

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 № 16
«Характеристика повторюваності лінгвістичних елементів у тексті»

Виконав:
Студент групи
Фес-32с
Бойко Кирило

Львів 2024

Завдання

використовуючи видану локально розміщену програму ++repetition3.3jar і онлайнпрограму (див. файл +Colab program_explanation.pdf), вивчити характеристику повторюваності для кількох природних і рандомних текстів, генерованих в попередніх лабораторних роботах.

Хід виконання роботи

Я обрав два природних тексти та два згенерованих текста для виконання лабораторної роботи

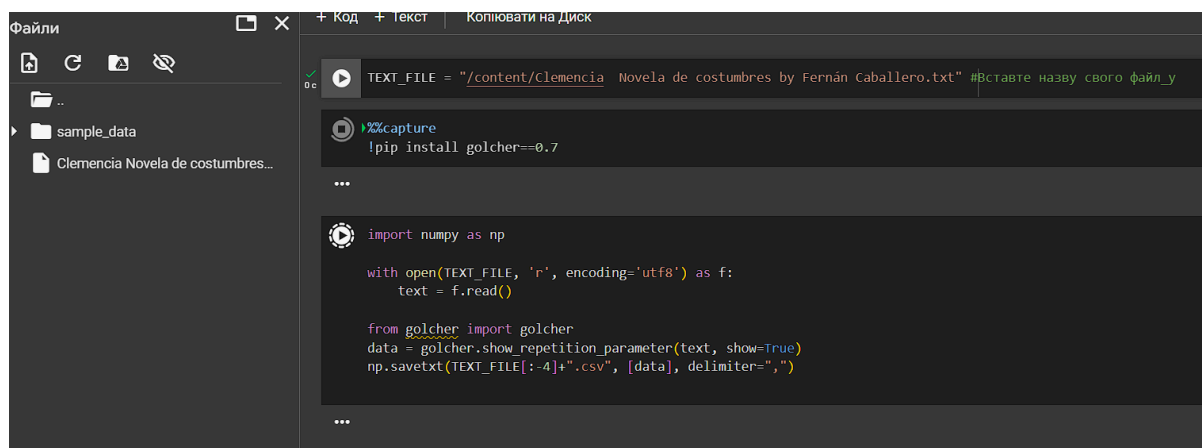
Clemencia Novela de costumbres by Fernán Caballero.txt

Monkey_M5_L200000_space0.2_eq_letters.txt

Monkey_M15_L200000_space0.2_eq_letters.txt

Людина-маятник.txt

Запустив програму, та завантажив туди всі обрані мною текстові файли



The screenshot shows a Jupyter Notebook interface. On the left, a file browser pane titled 'Файли' (Files) displays a directory structure with 'sample_data' and a file named 'Clemencia Novela de costumbres...'. The main area has tabs for '+ Код' (Code), '+ Текст' (Text), and 'Копіювати на Диск' (Copy to Disk). The 'Код' tab is active, showing a code cell with the following content:

```
TEXT_FILE = "/content/Clemencia Novela de costumbres by Fernán Caballero.txt" #Вставте назву свого файлу

%%capture
!pip install golcher==0.7

...

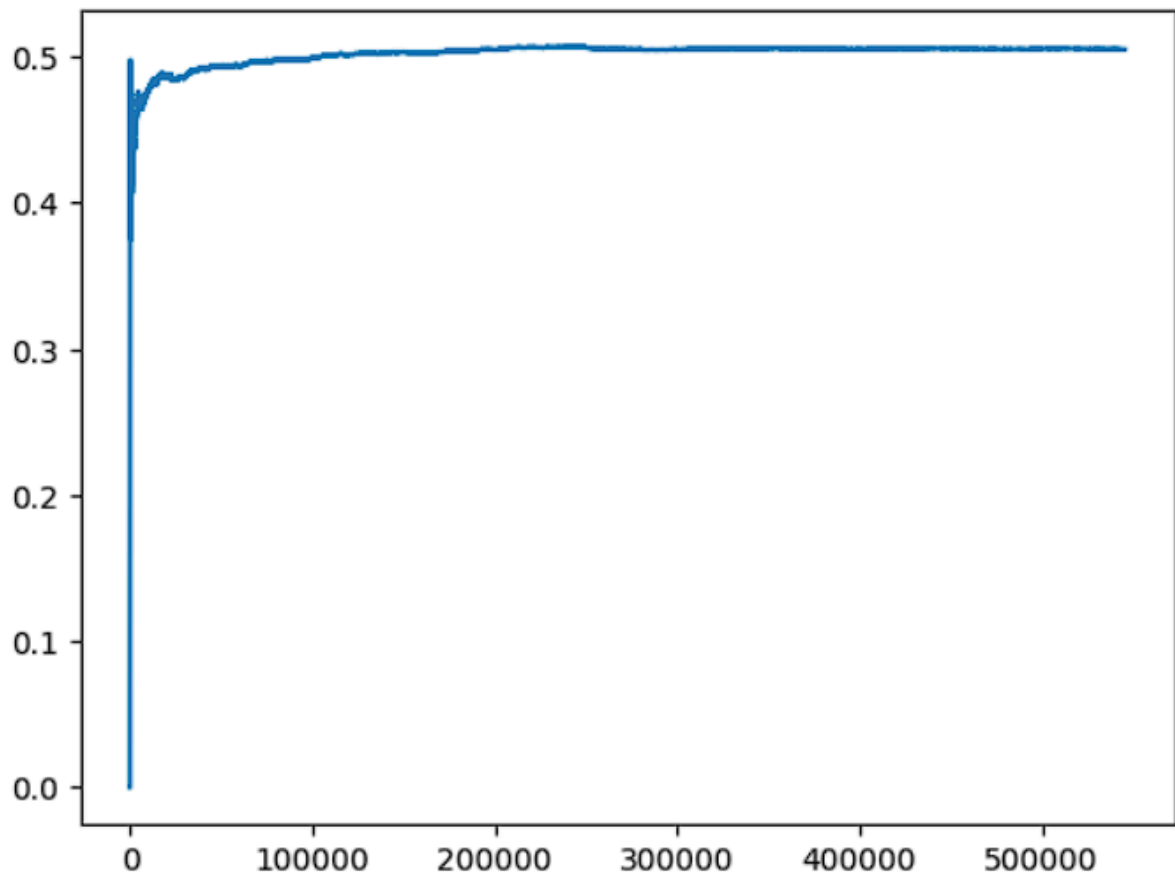
import numpy as np

with open(TEXT_FILE, 'r', encoding='utf8') as f:
    text = f.read()

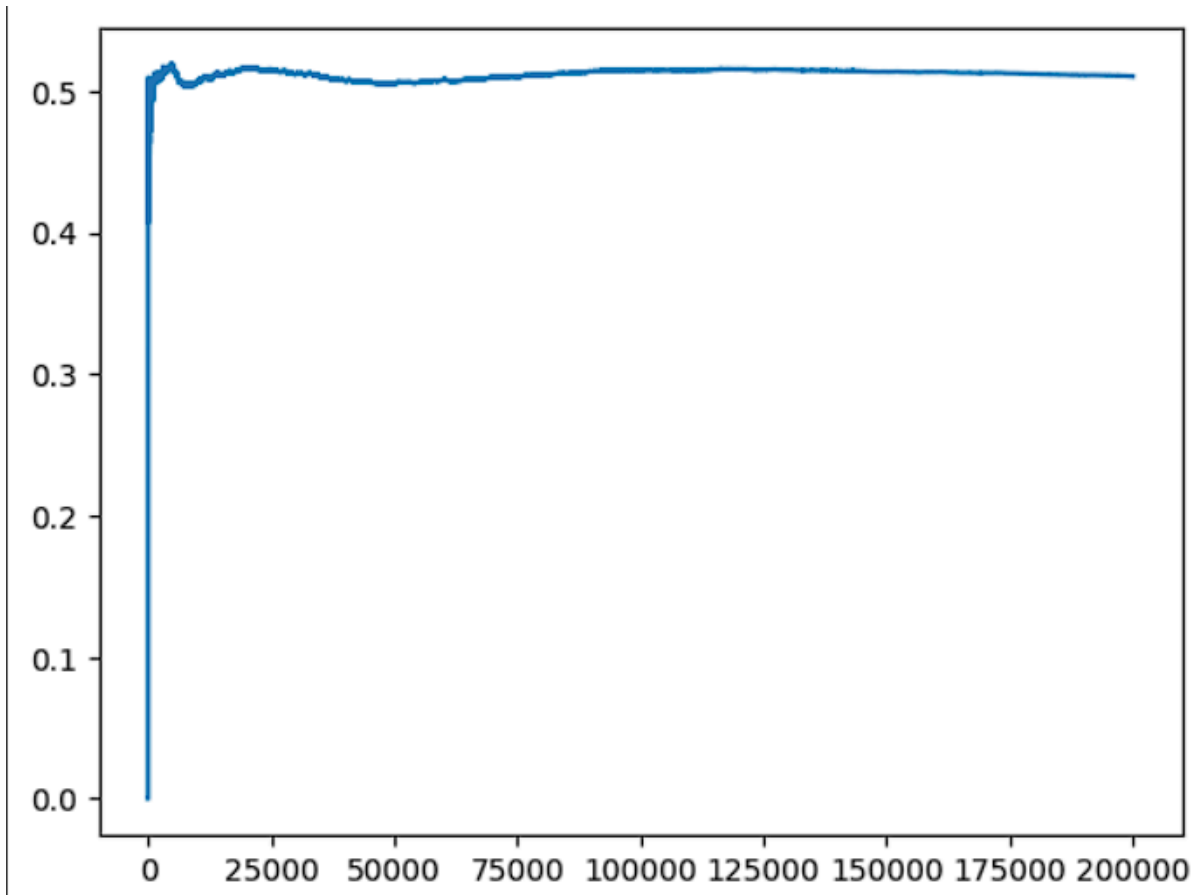
from golcher import golcher
data = golcher.show_repetition_parameter(text, show=True)
np.savetxt(TEXT_FILE[:-4]+".csv", [data], delimiter=",")

...
```

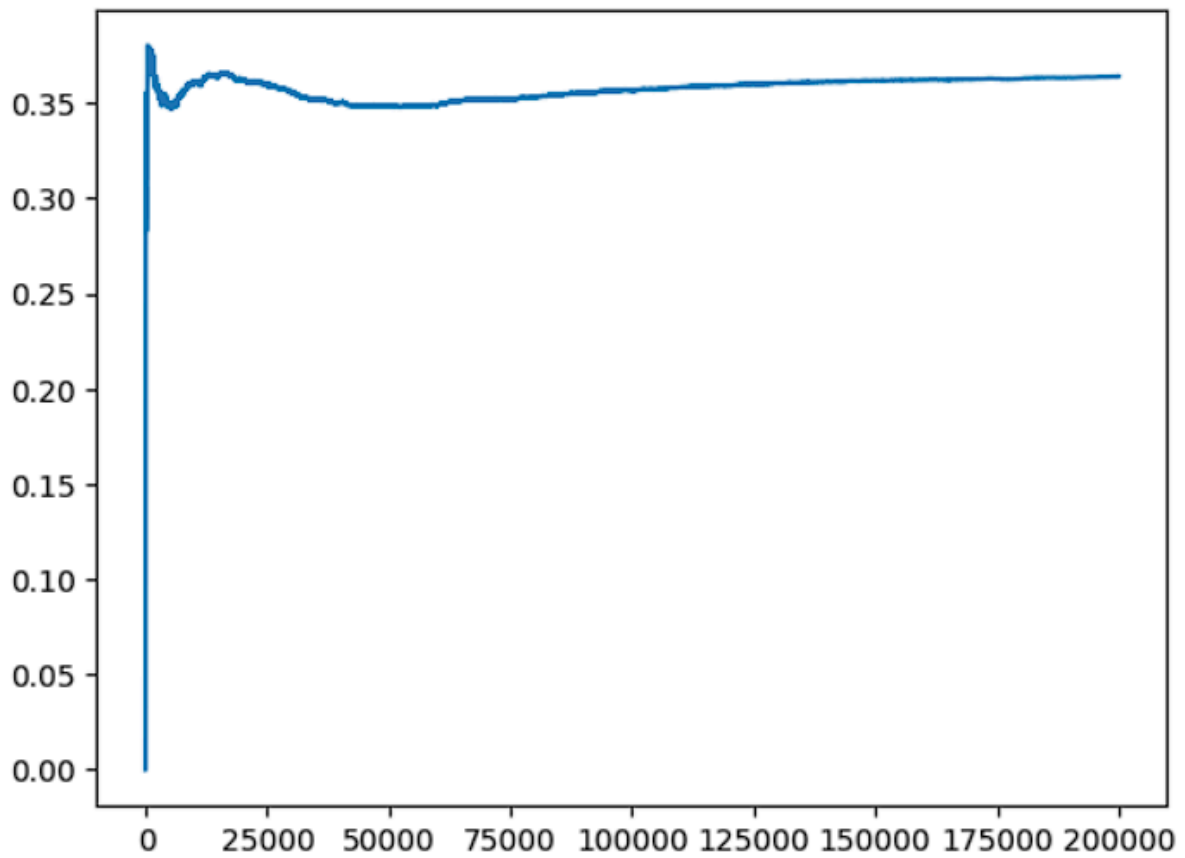
Результати



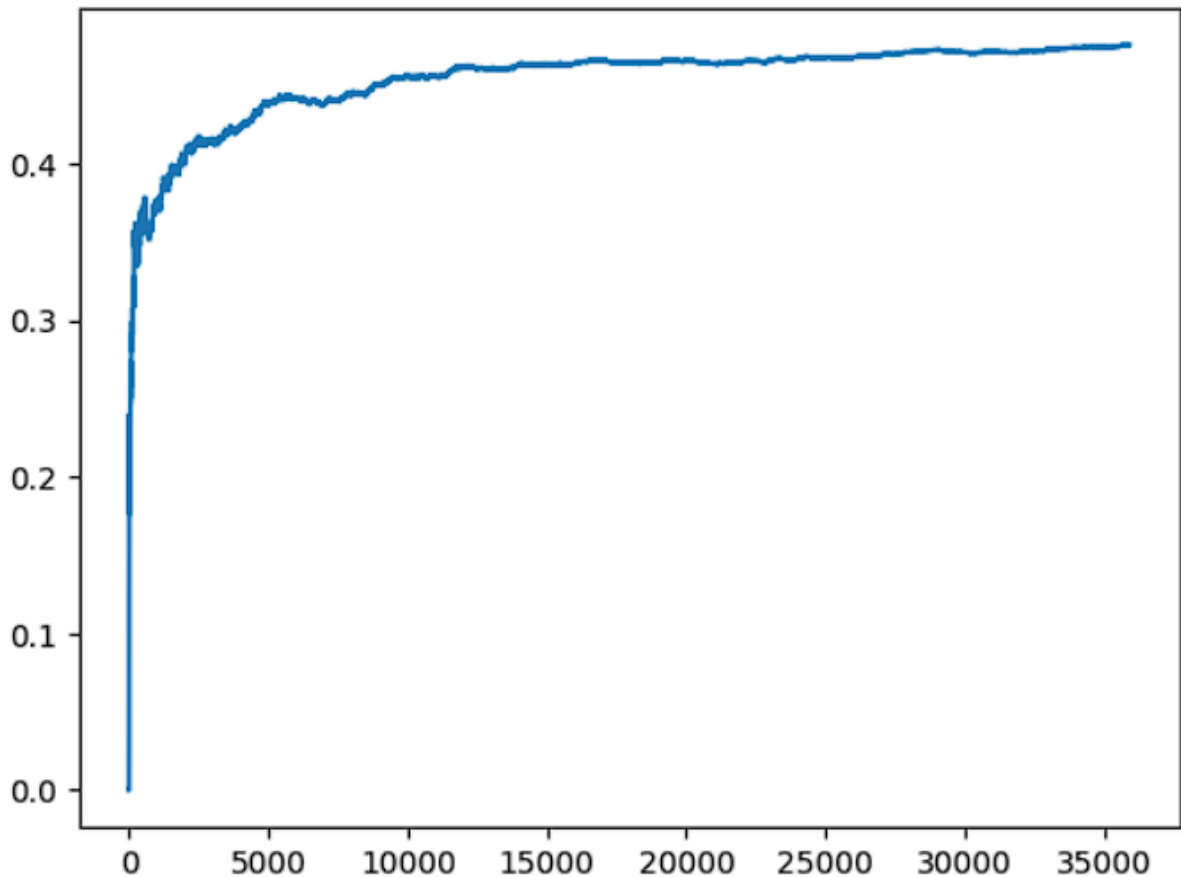
Графік $v(t)$ залежності параметра повторюваності v від довжини тексту t
Clemencia Novela de costumbres by Fernán Caballero.txt



Графік $v(t)$ залежності параметра повторюваності v від довжини тексту t
Monkey_M5_L200000_space0.2_eq_letters.txt



Графік $v(t)$ залежності параметра повторюваності v від довжини тексту t
Monkey_M15_L200000_space0.2_eq_letters.txt



Графік $v(t)$ залежності параметра повторюваності v від довжини тексту t
Людина-маятник.txt

На основі досліджень, я можу зробити такий висновок порівняння

Clemencia Novela de costumbres by Fernán Caballero:

- Графік повторюваності $v(t)$ для цього тексту показує поступове зростання параметра повторюваності зі збільшенням довжини тексту. Ця тенденція вказує на відносно високу послідовність лінгвістичних елементів, характерних для природного тексту. Відсутність різких коливань говорить про стабільну структуру, притаманну літературним творам.

Monkey_M5_L200000_space0.2_eq_letters:

- Для цього штучно згенерованого тексту графік демонструє більш нерівномірне зростання параметра $v(t)$ в порівнянні з природним текстом. Видно коливання на певних ділянках, що може свідчити про відсутність природної послідовності в тексті, характерної для випадково створених наборів символів. Параметр повторюваності менш стійкий.

Monkey_M15_L200000_space0.2_eq_letters:

- Порівняно з Monkey_M5, цей графік показує дещо більше вирівняну тенденцію в зростанні параметра $v(t)$. Це можна пояснити тим, що кількість імовірних

повторюваних елементів у випадковому тексті зростає, що зробило його більш однорідним, хоча рівень повторюваності все одно залишається нижчим за показники природного тексту.

Людина-маятник:

- Як і в першому випадку, текст "Людина-маятник" демонструє стабільне зростання повторюваності. Порівняно з випадковими текстами, цей графік має менше коливань і свідчить про наявність внутрішньої структури, яка часто зустрічається у природних текстах. Це підтверджує, що текст належить до категорії природних, із послідовним стилем.

На жаль, мені не вдалося запустити програму **++repetition 3.3jar**, оскільки при спробі запуску вона відразу закривається

Висновок: На цій лабораторній роботі я дослідив повторюваність лінгвістичних елементів у природних та випадково згенерованих текстах, використовуючи програму в googlecolab. Проведений аналіз показав, що природні тексти, такі як "Clemencia Novela de costumbres by Fernán Caballero" і "Людина-маятник", характеризуються стабільним зростанням параметра повторюваності без різких коливань. Це вказує на присутність певної структури та стилю. Натомість випадково згенеровані тексти мали більші коливання в значеннях $v(t)v(t)v(t)$, що є ознакою нестабільної структури і випадковості. Таким чином, можна зробити висновок, що природні тексти демонструють більш передбачувану та послідовну структуру повторюваності, що відрізняє їх від випадкових текстів.