

The Nepomuk-WebMiner Handbook

Jörg Ehrichs



The Nepomuk-WebMiner Handbook

Contents

1	Introduction	5
2	Dolphin Integration	6
3	Extractor Dialog	7
4	Service	8
5	Settings	9
5.1	Fetcher	9
5.1.1	General Settings	9
5.1.2	Plugin Selection	9
5.1.3	Background Service	10
5.2	Plugins	10
5.3	Filename Parsing	10
6	Credits and License	11

Abstract

This is the documentation for the KDE System Settings module that configures the Nepomuk Nepomuk-WebMiner service.

Chapter 1

Introduction

The Nepomuk-WebMiner finds additional metadata on the Internet (*via Python plugins*) and adds them to Nepomuk. It can retrieve data for any document/movie/TV show or music file.

Chapter 2

Dolphin Integration

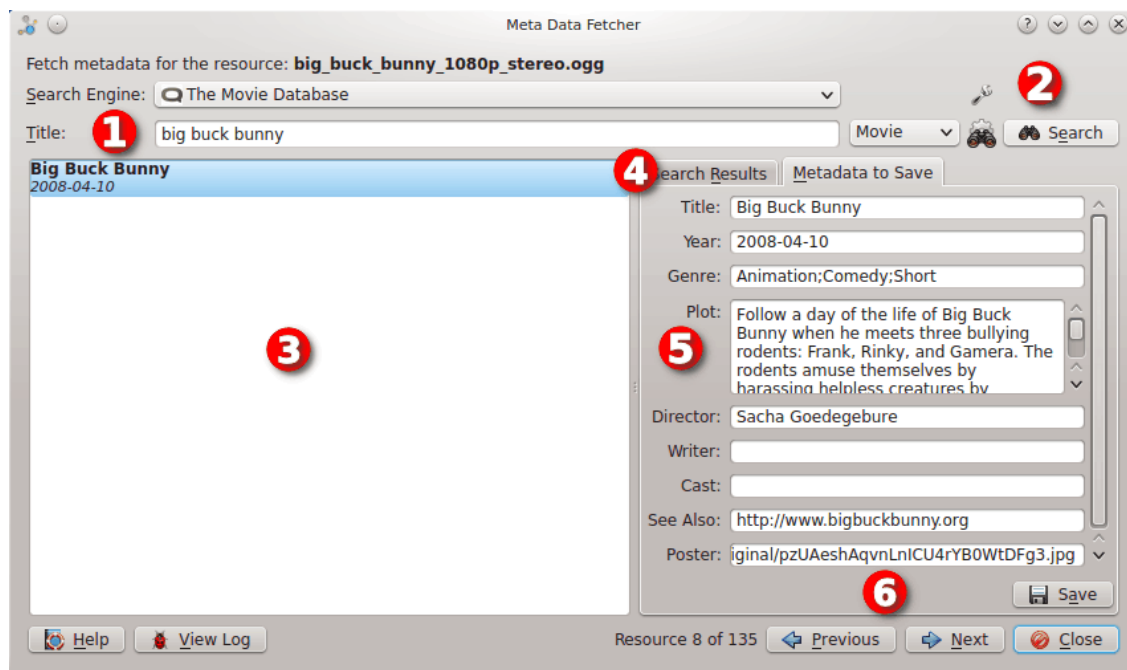
To start a manual extraction you can start the Nepomuk-WebMiner from Dolphin via the context menu. To do this right-click a file or a folder and select **Action** → **Desktop Search** → ...

Here you have 4 different options to fetch the metadata:

- **Normal:** This will show the extraction dialog but skip any file that already has some metadata attached to it.
- **Automatically:** This will run the extraction job in the background but skip any file that already has some metadata attached to it.
- **Force Download:** Will show the extraction dialog and also include files that already have metadata attached to it. This allows you to manually specify new search parameter and thus correct wrong metadata.
- **Automatically, Force Download:** This will run the extraction job in the background and also replace already any data from a previous extraction job.

Chapter 3

Extractor Dialog



The dialog allows to adjust the search parameter for each file. For each file the automatically extracted metadata can be seen in 1. Depending on the file type different search options are available. For a movie it is the title, for a TV show it is the name, season and episode, for publications it is the author and title.

Beside the search parameter it is possible to switch the used plugin for the lookup. By default the preferred plugin is selected.

Additional settings are available via the configure buttons as seen in 2.

Once everything is setup correctly the lookup can be started via the **Search** button.

All search results will be listed in the list view 3. Select the right entry you want fetch here.

Some more details about the currently selected entry is shown in tab 4. The show online link will open the web browser with the website that will be used for the data extraction. To start the extraction click on the **Fetch Metadata** button.

The results of the item extraction is shown in the tab 5. To save the metadata press the **Save** button.

To select a different file in the list of files the **Previous** and **Next** buttons in 6 can be used.

Chapter 4

Service

The background service is responsible for the metadata extraction for each new file on the disk. As soon as the Nepomuk fileindexer detects a new file the extractor is started and tries to fetch additional data from the internet.

NOTE

This will cause a lot of traffic especially when Nepomuk starts its initial indexing.

Chapter 5

Settings

5.1 Fetcher

The fetcher settings handle the general behavior of the Nepomuk-WebMiner.

5.1.1 General Settings

- **Download Banner/Poster/Cover:** If enabled the extractor looks for the matching poster/banner or cover of the item to fetch. The image will be saved into the application folder and added to the resource via Nepomuk.
- **Download Publication References:** This option allows the publication search to download all cited references of the current publication resource.

NOTE

As most publications tend to have many cited resources, this will take a lot of time. If the information about the cited resources are not needed or just for a few resources, it is better to disable this option. The cited references can be retrieved later on for a single resource if necessary.

- **Save Image in Resource Folder:** Instead of saving the image into the application folder the image is saved next to the file.

NOTE

Might lead to duplicated images if the same image for two different resources will be saved in different folders. This can happen for TV shows and the season posters, when the episodes are in two different folders.

5.1.2 Plugin Selection

Preferred Plugin: Defines which plugin will be used first to fetch the metadata from. Only if the first search failed the other plugins are queried to get the information.

5.1.3 Background Service

The service settings allow to **enable/disable** the background service. If the service is enabled the extractor will be run on each newly indexed file created by the Nepomuk fileindexerservice. When the service is disabled the metadata will only be fetched if the user requests it via the Dolphin context menu.

- **Fetch document information:** When this option is enabled the Service will automatically fetch additional information about any document that got indexed by Nepomuk.

IMPORTANT

This could expose private data to the used plugin search engine. The search will use data from the filename, any given metadata attached to the file and some possible title string from the first page of any pdf file.

If you are not sure you want this, leave this option disabled and call the metadata extractor manually on any document that should be used.

- **Fetch music information:** When this option is enabled the Service will automatically fetch additional information about any music file that got indexed by Nepomuk. If disabled you need to start the extractor manually for any music file.
- **Fetch video information:** When this option is enabled the Service will automatically fetch additional information about any video file that got indexed by Nepomuk. If disabled you need to start the extractor manually for any video file.
- **Use Folder Whitelist:** When this option is enabled only the folders specified by the whitelist are used for the metadata lookups. The list of the folders can be configured using **Customize Folders** button.

5.2 Plugins

In this tab all available plugins are shown. For more information about the plugin press the **Info** button. Some plugins allow to specify additional settings which can be seen via the **Configure** button.

5.3 Filename Parsing

Using this tab you can fine tune filename parsing during automatic metadata extraction. Use the examples to construct your own effective parsing regular expressions.

Chapter 6

Credits and License

This documentation is licensed under the terms of the [GNU Free Documentation License](#).
This program is licensed under the terms of the [GNU General Public License](#).