

# 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 - zapoznanie się z kodem doccano i jego dokumentacją, podział zadań, przygotowanie środowiska programistycznego, skrypty budujące i testujące  
25.03 - 31.03 - kontynuacja prac nad adaptacją doccano do naszych potrzeb, importowanie własnych modeli, pierwsze prace nad UI, przypisywanie modelu AI do projektu w doccano  
01.04 - 07.04 - prace nad UI, eksport danych do jednego z formatów plików, importowanie własnych modeli, implementacja automatycznego przesyłania informacji o adnotacji do modelu  
08.04 - 14.04 - implementacja statystyk uczenia, ukończenie prototypu aplikacji  
15.04 - 21.04 - możliwość podpięcia zewnętrznego modelu poprzez API/z pliku  
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

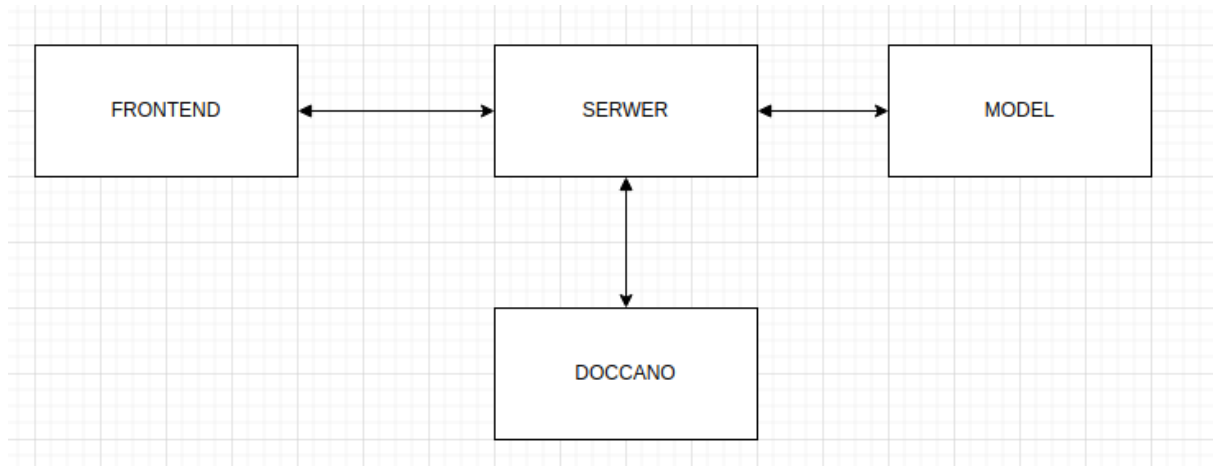
## Funkcjonalności

- 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

- Eksport danych do pliku (formaty .csv, .json, .jsonl)
- Wyświetlanie statystyk dotyczących modelu i procesu uczenia
- Umożliwienie użytkownikowi wyboru własnego modelu oraz korzystania z modelu domyślnego

## Stack technologiczny

Architektura mikroservisowa - komponenty jako oddzielne kontenery dockera.



### Frontend

- Vue.js

### API (Serwer)

- FastAPI

### Model

- FastAPI
- PyTorch