

Archive and Legacy

This section includes all Nutch 2.x material

This section includes all Pre Nutch 1.3 material.

- [Reference Section](#)
- [General Information](#)
- [Internal Nutch Documentation](#)
- [Development and Old Nutch 2.0](#)

Nutch 2.x

- [Nutch 2.X tutorial\(s\)](#)
- [Pre-Nutch 1.3 Plugin Resources](#)
- [Nutch <1.3 Tutorials](#)
- [Configuration](#)
- [Script Administration](#)

Reference Section

- [Frutch Wiki](#) – French Nutch Wiki
- [The Old Wiki](#)
- [Experiences with the Nutch search engine](#) author:Doug Cutting,"Video Lecture"
- [Instructions for running Bixo on EC2](#) (includes parts of Nutch)
- [Lucene](#)

General Information

- [OldFeatures](#) - Pre Nutch 1.3
- [Nutch_i18n](#)
- [OldFAQs](#)

Internal Nutch Documentation

- [NutchFileFormats](#) - some notes by [LarsAronsson](#), 30 June 2004

Development and Old Nutch 2.0

- [MultiLingualSupport](#) - *In development.*
- [InstallingWeb2](#)
- [Nutch2Architecture](#) – Discussions on the Nutch 2.0 architecture (old)
- [JavaDemoApplication](#) - A simple demonstration of how to use the Nutch APLin a Java application

Nutch 2.x

Nutch 2.X tutorial(s)

- [Nutch2Tutorial](#) – How to get Nutch 2.X to use HBase as persistence layer for Gora. This is the primary Nutch 2.X tutorial.
- [Setting up Nutch 2.x with Cassandra](#) - How to setup and run Nutch 2.x using Cassandra as storage.
- [How to map your Nutch 2.x Hbase table to Hive](#) - Sample query for Hive mapping.
- [Accumulo, Nutch, and Gora](#) - A step-by-step tutorial ⚠ Very Old
- [Nutch2Crawling](#) - A description of the crawling jobs and field to database mappings.
- [Nutch2Architecture](#) - A high level overview of the new architecture and design
- [Nutch2Roadmap](#) – Discussions on the architecture and features of Nutch 2.0
- [Build Nutch 2.0 in Eclipse](#) – How to setup your IDE environment comfortably.
- [ErrorMessagesInNutch2](#) – What they mean and suggestions for getting rid of them.
- [NutchConfigurationFiles-2.x](#) – Configuration files that are specific to Nutch-2.x
- [Understanding the columns/fields in Nutch 2.0 Webpage - Detailed article](#)
- [WorkingWithGoraSnapshots](#) - A step by step guide to working with Gora development code within your Nutch 2.x deployment
- [NutchRESTAPI](#) - A UML diagram and overview of the entire Nutch 2.X REST API.

Pre-Nutch 1.3 Plugin Resources

- [OldPluginCentral](#)
- [NutchOSGi](#)

Nutch <1.3 Tutorials

- [OldHadoopTutorial](#)

- [RunningNutchAndSolr](#) - How to configure Nutch to crawl, but post to Solr for search/index
- [Tutorial](#) – A Step-by-Step guide to getting Nutch up and running (≤ 1.2).
- [Tutorial](#) – A Step-by-Step installation guide for dummies: Nutch 0.9.
- [Nutch_-_The_Java_Search_Engine](#) (Builds on the basic tutorials. Includes index maintenance scripts)
- [RunNutchInEclipse](#) for v0.8
- [RunNutchInEclipse0.9](#) for v0.9 (Linux and Windows)
- [RunNutchInEclipse1.0](#) for v1.0 (Linux and Windows)

Configuration

- [Nutch v1.3 and Hadoop tutorial](#)
- [Upgrading Hadoop Version in Nutch](#) - Basic steps for upgrading Hadoop in Nutch.
- [Commandline](#) options for 0.7.x
- [Commandline](#) options for version 0.8
- [UpgradeFrom07To08](#)
- [Upgrading_from_0.8.x_to_0.9](#)
- [GettingNutchRunningWithUtf8](#) - For support of non-ASCII characters (Chinese, German, Japanese, Korean).
- [GettingNutchRunningWithResin](#) - Resin is a JSP/Servlet/EJB application server (alternative to tomcat).
- [GettingNutchRunningWithJetty](#)
- [GettingNutchRunningWithJboss](#)
- [GettingNutchRunningWithUbuntu](#)
- [GettingNutchRunningWithWindows](#)
- [GettingNutchRunningWithMacOsx](#)
- [GettingNutchRunningWithRedHatApplicationServer](#)
- [GettingNutchRunningWithDebian](#)
- [GettingNutchRunningWithSocksProxy](#)
- [CreateNewFilter](#) - for example to add a category metadata to your index and be able to search for it

Script Administration

- [Automating Fetches with Python](#) - How to automatic the Nutch fetching process using Python
- [Nutch_0.9_Crawl_Script_Tutorial](#)
- [CrossPlatformNutchScripts](#)
- [MonitoringNutchCrawls](#) - techniques for keeping an eye on a nutch crawl's progress.
- [Crawl](#) - script to crawl (and possible recrawl too)
- [IntranetRecrawl](#) - script to recrawl a crawl
- [Whole-Web Crawling incremental script](#) - crawled urls are searchable at each iteration after merging
- [MergeCrawl](#) - script to merge 2 (or more) crawls