

Laboratorium 1. Wprowadzenie do Pythona

Laboratorium to ma na celu zapoznanie się z podstawowymi elementami programów napisanych w Pythonie. Ukończenie wszystkich zadań pozwoli na zdobycie takich umiejętności jak odczyt i zapis do plików, stosowanie wyrażeń regularnych oraz wykorzystywanie bibliotek zewnętrznych.

Zadanie 1.1. Odczyt i zapis

Wczytywanie plików tekstowych, wyświetlanie ich zawartości oraz zapis plików to podstawowe umiejętności jakie będą używane w kolejnych laboratoriach.

- Wczytaj plik tekst `.txt`, który został zamieszczony na Eportalu.
- Wykorzystaj do tego `with open`.
- Wyświetl zawartość linia po linii.
- Do nowego pliku tekstowego `tekst_nieparzyste.txt` zapisz nieparzyste linie tekstu.

Dokumentacja:

- `with open`¹

Zadanie 1.2. Wyrażenia regularne

Operacje tekstowe takie jak podział na słowa, zdania, usuwanie określonych znaków, wyszukiwanie pewnych wzorców itp. są częste podczas przetwarzania treści języka naturalnego. Znajomość oraz biegłość w projektowaniu wyrażeń regularnych jest przydatną zdolnością.

1. Za pomocą wyrażenia regularnego podziel tekst wczytany w poprzednim zadaniu na zdania (*nie linie*).
2. Oczyszczyć zdania w taki sposób, aby nie zawierały podwójnych spacji. Wykorzystaj do tego kompilowane wyrażenie regularne.
3. Wylicz ilość słów w całym tekście.
4. Za pomocą wyrażenia regularnego podziel tekst wczytany w poprzednim zadaniu na słowa.
5. Oczyszczyć słowa w taki sposób, aby nie zawierały specjalnych znaków (zostawić znaki diakrytyczne). Wykorzystaj do tego kompilowane wyrażenie regularne.
6. Wylicz ilość zdań w całym tekście.
7. Do pliku JSON zapisz:
 - Ilość zdań
 - Lista oczyszczonych zdań
 - Ilość słów
 - Lista oczyszczonych słów

Dokumentacja:

- `re`²
- `re.compile`³
- `json`⁴

Zadanie 1.3. Wiatraki słów

Ostatnie zadanie ma na celu w sposób praktyczny przećwiczyć umiejętności korzystania z zewnętrznych bibliotek.

- Z utworzonego pliku JSON wczytaj oczyszczone słowa.
- Za pomocą komendy `pip` zainstaluj bibliotekę `wordcloud`.
- Utwórz chmurę w kształcie zadanym przez `maska.jpg` i zapisz do pliku `chmura.png`.
- Utwórz kolejną chmurę w kształcie zadanym przez `maska.jpg`, ale ze wskazaniem słów, które powinny zostać odrzucone tzw. **stop words** i zapisz do pliku `lepsza_chmura.png`.
- Obie chmury wykonaj bez linii konturowej.
- W komentarzu rozwiązania zapisz zaobserwowane różnice pomiędzy obrazami.

Dokumentacja:

- `pip`⁵
- `word cloud with mask`⁶

¹<https://docs.python.org/3/tutorial/inputoutput.html#reading-and-writing-files>

²<https://docs.python.org/3/library/re.html>

³<https://docs.python.org/3/library/re.html#re.compile>

⁴<https://docs.python.org/3/library/json.html>

⁵<https://docs.python.org/pl/3.7/installing/index.html>

⁶https://amueller.github.io/word_cloud/auto_examples/masked.html