

Konfiguracja wyników wyszukiwania

Wagi atrybutów

Wagi atrybutów są szczególnie istotne dla działania wyszukiwania ogólnego (mniej dla zaawansowanego) i pozwalają kontrolować to, jak bardzo dopasowanie wartości konkretnego atrybutu do szukanej frazy wpłynie na pozycję obiektu w wynikach wyszukiwania. Przykładowo, w domyślnej konfiguracji największą wagę mają atrybuty Tytuł i Twórca, więc po wyszukaniu słowa "Tadeusz", na początku listy wyników pojawią się obiekty, które zawierają to imię jako część tytułu lub autora, a dopiero później obiekty, dla których to imię występuje w innych atrybutach (np. Współtwórca), czy tylko w treści. W przypadku niektórych obiektów może się też zdarzyć, że szukane słowo występuje w treści lub dodatkowych atrybutach na tyle często, że tak silne dopasowanie przewyższy wpływ ustawionych wag i obiekt będzie miał lepszą pozycję niż obiekt dopasowany po tytule. Ustawianie wag m.in. pozwala kontrolować, jak często tego typu sytuacja występuje.

Wagi atrybutów ustawiane są w pliku serwera dLibra: *conf/se/searchWeights.properties*. Plik jest podzielony na sekcje:

1. Wagi poszczególnych atrybutów:

```
Title=25
Creator=20
```

W każdej linijce znajduje nazwa RDF atrybutu, oraz po znaku równości wartość przypisanej wagi. Im większa wartość, tym wyżej na liście wyników będą obiekty dopasowane po tym atrybucie. Ustawienie wagi na 0 spowoduje, że wyszukiwanie po takim atrybucie nie będzie możliwe.

2. Specjalna wartość dla metadanych w ogólności:

```
dlibra_metadata=15
```

Waga ta wpływa na wszystkie atrybuty, które nie zostały bezpośrednio zdefiniowane w poprzedniej sekcji. Pozwala na ustawienie priorytetu dopasowywania po metadanych względem wyszukiwania w treści.

3. Specjalna wartość dla wyszukiwania w treści:

```
dlibra_content=1
```

Waga ta wpływa na pozycję obiektów, w których udało się dopasować wyszukiwaną frazę do zawartości tekstowej plików z treścią obiektu.

4. Wagi w zależności od dopasowania zakresów dat:

```
date.match.perfect=100
date.match.inside=50
date.match.containing=20
date.match.partial=1
```

Atrybuty typu data pozwalają na przeszukiwanie ich po zakresach dat (np. przy pomocy formularza wyszukiwania zaawansowanego, lub po kliknięciu na wartość takiego atrybutu na stronie z opisem jakiegoś obiektu). Ta konfiguracja wpływa na to, że obiekty z przypisanym zakresem który jest dokładnie zgodny z wyszukiwanym zakresem (*perfect*, np. ten sam rok) znajdą się na początku listy wyszukiwania, dalej będą obiekty z zakresami zawierającymi się w wyszukiwanym zakresie (*inside*, np. jeden miesiąc w roku), zawierającymi wyszukiwany zakres (*containing*, np. dziesięciolecie zawierające rok) i pokrywające się z nim tylko częściowo (*partial*).

Listy stopwords

Stopwords to funkcja mechanizmu wyszukiwania, która pozwala ignorować niektóre słowa, które często występują w danym języku, ale nie niosą konkretnej treści, np. spójniki. Dzięki niej, indeksy wyszukiwawcze zajmują mniej miejsca i szybciej działają, a wyniki wyszukiwania są lepiej dopasowane do intencji zapytania. W razie potrzeby listy *stopwords* można dostosować do potrzeb biblioteki - są zapisane w plikach *conf/solr/main/conf/stopwords_**.txt* oraz *conf/solr/synonym/conf/stopwords_**.txt*, gdzie ** to dwuliterowy kod języka.