Cyfryzacja formularzy

Ułatw urzędnikom automatyczne tworzenie cyfrowych formularzy dla obywateli – aby skrócić czas załatwiania spraw i usprawnić korzystanie z cyfrowych rozwiązań administracji publicznej również na szczeblu samorządowym. Razem łatajmy dziury w cyfrowych procesach – jeden formularz at a time!

1. Wprowadzenie - opis organizacji, sytuacji i stanu aktualnego

Centralny Ośrodek Informatyki zajmuje się rozwijaniem usług cyfrowych, których głównym celem jest ułatwienie obywatelom załatwiania spraw urzędowych drogą elektroniczną. Kluczowym wyzwaniem w tym obszarze jest zamiana tradycyjnych, analogowych formularzy na ich w pełni cyfrowe odpowiedniki.

Obecnie urzędy udostępniają na swoich stronach internetowych oraz w Biuletynach Informacji Publicznej formularze w formatach takich jak PDF, DOC, DOCX, ODS, XLSX, a także skany w postaci plików graficznych JPG, JPEG, PNG. Udostępniane formularze zawierają niezbędne dane, które należy podać, aby załatwić sprawę oraz kluczowe informacje co do samego procesu załatwianej sprawy. Obywatel musi taki formularz wydrukować, wypełnić odręcznie, a następnie zanieść do urzędu lub odesłać w formie pliku bądź skanu. Ponieważ formularze te są projektowane z myślą o wypełnianiu ręcznym, zawierają specyficzne dla tej formy elementy i są pozbawione udogodnień, jakie oferują rozwiązania cyfrowe.

W praktyce oznacza to, że cyfryzacja każdego formularza wymaga wieloetapowej analizy. Należy przede wszystkim poznać proces, w ramach którego formularz realizuje jakieś postępowanie administracyjne, następnie zidentyfikować we wniosku niezbędne dane i zależności między nimi, określić elementy do uproszczenia i optymalizacji, oraz przeprojektować treść na wersję elektroniczną. Konieczne jest również zaplanowanie walidacji wprowadzanych przez użytkownika danych oraz w razie potrzeby integracji z systemami zewnętrznymi.

Proces ten jest niezwykle czasochłonny i kosztowny. Utrudnia go ogromna różnorodność dokumentów stosowanych przez urzędy oraz brak jednolitych standardów ich projektowania. Nawet formularze dotyczące tego samego postępowania administracyjnego potrafią znacząco różnić się zakresem wymaganych danych czy sposobem ich prezentacji w zależności od urzędu. W efekcie szybkie tworzenie formularzy elektronicznych jest mocno utrudnione, a większość z nich wymaga indywidualnego podejścia i oddzielnej analizy.

Dynamiczny rozwój sztucznej inteligencji stwarza możliwość automatyzacji tego procesu. Nowoczesne modele potrafią analizować dokumenty, rozpoznawać ich strukturę i poszczególne pola, a w rezultacie przekształcać nieustrukturyzowane treści w uporządkowane modele danych. Dzięki temu możliwe staje się znaczne skrócenie czasu potrzebnego na przygotowanie formularza elektronicznego na bazie formularza udostępnionego przez urząd oraz zminimalizowanie nakładu pracy manualnej.

2. Wyzwanie

COI poszukuje rozwiązania, które wykorzysta AI do automatycznej analizy formularzy w różnych formatach i generowania na ich podstawie plików zawierających dane w zdefiniowanej w wymaganiach postaci ustrukturyzowanej, będących podstawą tworzonych formularzy elektronicznych.

Tak przygotowane dane pozwolą na budowę nowoczesnych formularzy elektronicznych, które w przyszłości będą mogły być obsługiwane również przez agentów AI. Docelowo wypracowane rozwiązanie może znaleźć zastosowanie również w innych wspieranych sztuczną inteligencją projektach, realizowanych przez COI na rzecz cyfryzacji polskiej administracji publicznej.

---

3. Oczekiwany rezultat

3.1. Na potrzeby wyzwania zostały przygotowane:

I. Przykładowe pliki z formularzami Wniosków w kilku formatach, o zróżnicowanych poziomach złożoności i wielkości oraz dla wybranych formularzy odpowiadające im pliki JSON, będące wynikiem analizy tych formularzy i screeny przykładowej eUsługi;
II. Formularze w kilku formatach, o zróżnicowanych poziomach złożoności i wielkości, dla których rozwiązanie przygotowane w ramach tego Wyzwania powinno wygenerować pliki JSON - podlegające późniejszej ocenie;
III. Wymagania opisujące sposób projektowania i tworzenia formularzy elektronicznych, które projektowane rozwiązanie powinno uwzględnić podczas generowania wynikowych plików JSON oraz założenia techniczne konfiguracji formularzy;
IV. Opis słowników oraz danych inicjalnych, możliwych do użycia w formularzach;

3.2. Minimalny zakres (obowiązkowy) umożliwiający ocenę przygotowanego rozwiązania:

I. Usługa udostępniona przez API REST, która przyjmuje pliki formularzy we wskazanych formatach (PDF, DOC, DOCX, ODS, XLSX, a także skany w postaci plików graficznych JPG, JPEG, PNG) i zwraca strukturę danych w formacie JSON:
    - odpowiadającą zawartości zaimportowanego formularza (dane wejściowe stanowią formularze z punktu 3.1.II),
    - zgodną ze strukturą oczekiwaną przez aplikację (oczekiwany plik wynikowy znajduje się w danych wejściowych do wyzwania w punkcie 3.1.I),
    - spełniającą wymagania projektowania formularzy (np. Wydzielenie stron tematycznych w formularzu - punkt 3.1.III)
II. Dokumentacja architektury rozwiązania, szerzej opisana w punkcie 5. Wymagania techniczne.

3.3. Opcjonalne rozszerzenia (dodatkowo punktowane)

I. Skorzystanie z bazy słowników, danych inicjalnych i powiązanie pasujących słowników oraz danych inicjalnych z odpowiadającymi im polami w JSONie.
II. Zaproponowanie nowego słownika, jeżeli w bazie danych wejściowych brakuje odpowiedniego słownika, do zasilania wybranych pol w formularzu elektronicznym.
III. Określenie zależności między polami, sekcjami i stronami formularza. Wskazanie dynamicznych przepływów między polami, sekcjami i stronami formularza, zależnych od wyboru użytkownika na formularzu.
IV. Prosty frontend umożliwiający załadowanie plików z formularzami do usługi w interfejsie graficznym i przeglądanie wyjściowej struktury danych (jsona).
V. Aplikacja do wizualizacji wynikowych plików json – umożliwiająca zaprezentowanie formularzy eUsług powstałych na bazie plików json
VI. Wszelkie ułatwienia z perspektywy weryfikacji i walidacji danych (np. typ danych, długość pola, walidacja numeru telefonu, numeru PESEL itp.), jak również UX (np. przetwarzanie zbiorowe dokumentów – batch, makiety interfejsu rozwiązania).

Docelowym użytkownikiem rozwiązania są urzędnicy zajmujący się cyfryzacją formularzy e-usług publicznych, w szczególności przenoszeniem tradycyjnych formularzy papierowych do formy cyfrowej.

---

4. Wymagania formalne

Projekt przesyłany do oceny powinien zawierać:

- szczegółowy opis i tytuł projektu,
- prezentację w formacie PDF (maksymalnie 10 slajdów),
- wynikowe pliki JSON przedstawiające strukturę formularzy,
- film umieszczony w dostępnym, otwartym repozytorium (link), trwający maksymalnie 3 minuty i prezentujący projekt,
- repozytorium kodu

Dodatkowo może zawierać:

- zrzuty ekranu,
- linki do demonstracji,
- materiały graficzne lub inne elementy związane z projektem.

---

5. Wymagania techniczne

Język programowania (preferowane)

- Java 21 lub nowsza
- Python 3.14 lub nowszy

Biblioteki i frameworki

- wyłącznie rozwiązania open source, nie oparte na licencjach typu copyleft (wymagających upublicznienia całości lub części kodu) np. AGPL, GNU-GPL
- Brak użycia zależności takich jak zewnętrzne API, MCP, SaaS itp.
- W szczególności rozwiązanie nie może używać modeli LLM dostępnych wyłącznie przez API (ale można ich używać do pisania/testowania/projektowania kodu)
- brak wymogu zakupu licencji komercyjnych
- preferowane najnowsze stabilne wersje bibliotek i frameworków

REST API

Aplikacja musi udostępniać REST API umożliwiający przesłanie pojedynczego pliku będącego blankietem formularza.

Obsługiwane formaty wejściowe

- PDF (.pdf)
- DOC (.doc .docx)
- ODS (.ods)
- XLSX (.xlsx)
- JPEG (.jpg .jpeg)
- PNG (.png)

Wynik przetwarzania

W odpowiedzi aplikacja zwraca dokument JSON zawierający konfigurację formularza przygotowaną do importu w systemie docelowym.

Walidacja typu pliku

System musi akceptować wyłącznie obsługiwane formaty plików wraz z weryfikacją MIME type w celu potwierdzenie, że zawartość pliku pokrywa się z jego rozszerzeniem

Walidacja nazwy pliku

Nazwa przesyłanego pliku:

- nie może przekraczać 250 znaków
- nie może zawierać następujących znaków ; : > < / \ | ' * % $

Walidacja rozmiaru pliku

Maksymalny rozmiar przesyłanego pliku musi być:

- konfigurowalny
- definiowany poprzez parametry aplikacji

Przechowywanie plików

Dopuszczalne jest wyłącznie tymczasowe przetwarzanie pliku w pamięci operacyjnej lub przestrzeni tymczasowej wykorzystywanej podczas realizacji żądania.

Konteneryzacja

Rozwiązanie musi być dostarczone wraz z kompletną konfiguracją Docker umożliwiającą uruchomienie aplikacji w odizolowanym środowisku.

Wymagania dla obrazów Docker

Należy stosować następujące zasady:

- wykorzystywać oficjalne obrazy bazowe
- preferować lekkie dystrybucje (np. Alpine Linux)
- jawne wersjonowanie
- nie używać tagu latest
- uruchamiać proces aplikacji na dedykowanym użytkowniku posiadającym ograniczone uprawnienia
- udostępniać wyłącznie porty wymagane do działania aplikacji
- dostarczać sekrety oraz dane wrażliwe wyłącznie na etapie uruchamiania kontenera (runtime), np. poprzez zmienne środowiskowe lub mechanizmy zarządzania sekretami

Konfiguracja aplikacji

Aplikacja powinna umożliwiać konfigurację minimalnego niezbędnego zestawu parametrów uruchomieniowych, w tym między innymi:

- maksymalnego rozmiaru pliku
- parametrów związanych z uwierzytelnianiem JWT
- ewentualnych parametrów integracyjnych wymaganych przez rozwiązanie używanych do połączenia lub konfiguracji zasobów/bibliotek

Dokumentacja

Wraz z rozwiązaniem należy dostarczyć plik README.md zawierający:

- instrukcję budowania obrazu Docker
- instrukcję uruchomienia aplikacji
- opis konfiguracji
- opis REST API

6. Sposób testowania i/lub walidacji

Finaliści zostaną poproszeni o uruchomienie rozwiązania na podstawie nowych plików, które zostaną przekazane podczas obrony rozwiązania - aby zweryfikować jego poprawność.

---

7. Dostępne zasoby

Przygotowane materiały znajdują się na discordzie na kanale wyzwania Cyfryzacja formularzy/materiały-dodatkowe.

Na potrzeby wyzwania zostały przygotowane:

- Przykładowe pliki z formularzami Wniosków w kilku formatach, o zróżnicowanych poziomach złożoności i wielkości oraz dla wybranych formularzy odpowiadające im pliki JSON, będące wynikiem analizy tych formularzy i screeny przykładowej eUsługi;
- Formularze w kilku formatach, o zróżnicowanych poziomach złożoności i wielkości, dla których rozwiązanie przygotowane w ramach tego Wyzwania powinno wygenerować pliki JSON - podlegające późniejszej ocenie;
- Wymagania opisujące sposób projektowania i tworzenia formularzy elektronicznych, które projektowane rozwiązanie powinno uwzględnić podczas generowania wynikowych plików JSON oraz założenia techniczne konfiguracji formularzy;
- Opis słowników oraz danych inicjalnych, możliwych do użycia w formularzach;
- Tokeny do wykorzystania w pracy nad rozwiązaniem (dla pierwszych 20 zespołów)

---

8. Kryteria oceny

- Kompletność projektu (osiągnięcie kolejnych progów) — 30%:
    - Dostarczony wymagany minimalny zakres
    - Dostarczone opcjonalne rozszerzenia:
        - Punkt I (Słowniki i dane inicjalne)
        - Punkt II (Nowy słownik)
        - Punkt III (Dynamiczne zależności między obiektami)
        - Punkt IV (Frontend do importu plików)
        - Punkt V (Walidacje pól)
- Intuicyjność rozwiązania — 20%
- Oryginalność — 10%
- Jakość kodu — 20%
- Potencjał wdrożeniowy (w tym koszt rozwiązania) — 20%

---

9. Dodatkowe uwagi / kontekst wdrożeniowy

Najlepsze rozwiązanie może zostać wdrożone w ramach projektów zaplanowanych przez COI na najbliższe lata (Generator Formularzy/AI HUB/ Multiportal). W przypadku chęci

wykorzystania wypracowanego rozwiązania możliwe jest zaproszenie zwycięskiego zespołu do współpracy w toku prac nad wdrażanymi rozwiazaniami.

---

10. Kontakt

Mentorzy COI dostępni są podczas wydarzenia na stoisku COI oraz online, za pośrednictwem kanałów wskazanych przez organizatora (np. Discord – analogicznie do 2025).
