

Narzędzie do półautomatycznej adnotacji tekstu

Zespół nr. 5:

Rafał Budnik

Zuzanna Damszel

Ireneusz Okniński

Opis projektu:

Narzędzie ma na celu ułatwienie procesu adnotacji tekstu poprzez wykorzystanie techniki active learning, gdzie użytkownik interaktywnie adnotuje tekst, a model jest douczany na bieżąco. W ramach projektu użytkownik będzie mógł zaimportować dane w formacie .csv, .json i .jsonl, adnotować tekst, poprawiać zaproponowane adnotacje, zatrzymać i wznowić proces adnotacji, oraz otrzymać statystyki dotyczące modelu i procesu uczenia.

Harmonogram pracy

11.03 - 17.03 - wstępna koncepcja, zarys architektury

18.03 - 24.03 - research na temat Named Entity Recognition oraz Active Learning; prace nad modelem NER

25.03 - 31.03 - zmiana w koncepcji; opracowanie nowej koncepcji projektu

01.04 - 07.04 - zbudowanie wstępnego szkieletu aplikacji desktopowej, prace nad UI

08.04 - 14.04 - dodawanie własnych labeli do adnotacji, prace nad UI, ukończenie prototypu aplikacji

15.04 - 21.04 - eksport zaadnotowanych danych, integracja aplikacji z faktycznym modelem, wyświetlanie statystyk uczenia

22.04 - 28.04 - testowanie funkcjonalności/testowanie różnych modeli

29.04 - 05.05 - majówka: brak postępów

06.05 - 12.05 - prototyp nr 2, zgłoszenie projektu do JOSS

13.05 - 19.05 - poprawki

20.05 - 26.05 - poprawki

27.05 - 02.06 - finalna wersja projektu

Funkcjonalności

- Wczytanie zbioru danych przez użytkownika (formaty .csv, .json, .jsonl)
- Wypisanie adnotacji (labels), które mają być oznaczone w tekście
- Iteracyjne adnotowanie tekstu przez użytkownika oraz douczanie modelu
- Ręczna poprawa adnotacji zaproponowanych przez model
- Zapis zaadnotowanych danych do bazy plików
- Eksport danych do pliku (formaty .csv, .json, .jsonl)

- Wyświetlanie statystyk dotyczących modelu i procesu uczenia
- Umożliwienie użytkownikowi wyboru modelu z listy kilku modeli

Stack technologiczny

Aplikacja desktopowa

- Kivy

Model

- FastAPI
- PyTorch