

Nutch on windows without cygwin

Update : Nutch 0.9 (bundled with hadoop 0.10) works out of the box on windows systems without cygwin [after setting up nutch-default.xml] !

Nutch 0.9 (bundled with hadoop 0.10) [not recommended]

It is possible to run a simple Nutch instance on windows without cygwin!

This is intended for users of java who want to know how to use nutch without cygwin.

After configuring the hadoop.xml file for [Nutch on local filesystem](#), configuring log4j.properties, configuring folders and configuring plugins just as described in other tutorials,

some little patches where necessary to make nutch 0.8 with hadoop 0.11 cooperate: <http://files.pannous.de/org.rar>

Other combinations of versions might work without patches. To get to know nutch it can be useful to play with the sources.

After all exceptions have been eliminated we are able to use nutch from java:

CRAWL:

```
Crawl.main(new String[]{dirWithUrls, "-dir", indexDirToBeCreated});
```

SEARCH:

```
NutchBean bean = new NutchBean(configuration, path); Hits hits = bean.search(Query.parse("Google", configuration), 10);
```

These patches were necessary:

- eliminates spaces from the \$PATH variable ("for runChild in [TaskRunner](#) ")
- get rid of the LOG.warn(dir + " already exists."); inconsistency :

```
new File(index + "/crawldb/current").mkdirs(); new File(index + "/linkdb/current").mkdirs();
```

- fixing some [NoMethodFound](#) conflicts in fetcher package
- fixing one UTF8 / Text Classcast version conflict
- No hadoop services have to be started by hand whatsoever. But for you have to set
<name>mapred.job.tracker</name>
<value>local</value>

again: Other combinations of versions might work without patches.