

GettingNutchRunningWithResin

Running Nutch with Resin

These are the special tweaks I had to do to get Nutch to run with Resin2.

I followed the tutorial.

I did everything the tutorial said. In the command where I am asking nutch to crawl, I wanted to point it my own searcher dir (look below). So, I modified the command like so:

```
bin/nutch crawl urls -dir /home/paul/nutch-searcher.dir -depth 3 >& crawl.log
```

Xerces changes:

Resin does not use Xerces as its default xml parser and Nutch is happier with Xerces. So Resin should be told to use Xerces. You should add following lines to resin.conf in the element.

```
<system-property javax.xml.parsers.DocumentBuilderFactory="org.apache.xerces.jaxp.DocumentBuilderFactoryImpl"/>
<system-property javax.xml.parsers.SAXParserFactory="org.apache.xerces.jaxp.SAXParserFactoryImpl"/>
```

Logging

I was not using jdk1.4 logging yet, so I also added the following system property to see all of Nutch's logging. I configured the java1.4logging.conf as:

```
<system-property java.util.logging.config.files='/home/paul/java1.4logging.conf'/>
```

Another problem that comes up while using resin is that Nutch was not able to find searcher.dir

Not finding searcher dir

my search page looked like this:

```
500 Servlet Exception
java.lang.NullPointerException
    at net.nutch.searcher.NutchBean.init(NutchBean.java:82)
    ....
```

so in the logs it looked like:

```
050227 223521 10 creating new bean
050227 223521 10 opening segment indexes in /usr/local/resin-2.1.14/segments
```

if everything was good, your logs would look similar to:

```
050227 223317 10 creating new bean

050227 223317 10 opening merged index in /home/paul/nutch-searcher.dir/index
050227 223317 10 query request from 67.116.52.86

050227 223318 10 query: bhangra
050227 223318 10 searching for 20 raw hits
050227 223319 10 found resource common-terms.utf8 at file:/home/paul/www/WEB-INF/classes/common-terms.utf8
050227 223319 10 total hits: 4
```

so I modified nutch-site.xml like following:

```
{{{<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="nutch-conf.xsl"?>

<!-- Put site-specific property overrides in this file. -->

<nutch-conf>
<property>
<name>searcher.dir</name>
<value>/home/paul/nutch-searcher.dir</value>
<description>My path to nutch's searcher dir.</description>
</property>
</nutch-conf>
}}}}
```

Note: Same property exists in the nutch-default.xml, but you should not change it. Use nutch-site.xml to change properties for your specific installation. This suggestion is reiterated from the first comment in nutch-default.xml.

Resin3 should have similar issues, so one should be able to fix them in a similar manner. I have not tried it on Resin3 yet, but will soon.