Nazwa wyzwania

Odwołania2json
Od dokumentów do danych. Inteligentna transformacja odwołań KIO do kanonicznego modelu JSON

Krótki opis wyzwania

Każdego roku do systemu zamówień publicznych trafiają tysiące dokumentów zawierających kluczowe informacje o postępowaniach, wykonawcach, pełnomocnikach i czynnościach zamawiających. Najczęściej są to dokumenty tekstowe z kluczowymi, jednak ułożonymi w przeróżny sposób, informacjami z reguły na pierwszych trzech stronach. Celem wyzwania jest stworzenie rozwiązania, które automatycznie zamieni wielodokumentową sprawę administracyjną w spójny, ustrukturyzowany model danych JSON gotowy do dalszego wykorzystania przez systemy informatyczne administracji publicznej.

1. Wprowadzenie

Urząd Zamówień Publicznych odpowiada za funkcjonowanie systemu zamówień publicznych w Polsce oraz wspiera uczestników rynku w prawidłowym stosowaniu przepisów prawa. Jednym z elementów tego systemu są postępowania odwoławcze prowadzone przed Krajową Izbą Odwoławczą.

Sprawa odwoławcza nie jest pojedynczym dokumentem. Stanowi zbiór wielu powiązanych ze sobą dokumentów zawierających informacje o wykonawcach, zamawiających, pełnomocnikach, postępowaniach, płatnościach, czynnościach procesowych i innych elementach wymagających dalszego przetwarzania.

Obecnie większość takich dokumentów występuje w postaci plików PDF o zróżnicowanej strukturze. Część z nich powstaje elektronicznie, część jest wynikiem skanowania dokumentów papierowych. Informacje niezbędne do dalszej obsługi sprawy są rozproszone pomiędzy wieloma dokumentami i często zapisane w różny sposób.

Dla człowieka zrozumienie takiej dokumentacji jest stosunkowo proste. Dla systemów informatycznych stanowi jednak duże wyzwanie. Aby możliwa była automatyzacja dalszych procesów, dane muszą zostać najpierw przekształcone do postaci ustrukturyzowanej.

W praktyce oznacza to konieczność przejścia od modelu:

wiele dokumentów PDF → analiza treści → identyfikacja informacji → model danych JSON

Wyzwanie dotyczy zatem budowy rozwiązania, które pozwoli automatycznie przekształcać wielodokumentowe sprawy administracyjne w kanoniczny model danych możliwy do dalszego wykorzystania przez systemy administracji publicznej.

Najciekawsze rozwiązania mają szansę stać się podstawą pilotażowych wdrożeń realizowanych przez Urząd Zamówień Publicznych.

2. Wyzwanie

Zaprojektuj i zaimplementuj rozwiązanie, które:

1. przyjmie dokument PDF składający się na odwołanie: dokument z treścią odwolania, pełnomocnictwo, dokument potwierdzenia przelewu
2. rozpozna rodzaj dokumentu
3. wyodrębni informacje biznesowe istotne dla danego typu dokumentu,
4. wygeneruje JSON dla każdego dokumentu zgodny z https://github.com/Urzad-Zamowien-Publicznych/hacknation-2026/blob/main/oczekiwane-json.

Zestawy treningowe dostępne są na GitHub: https://github.com/Urzad-Zamowien-Publicznych/hacknation-2026.

3. Oczekiwany rezultat

Oczekiwanym rezultatem jest działająca aplikacja pobierająca plik PDF i zwracająca plik JSON w ustalonym formacie. Więcej szczegółów znajdziesz w naszym repozytorium GitHub:

https://github.com/Urzad-Zamowien-Publicznych/hacknation-2026

4. Wymagania formalne

Projekt przesyłany do oceny powinien zawierać:

- tytuł projektu,
- Link do GitHub działającego prototypu umożliwiający wgranie dokumentów i odczytanie rezultatów ekstrakcji informacji.

Dodatkowo uczestnicy mogą dostarczyć:

- dokumentację techniczną,
- wyniki własnych testów jakości.

5. Wymagania techniczne

Rozwiązanie musi:

- wykorzystywać wyłącznie komponenty open source (w tym otwarto-źródłowe modele językowe),
- umożliwiać uruchomienie w środowisku lokalnym (on-premise) bez GPU (LLM po API).

Dodatkowo premiowane będą rozwiązania:

- łatwo rozszerzalne o nowe klasy dokumentów / umożliwiające konfigurację nowych klas dokumentów bez modyfikacji kodu źródłowego,
- wspierające transparentność procesu ekstrakcji danych.

6. Sposób testowania i walidacji

Ocena zostanie przeprowadzona na dwóch zbiorach danych:

Zbiór treningowy

Udostępniony uczestnikom podczas realizacji wyzwania.

Zbiór testowy

Nieudostępniony uczestnikom. Zbiór testowy będzie zawierał realne sprawy odwoławcze przygotowane przez organizatora.

Podstawową miarą jakości będzie wskaźnik F1 obliczony na zbiorze testowym.

7. Dostępne zasoby

Organizator udostępni:

- zestawy dokumentów syntetycznych,
- specyfikację oczekiwanego modelu danych.

Szczegółowe dostępne na GitHub.

8. Kryteria oceny

Etap 1 - preselekcja:

1. Jakość ekstrakcji danych (F1 Score) – 100%

Ocena zgodności wyników z danymi referencyjnymi na nieznanym zbiorze testowym.

Etap 2 - wybór finalny:

1. Jakość ekstrakcji danych (F1 Score) – 50%

Ocena zgodności wyników z danymi referencyjnymi na nieznanym zbiorze testowym.

2. Elastyczność i możliwość rozbudowy rozwiązania – 30%

Łatwość dodawania nowych typów dokumentów i nowych klas danych bez konieczności przebudowy całego systemu.

3. Architektura i jakość techniczna – 20%

Przejrzystość projektu, modularność, utrzymanie i dokumentacja.

9. Dodatkowe uwagi i kontekst wdrożeniowy

Poszukiwane są rozwiązania, które mogą stanowić podstawę dalszych prac wdrożeniowych i pilotażowych w UZP.

Szczególną wartość będą miały rozwiązania umożliwiające budowę uniwersalnego narzędzia do pozyskiwania ustrukturyzowanych danych z dokumentów PDF.

10. Kontakt

Podczas wydarzenia uczestnicy będą mogli korzystać ze wsparcia mentorów technicznych oraz ekspertów merytorycznych z zakresu zamówień publicznych.

Szczegółowe kanały komunikacji oraz harmonogram dyżurów mentorów zostaną przekazane uczestnikom przed rozpoczęciem hackathonu.