,520		0,263	0,750	0,940	0,695			0,190							
л	0,789					0,395				0,620	0,282			0,190			0,714					0,132							0,282	0,113	0,376	
м	0,582	0,190	0,190			0,244				0,564	0,432		0,380	0,560		0,940	0,395	0,380	0,380			0,526									0,380	
н	1,973					0,676			0,751	0,658				0,263		0,451	0,714			0,169	0,940	0,376			0,560	0,380	0,380	0,190	0,263	0,750	0,357	
о		0,395	0,977	1,520		0,695	0,190	0,560	0,750	0,310		0,169	0,560	0,263	0,545	1,315	0,319	0,190	0,451	0,564	0,658	0,263		0,380	0,169	0,190	0,282	0,750		0,432		
п	0,582					0,338				0,282	0,263		0,190	0,132		0,190	0,371		0,921		0,380	0,190		0,380								
р	0,920	0,190	0,190	0,190		0,380	0,395		0,190	0,864	0,488		0,940	0,190	0,310	1,180			0,113	0,132	0,432		0,190	0,132		0,750			0,380	0,560		
с	0,310		0,225			0,319				0,263	0,225		0,263	0,113	0,190	0,940	0,169	0,310			1,860	0,190		0,750	0,190			0,714		0,939		
т	0,695		0,188			0,880			1,465	0,310		0,150		0,380	0,770			0,432	0,190	0,750	0,395		0,190		0,190			0,526	0,380	0,113		
у		0,190	0,545	0,560		0,282		0,750	0,169		0,380		0,150	0,270	0,169	0,113		0,560	0,270	0,432	0,244		0,940		0,560	0,560				0,380		
ф	0,560					0,380				0,190											0,190											
х	0,169	0,190				0,190				0,380	0,190			0,113		0,380	0,263		0,190		0,190											
ц						0,169				0,190	0,376																		0,113	0,940	0,150	
ч	0,338					0,282				0,526	0,244					0,560	0,560					0,150										
ш	0,560		0,380			0,940				0,244	0,560		0,940	0,190		0,560	0,150				0,380	0,750										
щ	0,560					0,750			0,190								0,470															
ь			0,190									0,413		0,750	0,750	0,244			0,263	0,190					0,560		0,380				0,190	
ю	0,190		0,560			0,560		0,190							0,750			0,190	0,560					0,190	0,750					0,190		
я			0,560	0,560	0,750	0,190		0,750				0,488	0,380	0,169	0,750			0,190	0,380	0,940			0,190		0,190					0,190		

3. Знаходження відносної частоти появи триграм українського алфавіту

- Аналогічно до першого завдання, створіть програму яка підрахує відносні частоти появи триграм української мови аналізуючи вміст текстів, що задаються.
- Для текстів, що були використані в першому завданні проаналізуйте отримані результати та в звіті наведіть:
 - таблицю з відносними частотами триграм, відсортовану за спаданням частоти
 - послідовність з 30-ти найбільш імовірних триграм
 - діаграму відсортовану по частотам появи 30-ти найбільш імовірних триграм

Для Dvигun.txt:

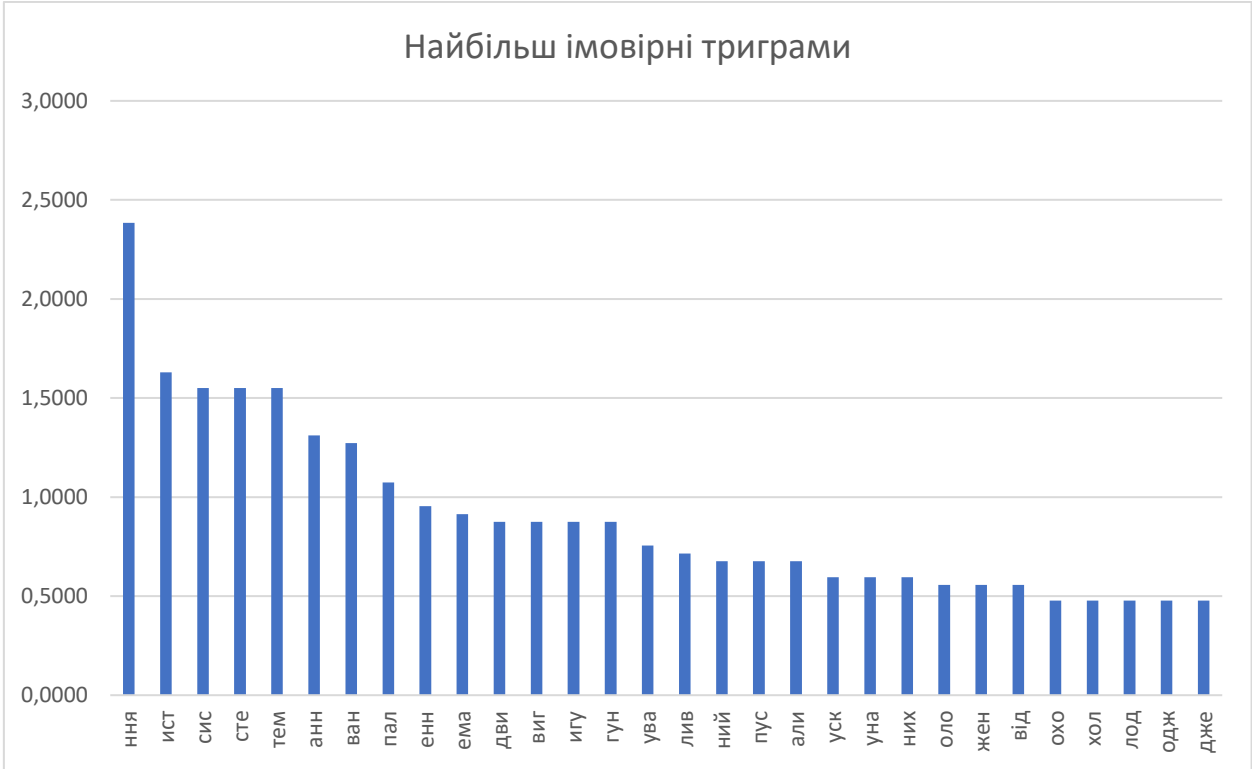
Аналіз триграм для файлу: Dvigun.txt				
Таблиця з триграмами, відсортована за спаданням частоти:				
ння: 2,3847%	ист: 1,6296%	сис: 1,5501%	сте: 1,5501%	тем: 1,5501%
анн: 1,3116%	ван: 1,2719%	пал: 1,0731%	енн: 0,9539%	ема: 0,9141%
дви: 0,8744%	виг: 0,8744%	игу: 0,8744%	гун: 0,8744%	ува: 0,7552%
лив: 0,7154%	ний: 0,6757%	пус: 0,6757%	али: 0,6757%	уск: 0,5962%
уна: 0,5962%	них: 0,5962%	оло: 0,5564%	жен: 0,5564%	від: 0,5564%
охо: 0,4769%	хол: 0,4769%	лод: 0,4769%	одж: 0,4769%	дже: 0,4769%
пов: 0,4769%	ові: 0,4769%	віт: 0,4769%	ітр: 0,4769%	лін: 0,4372%
тря: 0,4372%	ної: 0,4372%	скн: 0,3975%	ивн: 0,3975%	иво: 0,3577%
газ: 0,3577%	люв: 0,3577%	юва: 0,3577%	пра: 0,3577%	ого: 0,3180%
іль: 0,3180%	зап: 0,3180%	апа: 0,3180%	вип: 0,3180%	ипу: 0,3180%
бло: 0,3180%	лок: 0,3180%	цил: 0,3180%	или: 0,3180%	інд: 0,3180%
ндр: 0,3180%	рів: 0,3180%	ере: 0,3180%	ова: 0,3180%	чен: 0,3180%
кон: 0,3180%	але: 0,3180%	пор: 0,2782%	рян: 0,2782%	льн: 0,2782%
впу: 0,2782%	алю: 0,2782%	при: 0,2782%	для: 0,2782%	тал: 0,2782%
без: 0,2782%	ідп: 0,2782%	дпр: 0,2782%	рац: 0,2782%	аць: 0,2782%
цьо: 0,2782%	ьов: 0,2782%	ани: 0,2782%	азі: 0,2782%	зів: 0,2782%
лек: 0,2782%	ект: 0,2782%	ску: 0,2782%	оро: 0,2385%	роз: 0,2385%
под: 0,2385%	ьни: 0,2385%	опо: 0,2385%	заб: 0,2385%	езп: 0,2385%
сум: 0,2385%	умі: 0,2385%	міш: 0,2385%	іші: 0,2385%	ції: 0,2385%

ння: 2,3847%
ист: 1,6296%
сис: 1,5501%
сте: 1,5501%
тем: 1,5501%
анн: 1,3116%
ван: 1,2719%
пал: 1,0731%
енн: 0,9539%
ема: 0,9141%
дви: 0,8744%
виг: 0,8744%
игу: 0,8744%
гун: 0,8744%
ува: 0,7552%
лив: 0,7154%
ний: 0,6757%
пус: 0,6757%
али: 0,6757%
уск: 0,5962%
уна: 0,5962%
них: 0,5962%
оло: 0,5564%
жен: 0,5564%
від: 0,5564%
охо: 0,4769%
хол: 0,4769%
лод: 0,4769%
одж: 0,4769%
дже: 0,4769%

ння ист сис сте тем анн ван пал енн ема дви виг игу гун ува лив ний пус али
уск уна них оло жен від охо хол лод одж дже

[illegible]

Діаграма



Для Misto.txt

```
Аналіз триграм для файлу: Misto.txt

Таблиця з триграмами, відсортована за спаданням частоти:
ого: 1,1823%   сте: 0,6857%   він: 0,5675%   ати: 0,5675%   теп: 0,5439%
при: 0,5439%   епа: 0,5202%   пан: 0,5202%   ому: 0,5202%   йог: 0,4966%
ува: 0,4966%   від: 0,4729%   про: 0,4729%   іст: 0,4493%   вся: 0,4493%
ськ: 0,4256%   зна: 0,4256%   ння: 0,4256%   ста: 0,4020%   йом: 0,4020%
анн: 0,3547%   міс: 0,3310%   ити: 0,3310%   ось: 0,3310%   тьс: 0,3074%
ься: 0,3074%   пер: 0,3074%   чин: 0,3074%   рад: 0,2838%   оло: 0,2838%
трі: 0,2838%   тер: 0,2838%   ала: 0,2838%   уст: 0,2838%   стр: 0,2838%
чен: 0,2601%   над: 0,2601%   аді: 0,2601%   кою: 0,2601%   сто: 0,2601%
ова: 0,2601%   ить: 0,2601%   ере: 0,2601%   поч: 0,2601%   лас: 0,2601%
оди: 0,2365%   ійк: 0,2365%   ели: 0,2365%   кол: 0,2365%   оли: 0,2365%
ньо: 0,2365%   ьог: 0,2365%   нав: 0,2365%   опо: 0,2365%   вар: 0,2365%
пис: 0,2365%   ько: 0,2365%   ним: 0,2365%   роз: 0,2365%   ина: 0,2365%
зус: 0,2365%   зос: 0,2365%   дій: 0,2128%   том: 0,2128%   сти: 0,2128%
зав: 0,2128%   ват: 0,2128%   оми: 0,2128%   спо: 0,2128%   тис: 0,2128%
ися: 0,2128%   дру: 0,2128%   ови: 0,2128%   лос: 0,2128%   вчи: 0,2128%
нко: 0,1892%   уде: 0,1892%   сво: 0,1892%   буд: 0,1892%   най: 0,1892%
ила: 0,1892%   оро: 0,1892%   ами: 0,1892%   вон: 0,1892%   пок: 0,1892%
лик: 0,1892%   бул: 0,1892%   але: 0,1892%   пом: 0,1892%   вав: 0,1892%
вел: 0,1892%   иса: 0,1892%   пов: 0,1892%   мов: 0,1892%   адч: 0,1655%
дче: 0,1655%   енк: 0,1655%   ний: 0,1655%   сел: 0,1655%   ком: 0,1655%
аро: 0,1655%   тат: 0,1655%   кра: 0,1655%   под: 0,1655%   айо: 0,1655%
ало: 0,1655%   іда: 0,1655%   ові: 0,1655%   авс: 0,1655%   ино: 0,1655%
```


4. Криптоаналіз

Зашфрували фрагменти тексту: $a = 33$ $b = 21$

Вбіьгмі хТіпіа есТвбгТплеієжТогтгбсбТісТірибеяТгвгпимсгх.ТЩлТмяйлТ
ичі хТА лвсгс.ТСТіглпмгпіїТпигТіяа бишскТапгогТілдеуєсТРгбіас.Т
гжТвгмиеіпауТбсмиа фТ–ТпигТгмбйяк хау.ТїсТгвіагдТдсжря гхгзТмбйііТ
гпсбічс,ТА лвсгТпвійгскТїсмижея.ТПимТ гогТжгдяТшгдяхТа
скТігвбікдгТжТвбієбг:Т«ОімеідТіегшігдТяупеуегауТжгдяТгрлбгя іТресєі
іггеяТїсмижеяТпТєяьгпсбея,Твбірібсехгіщф,ТпТгьбггщуТвіаггогТмгрбгря
яТдгегмгогТдицсгіс».

Отриманні шифротексти від колеги:

АпіїщЯезм іщжюєабЯвЯезбу

ЯЛіхьяюзхіобЯшЯаубюсзузлшЯзкбвь.ЯУьїЯвщеіуїбжЯХщшающ,ЯизЯющсі
ЯуцгцїїбЯгцвзлЯвЯуіаіозицл ЯезвцщубюєЯфзрзЯуьмізчЯюпр
.ЯЩоіЯХщшаюЯуьклзуобжюєаб,Явщеіуїбгм ЯЛіхьяюзхіоб,ЯизЯпіЯїіЯу
оьсшжЯфзрзЯуькЯдшгц ,ЯбсщЯающїіЯфзрзЯашешної
пігЯкзЯащлзчЯалігюь.ЯКштЯїщлщрщжюєабЯеігісзїщю
Ямзозуьщ,ЯизЯвлздіЯфзрзЯезвцщю ю Яішкер ЯьЯающу юєЯшлзуш.

Літера	Відносна частота (%)
я	15.45
з	9.39
щ	6.97
і	5.76
ю	5.45
у	5.45
а	4.55
ї	4.24
б	4.24
ш	3.94
л	3.33
ь	3.33
в	2.73
е	2.42
о	2.42
р	2.12
є	1.82
х	1.82
с	1.82
к	1.82
ж	1.52
г	1.52
м	1.21
и	1.21
ц	1.21
ф	1.21
п	0.91
ч	0.61
ґ	0.61
д	0.61
т	0.30

Біграма	Відносна частота (%)
зя	2.532
яу	2.215
бя	1.899
аю	1.899
яа	1.899
ія	1.899
яе	1.582
ює	1.582
яв	1.582
уь	1.582
їщ	1.266
щя	1.266
ез	1.266
об	1.266
зу	1.266
ющ	1.266
яф	1.266
фз	1.266
зр	1.266
рз	1.266
лз	1.266
пі	0.949
жю	0.949
єа	0.949
аб	0.949
лі	0.949
іо	0.949
ья	0.949
еі	0.949
їб	0.949

Триграма	Відносна частота (%)
язз	1.3201
яуь	1.3201
яфз	1.3201
фзр	1.3201
зрз	1.3201
рзя	1.3201
жює	0.9901
юєа	0.9901
єаб	0.9901
абя	0.9901
ающ	0.9901
яиз	0.9901
изя	0.9901
щжю	0.6601
бярв	0.6601
ярв	0.6601
ялі	0.6601
ліх	0.6601
іхь	0.6601
хья	0.6601
ьяю	0.6601
аюз	0.6601
юзх	0.6601
зхі	0.6601
хіо	0.6601
іоб	0.6601
обя	0.6601
явщ	0.6601
вще	0.6601
щеі	0.6601
еіу	0.6601
іуї	0.6601
уїб	0.6601
яхщ	0.6601
хщш	0.6601

Порівняємо частоту появи літер нашого тексту та шифротексту.

Була зроблена програма, що просто перебирає всі a та b для поданого тексту на вхід.

```
Ключі a=31, b=29:  
Сцена починається з появи Мефістофеля у святковому одязі. Він запевняє Фауста, що таке вбрання разом з веселощами позбав  
лять його вічної туги. Але Фауст відмовляється, запевняючи Мефістофеля, що це не вилікує його від журби, яка стане його  
супутницею до самої смерті. Дух намагається переконати чоловіка, що зможе його позбавити нудьги і ставить умову.
```

Текст розшифровано.

5. Криптоаналіз довільного моноалфавітного шифру (додаткове завдання на +5 балів)

1. Згенеруйте випадкову моноалфавітну підстановку (таблицю відповідності) та зашифруйте за її допомогою достатньо довгий фрагмент тексту.
2. Виконайте криптоаналіз отриманого шифротексту, використовуючи частотний аналіз (в т.ч. частоти біграм, триграм та припущень).

Використана таблиця підстановки:

А	->	Д
Б	->	Е
В	->	Є
Г	->	Ж
Ґ	->	З
Д	->	И
Е	->	І
Є	->	Ї
Ж	->	Й
З	->	К
И	->	Л
І	->	М
Ї	->	Н
Й	->	О
К	->	П
Л	->	Р
М	->	С
Н	->	Т
О	->	У
П	->	Ф
Р	->	Х
С	->	Ц
Т	->	Ч
У	->	Ш
Ф	->	Щ
Х	->	Ь
Ц	->	Ю
Ч	->	Я
Ш	->	
Щ	->	А
Ь	->	Б
Ю	->	В
Я	->	Г
	->	Ґ

Вигляд зашифрованого тексту:

ЦтлГоужуГауотуГфукрдклрлГкГпутіо.ГЧуГешрлГієдГкиухуєітмГфдхшеп
л,ГгпмГилелрлцГаіГк-фмГруед,ГгпГшцмГцісmtдхлцчл,ГауГрл
ітбГфупмтядрлГтдшпл.ГКиухуєм,Гхшс'гтмГньтмГрлюгГшпхледеГтмйтлоГф
шь,ГауГтіГктдеГмаіГехлчел.ГЄутлГишйГкдецчлидрлцГкГчдпужуГедчбпує
ужуГфхлемчдттГГмГцчугрлГтіхшьусу,Гфучшфлє лГауямГеГкісрв.

Біграма	Відносна частота (%)	Триграма	Відносна частота (%)
лГ	2.564	Гтд	0.8724
дГ	2.184	тдГ	0.6171
уГ	1.909	Гфу	0.5958
Гт	1.656	цГГ	0.5533
мГ	1.519	Гті	0.5533
іГ	1.477	рлГ	0.5107
Гф	1.466	члГ	0.5107
Гє	1.445	Гчд	0.5001
Гч	1.414	плГ	0.4894
Гц	1.350	ужу	0.4681
тд	1.308	жүГ	0.4469
Гк	1.234	Гау	0.4469
Ге	1.066	тіГ	0.4469
оГ	1.044	Ггп	0.4256
чл	0.960	ГоГ	0.4256
ГГ	0.939	Гфх	0.4256
чд	0.939	Гиу	0.4256
шГ	0.897	ГмГ	0.3937
Ги	0.886	слГ	0.3830
ЄГ	0.855	ауГ	0.3724
рл	0.823	Геш	0.3724
пл	0.770	Гкд	0.3511
пд	0.749	Гчу	0.3192
ГГ	0.738	чдГ	0.3085
ьГ	0.728	Гпу	0.3085
ху	0.717	пдГ	0.3085
цч	0.707	лог	0.2979
жу	0.696	чуГ	0.2979
фу	0.686	Гцч	0.2979
дє	0.686	меГ	0.2873

Співставимо відносні частоти літер до попередніх даних з файлу Misto.txt

Misto.txt		Encrypted.txt	
Літера	Відносна частота (%)	Літера	Відносна частота (%)
	15,76	г	17,84
о	8,17	д	7,54
а	7,16	у	7,29
н	5,39	л	6,30
и	5,32	є	4,49
в	5,10	т	4,31
і	4,48	ч	4,31
с	4,21	м	4,08
т	4,19	п	3,64
р	3,70	і	3,59
е	3,56	ц	3,55
у	3,24	х	3,11
к	3,03	р	3,06
м	2,93	ш	2,91
п	2,77	и	2,51
д	2,76	с	2,25
л	2,57	ф	2,14
з	2,19	е	2,12
я	1,92	г	1,99
й	1,54	к	1,93
ь	1,34	б	1,49
г	1,29	о	1,41
б	1,20	я	1,33
ч	1,16	ь	1,32
ю	0,83	ж	1,22
ж	0,74		0,82
х	0,70	н	0,76
ш	0,65	а	0,72
ц	0,63	й	0,66
ї	0,57	ю	0,57
щ	0,43	в	0,52
є	0,39	ї	0,21
ф	0,09	щ	0,02

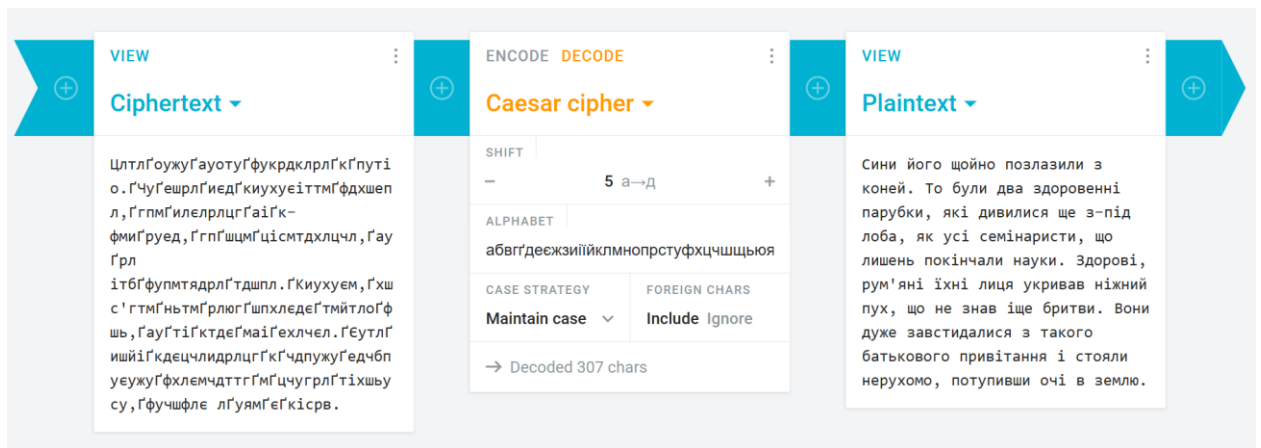
Можна помітити що «пробіл» відповідає літері «Г»

Припустимо що пробіл буде 34 елементом у алфавіті.

Таким чином $34 + 5 = 5$ літера = Г.

Відповідно шифр розгадано і якщо ми виконаємо зсув на 5 позицій вправо отримає розшифрований текст.

Це шифр Цезаря.



Як бачимо припущення було правильним.
Текст розшифровано

Висновок: у процесі виконання лабораторної роботи було проаналізовано ймовірнісні параметри появи літер української абетки в різних українських текстах. Досліджено частотність літер, біграм і триграм. Розглянуто роботу афінного шифру та здійснено його розшифрування за допомогою методів криптоаналізу. Для розробки програмного забезпечення використовувалась мова програмування C#. Було створено моноалфавітний шифр та виконано криптоаналіз шифротексту.