

W ramach naszego projektu, głównym wyzwaniem jest skuteczna identyfikacja i klasyfikacja nazwanych jednostek (NER) w tekstach. Nazwane jednostki obejmują różnorodne kategorie, takie jak osoby, miejsca, daty, organizacje, co stanowi kluczową informację w analizie tekstu. Obecnie brakuje precyzyjnych narzędzi NER, co prowadzi do utraty danych i komplikuje analizę tekstu, zwłaszcza w kontekście zróżnicowanych dziedzin.

Integracja Istniejących Modeli NER:

Głównym celem naszego rozwiązania jest skuteczna integracja istniejących modeli NER, które są wydajne i precyzyjne w identyfikacji nazwanych jednostek w różnych kontekstach.

Dopasowanie do Specyfiki Dziedziny:

Rozwiązanie powinno być elastyczne i dostosowujące się do różnych dziedzin, aby skutecznie obsługiwać teksty z różnych źródeł i branż.

Wybór modeli:

Do projektu zostaną wykorzystane trzy modele, dzięki którym można wykonać klasyfikacje nazwanych jednostek.

1. [Bert-base-NER](#)

Opis: model dostosowany do klasyfikacji nazwanych jednostek do znajdowania lokacji, organizacji, osób i innych różnorodnych rzeczy

Zasadność: model bardzo dobrze nadaje się do tego typu zadań

2. [SpaCy](#)

Opis: model o wytrzymałości przemysłowej odpowiedni do ekstrakcji informacji na dużą skalę.

Zasadność: model jest bardzo łatwy w użyciu i reklamuje się na szybki i efektywny

3. [ChatGPT](#)

Opis: jest łatwo dostępny i nie wymaga specjalistycznej wiedzy w dziedzinie modelowania języka naturalnego.

Zasadność: rozumie kontekst językowy i może przetwarzać bardziej skomplikowane pytania dotyczące imion i nazwisk w kontekście danego tekstu.