

OldPluginCentral

[OldPluginCentral](#) is a repository for pre-Nutch 1.3 plugin's. Looking back, it actually contains a wealth of Nutch plugin resources as well as tutorials for building plugins.

- [Plugin Tutorials](#)
- [Plugins that Come with Nutch \(0.9\)](#)
- [Additional Plugins in Dev Branch \(0.8\)](#)

Plugin Tutorials

- [WritingPluginExample-1.2](#) - Step-by-step example of how to write a plugin for the current development.
- [WritingPluginExample-0.9](#) - Step-by-step example of how to write a plugin for the 0.9 branch.
- [HowToMakeCustomSearch](#) - A custom plugin enabling us to search for the author of a website in our index by his email id. (N.B. This plugin is for Nutch release 1.0)
- [Example of writing a custom plugin](#) by Sujitpal

Plugins that Come with Nutch (0.9)

In order to get Nutch to use any of these plugins, you just need to edit your conf/nutch-site.xml file and add the name of the plugin to the list of plugin.includes.

- [clustering-carrot2](#) - Online Search Results Clustering using Carrot2's components.
- [creativecommons](#) - Support for crawling and searching Creative-Commons licensed content.
- [index-basic](#) - Adds url, content and anchor fields to the index.
- [index-more](#) - Adds date, content-length, contentType, primaryType and subtype fields to the index.
- [languageidentifier](#) - Adds a lang field to the index and allows you to query against it.
- [ontology](#) - Helps refine queries based on owl files.
- [parse-ext](#) - A wrapper that invokes external command to do real parsing job.
- [parse-html](#) - Parses HTML documents
- [parse-js](#) - Parses JavaScript
- [parse-mp3](#) - Parses MP3s
- [parse-zip](#) - Parses ZIP archives
- [parse-mspowerpoint](#) - Parses Microsoft Powerpoint files
- [parse-msword](#) - Parses MS Word documents
- [parse-msexcel](#) - Parses MS Excel documents
- [parse-pdf](#) - Parses PDFs
- [parse-rss](#) - Parses RSS feeds
- [parse-oo](#) - Parses [OpenOffice](#) files
- [parse-swf](#) - Parses Shockwave Flash
- [parse-rtf](#) - Parses RTF files
- [parse-text](#) - Parses text documents
- [protocol-file](#) - Retrieves documents from the filesystem
- [protocol-ftp](#) - Retrieves documents through ftp
- [protocol-http](#) - Retrieves documents through http
- [protocol-httpclient](#) - Retrieves documents through http and https
- [query-basic](#) - Runs queries against content, url and anchor fields
- [query-more](#) - Runs queries against date, content-length, contentType, primaryType and subType fields.
- [query-site](#) - Runs queries against site field
- [query-url](#) - Runs queries against url field.
- [urlfilter-prefix](#)
- [urlfilter-regex](#)

Additional Plugins in Dev Branch (0.8)

- [analysis-de](#)
- [analysis-fr](#)
- [lib-commons-httpclient](#)
- [lib-http](#)
- [lib-jakarta-poi](#)
- [lib-log4j](#)
- [lib-lucene-analyzers](#) - Lucene analyzers
- [lib-nekohtml](#) - automatic tag balancer
- [lib-parsems](#) - parse ms documents framework
- [parse-msexcel](#) - Parses MS Excel documents
- [parse-mspowerpoint](#) - Parses MS Powerpoint documents
- [parse-oo](#) - Parses Open Office and Star Office documents (Extentsions: ODT, OTT, ODH, ODM, ODS, OTS, ODP, OTP, SXW, STW, SXC, STC, SXI, STI)
- [parse-swf](#) - Parses Flash SWF files
- [microformats-reltag](#) - Adds [rel-tag](#) fields to the index and runs queries against them.
- [parse-zip](#)