

# Integrating Tika With Extracting Request Handler

## Integrating Tika With [Extracting Request Handler](#)

0. Because the online apache repositories do not have Tika-1.8-SNAPSHOT for now, we have to use local maven repository.

### 1. Checkout tika-trunk:

```
$ svn co https://svn.apache.org/repos/asf/tika/trunk/ tika-trunk
```

### 2. Build Tika

```
$ cd tika-trunk  
  
$ mvn install
```

### 3. Download tika-parsers dependencies:

```
$ cd tika-parsers  
  
$ mvn dependency:copy-dependencies
```

### 4. Checkout lucene-solr-4-10

```
$ svn checkout http://svn.apache.org/repos/asf/lucene/dev/branches/lucene_solr_4_10 lucene_solr_4_10
```

### 5. Modified the lucene\_solr\_4\_10/lucene/ivy-setting.xml by uncommenting line 45-52 and line 56:

Line 45-52:

```
<filesystem name="local-maven-2" m2compatible="true" local="true">  
  <artifact  
    pattern="\${local-maven2-dir}/[organisation]/[module]/[revision]/[module]-[revision].[ext]  " />  
  <ivy  
    pattern="\${local-maven2-dir}/[organisation]/[module]/[revision]/[module]-[revision].pom" />  
</filesystem>
```

Line 56:

```
<resolver ref="local-maven-2" />
```

### 6. Replace the lucene\_solr\_4\_10/solr/contrib/extraction/ivy.xml to the following ivy.xml:

```
<!--  
  
Licensed to the Apache Software Foundation (ASF) under one  
or more contributor license agreements. See the NOTICE file  
distributed with this work for additional information  
regarding copyright ownership. The ASF licenses this file  
to you under the Apache License, Version 2.0 (the  
"License"); you may not use this file except in compliance
```

with the License. You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing,  
software distributed under the License is distributed on an  
"AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY  
KIND, either express or implied. See the License for the  
specific language governing permissions and limitations  
under the License.

-->

```
<ivy-module version="2.0">
  <info organisation="org.apache.solr" module="extraction"/>
  <configurations defaultconfmapping="compile->master;test->master">
    <conf name="compile" transitive="false"/>
    <conf name="test" transitive="false"/>
  </configurations>
  <dependencies>
    <!-- Tika JARs -->
    <dependency org="org.apache.tika" name="tika-core" rev="1.8-SNAPSHOT" conf="compile"/>
    <dependency org="org.apache.tika" name="tika-parsers" rev="1.8-SNAPSHOT" conf="compile"/>
    <dependency org="org.apache.tika" name="tika-xmp" rev="1.8-SNAPSHOT" conf="compile"/>
    <!-- Tika dependencies - see http://tika.apache.org/1.3/gettingstarted.
html#Using_Tika_as_a_Maven_dependency -->
    <!-- When upgrading Tika, upgrade dependencies versions and add any new ones
        (except slf4j-api, commons-codec, commons-logging, commons-httpclient, geronimo-stax-api_1.0_spec,
jcip-annotations, xml-apis, asm)
        WARNING: Don't add netcdf / unidataCommon (partially LGPL code) -->
    <dependency org="org.gagravarr" name="vorbis-java-tika" rev="0.6" conf="compile"/>
    <dependency org="org.gagravarr" name="vorbis-java-core" rev="0.6" conf="compile"/>
    <dependency org="org.apache.james" name="apache-mime4j-core" rev="0.7.2" conf="compile"/>
    <dependency org="org.apache.james" name="apache-mime4j-dom" rev="0.7.2" conf="compile"/>
    <dependency org="org.apache.commons" name="commons-compress" rev="1.9" conf="compile"/>
    <dependency org="org.apache.pdfbox" name="pdfbox" rev="1.8.8" conf="compile"/>
    <dependency org="org.apache.pdfbox" name="fontbox" rev="1.8.8" conf="compile"/>
    <dependency org="org.apache.pdfbox" name="jempbox" rev="1.8.8" conf="compile"/>
    <dependency org="org.bouncycastle" name="bcmail-jdk15" rev="1.45" conf="compile"/>
    <dependency org="org.bouncycastle" name="bcprov-jdk15" rev="1.45" conf="compile"/>
```

```

<dependency org="org.apache.poi" name="poi" rev="3.11" conf="compile"/>

<dependency org="org.apache.poi" name="poi-scratchpad" rev="3.11" conf="compile"/>

<dependency org="org.apache.poi" name="poi-ooxml" rev="3.11" conf="compile"/>

<dependency org="org.apache.poi" name="poi-ooxml-schemas" rev="3.11" conf="compile"/>

<dependency org="org.apache.xmlbeans" name="xmlbeans" rev="2.6.0" conf="compile"/>

<dependency org="dom4j" name="dom4j" rev="${dom4j/dom4j}" conf="compile"/>

<dependency org="org.ccil.cowan.tagsoup" name="tagsoup" rev="1.2.1" conf="compile"/>

<dependency org="com.googlecode.mp4parser" name="isoparser" rev="1.0.2" conf="compile"/>

<dependency org="org.aspectj" name="aspectjrt" rev="1.8.0" conf="compile"/>

<dependency org="com.drewnoakes" name="metadata-extractor" rev="2.6.2" conf="compile"/>

<dependency org="de.l3s.boilerpipe" name="boilerpipe" rev="1.1.0" conf="compile"/>

<dependency org="rome" name="rome" rev="1.0" conf="compile"/>

<dependency org="jdom" name="jdom" rev="1.0" conf="compile"/>

<dependency org="com.googlecode.juniversalchardet" name="juniversalchardet" rev="1.0.3" conf="compile"/>

<dependency org="org.tukaani" name="xz" rev="1.5" conf="compile"/>

<dependency org="com.adobe.xmp" name="xmpcore" rev="5.1.2" conf="compile"/>

<dependency org="com.uwyn" name="jhighlight" rev="1.0" conf="compile"/>

<!-- Other ExtractingRequestHandler dependencies -->

<dependency org="com.ibm.icu" name="icu4j" rev="${com.ibm.icu/icu4j}" conf="compile"/>

<dependency org="xerces" name="xercesImpl" rev="${xerces/xercesImpl}" conf="compile"/>

<dependency org="org.slf4j" name="jcl-over-slf4j" rev="${org.slf4j/jcl-over-slf4j}" conf="test"/>

<exclude org="*" ext="*" matcher="regexp" type="${ivy.exclude.types}"/>

</dependencies>

</ivy-module>

```

## 7. Compile solr

Change the working directory to lucene\_solr\_4\_10/solr/

```
$ ant compile
```

## 8. Generate new sha1 files for the jars

```
$ ant jar-checksums
```

## 9. Done. Enjoy.