Hallucination Nation

Fine-tuning HerBERT-Large i porównanie z metodami regułowymi w anonimizacji danych

O PROJEKCIE

- **Glowny cel:** Konsorcjum rozwijające polskojęzyczny model PLLuM potrzebuje precyzyjnej anonimizacji dużych zbiorów tekstowych, ponieważ dotychczasowe metody nie radzą sobie ze złożonością języka polskiego, a zgodność z RODO wymaga usunięcia Danych Osobowych bez utraty sensu i struktury tekstów.

- **Przyklad anonimizacji:** Nazywam się {name} {surname}, mieszkam w {address}, a mój numer PESEL to {pesel}.

- **Kluczowe wyzwania:** Kluczowymi wyzwaniami są zmienność form zapisu danych wrażliwych, konieczność rozpoznawania kontekstu zamiast sztywnych reguł oraz brak wyspecjalizowanych modeli NER dla anonimizacji w języku polskim.

- **Output:** Komponent (biblioteka Python/skrypt) anonimujący tekst poprzez automatyczne podmienianie danych wrażliwych na odpowiednie tokeny zastępcze.

NASZE PODEJSCIE

{01}

RESEARCH

Przegląd literatury naukowej (Google Scholar), aby określić istniejące podejścia do anonimizacji

{02}

WSTEPNE PRZETWARZANIE

Wybor metryk oceny i rozszerzanie danych

{03}

PROGRAMOWANIE

Równoległy rozwój rozwiązań

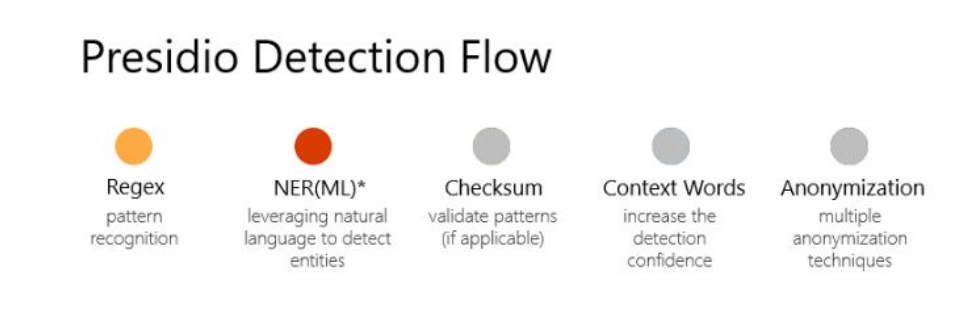
{04}

WALIDACJA I PODSUMOWANIE

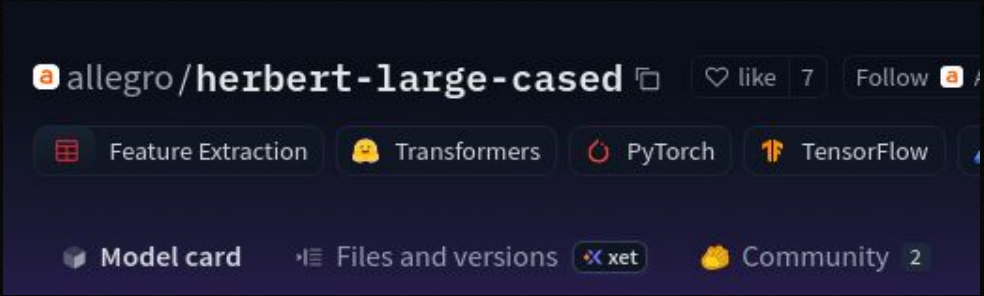
Walidacja i podsumowanie rezultatów

IMPLEMENTACJA

- **Metoda regulowa wspomagana AI-** rozwiazanie w oparciu o biblioteke Microsoft Presidio, z dodatkowymi regulami + deanonimizacja syntetyczna



- **fine-tuning HerBERT-Large do NER-** augmentacja danych + fine-tuning HerBERT'a



REKOMENDACJE

Podejscie regulowe

Niskie wymagania obliczeniowe – brak potrzeby trenowania modeli, szybka i tania inferencja. {01}

Wymaga ręcznego utrzymywania wzorców, co zmniejsza skalowalność. {02}

Oparte na regułach i słownikach, przez co ma ograniczoną zdolność rozumienia kontekstu. {03}

Fine-tuning HerBERT-Large do NER

Wymaga dużej liczby ręcznie anotowanych danych, co jest kosztowne i czasochłonne. {01}

Model jest ciężki obliczeniowo, przez co trening i inferencja mogą być wolne oraz wymagać mocnej infrastruktury. {02}

Bardzo wysoka jakość rozpoznawania encji dzięki dostosowaniu modelu do specyfiki języka polskiego. {03}