

GettingNutchRunningWithMacOsx

Running Nutch with Mac OSX

Downloading and setting up Tomcat

Download Tomcat (<http://tomcat.apache.org/>). The latest versions require J2SE 1.5 which can be downloaded from www.apple.com (Tiger users only). I downloaded apache-tomcat-5.5.12.tar.gz.

Open a terminal window and copy the file to /usr/local (cp apache-tomcat-5.5.12.tar.gz /usr/local) tar -zxvf apache-tomcat-5.5.12.tar.gz Start Tomcat (see below)

```
export JAVA_HOME=/System/Library/Frameworks/JavaVM.framework/Versions/1.5.0/Home (or /usr )
/usr/local/apache-tomcat-5.5.12/bin/startup.sh
```

You will see something like:

```
Using CATALINA_BASE:   /usr/local/apache-tomcat-5.5.12
Using CATALINA_HOME:   /usr/local/apache-tomcat-5.5.12
Using CATALINA_TMPDIR: /usr/local/apache-tomcat-5.5.12/temp
Using JRE_HOME:        /System/Library/Frameworks/JavaVM.framework/Versions/1.5.0/Home
```

Check that tomcat is running by opening <http://localhost:8080>. This should bring up Tomcat's Welcome Page.

Finally edit tomcat-users.xml which is in your Tomcat/conf Directory and add a 'manager' role.

Downloading and setting up Nutch

Download nutch-0.7.1.tar.gz or some other release and place the file somewhere in your Home directory. Expand the file using Stuffit Expander or the tar command. Open <http://localhost:8080> and click on the link 'Tomcat Manager' Click select WAR file to upload. Browse to the Nutch Directory and select the file 'nutch-0.7.1.war' which is located in the nutch root folder. Click 'Deploy' Check <http://localhost:8080/nutch-0.7.1/en/search.html>. You should see the Nutch Search Form.

Crawling

Note that the nutch command line tool (in our case nutch-0.7.1/bin/nutch) is not installed under the Tomcat web-application (\$CATALINA_HOME/webapps/nutch-0.7.1/WEB-INF/...). You can either leave it there or move it manually to your tomcat/webapps/nutch/WEB-INF/classes. In the first case you will have to do some classpath configuring or maintain two nutch-site.xml files (one for indexing and one for searching).

Using Terminal, cd to the directory where your bin/nutch is located. From here you can follow the instructions from the [tutorial](#).

Just like any other mac application the Terminal is scriptable which is a nice feature. The applescript below will start a crawl just by doubleclicking it's icon.

```
tell application "Terminal"
    if ((count of the window) = 0) or ~
        (the busy of window 1 = true) then
        tell application "System Events"
            keystroke "n" using command down
        end tell
    end if
    do script "cd Desktop/nutch-0.7.1" in window 1
    do script "export JAVA_HOME=/System/Library/Frameworks/JavaVM.framework/Versions/1.5.0/Home" in window 1
    do script "bin/nutch crawl -dir ~/nutch_index -depth 20 conf/link.txt" in window 1
end tell
```