

WritingPlugins

Nutch's plugin system is based on the one used in Eclipse 2.x. Plugins are central to how nutch works. All of the parsing, indexing and searching that nutch does is actually accomplished by various plugins.

In writing a plugin, you're actually providing one or more *extensions* of the existing *extension-points*. The core Nutch *extension-points* are themselves defined in a plugin, the [NutchExtensionPoints](#) plugin (they are listed in the [NutchExtensionPoints plugin.xml](#) file). Each *extension-point* defines an interface that must be implemented by the *extension*. The core extension points are:

- [OnlineClusterer](#) – An extension point interface for online search results clustering algorithms (from javadoc).
- [IndexingFilter](#) – Permits one to add metadata to the indexed fields. All plugins found which implement this extension point are run sequentially on the parse (from javadoc).
- [Ontology](#)
- [Parser](#) – Parser implementations read through fetched documents in order to extract data to be indexed. This is what you need to implement if you want Nutch to be able to parse a new type of content, or extract more data from currently parseable content.
- [HtmlParseFilter](#) – Permits one to add additional metadata to HTML parses (from javadoc).
- [Protocol](#) – Protocol implementations allow nutch to use different protocols (ftp, http, etc.) to fetch documents.
- [QueryFilter](#) – Extension point for query translation. Permits one to add metadata to a query (from javadoc).
- [URLFilter](#) – URLFilter implementations limit the URLs that nutch attempts to fetch. The [RegexURLFilter](#) distributed with Nutch provides a great deal of control over what URLs Nutch crawls, however if you have very complicated rules about what URLs you want to crawl, you can write your own implementation.
- [NutchAnalyzer](#) – An extension point that provides some language specific analyzers (see [MultiLingualSupport](#) proposal). *Since it is in development stage, it is not in released javadoc.*

Setup

Start by [downloading](#) the Nutch source code. Once you've got that make sure it compiles as is before you make any changes. You should be able to get it to compile by running ant from the directory you downloaded the source to. If you have trouble you can write to one of the [Mailing Lists](#).

Use the source code for the plugins distributed with Nutch as a reference. They're in `[!YourCheckoutDir]/src/plugin`.

Required Files

You'll need to create a directory inside of the plugin directory with the name of your plugin. Inside that directory you need the following:

- A `plugin.xml` file that tells nutch about your plugin.
- A `build.xml` file that tells ant how to build your plugin.
- The source code of your plugin.

Getting Nutch to Use a Plugin

In order to get Nutch to use a given plugin, you need to edit your `conf/nutch-site.xml` file and add the name of the plugin to the list of `plugin.includes`.

<<< See also: [WritingPluginExample](#)

<<< See also: [WritingPluginExample for version 0.8](#)

<<< See also: [HowToContribute](#)

<<< [PluginCentral](#)