

Міністерство освіти й науки України
Львівський національний університет імені Івана Франка
Факультет електроніки та комп'ютерних технологій
з предмета: Комп'ютерна лінгвістика

Звіт
про виконання лабораторної роботи № 10
**«Дослідження базових лінгвістичних властивостей рандомних символьних
послідовностей»**

Виконав:
Студент групи
Фес-32с
Бойко Кирило

Львів 2024

Завдання

1. Для досліджень використати випадкові та випадковізовані тексти, генеровані в лабораторній роботі №9.
2. Дослідити усереднений по всіх словах параметр R кластеризації слів у цих текстах і усереднений по всіх словах параметр γ кореляції слів у них.
3. Дослідити ті ж параметри для такої ж кількості природних текстів, написаних різними мовами.
4. Зробити висновки стосовно схожості або відмінності параметрів R і γ для природних і випадкових або випадковізованих текстів.

Хід виконання роботи

Запустив програму **+shuffler (words only)** та завантажив до неї перший текст

Clemencia Novela de costumbres by Fernán Caballero

Провів перевірку, кількість циклів випадковізації $N = 50\,000$

Парметри	Вхідний текст	Вихідний текст
20 19 64 11 26 43 85 76 97 18 68 30 65 43 68 20 17 25 97 20 83 58 75 14 92 1 43 35 45 2 75 70 32 28 16 27 92 37 55 76 90 27 1 41 29 88 5 70 23 86 2 24 59 23 94 62 43 51 19 26 99 51 31 96 43 98 97 38 27 42 28 79 69 21 30 56 57 34 85 5 96 73 12 58 1 52 11 10 34 54 83 41 19 30 77 39 96 5 98 30		

Завантажив текст мавпи Міллера з попередньої лабораторної роботи

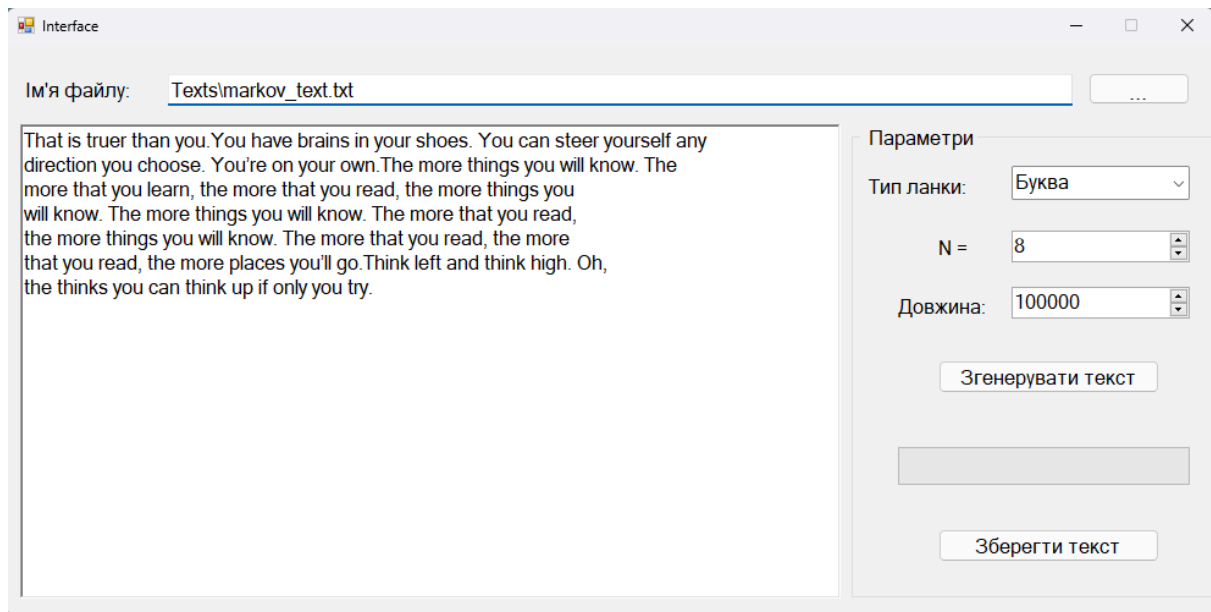
Парметри	Вхідний текст	Вихідний текст
38 17 28 84 32 27 61 97 19 46 33 45 63 34 81 34 30 23 16 55 81 2 90 86 10 67 15 20 6 16 26 31 17 47 69 74 75 83 37 66 47 81 55 50 60 72 90 34 73 12 57 94 8 76 7 37 26 42 74 64 12 42 91 3 21 3 28 17 65 7 59 46 86 49 8 1 46 41 26 38 77 80 46 93 59 31 24 41 40 33 49 34 10 72 55 6 89 52 30 58		

Відкрив програму **+Markov texts** з лабораторної роботи 27 та згенерував текст, з такими налаштуваннями:

$$N = 8$$

Тип ланки: букви

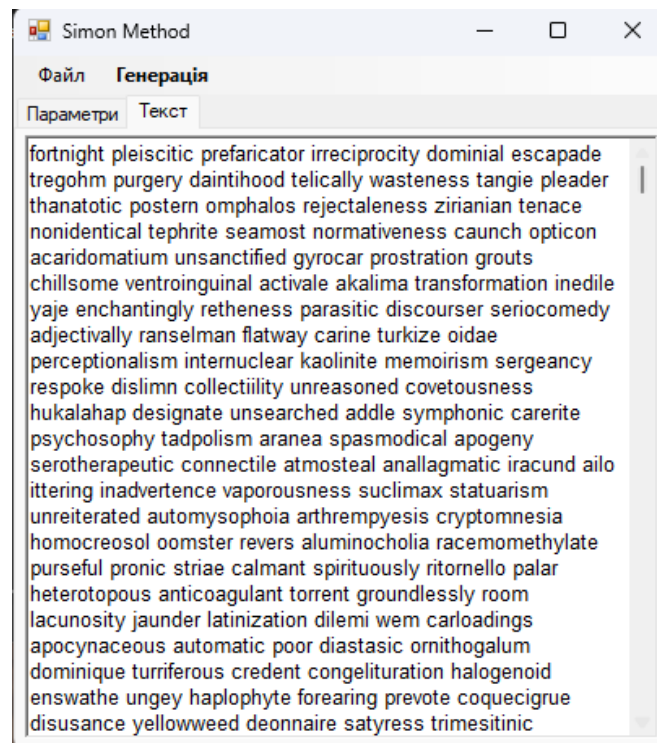
Довжина: 100000



Зберіг цей текст і завантажив його в поточну програму лабораторної роботи для подальшого аналізу

```
10 90 67 46 27 86 34 71 41 40 29 7 33 26 65 98 77 77 28 3 19 83 9 31 88 37 81 34 84 33
83 29 32 35 38 40 32 10 45 74 72 9 17 54 6 44 85 42 87 3 60 43 54 13 36 27 82 37 58 89
6 1 86 8 57 1 75 98 87 26 66 68 32 54 31 11 66 54 25 44 51 39 46 81 73 41 74 19 99 12
57 94 51 56 68 72 22 10 96 5
```

Згенерував новий текст в програмі **+Simon_3in1**



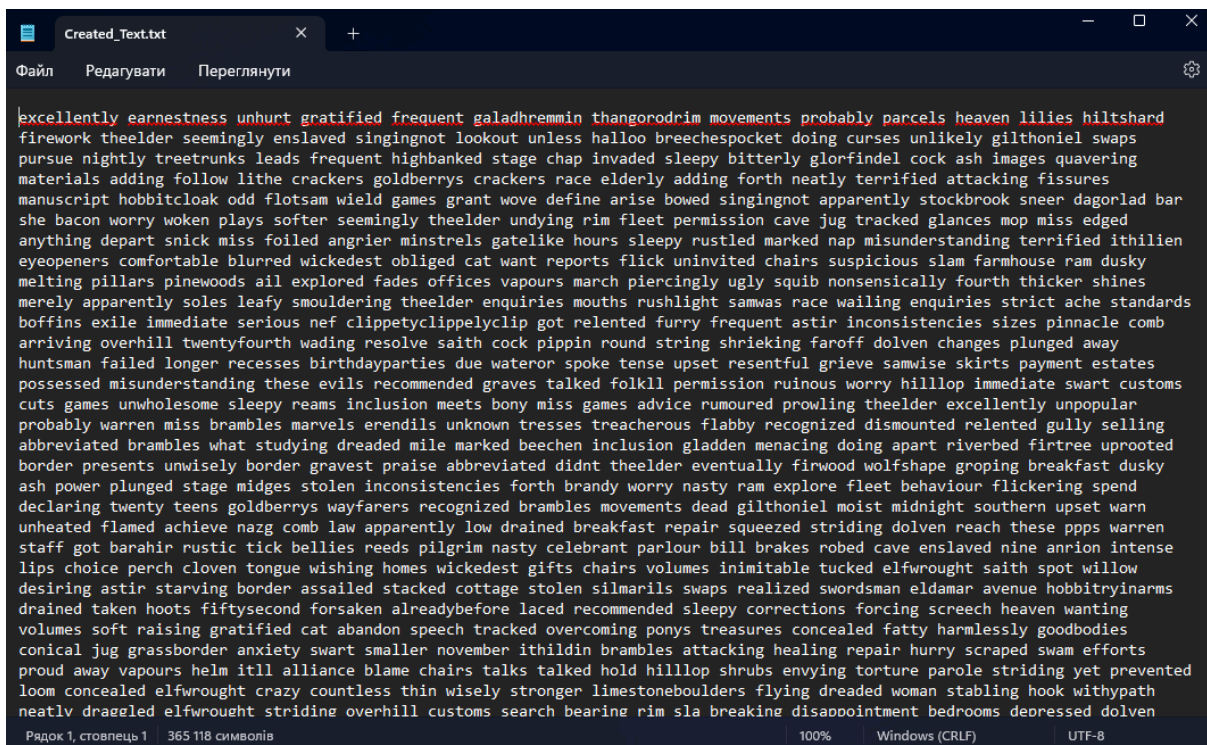
Результат дослідження:

Парметри	Вхідний текст	Вихідний текст
21 57 91 71 93 17 51 6 65 13 10 49 85 49 26 41 27 79 24 90 32 70 49 3 17 66 68 58 14 81		
60 55 9 14 18 39 13 50 91 51 36 40 61 54 50 30 38 44 22 3 46 87 34 85 66 51 8 88 55 28		
49 92 91 84 53 55 40 71 32 87 29 77 88 72 55 92 9 95 53 86 28 25 36 12 35 72 22 24 48		
67 31 24 94 64 3 43 49 77 1 95		

Згенерував текст за допомогою програми **+PolyaTexts** і завантажив його для подальшого дослідження

```
Enter the size of starting urn: 8000
Enter the text lenth: 50000
Enter Ro: 8
Enter Nu: 4
To input again press 2, to continue press 1: 1
Processing might take some time...

Finished in 0.06410717964172363
To view created txt array press 1
To view tokens press 2
To get .txt file press 3
Else to exit 3
Created!
-----
```



Created_Text.txt

Файл Редагувати Переглянути

excellently earnestness unhurt gratified frequent galadhrermin thangorodrim movements probably parcels heaven lilies hiltshard
firework theelder seemingly enslaved singingnot lookout unless halloo breechespocket doing curses unlikely gilthoniel swaps
pursue nightly treetrunks leads frequent highbanked stage chap invaded sleepy bitterly glorfindel cock ash images quavering
materials adding follow lithe crackers goldberrys crackers race elderly adding forth neatly terrified attacking fissures
manuscript hobbitcloak odd flotsam wield games grant wove define arise bowed singingnot apparently stockbrook sneer dagorlad bar
she bacon worry woken plays softer seemingly theelder undying rim fleet permission cave jug tracked glances mop miss edged
anything depart snick miss foiled angrier minstrels gatelike hours sleepy rustled marked nap misunderstanding terrified ithilien
eyeopeners comfortable blurred wickedest obliged cat want reports flick uninvited chairs suspicious slam farmhouse ram dusky
melting pillars pinewoods ail explored fades offices vapours march piercingly ugly squib nonsensically fourth thicker shines
merely apparently soles leafy smouldering theelder enquiries mouths rushlight samwas race wailing enquiries strict ache standards
boffins exile immediate serious nef clippetyclippelyclip got relented furry frequent astir inconsistencies sizes pinnacle comb
arriving overhill twentyfourth wading resolve saith cock pippin round string shrieking faroff dolven changes plunged away
hunterman failed longer recesses birthdayparties due wateror spoke tense upset resentful grieve samwise skirts payment estates
possessed misunderstanding these evils recommended graves talked folkil permission ruinous worry hilllop immediate swart customs
cuts games unwholesome sleepy reams inclusion meets bony miss games advice rumoured prowling theelder excellently unpopular
probably warren miss brambles marvels erendils unknown tresses treacherous flabby recognized dismounted relented gully selling
abbreviated brambles what studying dreaded mile marked beechen inclusion gladden menacing doing apart riverbed firtree uprooted
border presents unwisely border gravest praise abbreviated didnt theelder eventually firwood wolfshape groping breakfast dusky
ash power plunged stage midges stolen inconsistencies forth brandy worry nasty ram explore fleet behaviour flickering spend
declaring twenty teens goldberrys wayfarers recognized brambles movements dead gilthoniel moist midnight southern upset warn
unheated flamed achieve nazg comb law apparently low drained breakfast repair squeezed striding dolven reach these ppps warren
staff got barahir rustic tick bellies reeds pilgrim nasty celebrant parlour bill brakes robed cave enslaved nine anrion intense
lips choice perch cloven tongue wishing homes wickedest gifts chairs volumes inimitable tucked elfwrought saith spot willow
desiring astir starving border assailed stacked cottage stolen silmarils swaps realized swordsman eldamar avenue hobbitrynarms
drained taken hoots fiftysecond forsaken alreadybefore laced recommended sleepy corrections forcing screech heaven wanting
volumes soft raising gratified cat abandon speech tracked overcoming ponys treasures concealed fatty harmlessly goodbodies
conical jug grassborder anxiety swart smaller november ithildin brambles attacking healing repair hurry scraped swam efforts
proud away vapours helm itll alliance blame chairs talks talked hold hilllop shrubs envying torture parole striding yet prevented
loom concealed elfwrought crazy countless thin wisely stronger limestoneboulders flying dreaded woman stabling hook withpath
neatlv draaeled elfwrougth stridine overhill customs search bearing rim sla breakine disappointment bedrooms deopressed dolven

Рядок 1, стовпець 1 365 118 символів 100% Windows (CRLF) UTF-8

Парметри	Вхідний текст	Вихідний текст
	68 92 68 60 99 12 28 33 80 3 51 92 19 5 78 29 5 47 98 27 46 37 87 5 50 60 28 97 29 34 65 90 3 16 60 54 13 80 20 65 62 49 29 7 52 22 12 56 75 51 66 88 12 19 43 31 5 64 11 40 83 30 4 46 38 60 89 4 44 25 17 25 27 33 64 97 9 80 53 29 48 71 23 77 68 76 79 83 73 29 49 49 69 80 81 32 46 83 69 91	

Далі, я згенерувала ще 3 текста за допомогою програми **+monkey_texts**

Та провела дослідження з ними:

Парметри	Вхідний текст	Вихідний текст
	33 68 50 90 44 33 66 12 50 17 89 66 14 60 70 5 53 11 30 95 7 77 50 41 54 24 46 87 72 24 97 24 25 36 36 90 95 23 41 95 56 79 48 73 33 64 26 43 62 28 91 78 67 53 13 55 19 9 36 20 86 78 40 26 20 65 42 78 25 80 9 31 70 34 39 28 30 76 9 89 98 4 25 44 33 19 57 72 23 80 72 86 86 22 9 77 39 22 54 68	

Парметри	Вхідний текст	Вихідний текст
	57 37 26 72 77 20 96 57 70 4 64 29 65 84 41 26 38 89 70 28 54 13 93 41 92 94 16 26 12 89 74 53 51 33 74 99 30 74 76 16 10 19 65 64 87 23 56 70 46 76 75 33 84 11 65 45 43 85 80 82 4 71 45 80 29 12 77 32 10 42 96 64 14 55 19 36 47 30 53 70 39 46 79 36 15 42 68 41 67 29 56 44 93 93 12 39 74 85 35 76	

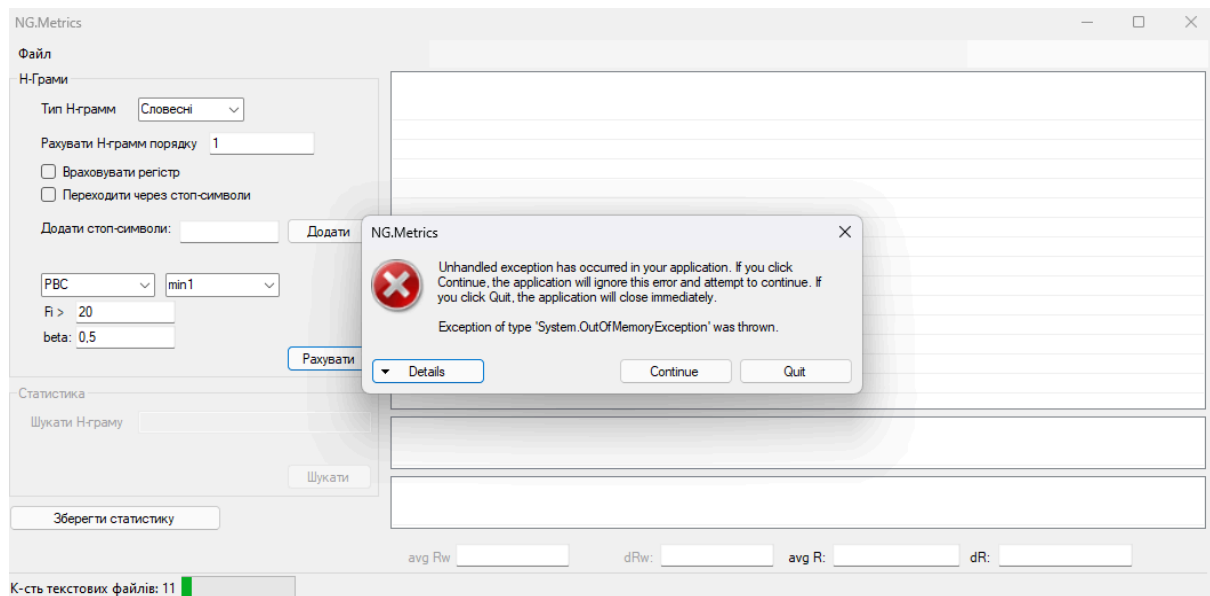
Парметри	Вхідний текст	Вихідний текст
	17 4 12 29 54 19 58 17 35 85 74 69 87 31 10 87 43 43 22 1 42 2 12 92 33 83 82 62 10 16 12 3 84 74 18 57 29 85 73 72 73 5 41 48 50 54 34 27 32 88 95 80 2 26 84 15 91 20 96 71 36 96 7 20 73 72 85 14 13 53 58 58 70 49 28 37 61 64 43 79 49 55 30 22 20 32 1 59 89 3 65 9 89 3 36 76 98 21 74 77	

```

Clemencia Novela de costumbres by Fernán Caballer...
Created_Text.txt
M30_L150000_space0_eq.txt
markov.txt
Monkey_M5_L200000_space0.2_eq_letters.txt
Monkey_M15_L100000_space0.2_log0.17letters.txt
Monkey_M15_L200000_space0.2_eq_letters.txt
simon.txt

```

На жаль, мені не вдалося виконати наступні завдання, оскільки програма давала збій у роботі



Висновок: У процесі виконання роботи було досліджено лінгвістичні характеристики випадкових та випадковизованих текстів за допомогою параметрів кластеризації RRR та кореляції γ . Для аналізу були згенеровані тексти за допомогою різних моделей, таких як текст мавпи Міллера, тексти Маркова, а також моделі Саймона і Полі. Отримані результати показали, що параметри кластеризації та кореляції в природних текстах суттєво відрізняються від тих, що характерні для випадкових текстів. Це підтверджує можливість використання цих параметрів для розрізнення різних типів текстів, що є важливим інструментом для лінгвістичних досліджень та аналізу текстової інформації.