
API JSA

dla referencyjnych baz instytucji

Spis treści

1.	1. Informacje ogólne.....	3
1.1.	1.1. Metryka dokumentu.....	3
1.2.	1.2. Historia zmian.....	3
1.3.	1.3. Słownik	3
2.	2. Założenia biznesowe i opis procesu	4
3.	3. Metody HTTP	5
4.	4. Protokół komunikacyjny	5
5.	5. Uwierzytelnianie i autoryzacja	5
6.	6. Format danych i komunikacja.....	5
6.1.	6.1. Komunikacja JSA -> Referencyjna baza instytucji.....	6
6.2.	6.2. Komunikacja Referencyjna baza instytucji -> JSA.....	6
7.	7. Scenariusze	7
7.1.1.	7.1.1. Zapytanie o wytypowanie dokumentów bazowych	7
7.1.2.	7.1.2. Przekazanie dokumentów bazowych	7
8.	8. Uwagi i przykłady.....	7

1. 1. Informacje ogólne

1.1. Metryka dokumentu

Właściciel	OPI PIB		
Autor			
Zatwierdził		Data Zatwierdzenia	
Wersja	1.0	Status	
Data utworzenia	10.10.2019	Data ostatniej modyfikacji	10.10.2019

1.2. Historia zmian

Lp	Autor	Data	Zakres	Opis
1	Ewa Dulińska / Małgorzata Wartacz / Daniel Karaś	10.10.2019	Utworzenie dokumentu	

1.3. Słownik

API – zestaw reguł definiujący komunikację pomiędzy programami komputerowymi.

Baza danych – baza zawierająca treści, które mają być przeszukiwane podczas badania antyplagiatowego w JSA, zgłoszonego w kontekście instytucji.

Baza instytucji – referencyjna baza porównawcza przygotowana przez Instytucję i powiązana z JSA poprzez REST API.

Dokumenty bazowe – dokumenty zwracane przez bazę instytucji, które zostały przez nią wytypowane jako materiał bazowy do badania antyplagiatowego danej pracy.

Instytucja – właściciel bazy danych oraz podmiot, w kontekście którego zgłaszane jest badanie antyplagiatowe.

JSA – Jednolity System Antyplagiatowy.

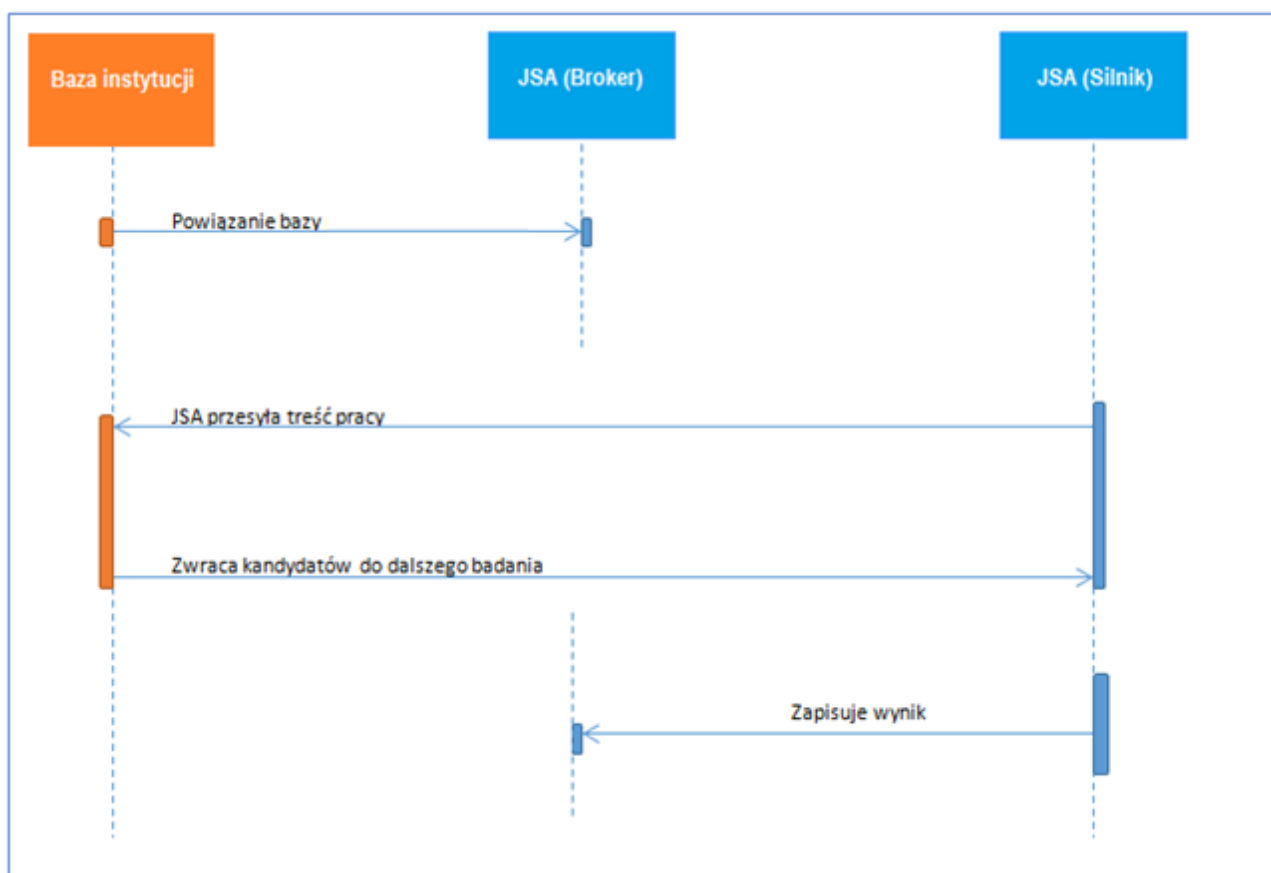
Nota – oznaczenie numeryczne będące reprezentacją mechanizmu przypisywania tekstom oceny stopnia podobieństwa. Wartość pola „score”, nieobligatoryjna.

UI – interfejs użytkownika, dostęp do systemu za pośrednictwem przeglądarki internetowej.

2. Założenia biznesowe i opis procesu

Biznesowe założenie JSA przewiduje możliwość podłączenia wewnętrznych zasobów instytucji do systemu JSA jako dodatkową bazę referencyjną w badaniu antyplagiatowym. Baza wykorzystywana jest tylko na potrzeby instytucji, która wskazuje ją w panelu administracyjnym JSA. Dane zebrane w bazie wykorzystywane są tylko podczas badań antyplagiatowych zgłoszonych w JSA w kontekście instytucji i nie są przekazywane żadnemu innemu podmiotowi korzystającemu z JSA. Komunikacja z JSA odbywa się przez API i zabezpieczona jest przy pomocy tokenu autoryzującego. Podczas całego procesu baza danych znajduje się na serwerze instytucji. Integracja bazy z JSA jest niezależna od typu czy struktury bazy danych.

Adres, pod którym znajduje się baza danych należy wskazać w panelu administracyjnym JSA w Ustawieniach w polu „Referencyjna baza instytucji”. Po prawidłowym powiązaniu bazy i założeniu nowego badania w JSA (poprzez UI lub API), system antyplagiatowy wysyłać będzie do bazy instytucji treść pracy w celu wytypowania tekstów, mogących być źródłem podobieństwa (dalej zwanych dokumentami bazowymi). Każdy z dokumentów bazowych może być opatrzony oceną stopnia podobieństwa (dalej zwaną notą), która wskaże stopień podobieństwa z tekstem badanego pliku. Dokumenty bazowe nie są zapisywane w całości po stronie bazy danych JSA. Zapisowi podlegają jedynie fragmenty wskazujące na wykryte podobieństwo, co związane jest z mechaniką przedstawiania źródła wykrytego podobieństwa.



Rysunek 1 Przepływ komunikatów JSA-instytucja

3. Metody HTTP

Do komunikacji pomiędzy JSA a serwerem instytucji wykorzystywana jest metoda POST.

1. dla przekazania treści badanej pracy z JSA do bazy instytucji,
2. dla odpowiedzi z serwera instytucji dokumentów bazowych do dalszego badania w JSA.

4. Protokół komunikacyjny

Adres końcówki REST dla bazy instytucji odpowiedzialnej za przyjmowanie i odpowiadanie na zapytania JSA należy wprowadzić w JSA w pole „Referencyjna baza instytucji” znajdujące się w panelu administracyjnym Ustawienia w zakładce Ogólne.

Wymiana danych pomiędzy bazą instytucji a JSA prowadzona jest w formacie JSON. Podstawowym elementem tego formatu jest pole, którego nazwa jest oddzielona od wartości pola dwukropkiem. Pola rozdziela przecinek. Kolejnym elementem jest obiekt JSON, zawierający się w nawiasach klamrowych. Tablice w formacie JSON są zapisywane z wykorzystaniem nawiasów kwadratowych.

5. Uwierzytelnianie i autoryzacja

Powiązanie bazy instytucji z JSA możliwe jest tylko poprzez protokół HTTPS. W celu zapewnienia poprawnej autentykacji należy stworzyć unikalny zestaw tokenów dla instytucji udostępniającej swoją bazę oraz dla JSA. Dostęp do tokenów ma tylko administrator JSA. Kod autoryzujący instytucję, czyli wartość pola „Kod autoryzacji instytucji” wygenerowany jest automatycznie i nie podlega edycji. Administrator musi nadać dla JSA unikalny token i wpisać go w pole „Kod autoryzacji JSA”.

Autentykacja odbywa się poprzez wymianę wyżej wymienionych tokenów w nagłówkach zapytań i odpowiedzi, w polu „**X-AUTH-TOKEN**”. System JSA tworząc zapytanie umieszcza w nagłówku „Kod autoryzacji JSA” natomiast w odpowiedzi zwrotnej (w nagłówku) oczekuje wartości „Kodu autoryzacji instytucji”.

UWAGA! Każdy token identyfikuje bezpośrednio instytucję. Zaleca się aby JSA także było rozpoznawane po stronie instytucji za pomocą unikalnego tokenu.

6. Baza instytucji

Baza instytucji powinna zostać przygotowana w podobny sposób jak wewnętrzne bazy referencyjne JSA (sugerowane narzędzie do indexu bazy to Elasticsearch). Każdy z dokumentów był zapisany w postaci ciągu znaków (kodowanych w UTF8) reprezentującego zawartość tekstu (plain text) opatrzony był co najmniej dwiema danymi: ID dokumentu oraz tytułem dokumentu.

7. Format danych i komunikacja

7.1. Komunikacja JSA -> Referencyjna baza instytucji (żądanie)

JSA przesyła do referencyjnej bazy instytucji zapytanie typu POST, którego treść (body) stanowi pełna treść badanej pracy w postaci ciągu znaków (kodowanych w UTF8) reprezentującego wyciągnięty tekst (plain text) badanego pliku. Na jego podstawie baza instytucji powinna wytypować dokumenty bazowe w postaci tekstowej. Będą one stanowić uzupełnienie zbioru referencyjnych baz systemu JSA (Baza aktów prawnych, Internet, ORPPD).

Nieobligatoryjne jest przygotowanie mechanizmu przypisywania tekstom oceny stopnia podobieństwa (dalej zwanej notą). Jest to oznaczenie numeryczne, które informować będzie o jakości podobieństwa wytypowanego dokumentu. Jeżeli instytucja zdecyduje się na uwzględnianie noty należy pamiętać o tym, aby sposób przyznawania not dokumentom był jednolity, tak, że wyższa wartość w polu „score” oznacza dokument bazowy o większej wartości badawczej.

Spośród nadesłanych dokumentów bazowych system JSA wybiera 30 dokumentów bazowych o najlepszych notach. Jeżeli noty nie zostaną przesłane to do badania zostanie wziętych pierwszych 30 dokumentów bazowych.

7.2. Komunikacja Referencyjna baza instytucji -> JSA (odpowieź)

JSA będzie oczekiwało na odpowiedź bazy instytucji przez określony czas. W przypadku niedostarczenia odpowiedzi zwrotnej, braku połączenia z bazą lub przekroczenia czasu połączenie zostanie zerwane (time-out połączenia po 90 sekundach), badanie otrzyma status „Błąd”, a następnie będzie procedowane zgodnie z tym statusem.

Wszystkie dane wskazane w odpowiedzi przez tę bazę będą widoczne w UI JSA w raporcie z badania w widoku Źródła podobieństwa.

W odpowiedzi na zapytanie ze strony JSA, referencyjna baza instytucji wysyła tablicę w formacie JSON zawierającą obiekty o następującym formacie:

Odpowiedź bazy instytucji

Nazwa	Typ	Wymagalność	Opis
metadata	(poniżej w JSON)	Tak	Metadane dokumentów bazowych
Text	String	Tak	Treść dokumentu
Score	String	Nie	Atrybut nieobligatoryjny przypisujący tekstom oceny stopnia podobieństwa

Metadane w JSON

Nazwa	Typ	Wymagalność	Opis
Id	String	Tak	Atrybut jednoznacznie identyfikujący zwrócony dokument w bazie
Title	String	Tak	Tytuł dokumentu
<dowolna> <i>np. Autor, Nazwisko</i>	String	Nie	Jedno lub więcej pól o dowolnej nazwie, zagnieżdżonych w polu „meta”, których wartości zostaną wyświetlone na raporcie w sekcji „Źródło podobieństwa”

UWAGA: Zwracana lista dokumentów bazowych powinna być zamknięta. W przypadku niezalezienia dokumentów bazowych należy zwrócić pustą listę.

8. Scenariusze

8.1.1. Zapytanie o wytypowanie dokumentów bazowych

1. JSA przesyła zapytanie na końcówkę REST bazy instytucji
2. Autoryzacja odbywa się poprzez przekazanie w HEADER tokenu JSA, w polu „X-AUTH-TOKEN”
3. Baza instytucji odczytuje treść pracy

8.1.2. Przekazanie dokumentów bazowych

1. Baza instytucji typuje dokumenty bazowe na podstawie przesłanej przez JSA treści pracy
2. Baza instytucji może opatrzyć teksty notą wskazaną w polu „score”
3. JSA odbiera od bazy instytucji zestaw tekstów wraz z metadanymi do dalszego badania.

9. Uwagi i przykłady

Przykład:

```
[
{
  "meta": {
    "id": "123",
    "title": "Tytuł1",
    "autor": "Imię1 Nazwisko1"
  },

```

```
"score": "7",
"text": "treść pracy"
}
{
"meta": {
"id": "456",
"title": "Tytuł2",
"autor": "Imię2 Nazwisko2"
},
"score": "13",
"text": "treść pracy2"
}
{
"meta": {
"id": "789",
"title": "Tytuł3",
"autor": "Imię3 Nazwisko3"
},
"score": "4",
"text": "treść pracy3"
}
]
```