

[EN] Configuring Search Results

Attribute Weights

Attribute weights are particularly important for the functioning of the general search (less so – for the advanced one) and make it possible to control how much the matching of the value of a particular attribute will influence the position of the object in the search results. For example, in the default configuration, attributes “Title” and “Creator” have the greatest weight, so after the “Tadeusz” word has been searched, objects which contain that name as a part of their title or author will appear at the beginning of the list, and objects which have that name in other attributes (for example, “Co-Creator”) or in their content will follow. In the case of some objects, the searched word may occur in their content or additional attributes so often that the strong match will take precedence over the set weights, and the object will have a better position than title-matched objects. The frequency of such an occurrence is one of the things which can be controlled by weight setting.

Weight attributes are set in the `conf/se/searchWeights.properties` file of the dLibra server. The file is divided into sections:

1. The weights of particular attributes:

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

Every line contains the RDF name of an attribute and, after the equal sign, the value of the assigned weight. The higher the value, the higher the objects matched by that attribute will appear on the list. If the value of an attribute is set to 0, it will not be possible to search by that attribute.

2. A special value for metadata in general:

```
dlibra_metadata=15
```

That weight has an impact on all attributes which have not been directly defined in the previous section. It makes it possible to set the priority of matching by metadata with respect to searching by content.

3. A special value for searching by content:

```
dlibra_content=1
```

That weight has an impact on the position of objects for which the searched phrase has been matched to the text of the files with object content.

4. Weights depending on the matching of date ranges:

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

Data-type attributes make it possible to search by date ranges (for example, with the use of the advanced search form or by clicking the value of such an attribute on a page with a description of an object). Thanks to that configuration, objects with an assigned range which corresponds precisely to the searched range (“*perfect*”, for example, the same year) will appear at the beginning of the search results list, followed by objects with ranges within the searched range (“*inside*”, for example, one month in the searched year), objects containing the searched range (“*containing*”, for example, a decade which contains the searched year), and objects with ranges which only partially overlap with the searched range (“*partial*”).

Stopwords Lists

Stopwords is a function of the search mechanism, which makes it possible to ignore certain words which are frequent in the given language but do not have any particular meaning, for example, connectives. Thanks to that function, search indexes take less space and operate faster, and search results are better matched to queries. In need, stopwords lists can be adjusted to the needs of a particular library. They are recorded in `conf/soir/main/conf/stopwords_**.txt` and `conf/soir/synonym/conf/stopwords_**.txt` files, where ** is a two-letter language code.