

# Raport

Damian Guzek & Kamil Sandak

Analiza tekstu - laboratorium 6 - zaawansowane programowanie

Nasz program składa się z 6 funkcji pozwalających na stwierdzenie w jakim stylu i o czym napisana jest dana książka:

- **pop\_100**
- **pop\_1000**
- **słowa\_stop**
- **liczby**
- **tresc**
- **najdluzszy**

Na MacOS program uruchamia się za pomocą komendy w terminalu:  
pc4:~ uzytkownik\$ **python3 program.py analizowany\_tekst.txt**

## OPIS FUNKCJI

**pop\_100** - oblicza jaki procent tekstu stanowi 100 najczęściej występujących wyrazów w tekście

**pop\_1000** - oblicza jaki procent tekstu stanowi 1000 najczęściej występujących wyrazów w tekście

**słowa\_stop** -zwraca jaki procent tekstu stanowią słowa, które nie determinują kontekstu tekstu

**liczby** - określa ile liczb znajduje się w tekście

**tresc** - wyznacz 10 najpopularniejszych wyrazów charakterystycznych dla danego tekstu

**najdluzszy** -zwraca zbiór składa się z wyrazów składających się z największej liczby liter

**Uwaga:** Z danych wejściowych zostały usunięte wstawki z programu gutenberga

## Eve's Diary - Mark Twain

<https://www.gutenberg.org/files/8525/8525-0.txt>

```
===== Raport=====
Nazwa załadowanego pliku:
twain.txt

100 najpopularniejszych wyrazów stanowi 65.75 % tekstu
1000 najpopularniejszych wyrazów stanowi 94.7 % tekstu
Słowa stop stanowią 61.16 % tekstu
Liczba liczb występujących w tekście: 6

10 najpopularniejszych wyrazów charakterystycznych dla twain.txt :
{'love', 'could', 'think', 'one', 'first', 'know', 'never', 'would', 'good', 'like'}

Najdłuższe wyrazy w twain.txt :
{'notwithstanding'}
===== Raport=====
```

## Garden and Forest Weekly

<http://www.gutenberg.org/cache/epub/32141/pg32141.txt>

```
===== Raport=====
Nazwa załadowanego pliku:
gardens.txt

100 najpopularniejszych wyrazów stanowi 46.35 % tekstu
1000 najpopularniejszych wyrazów stanowi 78.21 % tekstu
Słowa stop stanowią 41.79 % tekstu
Liczba liczb występujących w tekście: 645

10 najpopularniejszych wyrazów charakterystycznych dla gardens.txt :
{'mr', '50', '1', 'new', 'trees', 'white', '_', 'flowers', 'one', 'plants'}

Najdłuższe wyrazy w gardens.txt :
{'_pflanzenmischlinge'}
===== Raport=====
```

## The Koran

<http://www.gutenberg.org/cache/epub/2800/pg2800.txt>

```
===== Raport=====
Nazwa załadowanego pliku:
Koran.txt

100 najpopularniejszych wyrazów stanowi 59.59 % tekstu
1000 najpopularniejszych wyrazów stanowi 84.64 % tekstu
Słowa stop stanowią 51.78 % tekstu
Liczba liczb występujących w tekście: 5734

10 najpopularniejszych wyrazów charakterystycznych dla Koran.txt :
{'god', 'shall', 'day', 'hath', 'thee', 'lord', 'thou', 'ye', 'said', 'say'}

Najdłuższe wyrazy w Koran.txt :
{'gottesdienstlichen'}
===== Raport=====
```

## The Acts Of The General Assemblies of the Church of Scotland

<https://www.gutenberg.org/files/28957/28957-0.txt>

```
===== Raport=====
Nazwa załadowanego pliku:
law.txt

100 najpopularniejszych wyrazów stanowi 55.51 % tekstu
1000 najpopularniejszych wyrazów stanowi 81.58 % tekstu
Słowa stop stanowią 47.84 % tekstu
Liczba liczb występujących w tekście: 1987

10 najpopularniejszych wyrazów charakterystycznych dla law.txt :
{'us', 'god', '_', 'assembly', 'shall', 'kirk', 'ib', 'upon', 'mr', 'may'}

Najdłuższe wyrazy w law.txt :
{'unsatisfactoriness'}
===== Raport=====
```

## David Copperfield - Charles Dickens

<https://www.gutenberg.org/files/766/766-0.txt>

```
===== Raport=====
Nazwa załadowanego pliku:
dickens.txt

100 najpopularniejszych wyrazów stanowi 57.65 % tekstu
1000 najpopularniejszych wyrazów stanowi 84.38 % tekstu
Słowa stop stanowią 55.02 % tekstu
Liczba liczb występujących w tekście: 186

10 najpopularniejszych wyrazów charakterystycznych dla dickens.txt :
{'said', 'mr', 'could', 'micawber', 'one', 'little', 'upon', 'aunt', 'know', 'would'}

Najdłuższe wyrazy w dickens.txt :
{'retheguidingstarofmyexistence'}
===== Raport=====
```