Technical preview:

Extracting business terms and governance rules from PDF files

IBM Cloud Pak for Data 2.5.0

Contents

1. Overview

Business terms are used to standardize the definitions of business concepts in your enterprise. Governance rules describe the characteristics of the data to make assets compliant with corporate objectives. One of the ways to add new business terms and governance rules to your catalog is by extracting them from PDF files. The PDF file that you upload in one of your catalogs in IBM Cloud Pak for Data is analyzed in search for terms and rules candidates. Apart from terms and rules, also tags are extracted. Tags are metadata items used to simplify searching for assets.

When the terms and rules are extracted, you can decide which of them to add to the catalog. They are imported to the catalog as drafts and need to be approved in the workflow process.

To use this feature, you must have Watson Knowledge Catalog installed.

Note: This is a technology preview and is not yet supported for use in production environments.

2. Enabling extraction of terms and rules

To enable the extraction of terms and rules, you must add the feature flag `kh_enabled=true` to the URL. For example, when you are in the Organize > Data and AI governance > Business terms location, the URL with the feature flag is:

```
https://mycompany/gov/terms?kh_enabled=true
```

3. Extracting terms and rules

To extract terms and rules, complete the following steps:

1. Go to one of these locations:
    a. Organize > Data and AI governance > Business terms.
    b. Organize > Data and AI governance > Rules.
2. Click **Extract**.

3. Add the PDF file from which you want to extract terms or rules. The file size can’t exceed 8 MB.
4. Click **Extract terms** or **Extract rules**, and wait for the process to complete. It might take a while depending on the file size.
5. Review the list of extracted terms or rules, and select the ones that you want to add to the catalog. Click Next.
6. Select the category to which you want to add new terms or rules. You can create a new category.
7. Click **Add terms** or **Add rules**.
8. When the process if finished, click **View workflow** to see the new workflow task.
9. Follow the workflow process to publish the imported terms or rules. For more information, see the Workflow topic.