

Laboratorium 5 – Poszukiwanie bibliotek o określonej funkcjonalności

Filip Sobczak - Nr albumu: 152793

Wybrana dziedzina: Przetwarzanie języka naturalnego (NLP)

1. Biblioteki użyte w projekcie

1.1 spaCy

- **Przeznaczenie:** profesjonalne przetwarzanie języka naturalnego – tokenizacja, lematyzacja, rozpoznawanie nazwanych jednostek (NER), analiza składniowa.
- **Główne funkcje:**
 - Tokenizacja i POS tagging
 - NER (Named Entity Recognition)
 - Wsparcie wielu języków (w tym język angielski)
 - Obsługa modeli ML i integracja z TensorFlow, PyTorch

Dokumentacja: <https://spacy.io>

Repozytorium GitHub: <https://github.com/explosion/spaCy>

1.2 TextBlob

- **Przeznaczenie:** prostsze i bardziej przystępne narzędzie NLP dla początkujących – analiza sentymentu, wykrywanie języka, tłumaczenie, tagowanie.
- **Główne funkcje:**
 - Analiza sentymentu (polaryzacja, subiektywność)
 - Tagowanie części mowy
 - Tłumaczenie tekstu (w wersji wcześniejszej przez Google Translate – obecnie wymaga alternatywy)

Dokumentacja: <https://textblob.readthedocs.io/en/dev/>

Repozytorium GitHub: <https://github.com/sloria/TextBlob>

1.3 deep-translator

- **Przeznaczenie:** tłumaczenie tekstu między językami przy pomocy różnych silników (np. Google, DeepL, MyMemory).

- **Główne funkcje:**
 - Obsługuje wiele silników tłumaczenia
 - Tłumaczenie z i na wiele języków
 - Działa jako zamiennik dla nieaktywnego `TextBlob.translate()`

Dokumentacja: <https://deep-translator.readthedocs.io/>

Repozytorium GitHub: <https://github.com/nidhaloff/deep-translator>

2. Kluczowe zalety

Biblioteka	Zalety
spaCy	Wydajność, aktualne modele, dobra dokumentacja, szerokie wsparcie NLP
TextBlob	Łatwość użycia, szybka analiza sentymentu, dobre dla początkujących
deep-translator	Wiele źródeł tłumaczeń, prosty interfejs, wsparcie dla wielu języków

3. Ograniczenia

- **spaCy** – brak wbudowanego wsparcia dla języka polskiego w oficjalnych modelach.
- **TextBlob** – działa najlepiej z językiem angielskim. Analiza sentymentu i tłumaczenie nie działa poprawnie z językiem polskim.
- **TextBlob.translate()** – przestało działać z powodu ograniczeń Google Translate, dlatego użyto `deep-translator` jako zamiennika.
- **deep-translator** – wymaga połączenia z internetem do działania.

4. Podsumowanie

Wybrane biblioteki świetnie pokazują różne aspekty przetwarzania języka:

- `spaCy` – profesjonalne NLP,
- `TextBlob` – szybkie i proste operacje tekstowe,
- `deep-translator` – praktyczne tłumaczenia w wielu językach.

Repozytorium projektu zawiera przykładowy skrypt (`nlp_demo.py`) oraz ten raport.