

Solved problems

===P*roblem (exception) / *S*olution pairs===

P: Not a known field name: DEFAULT

S: Add plugin

```
<property>
<name>plugin.includes</name>
<value>query-basic|..... in nutch-default.xml
```

P: java.lang.NullPointerException at java.io.Reader.(Reader.java:61) ... at org.apache.nutch.analysis.CommonGrams.init(CommonGrams.java:152) at

S: the file common-terms.utf8 needs to be in the right directory (lib | classes?)

```
<property>
<name>analysis.common.terms.file</name>
<value>common-terms.utf8</value>
<description>The name of a file containing a list of common terms
that should be indexed in n-grams.
</description>
</property>
```

P: Bad mapred.job.tracker: local

S: if you want to run crawl without hdfs you can omit start-all.sh
just do " nutch crawl urlsdir "

P: ... getlocalpath [NullPointerException](#)

S: check mapred.local.dir and other tmp dirs in nutch-default.xml / hadoop-default.xml

P: extension point: org.apache.nutch.net.URLNormalizer does not exist

S: check your plugins + plugin.includes settings and add urlnormalizer-regex or urlnormalizer-(pass|regex|basic)

—
P: java.net.UnknownHostException "hostname"

S: add 127.0.0.1 "hostname" to the /etc/hosts file.

P: ...[null] [MalformedURLException]

S: add common-terms.utf8 to nutch dir

P: java.lang.ClassCastException: org.apache.hadoop.io.Text

S: wrong hadoop version / patch <http://files.pannous.de/org.rar>

P: java.lang.NoSuchMethodError: org.apache.hadoop.io.MapFile \$Writer.

S: wrong hadoop version / patch <http://files.pannous.de/org.rar>

P: [NullPointerException](#) when crawling :

S: add to nutch-site.xml:

```
<property>
<name>http.agent.name</name>
<value>NutchCVS</value>
<description>Our HTTP 'User-Agent' request header.</description>
</property>
```

P: java.io.IOException: config()

S: ignore it ! 😊

P: nutch crawl ... Job Failed!

S: manifold. set log4j.properties debug level ! log4j.rootLogger=ALL, stdout log4j.appender.stdout=org.apache.log4j.ConsoleAppender

P: No scoring plugins - at least one scoring plugin is required!

S: Add "scoring-opic" to <property> <name>plugin.includes</name>

P: ... java.net.SocketTimeoutException: Accept timed out

S: try using nutch without hdfs / check ports in hadoop file / RPC problems : start crawl without startall.sh ?

P: java.lang.NoClassDefFoundError xyz on windows

S: get rid of spaces in your classpath and path variables !
