OpenKM

Installation guide

Installation Guide

OpenKM allows to centralize all company's information in a single access point, guarantying the security of the data. In this way, only authorized users can have access to specific data. At the same time, it provides a complete audit service. OpenKM is a Web 2.0 application that works with Internet Explorer, Firefox, Safari and Opera. In addition, periodic backups avoid information losses.

This new version entails the following improvements: workflow (jBPM), a dashboard with information on the repository activities, push services (consults programmed by the user), support for WebDAV, download documents converted to PDF, information on the repository space used by the user, preview of videos and images, web administration, new languages, and the insertion of any document format (configurable), as well as the possibility of a default language.

However, one of the most relevant functions to mention is the indexing of the most common types of files: text, Office, Office 2007, OpenOffice, PDF, HTML, XML, MP3, JPEG, etc. Its powerful search engine indexes all documents making it easy to find any information.

- Software and hardware requirements
- OpenKM installation

  - Java installation
  - Download OpenKM
  - OpenKM configuration

    - JBoss mail configuration
    - jBPM configuration
    - Application configuration
    - Several OpenKM instances in the same JBoss 🔒
  - Configure JBoss service 🔒

    - Several JBoss in the same server 🔒
    - Several JBoss instances in the same server
    - Securing JBoss 🔒
    - Adding APR to JBoss 🔒
  - Changing OpenKM timeout session
  - Test OpenKM installation
  - Change default okmAdmin password
  - Create users and roles
  - Create password validator 🔒
- Notification and subscription messages
- OpenKM authentication 🔒

  - Active Directory 🔒
  - Open Directory 🔒
  - Central Authentication Service
- Repository configuration
- Third-party software integration

  - Apache 🔒
  - OCR
  - OpenOffice.org 🔒
  - SWFTools 🔒
  - Antivirus

- Acme CAD Converter
- Repository backup
  - Backup scripts 🔒
  - Backup restoring 🔒
  - Disaster recovering 🔒
- Repository mirroring 🔒
- WebDAV access
- Property Groups
- Property Groups definition
- Creating templates
  - PDF 🔒
  - OpenOffice 🔒
  - Text 🔒
- Debugging OpenKM
- Troubleshooting
- Error codes
- Using OpenKM with other databases
  - MySQL 🔒
  - PostgreSQL 🔒
  - MS SQLServer 🔒
  - Oracle 🔒
- Thesaurus
  - Thesaurus full example 🔒
- Automatic key extraction
  - Automatic key extraction full example 🔒
  - Creating automatic key extraction training files 🔒
- Experimental features

Software and hardware requirements

OpenKM can be installed in any operating system (Linux, Windows, Mac OS X, ...) since it is a multi-platform system. In order to install OpenKM you will need to install Java Development Kit 1.6. The Java Development Kit (JDK) 1.6 is available at Sun Developer Network [1]. If you install the JRE, OpenKM will not run properly.

OpenKM runs quite well in any computer with a common hardware configuration. It is recommended 2 GB RAM and a fast hard disk like SATA or better. A dual-core Intel based CPU with 3.20 GHz should be fine for a reduced community of users (30-50 concurrent users). Be sure to have enough disk space for document repository and its versions.

A Linux host is a good choice because this OS handle quite well heavy I/O applications like this, and it is very cheap :)

References

[1] http://java.sun.com/javase/downloads/index.jsp

OpenKM installation

This task is divided in the flowing steps:

- Java installation
- Download OpenKM
- OpenKM configuration
- Configure JBoss service
- Test OpenKM installation
- Change default okmAdmin password
- Create users and roles

Java installation

You should check if there is a previous java version installed in your system. You can check it running this command:

```
$ java -version
```

It will tell you the installed version, if any. Keep in mind that OpenKM needs JDK 1.6 (or higher) to run. If you use a Debian based Linux distribution, you can install the Java JDK this way:

```
$ sudo aptitude install sun-java6-jdk
```

In case of you have other JDK installed (like GCJ) you can select the right one this way:

```
$ sudo update-alternatives --config java
$ sudo update-alternatives --config javac
```

If you use a Windows based OS, download and install Java JDK from http://java.sun.com/javase/downloads/index.jsp.In other case visit http://java.sun.com for more information.

Download OpenKM

The files and further information about OpenKM can be found at http://www.openkm.com.The application files are hosted at SourceForge and OpenKM downloads are located at http://www.openkm.com/Download.html.

Download always the last release of OpenKM. Beta and release candidate (RC) are only recommended for testing and not are considered stable for production.

There is no migration support for beta or release candidates of the application. The migration procedure is only designed to work with stable product versions.

OpenKM configuration

Before running OpenKM for first time, you have to configure it properly. There are some modules that requires your attention:

- jBoss mail configuration
- jBPM configuration
- Application configuration

JBoss mail configuration

OpenKM is a J2EE application deployed in JBoss application server. Some parameters have a default configuration that fit common user requirements but others depends on specific user configuration.

> For specific JBoss issues, visit [1]
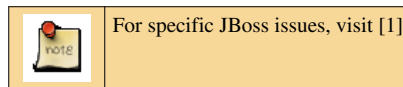
OpenKM sometimes send mail to users (because they have subscribed a document or folder, for example) and you have to check that this mail configuration is correct. If you edit the file *$JBOSS_HOME/server/default/deploy/mail-service.xml*:

```
$ vim $JBOSS_HOME/server/default/deploy/mail-service.xml
```

You will see something like this:

```xml
<mbean code="org.jboss.mail.MailService" name="jboss:service=OpenKM">
  <attribute name="JNDIName">java:/mail/OpenKM</attribute>
  <attribute name="User">nobody</attribute>
  <attribute name="Password">password</attribute>
  <attribute name="Configuration">
    <!-- A test configuration -->
    <configuration>
      <!-- Change to your mail server prototocol -->
      <property name="mail.store.protocol" value="pop3"/>
      <property name="mail.transport.protocol" value="smtp"/>
      <!-- Change to the user who will receive mail  -->
      <property name="mail.user" value="nobody"/>
      <!-- Change to the mail server -->
      <property name="mail.pop3.host" value="pop3.your-domain.com "/>
      <!-- Change to the SMTP gateway server -->
      <property name="mail.smtp.host" value="smtp.your-domain.com "/>
      <!-- The mail server port -->
      <property name="mail.smtp.port" value="25"/>
      <!-- Change to the address mail will be from -->
      <property name="mail.from" value="noreply@your-domain.com"/>
      <!-- Enable debugging output from the javamail classes -->
      <property name="mail.debug" value="false"/>
    </configuration>
  </attribute>
```

```
    <depends>jboss:service=Naming</depends>
</mbean>
```

Typically you only have to configure these parameters:

- **mail.smtp.host**: this is the host where is located your mail server. Can be localhost if you have a local mail server installed (like Postfix)
- **mail.from**: all the mails send by OpenKM will be from this mail. Can be in the form of noreply@your-domain.com.

But it depends on you own mail configuration. If you run intro troubles, set the parameter **mail.debug** to *true* and revise the JBoss log.

```
$ tail -f $JBOSS_HOME/server/default/log/server.log
```

For more info, read http://community.jboss.org/wiki/JavaMail.

Configuration for SMTP authentication

When we want to use a mail server via a secured SMTP with authentication we should some changes to mail-service.xml file (used an example using the SMTP server gmail.com).

```xml
<mbean code="org.jboss.mail.MailService" name="jboss:service=OpenKM">

  <attribute name="JNDIName">java:/mail/OpenKM</attribute>

  <attribute name="User">USUARIO@gmail.com</attribute>

  <attribute name="Password">PASSWORD</attribute>

  <attribute name="Configuration">

    <!-- A test configuration -->

    <configuration>

      <!-- Change to your mail server prototocol -->

      <property name="mail.store.protocol" value="pop3"/>

      <property name="mail.transport.protocol" value="smtp"/>


      <!-- Change to the user who will receive mail  -->

      <property name="mail.user" value="USUARIO@gmail.com"/>


      <!-- Change to the mail server  -->

      <property name="mail.pop3.host" value="pop.gmail.com"/>


      <!-- Change to the SMTP gateway server -->

      <property name="mail.smtp.host" value="smtp.gmail.com"/>

      <property name="mail.smtp.starttls.enable" value="true" />

      <property name="mail.smtp.socketFactory.class" value="javax.net.ssl.SSLSocketFactory" />


      <!-- The mail server port -->

      <property name="mail.smtp.port" value="465"/>

      <property name="mail.smtp.socketFactory.port" value="465" />


      <!-- Change to the address mail will be from  -->

      <property name="mail.from" value="USUARIO@gmail.com"/>

      <property name="mail.smtp.user" value="USUARIO@gmail.com"/>
```

```
        <property name="mail.smtp.password" value="PASSWORD"/>
        <property name="mail.smtp.auth" value="true"/>


        <!-- Enable debugging output from the javamail classes -->
        <property name="mail.debug" value="false"/>
    </configuration>
  </attribute>
  <depends>jboss:service=Naming</depends>
</mbean>
```

Adding PKI certificate

Some mail servers need you have installed PKI certificate on server, if you have not it installed you'll got some error like this:

```
ERROR [STDERR] javax.mail.MessagingException: Exception reading response;

nested exception is:

javax.net.ssl.SSLHandshakeException: sun.security.validator.ValidatorException: PKIX path building failed:

sun.security.provider.certpath.SunCertPathBuilderException: unable to find valid certification path to requested target
```

In this case must be installed the certificate.

- Download the file
- Execute application

```
java InstallCert serverName
```

- Select option '1', now has been generated a file called jssecacerts
- Overwrite jssecacerts to /usr/lib/jvm/java-6-sun/jre/lib/security/cacerts (the path depends on your OS server and jdk version)
- You must restart jboss to it takes some effect

For more information about installing certificate, read:

- http://blogs.oracle.com/gc/entry/unable_to_find_valid_certification
- http://www.grim.se/guide/jre-cert

For more info read:

- http://java.sun.com/products/javamail/javadocs/com/sun/mail/smtp/package-summary.html.
- jGuru: Fundamentals of the JavaMail API [2]

References

[1] http://jboss.org/community/docs/DOC-10376.
[2] http://java.sun.com/developer/onlineTraining/JavaMail/contents.html

JBPM configuration

jBPM configuration file is located at $JBOSS_HOME/server/default/deploy/OpenKM.war/WEB-INF/classes/jbpm.cfg.xml. If you edit this file, you will see something like this:

```
<string name="jbpm.mail.smtp.host" value="smtp.your-domain.com " />

<string name="jbpm.mail.from.address" value="noreply@your-domain.com" />

<string name="resource.mail.templates" value="jbpm.mail.templates.xml"/>

<bean name="jbpm.mail.address.resolver" class="com.openkm.workflow.OKMAddressResolver" singleton="true"/>
```

As in JBoss mail configuration, you have to edit a couple of properties:

- **jbpm.mail.smtp.host**: this is the host where is located your mail server. Can be localhost if you have a local mail server installed (like Postfix) Typically the same value of mail.smtp.host.
- **jbpm.mail.from.address**: all the mails send by OpenKM will be from this mail. Can be in the form of noreply@your-domain.com. Typically the same value of mail.from.

jBPM notify you when a task has been assigned to you or when he wants to send you a reminder. These mail templates are defined in the file jbpm.mail.templates.xml which is located at the same place of the previous one. Here you can define a more elegant mail message. Also you have to configure a parameter here:

- **BaseTaskListURL**: a value in the form of http://your-domain.com:8080/OpenKM.

For more info, read http://docs.jboss.com/jbpm/v3.2/userguide/html/mail.html.

Application configuration

OpenKM configuration file

OpenKM.cfg file is the main OpenKM configuration file. You can find this file in your server in *$JBOSS_HOME/OpenKM.cfg*.

A full listing of available configuration options can be found when logged into OpenKM as administrator role. See the Settings shortcut on the Administration shortcut located on the toolbar.

> Each time you make some change in the OpenKM.cfg file JBoss server must be restarted to take effect.

> Starting with OpenKM 5.1 these configuration properties will be located at database and will ease its management. To change the configuration properties, go to **Administration > Configuration**.

Changing uploading max file size

The default limit is 25MB. If you want to increase to 50MB:

```
max.file.size=50.
```

Changing repository.xml configuration filename and path

```
repository.config=repository.xml
```

Changing default repository home

By default repository is stored in *$JBOSS_HOME/repository* folder but you can change it with:

```
repository.home=repotest
```

Changing default connection role

By default connection role to OpenKM is called UserRole, you can change to other role name. This change need some changes too in web.xml file into OpenKM.war/WEB-INF

```
default.user.role=OtherRole
```

```xml
<security-constraint>
    <web-resource-collection>
      <web-resource-name>OpenKM Protected Area</web-resource-name>
      <!-- GWT -->
      <url-pattern>/es.git.openkm.frontend.Main/*</url-pattern>
      <url-pattern>/es.git.openkm.backend.Main/*</url-pattern>
      <!-- JSPs -->
      <url-pattern>/admin/*</url-pattern>
      <url-pattern>/mobi/*</url-pattern>
      <!-- Servlets -->
      <url-pattern>/RepositoryStartupServlet</url-pattern>
      <url-pattern>/Test</url-pattern>
      <url-pattern>/OKM*</url-pattern>
    </web-resource-collection>
    <auth-constraint>
      <role-name>AdminRole</role-name>
      <role-name>OtherRole</role-name>
    </auth-constraint>
    <user-data-constraint>
      <transport-guarantee>NONE</transport-guarantee>
    </user-data-constraint>
  </security-constraint>
  <login-config>
    <auth-method>FORM</auth-method>
    <realm-name>OpenKM</realm-name>
    <form-login-config>
      <form-login-page>/login.jsp</form-login-page>
      <form-error-page>/login.jsp?error=1</form-error-page>
```

```
    </form-login-config>
  </login-config>
  <security-role>
    <description>Admin user access</description>
    <role-name>AdminRole</role-name>
  </security-role>
  <security-role>
    <description>General user access</description>
    <role-name>OtherRole</role-name>
  </security-role>
```

Changing default admin role

By default connection role to OpenKM is called AdminRole, you can change to other role name. This change need some changes too in web.xml file into OpenKM.war/WEB-INF

```
default.admin.role=OtherAdminRole
```

```xml
<security-constraint>
    <web-resource-collection>
      <web-resource-name>OpenKM Protected Area</web-resource-name>
      <!-- GWT -->
      <url-pattern>/es.git.openkm.frontend.Main/*</url-pattern>
      <url-pattern>/es.git.openkm.backend.Main/*</url-pattern>
      <!-- JSPs -->
      <url-pattern>/admin/*</url-pattern>
      <url-pattern>/mobi/*</url-pattern>
      <!-- Servlets -->
      <url-pattern>/RepositoryStartupServlet</url-pattern>
      <url-pattern>/Test</url-pattern>
      <url-pattern>/OKM*</url-pattern>
    </web-resource-collection>
    <auth-constraint>
      <role-name>OtherAdminRole</role-name>
      <role-name>UserRole</role-name>
    </auth-constraint>
    <user-data-constraint>
      <transport-guarantee>NONE</transport-guarantee>
    </user-data-constraint>
  </security-constraint>
  <login-config>
    <auth-method>FORM</auth-method>
    <realm-name>OpenKM</realm-name>
    <form-login-config>
      <form-login-page>/login.jsp</form-login-page>
      <form-error-page>/login.jsp?error=1</form-error-page>
    </form-login-config>
  </login-config>
  <security-role>
```

```
    <description>Admin user access</description>
    <role-name>OtherAdminRole</role-name>
  </security-role>
  <security-role>
    <description>General user access</description>
    <role-name>UserRole</role-name>
  </security-role>
```

Change max results in UI

This option limits the search results in UI.

```
max.search.results=25
```

Since: OpenKM 4.0

Change access URL

By default OpenKM se the access URL to http://localhost:8080/OpenKM/index.jsp'", **but obviously you don't want to access your OpenKM installation only from localhost. This URL is used in mail notifications and the** copy-to-clipboard **feature.To change this default URL use this property:**

```
application.url=http://your-server.com/OpenKM/index.jsp
```

Since: OpenKM 4.0

Change principal adapter

OpenKM can handle user access using the JBoss DatabasePrincipalAdapter login module by default. OpenKM needs an available method for reading users and roles. The class DatabasePrincipalAdapter does this job. For more information take a look at OpenKM authentication and Active Directory

```
principal.adapter=com.openkm.core.DatabasePrincipalAdapter
```

Since: OpenKM 4.0

Switch OpenKM to demo mode

In demo mode some options are disabled by default.

```
system.demo=on
```

Since: OpenKM 4.0

Enabling OCR

To enable OCR you must put the files system path of OCR engine

```
system.ocr=/usr/local/bin/tesseract
```

Starting with OpenKM 5.1 you can select between 3 OCR engines:

- Tesseract 2.x
- Tesseract 3.x
- Cuneiform 0.7.x

See Third-party software integration: OCR for more info.

Setting default lang to OpenKM UI

OpenKM By default guess language depending on user language browser configuration. But you can set a language for all OpenKM users by default:

```
default.lang=es-ES
```

Force username to lowercase

You can forge all username be written in lower case. It's useful for example when you integrate with microsoft active directory where username is not case sensitive

```
system.login.username.lowercase=on
```

Since: OpenKM 4.0

Enable document preview

Several applications are used by OpenKM to generate the document preview: OpenOffice.org, pdf2swf and convert. Read Enable PDF to SWF conversion, Enable image preview and Enable OpenOffice.org service. Also take a look at Random errors in preview in case of problems.

Enable PDF to SWF conversion

To enable preview UI tab, OpenKM need to convert PDF files to SWF. And must have installed open office as a service too.

```
system.pdf2swf=/usr/bin/pdf2swf
```

> Starting from OpenKM 5.1 this configuration property has changed to **system.swftools.pdf2swf**.

Note: Read Third-party software integration: SWFTools for a more complete documentation.

Since: OpenKM 4.1

Enable PS to SWF conversion

To enable postscript document preview, OpenKM need to convert PS files to SWF using the **ps2pdf** utility from Ghostscript:

```
system.ghostscript.ps2pdf=/usr/bin/ps2pdf
```

Since: OpenKM 5.1.2

Enable image preview

To enable image preview, you need to install que ImageMagick *convert* utility and configure:

```
system.convert=/usr/bin/convert
```

> Starting from OpenKM 5.1 this configuration property has changed to **system.imagemagick.convert**.

Since: OpenKM 4.1

Enable OpenOffice.org integration

OpenOffice.org is used by OpenKM to convert files to PDF and generating the preview.

```
system.openoffice=on
```

Starting with OpenKM 5.0 this properties are changed to:

```
system.openoffice.path=/usr/lib/openoffice
system.openoffice.tasks=5
system.openoffice.port=2222
```

Where you should specify a OpenOffice.org installation path, the maximum number of conversion task before restart the service and a port where the OpenOffice.org conversion service will be attached. Only the first one property is mandatory.

Note 1: Read Third-party software integration: OpenOffice.org for a more up-to-date documentation.

Note 2: You can enhance OCR results configuring OpenOffice.org a dictionary. See OpenKM 5.1 OCR configuration for more info.

Since: OpenKM 4.0

Configuring e-mail

To configuring e-mail service you must enable subscription and notification properties. View a complete example in Notification and subscription messages.

```
subscription.message.subject=OpenKM – {0} – {1};
subscription.message.body=Document: <a href=\"{0}\">{1}</a>
notify.message.subject=OpenKM – NOTIFICATION – {0}
notify.message.body=Document: <a href=\"{0}\">{1}</a>
```

Since: OpenKM 4.1

Configuring password validation

To configuring password validation you have several properties. CompletePasswordValidator is default password validator class that comes by default with OpenKM. By default validator is not enabled. You can also create your own Create password validator 🔒 .

You can select which rules validation you want to enable, for example if you only enable validator.password.min.length property, the validator only will validate minimal length.

```
validator.password=com.openkm.validator.password.CompletePasswordValidator
validator.password.min.length=numeric value greater than 0
validator.password.max.length=numeric value greater than 0
validator.password.min.lowercase=numeric value greater than 0
validator.password.min.uppercase=numeric value greater than 0
validator.password.min.digits=numeric value greater than 0
validator.password.mini.special=numeric value greater than 0
validator.password.error.min.length=Your message error
validator.password.error.max.length=Your message error
```

```
validator.password.error.min.lowercase=Your message error
validator.password.error.max.uppercase=Your message error
validator.password.error.min.digits=Your message error
validator.password.error.min.special=Your message error
```

Since: OpenKM 4.1

Configuring chat service

By default chat and autologin are enabled. In order to enable or disable values can be "on" or "off".

```
chat.enabled=off
chat.autologin=off
```

Since: OpenKM 5.0

Configuring schedulers

There're some schedulers that OpenKM uses. Some internally like mail importer and getting repository news and others used by user interface like dashboard refreshing time and keepalive.

 Values are in minutes.
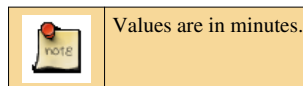
OpenKM inquires for repository news (by default each 24 hours)

```
schedule.repository.info=1440
```

OpenKM imports mails from imap server (by default set to 0, that is disabled)

```
schedule.mail.importer=0
```

 In OpenKM 4.1 and before, the default value was 60 and cannot be disabled.
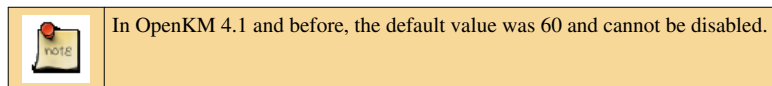
KeepAlive is used by user interface to mantaining browser connected to OpenKM although users stay sometime without using any operation. For it reason never might be lower that server timeout session (by default set to 15 minutes).

```
schedule.session.keepalive=15
```

Dashboard data is refreshing by default each 30 minutes.

```
schedule.dashboard.refresh=30
```

Since: OpenKM 4.1

Configuring wizard

You can configure a wizard each time any user uploading a document. Wizard can be a sequence of property groups, keywords (included thesausus) and categories. By default wizard is disabled.

Can be configured a property groups as a list of property groups name separed by ","

```
wizard.property.groups=okg:consulting,okg:technology
```

To enable keyword wizard value must be "on". Default value is "off".

```
wizard.keywords=on
```

To enable categories wizard values must be "on". Default value is "off".

```
wizard.categories=on
```

Since: OpenKM 5.0

Customizing application logo

You can replace the default OpenKM logo in the login page and reports.

```
logo.login
logo.mobi
logo.report
```

Recommended image size:

- logo.login: 316px × 74px
- logo.mobi: 161px × 38px
- logo.report: 150px × 35px

Also can set a message in the login page with:

```
logo.text
```

Since: OpenKM 5.1 (configuration stored into database)

Configuring WebDAV access

In recent OpenKM releases, WebDAV is disabled by default. If you want to enable it, set property **system.webdav.server** to on. The *okm:root* path may cause problems with some WebDAV clients. For this reason there is another configuration parameter **system.webdav.fix** which replace a path like *okm:root* with *okm_root*. See WebDAV access for more info.

Since: OpenKM 5.1

Disable document name mismatch check

By default OpenKM will check on checkin the document name to ensure you upload another version of the same document. You can disable this behavior with this configuration property:

```
system.document.name.mismatch.check=false
```

Since: OpenKM 5.0.1

Force keywords lowercase conversion

By default keywords are stored as is, but you can force a lowercase conversion using this configuration property:

```
system.keyword.lowercase
```

Since: OpenKM 5.1

Disable user assign on document creation

By default when an user creates a document or folder, he is added to the node with full permissions. You can disable this behavior this way:

```
user.assign.document.creation=off
```

Since: OpenKM 5.0.2

Improve OpenKM performance

If you experiment a slowndown of OpenKM when the number document grows, you can disable user quota or make use of the user items size cache.

To enable user items size cache:

```
user.item.cache=on
```

Starting from 5.1 **user.item.cache** is set to **on** by default.

To disable user quota, go to user profile and ser "Quota limit" to 0 (This field is in bytes). This feature of disabling quota is available from OpenKM 5.1.6.

Since: OpenKM 5.0.3

Change conversion cache home

In case you have little space in your JBoss partition, you can configure to store the conversion cache folder in another one. For this, use the **cache.home** configuration property. By default is set to $JBOSS_HOME/cache.

Since: OpenKM 5.1.8

Several OpenKM instances in the same JBoss

Manage several companies with an unique OpenKM instance can render in several security problems. Actually there is no way to make a relation between a group of user and a company. This means that when you manage document security all the users will be shown, even users in other companies. Also is not a good way because you can't configure especifically for each company.

The best way to accomplish this is having several OpenKM in one JBoss. The most complex and time-consuming configuration is related to JBoss. But configure an OpenKM instance is very straightforward and can take seconds. This implies some source code modifications, so read carefully Developer Guide before trying this:

> Replace # with an OpenKM installation id.
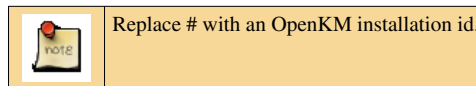
Files located at $OPENKM_PRJ

- **src/main/java/com/openkm/core/Config.java**

```java
public static String INSTALL = "#";
```

- **src/main/java/com/openkm/frontend/client/config/Config.java**

```java
private static String INSTALL = "#";
```

- **src/main/java/com/openkm/ws**

```
Disable WS deleting this folder.
```

- **src/main/resources/hibernate.cfg.xml**

```xml
<property name="hibernate.connection.datasource">java:/OKMWorkflow#DS</property>
```

- **src/main/webapp/WEB-INF/web.xml**

```xml
<display-name>OpenKM#</display-name>
```

- **src/main/webapp/WEB-INF/jboss-web.xml**

```xml
<security-domain>java:/jaas/OpenKM#</security-domain>
<context-root>/OpenKM#</context-root>
```

- **src/pom.xml**

```xml
<finalName>OpenKM#</finalName>
```

Files located at $JBOSS_HOME

- **OpenKM#.cfg**

```
Duplicate OpenKM configuratoin to meet the new context.
```

- **server/default/conf/login-config.xml**

```
Duplicate OpenKM entry to meet the new context.
```

- **server/default/deploy/openkm-ds.xml**

```
Duplicate OpenKM datasource configuration to meet the new context.
```

- Remove jcr-1.0.jar from OpenKM.war and copy to **server/default/lib**

Configure JBoss service

These instructions are related to install and run JBoss 4.2.3 as a service, which means that will be launched at system boot and will closed properly on system shutdown.

Running as a Linux Service

For security reasons you shouldn't run JBoss as **root**. It is better to create an user named openkm and run JBoss from him:

```
$ sudo adduser openkm
```

Create a file with the script:

```
$ sudo vim /etc/init.d/jboss
```

```
#!/bin/sh

### BEGIN INIT INFO
# Provides:          jboss
# Required-Start:    $remote_fs $syslog
# Required-Stop:     $remote_fs $syslog
# Default-Start:     2 3 4 5
# Default-Stop:      0 1 6
# Short-Description: Start and stop JBoss AS
# Description:       Enable JBoss AS service provided by daemon.
### END INIT INFO

ECHO=/bin/echo
TEST=/usr/bin/test
JBOSS_USER=openkm
JBOSS_IP=0.0.0.0
JBOSS_HOME=/home/openkm/jboss-4.2.3.GA
```

```
JBOSS_START_SCRIPT=$JBOSS_HOME/bin/run.sh
JBOSS_STOP_SCRIPT=$JBOSS_HOME/bin/shutdown.sh

$TEST -x $JBOSS_START_SCRIPT || exit 0
$TEST -x $JBOSS_STOP_SCRIPT || exit 0

start() {
    $ECHO -n "Starting JBoss"
    su - $JBOSS_USER -c "$JBOSS_START_SCRIPT -b $JBOSS_IP > /dev/null 2>
/dev/null &"
    $ECHO "."
}

stop() {
    $ECHO -n "Stopping JBoss"
    su - $JBOSS_USER -c "$JBOSS_STOP_SCRIPT -s $JBOSS_IP > /dev/null &"
    while [ "$(ps -fu $JBOSS_USER | grep java | grep jboss | wc -l)"
-gt "0" ]; do
        sleep 5; $ECHO -n "."
    done
    $ECHO "."
}

case "$1" in
    start)
        start
        ;;
    stop)
        stop
        ;;
    restart)
        stop
        sleep 30
        start
        ;;
    *)
        $ECHO "Usage: jboss {start|stop|restart}"
        exit 1
esac
exit 0
```

And make it executable:

```
$ sudo chmod 755 /etc/init.d/jboss
```

Now update the run-levels:

```
$ sudo update-rc.d jboss defaults
 Adding system startup for /etc/init.d/jboss ...
```

```
   /etc/rc0.d/K20jboss -> ../init.d/jboss
   /etc/rc1.d/K20jboss -> ../init.d/jboss
   /etc/rc6.d/K20jboss -> ../init.d/jboss
   /etc/rc2.d/S20jboss -> ../init.d/jboss
   /etc/rc3.d/S20jboss -> ../init.d/jboss
   /etc/rc4.d/S20jboss -> ../init.d/jboss
   /etc/rc5.d/S20jboss -> ../init.d/jboss
```

start jboss service:

```
$ sudo service jboss start
```

stop jboss service:

```
$ sudo service jboss stop
```

> If you are a Red Hat user use:
>
> ```
> $ chkconfig --add jboss
> ```
>
> Take a look at update-rc.d like command on Redhat Enterprise / CentOS Linux [1].

Also is a good idea to configure JBoss memory utilization. Edit the file *$JBOSS_HOME/bin/run.sh* and add a new parameter **JAVA_OPTS** where you can increase the system memory managed by the JVM (Java Virtual Machine):

```
DIRNAME=`dirname $0`
PROGNAME=`basename $0`
GREP="grep"
JAVA_OPTS="-Xmx1744m -XX:MaxPermSize=256m"
```

This example is for a system with 2 GB of RAM.

For more info, read http://jboss.org/community/docs/DOC-11566.

Configure on a Redhat / CentOS

Copy the jboss_init_redhat.sh located into $JBOSS_HOME/bin to /etc/init.d/jboss

```
$ cp bin/jboss_init_redhat.sh /etc/init.d/jboss
```

Make jboss script executable

```
$ chmod +x /etc/init.d/jboss
```

Modify default jboss script (add or modify some lines)

```
# chkconfig: - 64 36
# description: Jboss Start|Restart|Stop Application Server
# pidfile: /var/run/jboss.pid

....

JBOSS_HOME=${JBOSS_HOME:-"/home/openkm/jboss-4.2.3.GA"}

#define the user under which jboss will run, or use 'RUNASIS' to run as the current user
```

```
JBOSS_USER=${JBOSS_USER:-"openkm"}


#make sure java is in your path
JAVAPTH=${JAVAPTH:-"/usr/java/jdk1.6.0_27/bin"}


#configuration to use, usually one of 'minimal', 'default', 'all'
JBOSS_CONF=${JBOSS_CONF:-"default"}


#if JBOSS_HOST specified, use -b to bind jboss services to that address
JBOSS_HOST="0.0.0.0"
JBOSS_BIND_ADDR=${JBOSS_HOST:+"-b $JBOSS_HOST"}
```

Add jboss script as service

```
$ chkconfig --add jboss
```

If you see an error message like **service jboss does not support chkconfig**, edit /etc/init.d/jboss file and add the following lines right after "#!/bin/sh":

```
# chkconfig: 345 90 10
# description: Runs the JBoss Application Server
# processname: jboss
```

Change service level

```
$ chkconfig jboss --level 3 on
```

Run or stop jboss service

```
$ service jboss start
$ service jboss stop
```

Running as a Windows Service

You can install JBoss as a service in Windows platform by several ways:

First method

JBoss recomended is using jboss **native windows take** a look here http:/ / community. jboss. org/ wiki/ RunJBossAsAServiceOnWIndows.Must be downloaded **jboss native** in concordance with your architecture 32 bits or 64 bits http://labs.jboss.com/jbossweb/downloads/(jboss native).

- Uncompress files into your %JBOSS_HOME%
- If you want you can edit the service name editing the file %JBOSS_HOME%\bin\service.bat

```
set SVCNAME=JBAS50SVC
set SVCDISP=JBoss Application Server 5.0
set SVCDESC=JBoss Application Server 5.0.0 GA/Platform: Windows x86
```

- Then you must go to your windows console into %JBOSS_HOME%\bin\ folder and execute

```
service.bat install
```

- Now in your service control panel there's installed JBoss as a service. We tested it on several Windows and runs well.

> If you're executing **service.bat** on **Windows Vista / Windows 7** maybe you obtain **Access Denied error**. The command shell works slightly differently in Vista. By default when you run cmd.exe, it does not give administrator rights. To get around it, create a shortcut to cmd.exe on desktop, open Properties, Compatibility tab and check box for Administrator rights. Or simply execute the shortcut as administrator (popup menu with right click). Then you can execute **service.bat install**.

> Other windows service installation methods explained below we've not been tested directly by us.

Second method

- Download JavaService-2.0.10.zip from http://download.forge.objectweb.org/javaservice/
- Extract to C:\JavaService-2.0.10
- copy InstallJBoss.bat，UninstallJBoss.bat， JavaService.exe (JavaService.exe must be renames as JBossService.exe) to jboss bin directory. In our case C:\jboss-4.2.3.GA\
- Make sure JBOSS_HOME and JAVA_HOME are set. In our case: C:\jboss-4.2.3.GA and D:\jdk1.6.0 respectively
- Create a batch file install.bat in %JBOSS_HOME%/bin directory. Its contents should be as follows:

```
Rem Start Here


set JAVA_HOME=d:/jdk1.6.0
set javadll=%JAVA_HOME%\jre\bin\client\jvm.dll
set javatool=%JAVA_HOME%\lib\tools.jar
set javarun=%JBOSS_HOME%\bin\run.jar
set outlog=%JBOSS_HOME%\bin\stdout.log
set errlog=%JBOSS_HOME%\bin\stderr.log


JBossService.exe -install JBoss "%javadll%"
-Djava.class.path="%javatool%";"%javarun%" -Xms1024M -Xmx1024M
-XX:PermSize=64m -XX:MaxPermSize=128m -start org.jboss.Main -params -b
0.0.0.0 -stop org.jboss.Main -method systemExit -out "%outlog%" -err
"%errlog%" -current "%JBOSS_HOME%\bin" -manual


Rem End Here
```

- Make sure to set word wrap off in the file and do not forget the quotes. This will allow the Program

Files path to work correctly. You can change the location of stderr.log and stdout.log to other path if you prefer.

- Start a cmd window and cd to %JBOSS_HOME%\bin. run install.bat by typing install. This will install JBoss as a service.

To start the service type:

```
net start JBoss.
```

- JBoss should now be running as a service.
- To uninstall the JBoss service:

```
javaservice -uninstall jboss
```

Third method

Start a cmd window,To start the service type:

```
set JAVA_HOME=d:/jdk1.6.0
set JBOSS_HOME=C:/jboss-4.2.3.GA/
installJboss MSSQLSERVER -auto
```

MSSQLSERVER is a service that must be running before JBoss starts.

Fourth method

Is using Microsoft Resource KIT, you can get more information here:

- http://support.microsoft.com/kb/137890
- Windows XP Embedded SP2 Resource Kit [2]

Microsoft resource kit are several programs that covers specific functionalities, not present by default in Windows OS (really it could be a good question why it don't comes by default - other Microsoft policy? the question is that in these utilities are present service registration utility).

Mac OS X Server

Create a file called jboss.plist with values (some paths, adminserver user name could be different) save into /Users/adminuser/Library/LaunchAgents

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE plist PUBLIC "-//Apple Computer//DTD PLIST 1.0//EN"
                       "http://www.apple.com/DTDs/PropertyList-1.0.dtd">
<plist version="1.0">
  <dict>
    <key>Label</key>
    <string>openkm</string>

    <key>UserName</key>
    <string>adminuser</string>

    <key>GroupName</key>
    <string>staff</string>

    <key>ProgramArguments</key>
    <array>
      <string>/Users/adminserver/jboss-4.2.3.GA/bin/run.sh</string>
      <string>-c</string>
      <string>default</string>
      <string>-b</string>
      <string>0.0.0.0</string>
    </array>

    <key>RunAtLoad</key>
    <true/>
  </dict>
```

```
</plist>
```

Change the user owner to root and privilegees to 644

```
$ sudo chown root jboss.plist
$ sudo chmod 644 jboss.plist
```

Register as service

```
$ launchctl load <path-to-file>/jboss.plist
```

Start service

```
$ launchctl start openkm
```

Stop service

```
$ launchctl stop openkm
```

More information at Running JBoss As A Mac OS X Service [3].

JBoss with SSL

We encorage using apache proxy to port 80 or 443 and not expose directly jboss ports, due could be some security bugs doing it. For each people who decide exposing jboss ports there's some problem with IE and https / SSL jboss configuration. To solve it:

Edit the file $JBOSS_HOME/server/default/deploy/jboss-web.deployer/server.xml and uncomment

```
<Connector port="8443" protocol="HTTP/1.1" SSLEnabled="true"
           maxThreads="150" scheme="https" secure="true"
            keystorePass="XxXxXxXx"
           clientAuth="false" sslProtocol="TLS" />
```

You need to generate some keystorePass.

References

[1] http://www.cyberciti.biz/faq/rhel5-update-rcd-command/
[2] http://www.microsoft.com/downloads/details.aspx?FamilyID=F4A453E2-FCEF-49C2-87AC-CEBE233A5F8D&displaylang=en&
    displaylang=en|Microsoft
[3] http://community.jboss.org/wiki/RunningJBossAsAMacOSXService

Several JBoss in the same server

Using multiple IPs

If your server has several IP addresses assigned, the preferred solution is to use the -b command line option to bind each instance of JBoss AS to a particular IP address:

```
./run.sh -b 192.168.0.10
```

See also:

- Create virtual interface in Ubuntu [1]
- Add Virtual Interfaces in Linux [2]

Only one IP

Modify *$JBOSS_HOME/server/default/conf/jboss-service.xml* and uncomment the "Service Binding" section and select a "ServerName" value from **sample-bindings.xml**. See *$JBOSS_HOME/docs/examples/binding-manager/sample-bindings.xml* "ports-default" entries for all JBoss port properties.

Service	Default port	New Port	Description
jboss.remoting:type=Connector,name=DefaultEjb3Connector,handler=ejb3	3873	3973	EJB3
jboss:service=Naming	1098	1198	JNDI
jboss:service=WebService	8083	8183	web service
jboss:service=invoker,type=jrmp	4444	4544	
jboss:service=invoker,type=pooled	4445	4545	
jboss:service=HAJNDI	1100, 1101	1200, 1201	HA JNDI / clusters
jboss:service=invoker,type=pooledha	4448	4548	clusters
jboss:service=CorbaORB	19001	19101	Corba
jboss.jmx:name=SnmpAgent,service=trapd,type=logger	1162	1262	SNMP
jboss.jmx:name=SnmpAgent,service=snmp,type=adaptor	1161	1261	SNMP
jboss.mq:service=InvocationLayer,type=UIL2	8093	8193	JMS
jboss.mq:service=InvocationLayer,type=HTTP	8080	8180	JMS HTTP
jboss.mq:service=JMSProviderLoader,name=HAJNDIJMSProvider	1100	1200	HA JMS
jboss.web:service=WebServer	8080, 8443	8180, 8543	HTTP web server
jboss.messaging:service=Connector,transport=bisocket	4457	4557	JBoss Messaging

> By default there are several port mappings alredy defined which can be used: *ports-default*, *ports-01*, *ports-02* and *ports-03*.

To shutdown the different instances execute the **shutdown.sh** script from the *$JBOSS_HOME/bin* directory with extra arguments. If you have configured 2 instances of JBoss, one on port 1099 (the NamingService port) and another on 1199, then in order to shutdown the first JBoss instance use this command:

```
./bin/shutdown.sh -s jnp://localhost:1099
```

And for the second instance, use this command:

```
./bin/shutdown.bat -s jnp://localhost:1199
```

More info at:

- http://docs.jboss.org/jbossas/javadoc/4.0.2/org/jboss/Shutdown.java.html
- http://community.jboss.org/wiki/ConfiguringMultipleJBossInstancesOnOneMachine
- http://community.jboss.org/wiki/ConfigurePorts

References

[1] http://ubuntuforums.org/showthread.php?t=555319
[2] http://lazysystemadmin.blogspot.com/2010/05/add-virtual-interfaces-in-linux-quick.html

Several JBoss instances in the same server

You can have several JBoss instances in the same server, and only need to modify some default JBoss ports.

- Edit *$JBOSS_HOME/server/default/deploy/jbossweb-tomcat55.sar/server.xml* and change these ports: 8080, 8009 and 8443.
- Edit *$JBOSS_HOME/server/default/conf/jboss-service.xml* and change these ports: 8083, 4444, 4445,4446, 1098 and 1099.
- Edit *$JBOSS_HOME/server/default/deploy/jms/uil2-service.xml* and change this port: 8093.
- Edit *$JBOSS_HOME/server/default/deploy/ejb3.deployer/META-INF/jboss-service.xml* and change this port: 3873.
- Edit *$JBOSS_HOME/bin/shutdown.sh* and add to the last "localhost" line the new HTTP port.

```
# Execute the JVM
exec "$JAVA" \
    $JAVA_OPTS \
    -classpath $JBOSS_CLASSPATH \
    org.jboss.Shutdown localhost 8081 "$@"
```

> The simpliest way is increasing the default port by +1, keeping in mind the pairs 4444-4446 and 1098-1099 where your have to add +3,+2 because are consecutives.

Following this advice the news port will be:

- 8080 -> 8081 (Put this one in the shutdown.sh script)
- 8009 -> 8010
- 8443 -> 8444
- 8083 -> 8084
- 4444 -> 4447 (+3 because consecutive)
- 4445 -> 4448 (+3 because consecutive)
- 4446 -> 4449 (+3 because consecutive)
- 1098 -> 1100 (+2 because consecutive)
- 1099 -> 1101 (+2 because consecutive)
- 8093 -> 8094
- 3873 -> 3874

More info about this topic at:

- http://community.jboss.org/wiki/ConfigurePorts.

- http://community.jboss.org/wiki/VersionOfTomcatInJBossAS

Securing JBoss

When you run JBoss, OpenKM is accessible through the 8080 port. The bad news is that JBoss web console and JMX console is accessible and everybody can read and modify JBoss configuration. This is particularity dangerous if the OpenKM is accessible from Internet.

JMX Console

Edit the configuration file:

```
$ vim $JBOSS_HOME/server/default/deploy/jmx-console.war/WEB-INF/web.xml
```

Uncomment the block <security-constraint>. Edit this other configuration file:

```
$ vim $JBOSS_HOME/server/default/deploy/jmx-console.war/WEB-INF/jboss-web.xml
```

Uncomment the block defined by <security-domain>. Edit the users file:

```
$ vim $JBOSS_HOME/server/default/conf/props/jmx-console-users.properties
```

Change the password of the admin user.

Web Console

Edit the configuration file:

```
$ vim $JBOSS_HOME/server/default/deploy/management/console-mgr.sar/web-console.war/WEB-INF/web.xml
```

Uncomment the block <security-constraint>. Edit this other configuration file:

```
$ vim $JBOSS_HOME/server/default/deploy/management/console-mgr.sar/web-console.war/WEB-INF/jboss-web.xml
```

Uncomment the block defined by <security-domain>. Edit also this file:

```
$ vim $JBOSS_HOME/server/default/conf/login-config.xml
```

Change the path of the files web-console-users.properties and web-console-roles.properties (add props/ at the beginning of the file name)

```
<module-option name="usersProperties">props/web-console-users.properties
</module-option>
<module-option name="rolesProperties">props/web-console-roles.properties
</module-option>
```

And copy the files to the new location:

```
$ cp
$JBOSS_HOME/server/default/deploy/management/console-mgr.sar/web-console.war/WEB-INF/clas
 $JBOSS_HOME/server/default/conf/props
```

Now all the authentication files are located in the same folder. Edit the users file:

```
$ vim $JBOSS_HOME/server/default/conf/props/web-console-users.properties
```

Change the password of the admin user. The last step is restart JBoss:

```
$ /etc/init.d/jboss restart
```

For more info, visit:

- http://jboss.org/community/docs/DOC-12190.
- http://www.techienuggets.com/Detail?tx=9

Adding APR to JBoss

Out-of-the-box, JBoss AS 4.2.3 is a 100% pure Java application server, and embeds Tomcat 6 as its servlet container. By default, it will use Java to handle HTTP connections. With a few simple steps you can add the Apache Portable Runtime (APR) to achieve native HTTP processing performance.

If you run JBoss, you can notice certain console output that indicates no APR presence:

```
...
 17:33:23,852 INFO  [AprLifecycleListener] The Apache Tomcat Native
library which allows optimal performance in production environments was
 not found on the java.library.path:
/usr/java/java-6-sun-1.6.0.24/jre/lib/i386/server:/usr/java/java-6-sun-1.6.0.24/jre/lib/a
 17:33:23,973 INFO  [Http11Protocol] Initializing Coyote HTTP/1.1 on
http-127.0.0.1-8080
 17:33:23,974 INFO  [AjpProtocol] Initializing Coyote AJP/1.3 on
ajp-127.0.0.1-8009
 ...
 17:33:30,162 INFO  [Http11Protocol] Starting Coyote HTTP/1.1 on
http-127.0.0.1-8080
 17:33:30,181 INFO  [AjpProtocol] Starting Coyote AJP/1.3 on
ajp-127.0.0.1-8009
 17:33:30,202 INFO  [Server] JBoss (MX MicroKernel) [4.2.3.GA (build:
SVNTag=JBoss_4_2_3_GA date=200807181439)] Started in 23s:669ms
```

Stop JBoss and download JBoss Native from http://www.jboss.org/jbossweb/downloads.It is very important that you choose the right native-built libraries for your operating system. Extract the archive into $JBOSS_HOME:

```
$ cd jboss-4.2.3.GA
$ tar xzvf jboss-native-2.0.8-linux2-x64-ssl.tar.gz
```

Now you have a folder $JBOSS_HOME/bin/native with some dynamic libraries. Re-run JBoss AS, which will now detect the presence of the native libraries:

```
$ ./bin/run.sh
```

Notice certain console output that indicates the presence of the APR:

```
...
17:35:39,240 INFO  [AprLifecycleListener] Loaded Apache Tomcat Native library 1.1.18.
17:35:39,241 INFO  [AprLifecycleListener] APR capabilities: IPv6 [true], sendfile [true], accept filters [false], random [true].
17:35:39,511 INFO  [Http11AprProtocol] Initializing Coyote HTTP/1.1 on http-127.0.0.1-8080
17:35:39,512 INFO  [AjpAprProtocol] Initializing Coyote AJP/1.3 on ajp-127.0.0.1-8009
...
```

```
17:35:44,525 INFO  [Http11AprProtocol] Starting Coyote HTTP/1.1 on http-127.0.0.1-8080

17:35:44,544 INFO  [AjpAprProtocol] Starting Coyote AJP/1.3 on ajp-127.0.0.1-8009

17:35:44,566 INFO  [Server] JBoss (MX MicroKernel) [4.2.3.GA (build: SVNTag=JBoss_4_2_3_GA date=200807181439)] Started in 18s:462ms
```

In case it is not working edit the jboss-log4j.xml corresponding to your configuration add:

```
<category name="org.apache.catalina.core">
    <priority value="DEBUG"/>
</category>
```

Restart JBoss AS and look in server.log file and look for the exception near *org.apache.catalina.core.AprLifecycleListener.*

See also:

• http://anonsvn.jboss.org/repos/jbossnative/trunk/

Changing OpenKM timeout session

To change OpenKM timeout session you must add in web.xml file it (the number indicates the timeout in minutes). The web.xml file is located into OpenKM.war file (into WEB-INF folder). The OpenKM.war file is located in *$JBOSS_HOME/server/default/deploy* folder.

```
<session-config>
  <session-timeout>30</session-timeout>
</session-config>
```

Test OpenKM installation

Once you have configured OpenKM and JBoss you can start them. You can run JBoss from the command line from the $JBOSS_HOME folder:

```
$ ./bin/run.sh
```

This command will launch JBoss application server and start the OpenKM application. By default JBoss only listen at localhost, so OpenKM only will be accessible from http://localhost:8080/OpenKM. You will see a login screen like this:

You can log into OpenKM with okmAdmin user (default password is "admin"). If you want OpenKM to be accessible from others computers in your network, try to start JBoss with the -b parameter. This parameter means "bind" and with this command:

```
$ ./bin/run.sh –b 0.0.0.0
```

JBoss will bind to all network interfaces of the computer. Now OpenKM can be accessed from another computer at http://your-domain.com:8080/OpenKM.

	Don't close the terminal or JBoss will shutdown and OpenKM application will stop.

If you have configured JBoss service in your Debian / Ubuntu box, you can launch it this way:

```
$ /etc/init.d/jboss start
```

You can watch the server log to see OpenKM messages.

```
$ tail –f $JBOSS_HOME/server/default/log/server.log
```

Change default okmAdmin password

Once you have log into OpenKM you can see the application desktop. At the bottom-left you can see a red-and-white message which mean that you has logged as the administrator user and you must take care of your actions because can destroy the repository.
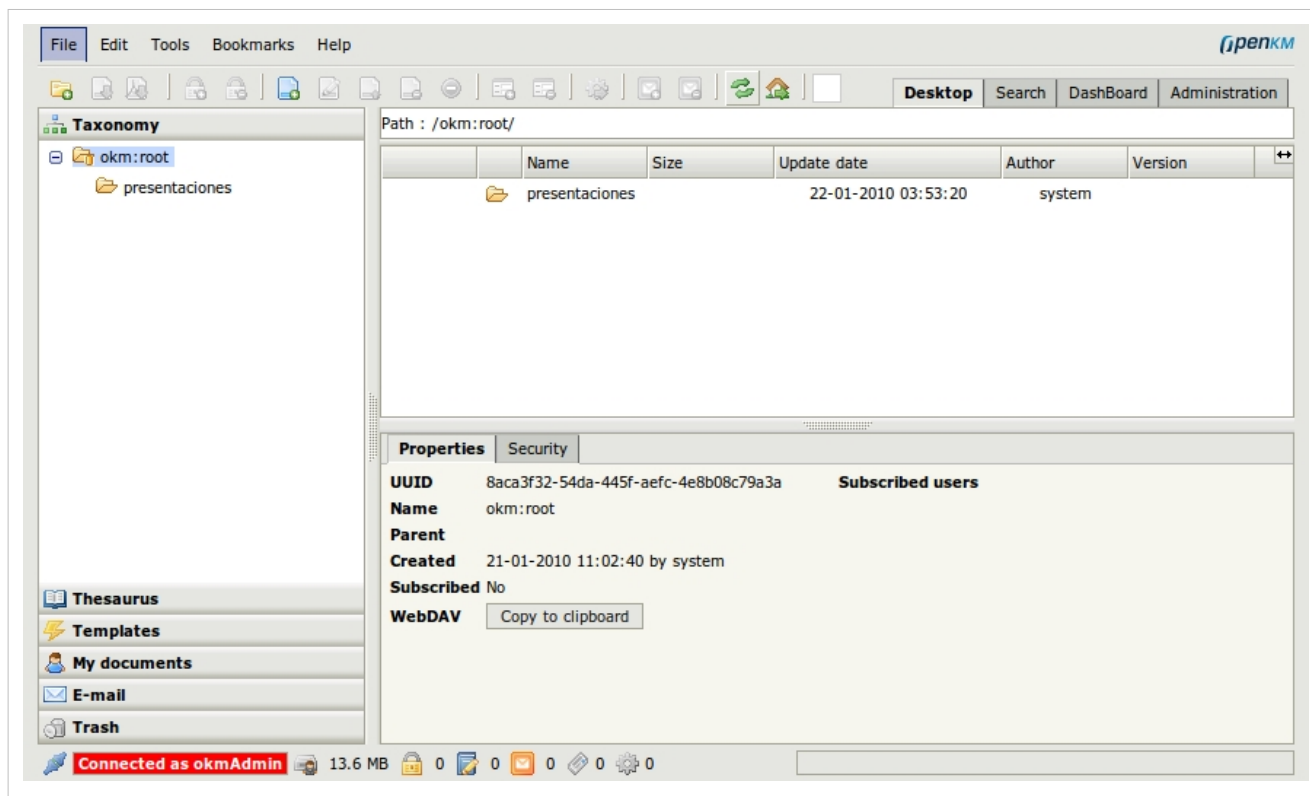
In the upper-right of the screen there is a tab called "Administration". This tab will be shown only if you log into OpenKM as the administration user. Click on this tab and you will see the OpenKM administration.
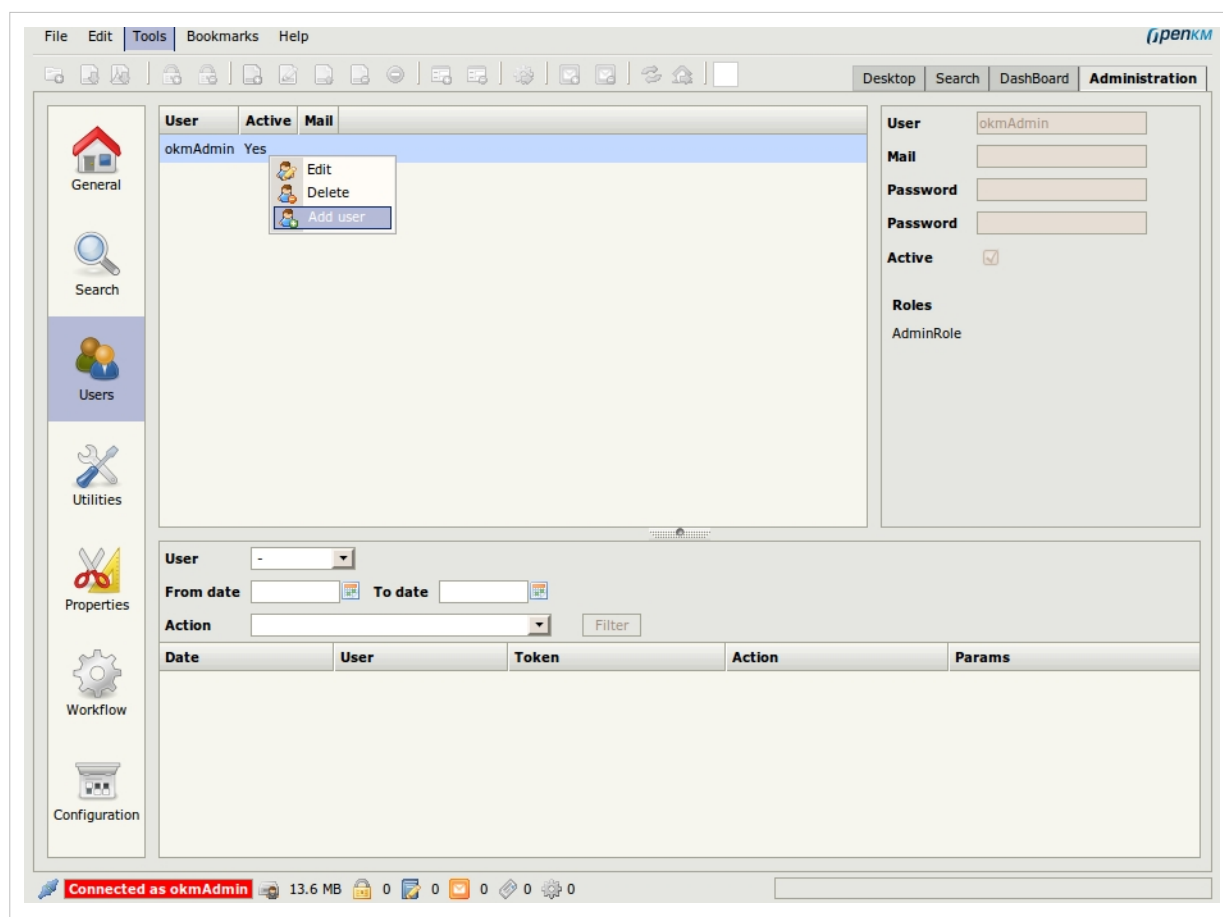
To manage users, click on Users icon. To modify the default okmAdmin password you have to right-click in the row and select Edit from the contextual menu. Here you can see the detail of the user modification form:

Here you can set an email and change the default password. Annotate the new password and keep it in a secured media.

Be careful and don't delete the AdminRole or deactivate the okmAdmin user because you won't be able to manage OpenKM.

Create users and roles

Log into your OpenKM installation at http://localhost:8080/OpenKM as okmAdmin user (default password is "admin"). Click on Administrator tab and click in User icon. Right-click to show the contextual menu an select the "Add user" option.
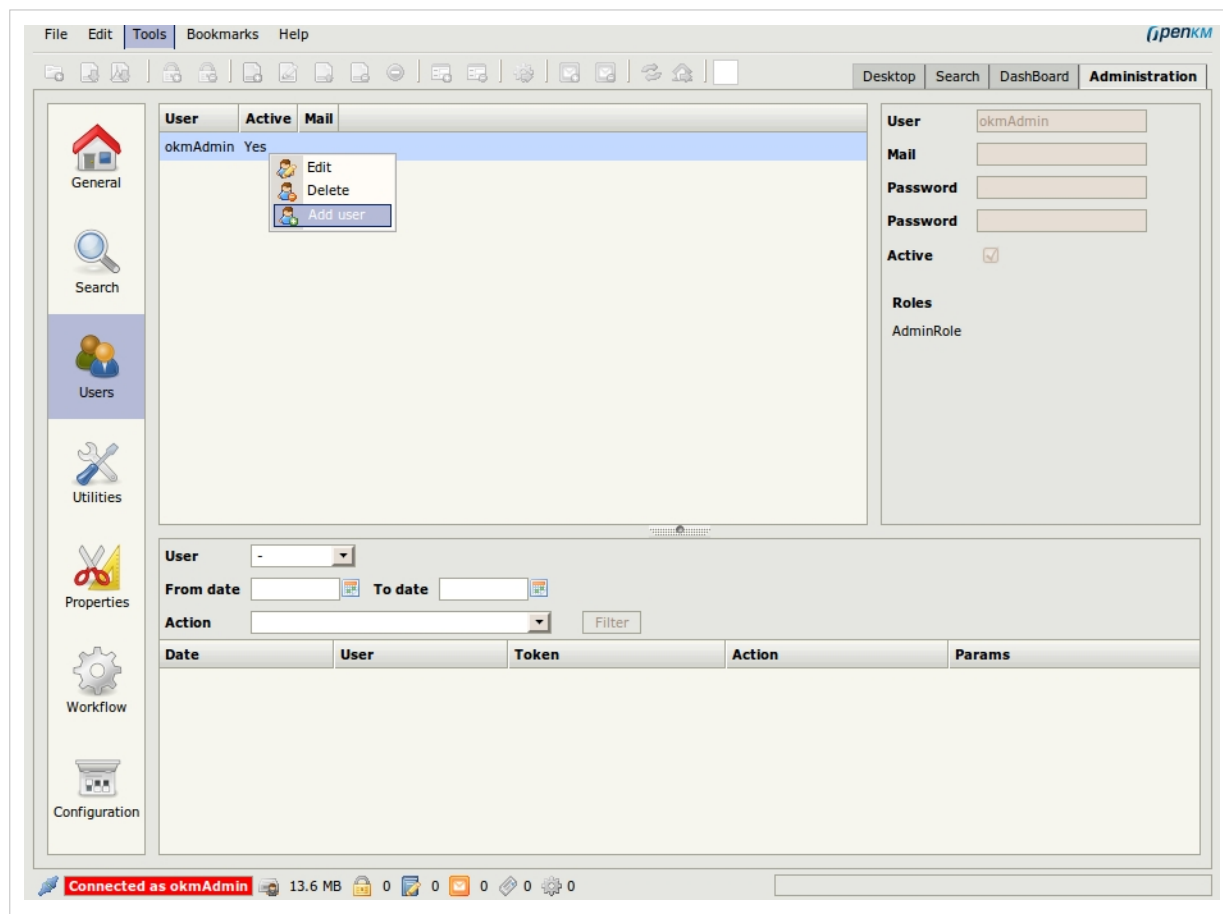
Set the user id, an mail account and a safety password. Every user in OpenKM need to be in the UserRole role. If you create an user and don't assign him the UserRole role, he won't be able to log into the application.

Additionally, you can create more roles to manage the repository security. An OpenKM user can be assigned to several roles.

If you assign an user the AdminRole role, this user will access to all the documents in OpenKM no matter the permissions they have.

Create password validator

You can create your own password validation in OpenKM. By default OpenKM has password validation disabled and there's used the **NoPasswordValidator** class. Also there's some default OpenKM password validation class called **CompletePasswordValidator** class.

Enabling **CompletePasswordValidator** class can configure minimum and maximum password length, number of lowercase and uppercase characters, number of numeric characters and number of special characters that appearing in password. You can see more information at Application configuration.

You can also create your own password class that must implements the PasswordValidator interface:

```java
public interface PasswordValidator {

  /**
   * Validate
   *
   * @param password
   * @throws ValidatorException
   */
  public void Validate(String password) throws ValidatorException;


}
```

CompletePasswordValidator class:

```java
/**
 * Complex password validator
 */
public class CompletePasswordValidator implements PasswordValidator {
  @SuppressWarnings("unused")
  private static Logger log =
LoggerFactory.getLogger(CompletePasswordValidator.class);

  @Override
  public void Validate(String password) throws ValidatorException {
    validateLength(password);
    checkLowerCase(password);
    checkUpperCase(password);
    checkDigits(password);
    checkSpecial(password);
  }

  /**
   * Validate length
   */
  private void validateLength(String password) throws
ValidatorException {
    if (Config.VALIDATOR_PASSWORD_MIN_LENGTH > 0 &&
        password.length() < Config.VALIDATOR_PASSWORD_MIN_LENGTH) {
```

```java
        throw new
ValidatorException(Config.VALIDATOR_PASSWORD_ERROR_MIN_LENGTH);
    }

    if (Config.VALIDATOR_PASSWORD_MAX_LENGTH > 0 &&
        password.length() > Config.VALIDATOR_PASSWORD_MAX_LENGTH) {
      throw new
ValidatorException(Config.VALIDATOR_PASSWORD_ERROR_MAX_LENGTH);
    }
  }

  /**
   * Validate lowercase characters
   */
  private void checkLowerCase(String password) throws
ValidatorException {
    int count = 0;

    if (Config.VALIDATOR_PASSWORD_MIN_LOWERCASE > 0) {
      for (int i=0; i<password.length(); i++) {
        if (Character.isLowerCase(password.charAt(i))) {
          count ++;
        }
      }

      if (Config.VALIDATOR_PASSWORD_MIN_LOWERCASE > count) {
        throw new
ValidatorException(Config.VALIDATOR_PASSWORD_ERROR_MIN_LOWERCASE);
      }
    }
  }

  /**
   * Validate uppercase characters
   */
  private void checkUpperCase(String password) throws
ValidatorException {
    int count = 0;

    if (Config.VALIDATOR_PASSWORD_MIN_UPPERCASE > 0) {
      for (int i=0; i<password.length(); i++) {
        if (Character.isUpperCase(password.charAt(i))) {
          count ++;
        }
      }

      if (Config.VALIDATOR_PASSWORD_MIN_UPPERCASE > count) {
```

```java
      throw new
ValidatorException(Config.VALIDATOR_PASSWORD_ERROR_MIN_UPPERCASE);
      }
    }
  }

  /**
   * Validate digits
   */
  private void checkDigits(String password) throws ValidatorException {
    int count = 0;

    if (Config.VALIDATOR_PASSWORD_MIN_DIGITS > 0) {
      for (int i=0; i<password.length(); i++) {
        if (Character.isDigit(password.charAt(i))) {
          count ++;
        }
      }

      if (Config.VALIDATOR_PASSWORD_MIN_DIGITS > count) {
        throw new
ValidatorException(Config.VALIDATOR_PASSWORD_ERROR_MIN_DIGITS);
      }
    }
  }

  /**
   * Validate special characters
   */
  private void checkSpecial(String password) throws ValidatorException
{
    int count = 0;

    if (Config.VALIDATOR_PASSWORD_MIN_SPECIAL > 0) {
      for (int i=0; i<password.length(); i++) {
        if (!Character.isLetterOrDigit(password.charAt(i)) &&
          !Character.isWhitespace(password.charAt(i))) {
          count ++;
        }
      }

      if (Config.VALIDATOR_PASSWORD_MIN_SPECIAL > count) {
        throw new
ValidatorException(Config.VALIDATOR_PASSWORD_ERROR_MIN_SPECIAL);
      }
    }
  }
```

```
}
```

Notification and subscription messages

When you are subscribed to a document, any change in the document will be notified by email. These email messages are created using some defined templates. You can modify these default templates to create messages adapted to your company style. This modification is done editing the file OpenKM.cfg.

> **In OpenKM 5.1 Velocity will be replaced by FreeMarker.**

These are the default values:

notification.message.subject	OpenKM - NOTIFICATION - ${documentName}
notification.message.body	\Document: \\${documentPath}\\ \User: \${userId}\ \Message: \${notificationMessage}\
subscription.message.subject	OpenKM - ${eventType} - ${documentPath}
subscription.message.body	\Document: \\${documentPath}\\ \User: \${userId}\ \Event: \${eventType}\ \Comment: \${subscriptionComment}\
subscription.twitter.status	OpenKM - ${documentUrl} - ${documentPath} - ${userId} - ${eventType}

And these are the substitution variables used:

Notification message subject & body

- ${documentUrl} - Document URL
- ${documentPath} - Document path
- ${documentName} - Document name
- ${userId} - User id
- ${notificationMessage} - Notification message

Subscription message subject & body & twitter

- ${documentUrl} - Document URL
- ${documentPath} - Document path
- ${documentName} - Document name
- ${userId} - User id
- ${eventType} - Event type
- ${subscriptionComment} - Subscription comment

> Take care of the OpenKM.cfg format: this is a Java properties file and has some limitations, for example, this file have to be coded in ISO-8859-1 format. Also single quotes (') should be protected with another single quote (").

More info on this can be found at:

- http://java.sun.com/j2se/1.5.0/docs/api/java/text/MessageFormat.html
- http://www.w3schools.com/tags/ref_entities.asp

If you want to elaborate a more complex template, the limitations of Java Properties are a disadvantage. In this case you can set the **subscription.message.subject** property to a template filename (for example, ComplexSubcriptionSubject.ftl). The template file should have the **.ftl** extension and needs to be locate at

$JBOSS_HOME directory.

> FreeMarker does not allow you to render null values (will throw an exception). If you expect a value to be null you can also use the ! operator: ${content.title!}. This is the equivalent to ${content.title!*}*.

More info about FreeMarker templates at FreeMarker User Guide [1].

OpenKM 5.0.x

These are the default values:

notification.message.subject	OpenKM - NOTIFICATION - $documentName
notification.message.body	Document: $documentPath User: $userId Message: $notificationMessage
subscription.message.subject	OpenKM - $eventType - $documentPath
subscription.message.body	Document: $documentPath User: $userId Event: $eventType Comment: $subscriptionComment
subscription.twitter.status	OpenKM - $documentUrl - $documentPath - $userId - $eventType

And these are the substitution variables used:

Notification message subject & body

- $documentUrl - Document URL
- $documentPath - Document path
- $documentName - Document name
- $userId - User id
- $notificationMessage - Notification message

Subscription message subject & body & twitter

- $documentUrl - Document URL
- $documentPath - Document path
- $documentName - Document name
- $userId - User id
- $eventType - Event type
- $subscriptionComment - Subscription comment

> Take care of the OpenKM.cfg format: this is a Java properties file and has some limitations, for example, this file have to be coded in ISO-8859-1 format. Also single quotes (') should be protected with another single quote (").

More info on this can be found at:

- http://java.sun.com/j2se/1.5.0/docs/api/java/text/MessageFormat.html
- http://www.w3schools.com/tags/ref_entities.asp

If you want to elaborate a more complex template, the limitations of Java Properties are a disadvantage. In this case you can set the **subscription.message.subject** property to a template filename (for example, ComplexSubcriptionSubject.vm). The template file should have the **.vm** extension and needs to be locate at *$JBOSS_HOME* directory.

More info about Velocity templates at Apache Velocity User Guide [2].

OpenKM 4.0 and older

These are the default values:

notify.message.subject	OpenKM - NOTIFICATION - {0}
notify.message.body	Document: {1}
User: {3}
Message:{4}
subscription.message.subject	OpenKM - {0} - {1}
subscription.message.body	Document: {1}
User: {3}
Event:{4}
Comment: {5}

And these are the substitution variables used:

NOTIFY MESSAGE SUBJECT	NOTIFY MESSAGE BODY
{0} - document path	{0} - document URL
{1} - document name	{1} - document path
	{2} - document name
	{3} - user id
	{4} - message

SUBSCRIPTION MESSAGE SUBJECT	SUBSCRIPTION MESSAGE BODY
{0} - event type	{0} - document URL
{1} - document path	{1} - document path
{2} - document name	{2} - document name
	{3} - user id
	{4} - event type

References

[1] http://freemarker.sourceforge.net/docs

[2] http://velocity.apache.org/engine/releases/velocity-1.6.2/user-guide.html

OpenKM authentication

Authentication (from Greek: αυθεντικός; real or genuine, from authentes; author) is the act of establishing or confirming something (or someone) as authentic, that is, that claims made by or about the subject are true. This might involve confirming the identity of a person, the origins of an artifact, or assuring that a computer program is a trusted one. This task is addressed by JAAS.

JAAS uses a service provider approach to its authentication features, meaning that it is possible to configure different login modules for an application without changing any code. The application remains unaware of the underlying authentication logic. It's even possible for an application to contain multiple login modules, somewhat akin to a stack of authentication procedures.

> Read Debugging JAAS configuration to learn how to debug a problematic JAAS configuration.

OpenKM relies the authentication on the standard JAAS implemented in JBoss application server. JBoss comes with some interesting modules which can be used to authenticate against a plain-text file, a database or an LDAP, for example. On recent versions, OpenKM uses the DatabaseServerLoginModule class to manage authentication.

> The JBoss security is configured in the file *$JBOSS_HOME/server/default/conf/login-config.xml*.

Also remember the principal.adapter configuration option. OpenKM need this configuration to create a list of users and roles available in the changing permissions dialog. This is done by the **DatabasePrincipalAdapter** class. This is an implementation of the **com.openkm.principal.PrincipalAdapter** interface:

```java
public interface PrincipalAdapter {
    /**
     * Method to retrieve all users from a authentication source.
     *
     * @return A Collection with all the users.
     * @throws PrincipalAdapterException If any error occurs.
     */
    public Collection<String> getUsers() throws PrincipalAdapterException;

    /**
     * Method to retrieve all roles from a authentication source.
     *
     * @return A Collection with all the roles.
     * @throws PrincipalAdapterException If any error occurs.
     */
    public Collection<String> getRoles() throws PrincipalAdapterException;

    /**
     * Method to retrieve the mail from a list of users.
     *
     * @param users A list of users.
     * @return A list of user emails.
     * @throws PrincipalAdapterException If any error occurs.
     */
```

```
    public Collection<String> getMails(Collection<String> users) throws
PrincipalAdapterException;
}
```

Roles

In OpenKM are defined by default two roles **AdminRole** and **UserRole**.

UserRole is mandatory for all users, because is internally used by OpenKM for connection grant purpose. Without this grant users can not connect to OpenKM and you'll get a 403 status code error.

You can set AdminRole to any user, and it'll get administrator privileges seeing any folder and making any operation without retrictions. Users with AdminRole have access to administrator tab in UI.

Plain-text file

This is the simplest security configuration. This was the default authentication method in older OpenKM versions. It is achieved using the JBoss UsersRolesLoginModule login module. User are stored in the file *$JBOSS_HOME/server/default/conf/props/openkm-users.properties* in this form:

```
user1=pass1
user2=pass2
...
```

The password in not encrypted. The roles are in the file *$JBOSS_HOME/server/default/conf/props/openkm-roles.properties* in this form:

```
user1=UserRole,Rol1,Rol2,...
user1=UserRole,Rol1,Rol2,...
...
```

This is the JBoss configuration for this method:

```xml
<application-policy name = "OpenKM">
   <authentication>
     <login-module code="org.jboss.security.auth.spi.UsersRolesLoginModule" flag = "required">
        <module-option name="usersProperties">props/openkm-users.properties</module-option>
        <module-option name="rolesProperties">props/openkm-roles.properties</module-option>
     </login-module>
     <login-module code="org.jboss.security.ClientLoginModule" flag="required" />
   </authentication>
</application-policy>
```

The principal.adapter should be set to es.git.openkm.principal.UsersRolesPrincipalAdapter.

Database

This is the default security configuration for recent OpenKM version. Is a good option because simplifies user and role management: now user and roles can be managed from OpenKM administration. This module connect to the database using a data-source.

```xml
<application-policy name = "OpenKM">

  <authentication>

    <login-module code="org.jboss.security.auth.spi.DatabaseServerLoginModule" flag = "required">

      <module-option name="dsJndiName">java:/OKMAuthDS</module-option>

      <module-option name="principalsQuery">select usr_pass as PASSWD from users where usr_id=? and
usr_active='true'</module-option>

      <module-option name="rolesQuery">select ur_role as ROLEID, 'Roles' from user_role where ur_user=?</module-option>

      <module-option name="hashAlgorithm">md5</module-option>

      <module-option name="hashEncoding">hex</module-option>

    </login-module>

  </authentication>

</application-policy>
```

The principal.adapter should be set to *com.openkm.principal.DatabasePrincipalAdapter*, which is the default value.

LDAP (Active Directory, Open Directory)

You can get LDAP integration through the LdapExtLoginModule login module.

```xml
<application-policy name="OpenKM">

  <authentication>

    <login-module code="org.jboss.security.auth.spi.LdapExtLoginModule" flag="required" >

      <module-option name="java.naming.provider.url">ldap://my-company.com:389</module-option>

      <module-option name="bindDN">cn=My_adm_account,ou=Admin Accounts,dc=my-company,dc=br</module-option>

      <module-option name="java.naming.security.authentication">simple</module-option>

      <module-option name="bindCredential">My_adm_account_password</module-option>

      <module-option name="baseCtxDN">ou=Users Accounts,dc=my-company,dc=com</module-option>

      <module-option name="baseFilter">(sAMAccountName={0})</module-option>

      <module-option name="rolesCtxDN">ou=Users Accounts,dc=my-company,dc=com</module-option>

      <module-option name="roleFilter">(sAMAccountName={0})</module-option>

      <module-option name="roleAttributeID">memberOf</module-option>

      <module-option name="roleAttributeIsDN">true</module-option>

      <module-option name="roleNameAttributeID">cn</module-option>

      <module-option name="roleRecursion">-1</module-option>

      <module-option name="searchScope">SUBTREE_SCOPE</module-option>

      <module-option name="defaultRole">UserRole</module-option>

    </login-module>

  </authentication>

</application-policy>
```

Here are some configuration comments:

- **bindDN**: This is some DN with read/search permissions on the baseCtxDN and rolesCtxDN.
- **bindCredential**: The password for the bindDN.
- **baseCtxDN**: The fixed DN of the context to start the user search from.

- **rolesCtxDN**: The fixed DN of the context to search for user roles.

Don't forget the <module-option name="defaultRole">UserRole</module-option> (adds this role to every authenticated user, because only users with that role are allowed to access OpenKM).

See also:

- Active Directory 🔒
- Open Directory 🔒
- Testing LDAP configuration 🔒
- Forum: Usuario administrador LDAP v.4 [1]

Changes from version 3.0. to 4.0

- **UserRol** now is called **UserRole**
- **AdminRol** now is called **AdminRole**

More information

More information about JASS and other login modules can be found at:

- JBoss and JAAS debug [2]
- JBoss Community: SecurityFAQ [3]
- JBoss Community: JBossSX [4]
- JBoss Community: LdapExtLoginModule [5]
- JBoss Community: OSXOpenDirectoryLoginConfig.xml [6]

References

[1] http://forum.openkm.com/viewtopic.php?f=13&t=3535

[2] http://primalcortex.wordpress.com/2007/11/28/jboss-and-jaas-debug/

[3] http://community.jboss.org/wiki/SecurityFAQ

[4] http://community.jboss.org/wiki/JBossSX

[5] http://community.jboss.org/wiki/LdapExtLoginModule

[6] http://community.jboss.org/wiki/OSXOpenDirectoryLoginConfigxml

Active Directory

To configure Active Directory we must make some changes in OpenKM.cfg configuration file and in login-config.xml file that can be found at *$JBOSS_HOME/server/default/conf*. For both changes you need to restart JBoss server.

OpenKM.cfg file example (you must change **192.168.0.6, Administrador, password and weyler** values to your active directory values)

```
system.login.lowercase=on
principal.adapter=com.openkm.principal.LdapPrincipalAdapter

principal.ldap.server=ldap://192.168.0.6
principal.ldap.security.principal=CN=Administrator,cn=users,dc=weyler,dc=local
principal.ldap.security.credentials=password

principal.ldap.user.search.base=cn=users,dc=weyler,dc=local
principal.ldap.user.search.filter=(objectclass=person)
principal.ldap.user.attribute=cn

principal.ldap.role.search.base=cn=users,dc=weyler,dc=local
principal.ldap.role.search.filter=(objectclass=group)
principal.ldap.role.attribute=cn

principal.ldap.mail.search.base=cn={0},cn=users,dc=weyler,dc=local
principal.ldap.mail.search.filter=(objectclass=person)
principal.ldap.mail.attribute=mail

principal.ldap.users.by.role.search.filter=(&(objectClass=group)(cn={0}))
principal.ldap.roles.by.user.search.filter=(&(objectClass=group)(cn={0}))
```

Starting with **OpenKM 5.0.4** we added more "users by role" and "roles by user" configuration properties:

```
principal.ldap.users.by.role.search.base=cn={0},cn=users,dc=weyler,dc=local
principal.ldap.users.by.role.search.filter=(objectclass=group)
principal.ldap.users.by.role.attribute=member

principal.ldap.roles.by.user.search.base=cn={0},cn=users,dc=weyler,dc=local
principal.ldap.roles.by.user.search.filter=(objectClass=person)
principal.ldap.roles.by.user.attribute=memberOf
```

OpenKM 4.1 and older

```
principal.adapter=es.git.openkm.principal.LdapPrincipalAdapter
principal.ldap.user.atribute=cn
principal.ldap.role.atribute=cn
principal.ldap.mail.atribute=mail
```

In case of Active directory (windows) it's important that all users login be in lower case, for it purpose we enable

```
system.login.lowercase=on
```

property in OpenKM.cfg. The reason is so simply, Windows not makes any difference between upper or lower case validating user name credentials.

login-config.xml file example (you must change **192.168.0.6, Administrador, password and weyler** values to your active directory values)

```xml
<application-policy name="OpenKM">
  <authentication>
    <login-module code="org.jboss.security.auth.spi.LdapExtLoginModule" flag="required" >
      <module-option name="java.naming.provider.url">ldap://192.168.0.6</module-option>
      <module-option name="bindDN">CN=Administrador,cn=users,dc=weyler,dc=local</module-option>
      <module-option name="java.naming.security.authentication">simple</module-option>
      <module-option name="bindCredential">password</module-option>
      <module-option name="baseCtxDN">cn=users,dc=weyler,dc=local</module-option>
      <module-option name="baseFilter">(sAMAccountName={0})</module-option>
      <module-option name="rolesCtxDN">cn=users,dc=weyler,dc=local</module-option>
      <module-option name="roleFilter">(member={1})</module-option>
      <module-option name="roleAttributeID">cn</module-option>
      <module-option name="roleAttributeIsDN">false</module-option>
      <module-option name="roleRecursion">2</module-option>
      <module-option name="searchScope">ONELEVEL_SCOPE</module-option>
      <module-option name="defaultRole">UserRole</module-option>
      <module-option name="allowEmptyPasswords">false</module-option>
    </login-module>
  </authentication>
</application-policy>
```

If you want to restrict the user who can log into OpenKM, you should change these two property in OpenKM.cfg:

```
principal.ldap.user.search.filter=(&(objectclass=user)(memberOf=CN=UserRole,CN=users,DC=weyler,DC=local))
principal.ldap.role.search.filter=(&(objectclass=group)(memberOf=CN=OpenKM,CN=users,DC=weyler,DC=local))
```

This means that only users within the UserRole groups will be shown as valid OpenKM users, and only roles which are included in the OpenKM group will be shown in OpenKM.

Also add this option one in login-config.xml:

```xml
<module-option name="baseFilter">(&amp;(sAMAccountName={0})(memberOf=CN=UserRole,CN=users,DC=weyler,DC=local))</module-option>
```

And remove this one:

```xml
<module-option name="defaultRole">UserRole</module-option>
```

All this means that only users member of the UserRole groups are able to log into OpenKM.

Enable debug at login process

It's good practice enable login debug when you make any change in autentication mechanism. Edit the file /server/default/conf/jboss-log4j.xml and add the category (remember you must restart jboss to it takes effect):

```xml
<category name="org.jboss.security">
    <priority value="TRACE" class="org.jboss.logging.XLevel"/>
</category>
```

or

```xml
<category name="org.jboss.security">
    <priority value="TRACE" class="org.jboss.logging.XLevel"/>
    <appender-ref ref="SECURITY_F"/>
</category>


<appender name="SECURITY_F" class="org.jboss.logging.appender.DailyRollingFileAppender">
    <param name="Append" value="true"/>
    <param name="DatePattern" value="'.'yyyy-MM-dd"/>
    <param name="File" value="${jboss.server.home.dir}/log/jboss.security.log"/>
    <layout class="org.apache.log4j.PatternLayout">
        <param name="ConversionPattern" value="%d{ABSOLUTE} %-5p [%c] %m%n"/>
    </layout>
</appender>
```

More info at JBoss and JAAS debug [2].

Active directory utilities

We recommend take a look at these tools:

- Apache Directory Studio [1]
- Active Directory Explorer Utility [2]

See also:

- Testing LDAP configuration
- Forum: Usuario administrador LDAP v.4 [1]
- JBoss and JAAS debug [2]
- LDAP authentication using LDAPExtUserModuleImpl is case-inse [3]
- LdapExtLoginModule [5]
- Problems with LdapExtLoginModule [4]

References

[1] http://directory.apache.org/studio/
[2] http://technet.microsoft.com/en-us/sysinternals/bb963907.aspx
[3] http://community.jboss.org/message/427398
[4] http://community.jboss.org/thread/159069

Open Directory

Here are some tips on OS X Open Directory:

- **Create user**: dscl . -create /Users/openkm
- **Assign shell**: dscl . -create /Users/openkm UserShell /bin/bash
- **Assign user name**: dscl . -create /Users/openkm RealName "OpenKM user"
- **Search UID**: dscl . -search /Users uid 503
- **Assign UID**: dscl . -create /Users/openkm UniqueID 503
- **Assign user home**: dscl . -create /Users/openkm NFSHomeDirectory /opt/openkm
- **Assign password**: dscl . -passwd /Users/openkm xxxxxx

More on OpenDirectory and the dscl Tool at Introduction to Porting UNIX/Linux Applications to Mac OS X [1].

See also:

- Testing LDAP configuration
- Forum: Usuario administrador LDAP v.4 [1]
- JBoss and JAAS debug [2]

References

[1] http://developer.apple.com/mac/library/documentation/Porting/Conceptual/PortingUnix/intro/intro.html

Central Authentication Service

The Central Authentication Service (CAS) is a single sign-on protocol for the web. Its purpose is to permit a user to access multiple applications while providing their credentials (such as userid and password) only once. It also allows web applications to authenticate users without gaining access to a user's security credentials, such as a password. The name CAS also refers to a software package that implements this protocol.

Once you have the CAS server up and running it is easy to set SSO in OpenKM. I wanted to keep basic authentication for the WebDav part that is why the filter mapping is complicated. I did the following changes in OpenKM:

1. In server/default/deploy/OpenKM.war/WEB-INF/web.xml:

```xml
<!-- CAS SSO-->
<!-- Facilitates CAS single sign-out -->
<listener>
  <listener-class>org.jasig.cas.client.session.SingleSignOutHttpSessionListener</listener-class>
</listener>


<!-- Following is needed only if CAS single-sign out is desired -->
<filter>
  <filter-name>CAS Single Sign Out Filter</filter-name>
  <filter-class>org.jasig.cas.client.session.SingleSignOutFilter</filter-class>
</filter>


<!-- Only 2 CAS filters are required for JAAS support -->
<context-param>
  <param-name>service</param-name>
```

```xml
    <param-value>https://openkm.chronossoftware.com/OpenKM</param-value>
</context-param>
<context-param>
    <param-name>casServerLoginUrl</param-name>
    <param-value>https://openkm.chronossoftware.com/cas/login</param-value>
</context-param>


<filter>
    <filter-name>CASWebAuthenticationFilter</filter-name>
    <filter-class>org.jasig.cas.client.jboss.authentication.WebAuthenticationFilter</filter-class>
</filter>
<filter>
    <filter-name>CASAuthenticationFilter</filter-name>
    <filter-class>org.jasig.cas.client.authentication.AuthenticationFilter</filter-class>
</filter>


<!-- CAS client filter mappings -->
<!-- The order of the following filters is vitally important -->
<filter-mapping>
    <filter-name>CAS Single Sign Out Filter</filter-name>
    <url-pattern>*.jsp</url-pattern>


    <!-- GWT -->
    <url-pattern>/frontend/*</url-pattern>


    <!-- JSPs -->
    <url-pattern>/admin/*</url-pattern>
    <url-pattern>/mobile/*</url-pattern>
    <url-pattern>/mobile-nt/*</url-pattern>


    <!-- Servlets -->
    <url-pattern>/RepositoryStartup</url-pattern>
    <url-pattern>/TextToSpeech</url-pattern>
    <url-pattern>/Test</url-pattern>
    <url-pattern>/frontend/*</url-pattern>
    <url-pattern>/extension/*</url-pattern>
    <!--url-pattern>/*</url-pattern-->
</filter-mapping>
<filter-mapping>
    <filter-name>CASWebAuthenticationFilter</filter-name>
    <url-pattern>*.jsp</url-pattern>


    <!-- GWT -->
    <url-pattern>/frontend/*</url-pattern>


    <!-- JSPs -->
    <url-pattern>/admin/*</url-pattern>
```

```xml
    <url-pattern>/mobile/*</url-pattern>
    <url-pattern>/mobile-nt/*</url-pattern>


    <!-- Servlets -->
    <url-pattern>/RepositoryStartup</url-pattern>
    <url-pattern>/TextToSpeech</url-pattern>
    <url-pattern>/Test</url-pattern>
    <url-pattern>/frontend/*</url-pattern>
    <url-pattern>/extension/*</url-pattern>


    <!--url-pattern>/*</url-pattern-->
  </filter-mapping>
  <filter-mapping>
    <filter-name>CASAuthenticationFilter</filter-name>
    <url-pattern>*.jsp</url-pattern>


    <!-- GWT -->
    <url-pattern>/frontend/*</url-pattern>


    <!-- JSPs -->
    <url-pattern>/admin/*</url-pattern>
    <url-pattern>/mobile/*</url-pattern>
    <url-pattern>/mobile-nt/*</url-pattern>


    <!-- Servlets -->
    <url-pattern>/RepositoryStartup</url-pattern>
    <url-pattern>/TextToSpeech</url-pattern>
    <url-pattern>/Test</url-pattern>
    <url-pattern>/frontend/*</url-pattern>
    <url-pattern>/extension/*</url-pattern>


    <!--url-pattern>/*</url-pattern-->
  </filter-mapping>
  <!-- /CAS SSO -->
```

2. In server/default/conf/login-config.xml:

```xml
<application-policy name="OpenKMWebDav">

  <authentication>

    <login-module code="org.jboss.security.auth.spi.LdapExtLoginModule" flag="required" >

      <module-option name="java.naming.provider.url">ldap://dokumentum2.chronossoftware.com:389</module-option>

      <module-option name="java.naming.security.authentication">simple</module-option>

      <module-option name="baseCtxDN">ou=Users,dc=igyuk,dc=hu</module-option>

      <module-option name="baseFilter">(cn={0})</module-option>

      <module-option name="rolesCtxDN">ou=Groups,dc= igyuk,dc=hu </module-option>

      <module-option name="roleFilter">(uniqueMember={1})</module-option>

      <module-option name="roleAttributeID">cn</module-option>

      <module-option name="roleNameAttributeID">cn</module-option>
```

```xml
      <module-option name="roleAttributeIsDN">true</module-option>

      <module-option name="roleRecursion">2</module-option>

      <module-option name="searchScope">ONELEVEL_SCOPE</module-option>

      <module-option name="allowEmptyPasswords">False</module-option>

      <module-option name="defaultRole">UserRole</module-option>

    </login-module>

  </authentication>

</application-policy>


<application-policy name="OpenKM">

  <authentication>

    <login-module code="org.jasig.cas.client.jaas.CasLoginModule" flag="required">

      <module-option name="ticketValidatorClass">org.jasig.cas.client.validation.Saml11TicketValidator</module-option>

      <module-option name="casServerUrlPrefix">https://dokumentum2.chronossoftware.com/cas</module-option>

      <module-option name="service">https://dokumentum2.chronossoftware.com/OpenKM</module-option>

      <module-option name="defaultRoles">UserRole</module-option>

      <module-option name="roleAttributeNames">groupMembership</module-option>

      <module-option name="principalGroupName">CallerPrincipal</module-option>

      <module-option name="roleGroupName">Roles</module-option>

      <module-option name="cacheAssertions">true</module-option>

      <module-option name="tolerance">20000</module-option>

      <module-option name="cacheTimeout">480</module-option>

    </login-module>

  </authentication>

</application-policy>
```

3. In repository.xml I changed the "OpenKM" to "OpenKMWebDav" at

```xml
<Security appName="OpenKM">
```

4. Add the certificate of the CAS server to cacerts using the instructions from JBoss mail configuration#Adding PKI certificate.

5. Copy these 2 JAR files from CAS client destribution to $JBOSS_HOME/server/default/deploy/OpenKM.war/WEB-INF/lib:

- cas-client-core-3.2.0.jar
- cas-client-integration-jboss-3.2.0.jar

Testing LDAP configuration

LDAP configuration maybe a little tricky sometimes. For this reason is very important following certain rules. First of all you need to configure JBoss LDAP authentication. If after that you can log into OpenKM, go ahead configuring OpenKM LDAP integration.

OpenKM is an J2EE that uses an Application Server called JBoss. In J2EE web applications, autentication and authorization are task delegated to the Application Server. This can be achieved because OpenKM use JAAS [1] and you need to configure JBoss to handle your users and passwords. JBoss has many predefined Login Modules which handles different the user and password storage, like RDBMS or LDAP. These Login Modules are configured at **$JBOSS_HOME/server/default/conf/login-config.xml** file.

Once the user is logged into OpenKM, the application also need to know which users and roles are defined. In the case of the default database based authentication, you can even manage these users, passwords and roles from OpenKM Administration. But if you use another authentication and authorization backend like LDAP, you only will see these users and their password but no modification is allowed.

In OpenKM you need to configure these LDAP queries:

- Get all users
- Get all roles
- Get email from a user
- Get users in a role
- Get roles from a user

Every query needs three configuration properties:

- Search base
- Filter
- Attribute

So, the "get all users" query is composed by:

- principal.ldap.user.search.base
- principal.ldap.user.search.filter
- principal.ldap.user.attribute

> Although OpenKM 5.1.x uses database for storing configuration properties, to simplify the usage this tool read these properties from OpenKM.cfg.

To build these queries, I recommend using the precious Apache Directory Studio [1] tool. It is multi-platform and works pretty well.

As you can imagine, configure all these properties correctly is complex. For this reason we have developed a tool for testing OpenKM configuration. This tool can be downloaded from http:/ / www. openkm. com/ download/ OpenKM-5.1-LDAP.zip.

> With these configuration parameters you should be able to configure your LDAP. But in some cases, the user location is split in two or more branches. In this case you will need to develop a custom LDAP adapter which meets your particular requirements.

References

[1] http://en.wikipedia.org/wiki/Java_Authentication_and_Authorization_Service

Repository configuration

OpenKM uses Apache Jackrabbit to handle the document repository. From the Jackrabbit site:

Apache Jackrabbit is a fully conforming implementation of the Content Repository for Java Technology API (JCR). A content repository is a hierarchical content store with support for structured and unstructured content, full text search, versioning, transactions, observation, and more. Typical applications that use content repositories include content management, document management, and records management systems.

This means that if you configure an OpenKM repository, you are configuring a Jackrabbit repository.Jackrabbit offers several repository configurations: it can be stored in the local filesystem or in a remote database, or even in the AWS (Amazon Web Service) cloud.

Configuration parameters

The repository configuration file, typically called repository.xml, specifies global options like security, versioning and clustering settings. A default workspace configuration template is also included in the repository configuration file. The top-level structure of the repository configuration file is shown below:

```
<!DOCTYPE Repository
        PUBLIC "-//The Apache Software Foundation//DTD Jackrabbit
1.4//EN"
        "http://jackrabbit.apache.org/dtd/repository-1.4.dtd">
<Repository>
    <FileSystem .../>
    <Security .../>
    <Workspaces .../>
    <Workspace .../>
    <Versioning .../>
    <SearchIndex .../>    <!-- optional -->
    <DataStore .../>      <!-- optional -->
</Repository>
```

The repository configuration elements are:

- **FileSystem**: The virtual file system used by the repository to store things like registered namespaces and node types.
- **Security**: Authentication and authorization configuration.
- **Workspaces**: Configuration on where and how workspaces are managed.
- **Workspace**: Default workspace configuration template.
- **Versioning**: Configuration of the repository-wide version store.
- **SearchIndex**: Configuration of the search index that covers the repository-wide /jcr:system content tree.
- **DataStore**: Data store configuration.

Bean configuration elements

Most of the entries in the configuration file are based on the following generic JavaBean configuration pattern. Such configuration specifies that the repository should use an instance of the specified class with the specified properties for the named functionality.

```
<ConfigurationElement class="fully.qualified.ClassName">
    <param name="property1" value="...">
    <param name="property2" value="...">
<ConfigurationElement>
```

Configuration variables

Jackrabbit supports configuration variables of the form ${name}. These variables can be used to avoid hardcoding specific options in the configuration files. The following variables are available in all Jackrabbit versions:

- **${rep.home}**: Repository home directory.
- **${wsp.name}**: Workspace name. Only available in workspace configuration.
- **${wsp.home}**: Workspace home directory. Only available in workspace configuration.

Security configuration

The security configuration element is used to specify authentication and authorization settings for the repository. The structure of the security configuration element is:

```
<Security appName="Jackrabbit">
    <SecurityManager .../> <!-- optional, available since 1.5 -->
    <AccessManager .../>      <!-- mandatory until 1.4, optional since 1.5 -->
    <LoginModule .../>        <!-- optional -->
</Security>
```

By default Jackrabbit uses the Java Authentication and Authorization Service (JAAS) to authenticate users who try to access the repository. The appName parameter in the <Security/> element is used as the JAAS application name of the repository.

If JAAS authentication is not available or (as is often the case) too complex to set up, Jackrabbit allows you to specify a repository-specific JAAS LoginModule that is then used for authenticating repository users. The default SimpleLoginModule class included in Jackrabbit implements a trivially simple authentication mechanism that accepts any username and any password as valid authentication credentials.

Once a user has been authenticated, Jackrabbit will use the configured AccessManager to control what parts of the repository content the user is allowed to access and modify. The default SimpleAccessManager class included in Jackrabbit implements a trivially simple authorization mechanism that grants full read access to all users and write access to everyone except anonymous users. The slightly more advanced SimpleJBossAccessManager class was added in Jackrabbit 1.3 (see JCR-650). This class is designed for use with the JBoss Application Server, where it maps JBoss roles to Jackrabbit permissions.

Workspace configuration

A Jackrabbit repository contains one or more workspaces that are each configured in a separate workspace.xml configuration file. The Workspaces element of the repository configuration specifies where and how the workspaces are managed. The repository configuration also contains a default workspace configuration template that is used to create the workspace.xml file of a new workspace unless more specific configuration is given when the workspace is created. See the createWorkspace methods in the JackrabbitWorkspace interface for more details on workspace creating workspaces.

The workspace settings in the repository configuration file are:

```xml
<Workspaces rootPath="${rep.home}/workspaces"
            defaultWorkspace="default"
            configRootPath="..." <!-- optional -->
            maxIdleTime="..."/>   <!-- optional -->
<Workspace .../>   <!-- default workspace configuration template -->
```

The following global workspace configuration options are specified in the Workspaces element:

- **rootPath**: The native file system directory for workspaces. A subdirectory is automatically created for each workspace, and the path of that subdirectory can be used in the workspace configuration as the ${wsp.path} variable.
- **defaultWorkspace**: Name of the default workspace. This workspace is automatically created when the repository is first started.
- **configRootPath**: By default the configuration of each workspace is stored in a workspace.xml file within the workspace directory within the rootPath directory. If this option is specified, then the workspace configuration files are stored within the specified path in the virtual file system (see above) configured for the repository.
- **maxIdleTime**: By default Jackrabbit only releases resources associated with an opened workspace when the entire repository is closed. This option, if specified, sets the maximum number of seconds that a workspace can remain unused before the workspace is automatically closed.

The workspace configuration template and all workspace.xml configuration files have the following structure:

```xml
<Workspace name="${wsp.name}">
    <FileSystem .../>
    <PersistenceManager .../>
    <SearchIndex .../>          <!-- optional -->
    <ISMLocking .../>           <!-- optional, available since 1.4 -->
</Workspace>
```

The workspace configuration elements are:

- **FileSystem**: The virtual file system passed to the persistence manager and search index.
- **PersistenceManager**: Persistence configuration for workspace content. For more info, read http://wiki.apache. org/jackrabbit/PersistenceManagerFAQ.
- **SearchIndex**: Configuration of the workspace search index.
- **ISMLocking**: Locking configuration for concurrent access to workspace content.

> To modify the configuration of an existing workspace, you need to change the workspace.xml file of that workspace. Changing the <Workspace/> element in the repository configuration file will not affect existing workspaces.

Third-party software integration

In order o extend OpenKM functionalities, it can be integrated with some external software which improves the OpenKM user experience adding new features to the application. We are working to expand this list of applications, so stay tuned!

- Apache 🔒
- OCR
- OpenOffice.org
- SWFTools 🔒
- Antivirus
- Acme CAD Converter

Third-party software integration: Apache

Expose OpenKM directly from JBoss can be dangerous if you need the application to be accessed from Internet. Also this 8080 may be closed by a firewall. For these reasons, is a good idea expose your OpenKM installation through the standard web port 80. In the following steps we explain how to configure Apache to handle these request and forward to JBoss application server using the AJP13 protocol.

From the Apache documentation: The AJP13 protocol is packet-oriented. A binary format was presumably chosen over the more readable plain text for reasons of performance. The web server communicates with the servlet container over TCP connections. To cut down on the expensive process of socket creation, the web server will attempt to maintain persistent TCP connections to the servlet container, and to reuse a connection for multiple request/response cycles.

Debian / Ubuntu

The first thing in to install the required Apache software. From Debian / Ubuntu you can install Apache with a single command:

```
$ sudo aptitude install apache2
```

Edit the file called /etc/apache2/apache2.conf and configure a ServerName to prevent warnings in the Apache startup process:

```
ServerRoot "/etc/apache2"
ServerName "your-domain.com"
```

Enable the proxy module, needed to forward petitions to JBoss:

```
$ sudo a2enmod proxy_ajp
```

Now create the configuration file /etc/apache2/sites-available/openkm.conf with this content:

```
<VirtualHost *>
    ServerName openkm.your-domain.com
    RedirectMatch ^/$ /OpenKM
    <Location /OpenKM>
        ProxyPass ajp://127.0.0.1:8009/OpenKM
        ProxyPassReverse http://openkm.your-domain.com/OpenKM
```

```
    </Location>
    ErrorLog /var/log/apache2/your-domain.com-error.log
    CustomLog /var/log/apache2/your-domain.com-access.log combined
</VirtualHost>
```

The VirtualHost ServerName must be other than ServerName in the main Apache configuration. Enable this site configuration:

```
$ sudo a2ensite openkm.conf
```

> If after restart Apache you see a warning like:
>
> ```
> [warn] NameVirtualHost *:80 has no VirtualHosts
> ```
>
> you have to change the virtual host definition from <VirtualHost *> to <VirtualHost *:80>

If you don't want to show the **/OpenKM** context, try this as your VirtualHost configuration:

```
<VirtualHost *>
  ServerName openkm.your-domain.com
  ProxyPass / ajp://localhost:8009/OpenKM/
  ProxyPassReverse / http://openkm.your-domain.com/OpenKM/
  ErrorLog /var/log/apache2/your-domain.com-error.log
  CustomLog /var/log/apache2/your-domain.com-access.log combined

  RewriteEngine on
  RewriteCond %{REQUEST_URI} /OpenKM*
  RewriteRule ^(/.*)$ http://localhost:8080$1 [P]

  RequestHeader edit X-GWT-Module-Base
^(http://openkm.your-domain)/frontend/(.*)$ $1/OpenKM/frontend/$2
</VirtualHost>
```

You need also set the configuration parameter **system.apache.request.header.fix** to on (true). Note than in OpenKM 5.1 the configuration is stored in database and the OpenKM.cfg configuration only make sense in the database (Hibernate) configuration parameters.

> If you see an error like:
>
> ```
> Invalid command 'RewriteEngine', perhaps misspelled or defined by a module not included in the server configuration
> ```
>
> you need to enable this Apache module:
>
> ```
> $ sudo a2enmod rewrite
> $ sudo a2enmod proxy_http
> $ sudo a2enmod headers
> ```

You have to enable explicitly the proxy access editing the Apache configuration file */etc/apache2/mods-available/proxy.conf*:

```
<IfModule mod_proxy.c>
  #turning ProxyRequests on and allowing proxying from all may allow
  #spammers to use your proxy to send email.
```

```
  ProxyRequests Off

  <Proxy *>
    AddDefaultCharset off
    Order deny,allow
    Allow from all
    Deny from all
    #Allow from .example.com
  </Proxy>

  # Enable/disable the handling of HTTP/1.1 "Via:" headers.
  # ("Full" adds the server version; "Block" removes all outgoing Via:
headers)
  # Set to one of: Off | On | Full | Block

  ProxyVia On
</IfModule>
```

Finally restart Apache:

```
$ sudo /etc/init.d/apache2 restart
```

Now you can access your OpenKM installation from http://openkm.your-domain.com/. Another advantage of using Apache is that you can log OpenKM access and generate web statistics.

SSL in Debian / Ubuntu

```
$ sudo mkdir /etc/apache2/ssl
$ sudo /usr/sbin/make-ssl-cert /usr/share/ssl-cert/ssleay.cnf /etc/apache2/ssl/apache.pem
$ sudo a2enmod ssl
```

Ensure ports 443 is listen in /etc/apache2/ports.conf

Do the same task related to port 80 but changing the configuration file /etc/apache2/sites-available/openkm.conf with this content:

```
<VirtualHost *:443>
    ServerName openkm.your-domain.com
    RedirectMatch ^/$ /OpenKM
    <Location /OpenKM>
        ProxyPass ajp://127.0.0.1:8009/OpenKM
        ProxyPassReverse http://openkm.your-domain.com/OpenKM
    </Location>
    ErrorLog /var/log/apache2/openkm.your-domain.com-error.log
    CustomLog /var/log/apache2/openkm.your-domain.com-access.log
combined

    SSLEngine on
    SSLCertificateFile /etc/apache2/ssl/apache.pem
</VirtualHost>
```

Red Hat / CentOS

Here you can use the yum application manager to install Apache:

```
$ sudo yum install httpd
```

Now edit the file /etc/httpd/conf.d/proxy_ajp.conf and add:

```
ProxyPass /OpenKM ajp://localhost:8009/OpenKM
```

After that, restart Apache to make effective this configuration.

```
$ sudo /etc/init.d/httpd restart
```

SSL in Red Hat / CentOS

```
$ sudo yum install mod_ssl openssl crypto-utils
```

Generate private keys (for more information visit Apache HTTP Secure Server Configuration [1])

```
$ genkey your-domain.com
```

> Normally you want to generate self-certificate and not sending to Certify Authority. Is good practice put some password on private key, but in this case each time you restart apache service it'll be demanded.

During the process will be generated two files at

```
SSLCertificateFile /etc/pki/tls/certs/your-domain.com.cert
SSLCertificateKeyFile /etc/pki/tls/private/your-domain.com.key
```

Must modify the /etc/httpd/conf.d/openkm.conf file

```
<VirtualHost *:443>
  ServerName your-domain.com
  ProxyPass / ajp://localhost:8009/OpenKM/
  ProxyPassReverse / http://your-domain.com/OpenKM/
  ErrorLog /var/log/your-domain.com-error.log
  CustomLog /var/log/your-domain.com-access.log combined

  RewriteEngine on
  RewriteCond %{REQUEST_URI} /OpenKM*
  RewriteRule ^(/.*)$ http://localhost:8080$1 [P]
  ProxyPassReverse / http://localhost:8080/

  SSLEngine on
  SSLCertificateFile /etc/pki/tls/certs/your-domain.com.cert
  SSLCertificateKeyFile /etc/pki/tls/private/your-domain.com.key
</VirtualHost>
```

Finally must modify SSLCertificateFile and SSLCertificateKeyFile values in file /etc/httpd/conf.d/ssl.conf

```
SSLCertificateFile /etc/pki/tls/certs/your-domain.com.cert
SSLCertificateKeyFile /etc/pki/tls/private/your-domain.com.key
```

Now you can access your OpenKM installation from http://openkm.your-domain.com/. Another advantage of using Apache is that you can log OpenKM access and generate web statistics.

Mac OS X

Edit the file called /etc/apache2/apache2.conf and configure a ServerName, enable proxy modules and mod_proxy:

```
ServerRoot "/usr"
ServerName "your-domain.com"
```

```
LoadModule proxy_module libexec/apache2/mod_proxy.so
LoadModule proxy_ajp_module libexec/apache2/mod_proxy_ajp.so
```

```
<IfModule mod_proxy.c>
      ProxyRequests Off
      <Proxy "*">
             AddDefaultCharset off
             Allow from all
             Deny from all
             Order Deny,Allow
      </Proxy>
      ProxyVia On
</IfModule>
```

Now create the configuration file /etc/apache2/sites/openkm.conf with this content:

```
<VirtualHost *:80>
   ServerName openkm.your-domain.com
   RedirectMatch ^/$ /OpenKM
   <Location /OpenKM>
       ProxyPass ajp://127.0.0.1:8009/OpenKM
       ProxyPassReverse http://openkm.your-domain.com/OpenKM
   </Location>
   ErrorLog /var/log/apache2/your-domain.com-error.log
   CustomLog /var/log/apache2/your-domain.com-access.log combined
</VirtualHost>
```

Finally restart apache

More info

For more info, visit:

- Apache Module mod_proxy [2]
- Apache Module mod_proxy_ajp [3]
- Apache Module mod_rewrite [4]
- mod_rewrite, a beginner's guide [5]
- Using Apache Virtual Hosts and ProxyPass Together [6]
- Apache configuration for GWT applications [7]

References

[1] http://www.linuxtopia.org/online_books/rhel5/rhel5_administration/rhel5_s1-httpd-secure-server.html

[2] http://httpd.apache.org/docs/2.2/mod/mod_proxy.html

[3] http://httpd.apache.org/docs/2.2/mod/mod_proxy_ajp.html

[4] http://httpd.apache.org/docs/2.2/mod/mod_rewrite.html

[5] http://www.workingwith.me.uk/articles/scripting/mod_rewrite

[6] http://www.thebuzzmedia.com/using-apache-virtual-hosts-and-proxypass-together/

[7] http://pgt.de/2011/01/27/apache-configuration-for-gwt-applications

Third-party software integration: OCR

Tesseract is an Open Source OCR engine adopted by Google. It works really well. The OCR natively can read TIFF documents and has hight ratio of recognition with images 300 dpi of resolution and converted to lineart (1 bit color).

If you are using a computer with Debian / Ubuntu, the installation simplifies a lot:

```
$ aptitude install tesseract-ocr
```

And

```
$ aptitude install tesseract-ocr-eng
```

If you want to add support for english language. You can also download Windows executables for tesseract-2.04 at http://code.google.com/p/tesseract-ocr/downloads/list.

Now you have to tell OpenKM to use this OCR application. Edit the file OpenKM.cfg:

```
$ vim OpenKM.cfg
```

And set the system.ocr property to the path of the tesseract executable:

```
system.ocr=/usr/local/bin/tesseract
```

For more info, go to http://code.google.com/p/tesseract-ocr/and Tesseract - Summary & first experiences [1].

There is also another interesting free OCR application called OCRopus. It has many improvements over Tesseract but is on early development stage. Last released version (0.3.1) is quite usable and works very well but have to be compiled and actually is a difficult task. Visit http://code.google.com/p/ocropus/for more info.

Compile from source code

You can download the source code from http://code.google.com/p/tesseract-ocr/ and compile yourself. Also download the language files you need and uncompress them in the same folder of the application.

```
$ sudo aptitude install build-essential libtiff4-dev
$ wget http://tesseract-ocr.googlecode.com/files/tesseract-2.04.tar.gz
$ tar xzvf tesseract-2.04.tar.gz
$ cd tesseract-2.04
$ ./configure --prefix=/opt/tesseract
$ make
$ wget http://tesseract-ocr.googlecode.com/files/tesseract-2.00.eng.tar.gz
$ tar xzvf tesseract-2.00.eng.tar.gz
$ sudo make install
```

The executable should be located at /opt/tesseract/bin/tesseract. More info about compilation at:

- http://code.google.com/p/tesseract-ocr/wiki/ReadMe
- http://code.google.com/p/tesseract-ocr/wiki/FAQ

OpenKM 5.1 OCR configuration

Starting with OpenKM 5.1 you can choose between several OCR engines:

OCR Engine	Text Extractor	Image Formats	Default program arguments
Tesseract 2.x	com.openkm.extractor.Tesseract2TextExtractor	TIFF	Config.SYSTEM_OCR ${fileIn} ${fileOut}
Tesseract 3.x	com.openkm.extractor.Tesseract3TextExtractor	TIFF PNG JPG GIF	Config.SYSTEM_OCR ${fileIn} ${fileOut}
Cuneiform	com.openkm.extractor.CuneiformTextExtractor	TIFF PNG JPG GIF	Config.SYSTEM_OCR ${fileIn} -o ${fileOut}
Abby	com.openkm.extractor.AbbyTextExtractor	TIFF PNG JPG GIF	Config.SYSTEM_OCR ${fileIn} -o ${fileOut}

So, if you want to pass a command line parameter to your tesseract executable, you should use this configuration:

```
system.ocr=/usr/bin/tesseract -l esp
```

In this OpenKM version you can also use an OpenOffice.org dictionary to enhance the OCR process. You can find these language specific dictionaries at OpenOffice.org Dictionary Repository [2]. After download, set this configuration property with the path to the dictorionay file:

```
system.openoffice.dictionary=/path/to/dictionary.(oxt|zip)
```

Software required

You can enable any of these text extractors adding it in the **textFilterClasses** param of the **SearchIndex** section in your repository.xml file.

You can download Tesseract 3 for Windows from tesseract-ocr Google Code [3]. To install Tesseract 3 in Ubuntu, add the PPA and install Tesseract OCR 3.0 SVN:

```
$ sudo add-apt-repository ppa:alex-p/notesalexp
$ sudo apt-get update
$ sudo apt-get install tesseract-ocr tesseract-ocr-eng
```

> You must add the PPA, install the latest Tesseract and then disable the PPA as it contains a lot of bleeding edge packages!

```
$ sudo add-apt-repository -r ppa:alex-p/notesalexp
```

Starting with OpenKM 5.1 we offer integration with Cognitive OpenOCR (Cuneiform) [4]. This OCR engine make a very good job improving Tesseract conversion ratios.

You can grab binaries from http://pkgs.org/package/cuneiform.

External links

- Tesseract OCR Google Groups [5]
- First Interactions with Tesseract OCR on Ubuntu Linux [6]

References

[1] http://www.win.tue.nl/~aeb/linux/ocr/tesseract.html
[2] http://extensions.services.openoffice.org/en/dictionaries
[3] http://code.google.com/p/tesseract-ocr/
[4] http://en.openocr.org/
[5] http://groups.google.com/group/tesseract-ocr
[6] http://triviaatwork.blogspot.com/2009/08/first-interactions-with-tesseract-ocr.html

Third-party software integration: OpenOffice.org

OpenKM can convert some document types to PDF. This is a great help if need to read an Microsoft Office / OpenOffice.org document and you don't have the software installed in the computer.

> OpenKM has been roughly tested with **OpenOffice.org 3.2.1**, but we have also configured with **OpenOffice.org 3.3.2** and **LibreOffice 3.3.2** and seems to work pretty well, but for now we prefer OpenOffice.org 3.2.1 for production environments.

In a recent Ubuntu or Debian release, you can install it simply as:

```
$ sudo aptitude install openoffice.org
```

Recently has been released OpenOffice.org 3.3.0 and is not fully tested unde OpenKM. You can download the previous stable from OpenOffice Archive Servers [1].

OpenKM 5.1

You can also configure OpenKM to use a remote server for OpenOffice document conversion.

```
system.openoffice.server=http://localhost:8080/converter/convert
```

> This feature is under development and actually not released for public testing.

OpenKM 5.0

In this OpenKM some things changes. You can configure OpenOffice.org listen port and a maximun conversion tasks:

```
system.openoffice.path=/usr/lib/openoffice
system.openoffice.tasks=5
system.openoffice.port=2222
```

In case of LibreOffice and Linux, the **system.openoffice.path** property should be set to */usr/lib/libreoffice*. Note that **system.openoffice.tasks** and **system.openoffice.port** have already a default value and is not needed to be set.

> Starting from OpenKM 5.0 you don't need to start OpenOffice.org as a service because OpenKM will take case of this. You only need to set the **system.openoffice** configuration property in OpenKM.cfg to the path of a working OpenOffice.org installation.

Linux

```
system.openoffice.path=/usr/lib/openoffice
```

Windows

```
system.openoffice.path=C:/Program Files/OpenOffice.org 3
```

or

```
system.openoffice.path=C:\\Program Files\\OpenOffice.org 3
```

OpenKM 4.x and older versions

You need an OpenOffice.org installation in the OpenKM server, and also this OpenOffice.org application has to be running in server mode (also known as headless). In Debian / Ubuntu, depending of you OpenOffice.org version you will have to install an X11 virtual server or not:

```
$ apt-get install xvfb
```

And start it using this command:

```
$ xvfb-run /usr/lib/openoffice/program/soffice -headless -accept="socket,host=127.0.0.1,port=8100;urp;" -nofirststartwizard
```

From OpenOffice.org 2.3, it is not necessary the X11 virtual server but you should install these packages:

```
$ aptitude install openoffice.org-headless openoffice.org-java openoffice.org
```

But before of this, you must enable a couple of repositories:

```
 deb http://en.archive.ubuntu.com/ubuntu/ hardy-updates universe
 deb http://en.archive.ubuntu.com/ubuntu/ hardy-updates multiverse
```

This script simplifies the start process (For security reasons, you should no start OpenOffice.org as root):

```
#!/bin/sh
unset DISPLAY
/usr/lib/openoffice/program/soffice
"-accept=socket,host=localhost,port=8100;urp;StarOffice.ServiceManager"
 -nologo
 -headless -nofirststartwizard
```

OpenOffice.org will listen at port 8100, so you can check that the application has started running this:

```
$ netstat -putan | grep 8100
```

Also you can configure OpenOffice.org as a service with this script:

```
#!/bin/bash
# openoffice.org headless server script
#
# chkconfig: 2345 80 30
# description: headless openoffice server script
# processname: openoffice
#
# Author: Vic Vijayakumar
```

```
# Modified by Paco Avila and Federico Ch. Tomasczik
#
SOFFICE=/usr/bin/soffice
PIDFILE=/var/run/openoffice-server.pid
set -e
case "$1" in
    start)
        if [ -f $PIDFILE ]; then
            echo "OpenOffice headless server has already started."
            sleep 5
            exit
        fi
        echo "Starting OpenOffice headless server"
        $SOFFICE -headless -nologo -nofirststartwizard
-accept="socket,host=127.0.0.1,port=8100;urp" & > /dev/null 2>&1
        touch $PIDFILE
        ;;
    stop)
        if [ -f $PIDFILE ]; then
            echo "Stopping OpenOffice headless server."
            killall -9 soffice && killall -9 soffice.bin
            rm -f $PIDFILE
            exit
        fi
        echo "Openoffice headless server is not running."
        exit
        ;;
    *)
        echo "Usage: $0 {start|stop}"
        exit 1
esac
exit 0
```

Change the permissions to this file:

```
$ chmod 0755 /etc/init.d/openoffice
```

Install openoffice init script links:

For Debian based distros like Ubuntu

```
$ update-rc.d openoffice defaults
```

For RedHat based distros like CentOS

```
$ chkconfig --add openoffice
```

And this script will launch OpenOffice.org on every system reboot. Also you can launch it manually this way:

```
$ /etc/init.d/openoffice start
```

More info at:

- OpenOffice Installation Instructions for Windows [2]
- http://www.artofsolving.com/node/10
- http://www.oooforum.org/forum/viewtopic.phtml?t=11890
- http://code.google.com/p/openmeetings/wiki/OpenOfficeConverter

Windows

Install Windows Server 2003 Resource Kit Tools [3] in your computer for example at c:\tools

Create a openoffice service **cd c:\tools**

```
instsrv openoffice c:\tools\srvany.exe
```

Modify sy**stem registry using regedit, got to HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\openoffice**

Add key **Parameters**

Into Parameters add mew parameters key **Application** as **REG_SZ** (String value) with your server soffice.exe path, for example **C:\Archivos de programa\OpenOffice.org 3\program\soffice.exe**

Into Parameter add new parameters key AppParameters as REG_SZ with value **-nofirststartwizard -nologo -headless -accept=socket,host=localhost,port=8100;urp;StarOffice.ServiceManager**

To start service type:

```
net start openoffice
```

To ensure service is well installed:

```
netstat –anp tcp
```

To delete the service type:

```
sc delete openoffice
```

More info at:

- http://yaib.eu/2008/10/30/run-openofficeorg-3-as-windows-service/
- http://support.microsoft.com/kb/q137890/
- http://www.iopus.com/guides/srvany.htm

References

[1] http://archive.services.openoffice.org/pub/openoffice-archive/stable/3.2.1/

[2] http://support.etouch.net/cm/wiki/?id=34626

[3] http://go.microsoft.com/fwlink/?LinkId=4544

Third-party software integration: SWFTools

To enable preview UI tab, OpenKM need to convert pdf files to swf.

> Starting from OpenKM 5.1 this configuration property has changed to **system.swftools.pdf2swf**.

Linux

Ubuntu

Swftools can be installed running in terminal:

```
$ sudo aptitude install swftools
```

Enable in OpenKM.cfg:

```
system.pdf2swf=/usr/bin/pdf2swf
```

> In Ubuntu 10.04 is possible you need to add this line to **/etc/apt/sources.list**
> ```
> deb [1] lucid partner
> ```

> In Ubuntu 10.10 this package is not included, so you need to compile it or grab from [2]
> ```
> $ sudo add-apt-repository ppa:ella-animation/dev
> $ sudo aptitude update
> $ sudo aptitude install swftools
> ```

Red hat

Better if you compile version 0.9.x

```
$ yum install zlib-devel libjpeg-devel giflib-devel freetype-devel gcc gcc-c++
$ wget http://www.swftools.org/swftools-0.9.1.tar.gz
$ tar xzf swftools-0.9.1.tar.gz
$ cd swftools-0.9.1
$ ./configure --prefix=/usr/local
$ make
$ make install
```

Debian

This package is not include in the Debian repositories, so you need to compile from source:

```
$ aptitude install build-essential libgif-dev xpdf libfreetype6 libfreetype6-dev libjpeg62 libjpeg8 libjpeg8-dev
$ wget http://www.swftools.org/swftools-0.9.1.tar.gz
$ tar xzf swftools-0.9.1.tar.gz
$ cd swftools-0.9.1
$ ./configure --prefix=/usr/local
$ make
$ make install
```

Or you can try to install the Ubuntu package. For Debian Squeeze I recommend the Ubuntu Lucid package.

Improving performance

For better performance, use this configuration. Create file pdf2swf_wrapper.sh

```
#!/bin/bash


/usr/bin/pdf2swf $* -f -t -G -s storeallcharacters
```

The last 5 parameters specifies that:

- -f : Fonts should be embedded, improves searchability in the document
- -t : Inserts a stop between each frame, improves stability
- -G : Makes the document smaller and faster to render
- -s storeallcharacters : Stores all character information about the texts in the document, improves searchability

Make it executable:

```
$ chmod +x pdf2swf_wrapper.sh
```

And don't forget to change the OpenKM.cfg file:

```
system.pdf2swf=/path/to/pdf2swf_wrapper.sh
```

> Take a look at pdf2swf wiki [3] for complete parameters reference, and remember that JBoss must be restarted after every OpenKM.cfg modification.

See also Troubleshooting: Some PDF file not shown in preview.

Windows

Donwload last swftools http://www.swftools.org/download.html and install it.

Enable in OpenKM.cfg (put this character to indicate path separator "/" not "\")

```
system.pdf2swf=c:/program files/swftools/pdf2swf.exe
```

Improving performance

For better performance in conversion, create file pdf2swf_wrapper.bat

```
c:\path\to\pdf2swf %1 %2 %3 -f -t -G -s storeallcharacters
```

The last 5 parameters specifies that:

- -f : Fonts should be embedded, improves searchability in the document
- -t : Inserts a stop between each frame, improves stability
- -G : Makes the document smaller and faster to render
- -s storeallcharacters : Stores all character information about the texts in the document, improves searchability

And don't forget to change the OpenKM.cfg file:

```
system.pdf2swf=c:/path/to/pdf2swf_wrapper.bat
```

> Take a look at pdf2swf wiki [3] for complete parameters reference, and remember that JBoss must be restarted after every OpenKM.cfg modification.

See also Troubleshooting: Some PDF file not shown in preview.

References

[1] http://archive.canonical.com/ubuntu

[2] https://launchpad.net/~ella-animation/+archive/dev:

[3] http://wiki.swftools.org/index.php/Pdf2swf

Third-party software integration: Antivirus

OpenKM can check if a submitted document is infected. It works with an Open Source antivirus software called ClamAV. Edit OpenKM.cfg and add this line:

```
system.antivir=/path/to/clamscan
```

This screenshot shows an error message from OpenKM because the submitted document is infected by a virus:
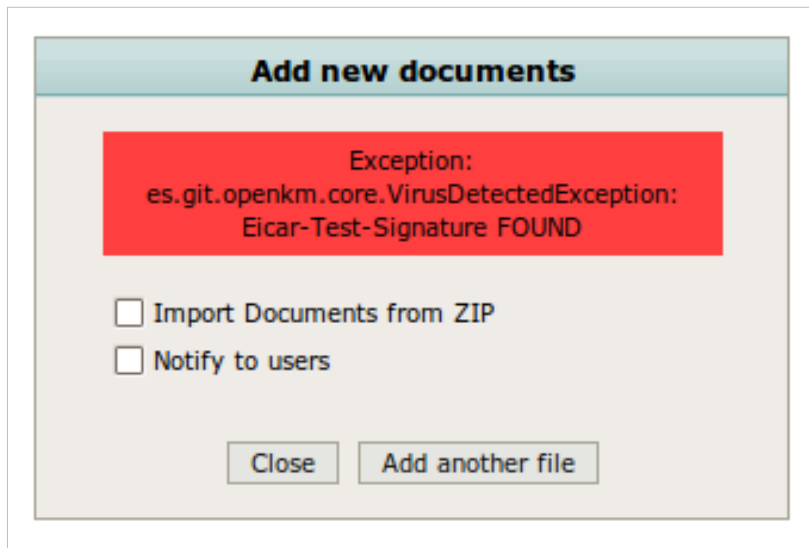
To install ClamAV on Debian / Ubuntu distribution:

```
$ sudo aptitude install clamav
```

To install ClamAV in Centos 5.2 you need more work. First create a file named */etc/yum.repos.d/dag.repo* with this content:

```
[dag]
name=Dag RPM Repository for Red Hat Enterprise Linux
baseurl=http://apt.sw.be/redhat/el$releasever/en/$basearch/dag/
gpgcheck=1
gpgkey=http://dag.wieers.com/packages/RPM-GPG-KEY.dag.txt
enabled=1
```

Now install the program as root:

```
$ yum install clamd.i386
```

Start the daemon:

```
$ /etc/init.d/clamd start
```

And update the virus database:

```
$ freshclam
```

Repository backup

OpenKM backup depends on repository configuration. By default, all documents are stored in $JBOSS_HOME/repository. This default folder can be changed using the property repository.home in the file OpenKM.cfg. If you want to make a backup, you should keep a copy of this folder. Is also recommended to backup files located at *$JBOSS_HOME/server/default/data/hypersonic* because there are several databases which handle important data:

- **OKMActivity**: Stores the user activity log. All the actions performed by an user in OpenKM are logged. Only the administrator user can access this log and search for specific actions and users. This database can be very big if there are many users and make an intensive use of the application.

- **OKMAuth**: This database is used to manage authentication. All the users, roles and passwords are here, so take care of it. If this database is damaged, you won't be able to log into OpenKM.

- **OKMDashboard**: Is used to keep the visited items of the dashboard.

- **OKMWorkflow**: Is used to store the process definitions and process instance data. Also is used to manage task and node transitions.

> It is recommended to stop JBoss before start the backup to prevent file modifications in the middle of the process because can make the backup useless.

The backup process should have these steps:

- Stop JBoss
- Backup important folders
- Start JBoss

In Unix / Linux you can backup the repository folder running these commands from $JBOSS_HOME. First stop JBoss:

```
$ ./bin/shutdown.sh -S
```

Once JBoss is stopped (you can see the JBoss log to ensure it has stopped) , make the backup:

```
$ tar czf repository.tgz repository
```

To check the backup, rename the repository folder:

```
$ mv repository repository-old
```

Restore the backup:

```
$ tar xzf repository.tgz
```

And start JBoss:

```
$ ./bin/run.sh
```

See also Backup scripts.

In Windows could be something like this:

- Stop JBoss

```
c:\jboss\bin\shutdown.bat -S
```

- Make backup

```
xcopy c:\jboss-4.2.3.GA\*.* e:\Backup\ /s/e
```

- Start JBoss again

```
c:\jboss-4.2.3.GA\bin\run.bat -b 0.0.0.0
```

The complete script may be like this one:

```
 c:\jboss\bin\shutdown.bat -S
 xcopy c:\jboss-4.2.3.GA\*.* e:\Backup\ /s/e
 c:\jboss-4.2.3.GA\bin\run.bat -b 0.0.0.0
```

There's a little problem in default shutdown.bat and run.bat script because makes a pause at ends, you might delete the final line in both scripts

```
if "%NOPAUSE%" == "" pause
```

> Due to Jackrabbit flexibility you can configure your OpenKM installation to store the documents in a database, for example. In this case you have to modify the backup procedure. Anyway the $JBOSS_HOME/repository folder should be backup because it contains repository meta data.

There is an alternative method for backups: you can export the whole repository from OpenKM administration. The folder structure will be re-created in the local filesystem, and all document on them but you will lose the document history and other metadata like Property Groups.

> Ensure you have enough free disk space before perform this operation.

Backup scripts

These backup scripts use rsync to minimize network load and creates incremental backups, preserving last four backups. For more info, read http://www.mikerubel.org/computers/rsync_snapshots/.

To install the cron job, run:

```
$ sudo crontab -e
```

And add these lines according to your personal configuration:

```
MAILTO=nomail@openkm.com
@weekly /root/backup.sh | tee -a /root/backup.log
```

Or, if you want to separate log reports by date:

```
MAILTO=nomail@openkm.com
@weekly /root/backup.sh | tee /root/backup.$(date
+\%Y.\%m.\%d_\%H.\%M.\%S).log
```

> MAILTO may also be used to direct mail to multiple recipients by separating recipient users with a comma.

For more information take a look at Crontab quick reference [1]

Remote backup with rsync

```bash
#!/bin/bash
#
## BEGIN CONFIG ##
HOST=$(uname -n)
FILES="/home/openkm"
BACKUP_DIR="/mnt/backup"
## END CONFIG ##
echo -e "### BEGIN: $(date +"%x %X") ###\n"

# Stop JBoss
/etc/init.d/jboss stop
while [ "$(ps -ef | grep java | grep jboss | wc -l)" -gt "0" ]; do
  sleep 5; echo ".";
done

# Copy to backup server
rsync -apzhR --stats --delete --exclude=*~ --delete-excluded $FILES
backup@server:$BACKUP_DIR/$HOST

# Start JBoss
/etc/init.d/jboss start
echo -e "\n### END: $(date +"%x %X") ###"
```

Remote backup with rsync and rotation

```bash
#!/bin/bash
#
## BEGIN CONFIG ##
HOST=$(uname -n)
FILES="/home/openkm"
BACKUP_DIR="/mnt/backup"
## END CONFIG ##
echo -e "### BEGIN: $(date +"%x %X") ###\n"

# Stop JBoss
/etc/init.d/jboss stop

# Copy to backup server
ssh backup@server "cd $HOST; rm -rf backup.3; mv backup.2 backup.3; mv
backup.1 backup.2; mv backup.0 backup.1"
rsync -apzhR --stats --delete --exclude=*~ --delete-excluded
--link-dest="$BACKUP_DIR/$HOST/backup.1" \
$FILES backup@server:$BACKUP_DIR/$HOST/backup.0

# Start JBoss
/etc/init.d/jboss start
echo -e "\n### END: $(date +"%x %X") ###"
```

Local backup with rsync and rotation

```bash
#!/bin/bash
#
## BEGIN CONFIG ##
HOST=$(uname -n)
FILES="/home/openkm"
BACKUP_DIR="/mnt/backup"
## END CONFIG ##
echo -e "### BEGIN: $(date +"%x %X") ###\n"

# Stop JBoss
/etc/init.d/jboss stop

# Calculate snapshot
LAST_SNAPSHOT=`ls -ltr $BACKUP_DIR | tail -1 | awk {'print $8'} | cut
-d . -f 2`
NEW_SNAPSHOT=$((LAST_SNAPSHOT+1))

# Copy to backup server
rsync -apzhR --stats --delete --exclude=*~
--exclude="$JBOSS_HOME/cache" --delete-excluded \
--link-dest="$BACKUP_DIR/$HOST/backup.$LAST_SNAPSHOT" $FILES
```

```
"$BACKUP_DIR/$HOST/backup.$NEW_SNAPSHOT"


# Start JBoss
/etc/init.d/jboss start
echo -e "\n### END: $(date +"%x %X") ###"


# Status
echo "==============================";
du -hs $BACKUP_DIR
echo "******************************";
du -hs --time $BACKUP_DIR/*
echo "==============================";
df -h | grep "$BACKUP_DIR"
echo "==============================";
```

Remote backup with rdiff-backup

> rdiff-backup from Debian throws a warning due to a deprecated method [os.popen2 is deprecated [2]], which can be hidden following the steps at [How to shut up Python deprecation warnings [3]].

```
#!/bin/bash
#
## BEGIN CONFIG ##
HOST=$(uname -n)
FILES="/home/openkm"
BACKUP_DIR="/mnt/backup"
## END CONFIG ##
echo -e "### BEGIN: $(date +"%x %X") ###\n"


# Stop JBoss
/etc/init.d/jboss stop


# Backup and purge old backups
rdiff-backup --remove-older-than 30B backup@server::$BACKUP_DIR/$HOST
rdiff-backup -v 3 --print-statistics --exclude /media --exclude /mnt \
  --include $FILES --exclude '**' / backup@server::$BACKUP_DIR/$HOST


# Start JBoss
/etc/init.d/jboss start
echo -e "\n### END: $(date +"%x %X") ###"


# Status
echo "==============================";
rdiff-backup --list-increment-sizes backup@server::$BACKUP_DIR/$HOST
echo "==============================";
```

More info about rdiff-backup:

- rdiff-backup home page [4]
- rdiff-backup wiki [5]

Local backup to USB disk

USB disk mount point can be defined in */etc/fstab* as:

```
/dev/sdb1          /mnt/backup      ext4    defaults        0        0
```

```bash
#!/bin/bash
#
## BEGIN CONFIG ##
HOST=$(uname -n)
MYSQL_PASS=""
OPENKM_HOME="/home/openkm"
JBOSS_HOME="$OPENKM_HOME/jboss-4.2.3.GA"
DATABASE_EXP="$OPENKM_HOME/db"
BACKUP_DIR="/mnt/backup"
## END CONFIG ##

if [ $(id -u) != 0 ]; then echo "You should run this script as root";
exit; fi

echo -e "### BEGIN: $(date +"%x %X") ###\n"
rm -rf $DATABASE_EXP
mkdir -p $DATABASE_EXP

# Mount disk
if mount | grep "$BACKUP_DIR type" > /dev/null; then
  echo "$BACKUP_DIR already mounted";
else
  mount $BACKUP_DIR;
fi

# Stop JBoss
/etc/init.d/jboss stop

# Clean logs
echo "Clean JBoss temporal files."
rm -rf $JBOSS_HOME/server/default/log
rm -rf $JBOSS_HOME/server/default/tmp
rm -rf $JBOSS_HOME/server/default/work

# Backup de MySQL
if [ -n "$MYSQL_PASS" ]; then
  MYSQL_DBS=$(mysqlshow -h localhost -u root -p$MYSQL_PASS | awk '(NR >
2) && (/[a-zA-Z0-9]+[ ]+[|]/) && ($2 != "mysql") && ($2 != "test") &&
($2 != "information_schema") { print $2 }');
```

```
  for DB in $MYSQL_DBS ; do
    echo "* Backuping MySQL data from $DB...";
    mysqldump -h localhost -u root -p$MYSQL_PASS $DB >
$DATABASE_EXP/mysql_$DB.sql
  done
  echo "-----------------------------------";
fi

# Backup and purge old backups
rdiff-backup --remove-older-than 30B $BACKUP_DIR/$HOST
rdiff-backup -v 3 --print-statistics --exclude /media --exclude /mnt \
  --include $OPENKM_HOME --exclude '**' / $BACKUP_DIR/$HOST

# Start JBoss
/etc/init.d/jboss start
echo -e "\n### END: $(date +"%x %X") ###"

# Status
echo "================================";
rdiff-backup --list-increment-sizes $BACKUP_DIR/$HOST
echo "********************************";
df -h | grep "$BACKUP_DIR"
echo "================================";

# Umount disk
sync
umount $BACKUP_DIR
```

Be sure to have the last JBoss startup script, otherwise use this piece of code to ensure that JBoss is stopped before backuping the files:

```
while [ "$(ps -ef | grep java | grep jboss | wc -l)" -gt "0" ]; do
  sleep 5; echo ".";
done
```

References

[1] http://adminschoice.com/crontab-quick-reference

[2] http://bugs.debian.org/cgi-bin/bugreport.cgi?bug=590107

[3] http://www.linuxinsight.com/how-to-shut-up-python-deprecation-warnings.html

[4] http://rdiff-backup.nongnu.org/

[5] http://wiki.rdiff-backup.org/wiki/index.php/Main_Page

Backup restoring

In cause of a hardware failure you may need to restore a backup. Backups are very important for OpenKM because you store a lot of documentation on it.

First of all, we need to locate the right backup. In case of rdiff-backup you can list with:

```
$ rdiff-backup --list-increment-sizes /path/to/backup
Sun Sep 11 05:00:19 2011        4.25 GB          4.25 GB   (current mirror)
Sun Sep  4 00:00:18 2011        13.3 MB          4.26 GB
Sun Aug 28 00:00:13 2011         674 MB          4.92 GB
Sun Aug 21 00:00:14 2011        5.50 MB          4.93 GB
Sun Aug 14 00:00:16 2011        1.75 MB          4.93 GB
Sun Aug  7 00:00:12 2011         288 KB          4.93 GB
Sun Jul 31 00:00:13 2011        43.0 KB          4.93 GB
Fri Jul 29 10:36:39 2011        5.56 KB          4.93 GB
```

Then we have decided to restore the backup of made on *Sun Aug 28 00:00:13 2011*:

```
$ rdiff-backup --restore-as-of 2011-08-28 /path/to/backup /path/to/destination
```

The **--restore-as-of** parameter accepts several formats. See **rdiff-backup** documentation for more info.

Inside of **/path/to/destination** you should see a directory **/home/openkm**, and inside it a couple of directories:

- **db**: The backup of the database.
- **jboss-4.2.3.GA**: The backup of the JBoss installation.

> The default OpenKM installation uses an embedded database called Hipersonic. So, the **db** directory may not exists, because all the needed info is under the **jboss-4.2.3.GA** directory. Refer to Using OpenKM with other databases for more info.

In case you have a MySQL database configured, these are the steps:

```
DROP DATABASE IF EXISTS okm_repo_bak;
DROP DATABASE IF EXISTS okm_app_bak;

CREATE DATABASE okm_repo_bak DEFAULT CHARACTER SET utf8 DEFAULT COLLATE
 utf8_bin;
CREATE DATABASE okm_app_bak DEFAULT CHARACTER SET utf8 DEFAULT COLLATE
utf8_bin;

GRANT ALL ON okm_repo_bak.* TO openkm@localhost WITH GRANT OPTION;
GRANT ALL ON okm_app_bak.* TO openkm@localhost WITH GRANT OPTION;
```

> We have already created a **openkm** user when the database was created on OpenKM installation. If you want to restore the backuped databases in another MySQL installation you will need to create this user. Refer to MySQL-OpenKM 5.0 for more info.

Now restore the previous backup to these new databases:

> Be careful and don't restore the data into the wrong database because you can damage your current OpenKM installation.

```
$ mysql -h localhost -u openkm -p okm_repo_bak < db/mysql_okm_repo.sql
$ mysql -h localhost -u openkm -p okm_app_bak < db/mysql_okm_app.sql
```

Copy the **jboss-4.2.3.GA** to **/home/openkm/jboss-4.2.3.GA_bak** and modify these configuration files to use the new backuped databases:

- repository.xml
- server/default/deploy/openkm-ds.xml

Stop the current JBoss (from **/home/openkm/jboss-4.2.3.GA**) and start the new one (**/home/openkm/jboss-4.2.3.GA_bak**). Log into the restored OpenKM and make a **Repository export** from **Administration**. Once completed, stop the backuped JBoss and start the original one. Then go to **Administration > Repository import** and import the desired content.

> In case of a complete restore from backup, you don't need to create the **okm_repo_bak** and **okm_app_bak** database nor modify and start the backuped JBoss.

Disaster recovering

In case of hardware failure maybe your OpenKM repository became corrupted. In this case you should restore from the last backup (See Backup restoring). But what happen if the backup is missing or even also corrupted? Everything is lost? Well, not exactly. Depending on the kind of disaster you can recover at least the document content.

> This recovering process will only work if OpenKM is configured to store the document content in a File DataStore.

By default, OpenKM is configured to use a File DataStore. This means thar every document content is stored in the filesystem. This give us very good performance when retrieving document content. But also has a hidden benefit. If your OpenKM installation has been damaged but you can read these DataStore file, you can recover these document contents.

To help up with the process, we have created a simple Bash script, called **rescue.sh**:

```bash
#!/bin/bash
RESCUE="rescue"
mkdir -p $RESCUE

for DOC in $(find repository/repository/datastore -type f); do
  FILE=$(basename $DOC)
  MIME=$(file -i $DOC | cut -d' ' -f2 | cut -d';' -f1)

  if [ $MIME == "text/plain" ]; then
    EXT="txt"
  elif [ $MIME == "application/pdf" ]; then
    EXT="pdf"
  elif [ $MIME == "text/rtf" ]; then
    EXT="rtf"
  elif [ $MIME == "application/vnd.ms-office" ]; then
    EXT="doc"
  elif [ $MIME == "application/vnd.oasis.opendocument.text" ]; then
```
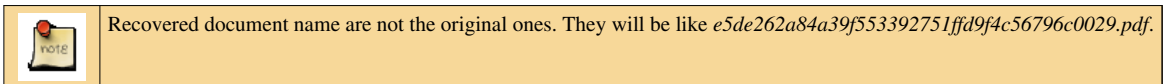
```
    EXT="odt"
  elif [ $MIME == "image/gif" ]; then
    EXT="gif"
  elif [ $MIME == "image/png" ]; then
    EXT="png"
  elif [ $MIME == "image/jpeg" ]; then
    EXT="jpg"
  elif [ $MIME == "image/tiff" ]; then
    EXT="tif"
  elif [ $MIME == "application/zip" ]; then
    EXT="zip"
  elif [ $MIME == "application/x-dosexec" ]; then
    EXT="exe"
  elif [ $MIME == "application/octet-stream" ]; then
    EXT="bin"
  else
    EXT=$MIME
  fi

  cp -v $DOC $RESCUE/$FILE.$EXT
done
```

You need to run this script from **$JBOSS_HOME** directory. All the recovered documents will be copied to the **$JBOSS_HOME/rescue** directory. This script is only a guide and should be improved to support more MIME types.

> Recovered document name are not the original ones. They will be like *e5de262a84a39f553392751ffd9f4c56796c0029.pdf*.

If you have a functional OpenKM installation but want to recover deleted documents from a damaged backup you can also make use of this script. After you have executed the **rescue.sh** script you will have a lot of file under the **rescue** directory. To remove the document already have in the current OpenKM installation, copy the rescue directory to the working OpenKM installation and execute this script, called **prune.sh**:

```
#!/bin/bash
RESCUE="rescue"

for DOC in $(find repository/repository/datastore -type f); do
  FILE=$(basename $DOC)

  if [ -f $RESCUE/$FILE.* ]; then
    rm $RESCUE/$FILE.*
  fi
done
```

Repository mirroring

To enhance OpenKM availability you can have two instances of the application running in different servers. If the principal server gets down due to a hardware failure you can switch to the mirrored server and keep working.
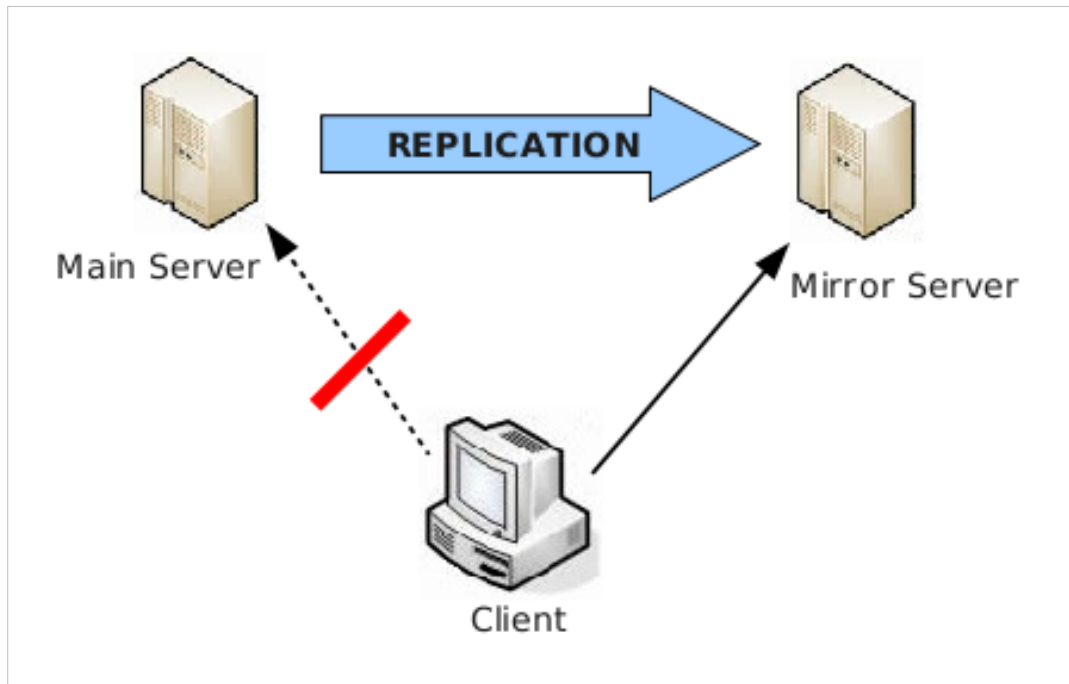
The replication only is done in one direction: from main server to mirror server. Mirror server can be configured to be in read-only mode so users can't add or modify documents. This behavior can be achieved using the system.readonly configuration parameter.

The follow script will propagate the repository changes from main server to mirror server:

```bash
#!/bin/bash
echo -e "### BEGIN: $(date +"%x %X") ###\n"

# Stop local JBoss
/home/openkm/OpenKM/secure_stop.sh

# Stop remote JBoss
ssh root@192.168.1.101 '/home/openkm/OpenKM/secure_stop.sh'

# Sync OpenKM repositories
rsync -rahe 'ssh -p 22022' --stats --delete --exclude OpenKM/OpenKM.cfg
 --exclude OpenKM/server/default/log /home/openkm/OpenKM
root@192.168.1.101:/home/openkm

# Start local JBoss
/etc/init.d/jboss start

# Start remote JBoss
ssh root@192.168.1.201 '/etc/init.d/jboss start'
```

```
echo -e "\n### END: $(date +"%x %X") ###"
```

This script should be executed - by user root - from the main server, and can be scheduled to be executed every day using Linux cron utility, for example. The script uses rsync to minimize network load and only modified or added documents will be transferred. The whole process can take a few minutes, depending on your repository activity.

Note that both OpenKM are installed at the openkm user home (/home/openkm) for simplicity.

And this is the secure_stop.sh script:

```
#!/bin/bash

# Stop JBoss
/etc/init.d/jboss stop
while [ "$(ps -ef | grep java | grep jboss | wc -l)" -gt "0" ]; do
   sleep 5; echo ".";
done
```

WebDAV access

Below you there is a list of WebDAV clientes grouped by Operating System. Once you have installed any of these tools, you can access to OpenKM repository with this URL:

```
http://your-domain.com/OpenKM/repository/default
```

To enhance compatibility, starting with OpenKM 5.0.4 when an WebDAV path is like **/okm:root/myfile.txt** it is translated to **/okm_root/myfile.txt** because some WebDAV clients does not like character : in the path. Starting with OpenKM 5.1 we have added a new configuration parameter called **system.webdav.fix** to make this behavior configurable.

	Since OpenKM 5.0 WebDAV is disabled by default. If you want to enable it, add
	`system.webdav≠on`
	to your OpenKM.cfg configuration file. Starting with OpenKM 5.1 this property will be renamed to:
	`system.webdav.server≠on`

Windows

From Windows you can access OpenKM repository exposed via WebDAV using one o this tools:

* http://www.netdrive.net/(Free for home use but does not handle document versioning properly)
* http://www.webdrive.com/(The recommended option and create a new document version when you save the modifications)

	This page from Greenbytes [1] is an excellent resource to analyze & fix problems with your Windows WebDAV client.

You can try connecting directly using windows networking:

* open networking
* connect as network resource
* write url http://host:8080/OpenKM/repository/default/okm:root

- write user and password

Connection from Windows without third-party software

- Open the Registry Editor utility (regedit.exe).
- Into **HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\WebClient\Parameters** look for **TypeUseBasicAuth** and replace the default value (1) by 2 allowing BasicAuth through NON SSL connections.
- Then map your drive using this address: \\localhost:8080/OpenKM/repository/default/okm_root/

See also: Forum: OpenKM + Webdav + Windows 7 => finally working [2].

Windows slow WebDAV performance

If you have a slow webdav connection in windows 7 please follow the instructions in Fix Slow WebDAV Performance in Windows 7 [3]. *It turns out to be a windows issue, not OpenKM*.

Furthermore in Windows 7/Vista you need to add your certificate (if its not issued by a verified CA) into windows if you plan to use Webdav over SSL (and,like me, do not compensate with digest auth):

1. Get the certificate from the server administrator in CER format or download it from Firefox.
2. Start windows Certificate Manager: in "cmd" type certmgr.msc
3. Select on the left "Trusted Root Certification Authorities"
4. In the menu "Actions" -> "All Tasks" select "Import" and then "Next"
5. Browse for the location of certificate file you saved earlier on your computer and click "Open".
6. You will get a few more warning messages. Just say to go through with it until the certificate is accepted.

Then you can map webdav content as a network drive without any other software. The folder should be: https://server:port/OpenKM/repository/default/

To make sure that everything is ok with certificates, open the address using IE. If no warning pops up then its a go.

This allows you to read/view files. If you need write access over a network drive you need to disable File Locking (windows does not support webdav file lock).

Note: For Windows XP there is a fix (KB907306).

See also: Forum: OpenKM + Webdav + Windows 7 => finally working [2].

Mac

From OS X you can use:

- http://cyberduck.ch/
- http://www.webdrive.com/(Yes, there is a version for Macintosh OS/X)

Linux

From Linux you can use:

- http://dav.sourceforge.net/
- http://savannah.nongnu.org/projects/davfs2

For example, in Ubuntu or any Debian-based Linux distro you can install this utility this way:

```
$ sudo aptitude install davfs2
```

Make a mount point:

```
$ mkdir ~/okmdav
```

And mount it:

```
$ sudo mount.davfs "http://localhost:8080/OpenKM/repository/default" ~/okmdav
```

See also:

- Ubuntu: Mounting remote filesystem using davfs2 (FUSE) [4]

References

[1] http://greenbytes.de/tech/webdav/webdav-redirector-list.html

[2] http://forum.openkm.com/viewtopic.php?f=4&t=5183&p=12123

[3] http://oddballupdate.com/2009/12/18/fix-slow-webdav-performance-in-windows-7/

[4] http://sysblogd.wordpress.com/2007/09/09/ubuntu-mounting-remote-filesystem-using-davfs2-fuse/

Property Groups

This is easy and powerful for building custom properties. You can user this feature to add information to documents. For example, if you store the invoices in a folder you can add a property group to set the client name and the quantity. You will see this important data without opening the documents. Also you can search by these custom properties later.

New feature in OpenKM 4.1
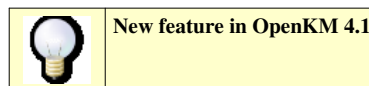
Refer to Property Groups definition where you can see the new way to describe Property Groups.

To try this test login as okmAdmin and go to Administration → Utilities → Register property groups. Set full path with the location of the local file *PropertyGroups.cnd* and push the register button. The properties will be registered.

Register property groups

Property Group definition path /home/openkm/PropertyGroups.cnd   Register

REGISTERED GROUPS CANNOT NOT BE UNREGISTERED. They can always be updated but they cannot be deleted after registration !!! Try it out on a test installation first.

The users can assign a property group to a document using these toolbar buttons:

- Add property group →
- Remove property group →

And the properties will be shown in the information area. You can edit the values of the properties using the Change button. Then you can change these values. Also can remove a property group from a document by clicking the Delete button.

OpenKM 4.0 and older

You need to create some file, for example **PropertyGroup.cnd** and insert some values in other two files **PropertyGroupBundle.properties** and **PropertyGroupValues.properties**.

- $JBOSS_HOME/PropertyGroup.cnd
- $JBOSS_HOME/PropertyGroupBundle.properties
- $JBOSS_HOME/PropertyGroupValues.properties

About files:

- **PropertyGroup.cnd**: There are general definitions on each parameter (for example the parameter okp:technology.type)

```
<okm = 'http://www.openkm.org/1.0'>
<okg = 'http://www.openkm.org/group/1.0'>
<okp = 'http://www.openkm.org/group/property/1.0'>


[okg:technology] mixin
- okp:technology.type (string) = '' multiple autocreated
- okp:technology.language (string) = '' autocreated
- okp:technology.description (string) = '' autocreated
- okp:technology.comment (string) = '' autocreated


[okg:consulting] mixin
- okp:consulting.type (string) = '' autocreated
- okp:consulting.comment (string) = '' autocreated
```

- **PropertyGroupValues.properties**: In this file you can see the properties values, for example for parameter (okp\:technology.type=4,one,two,three) has 3 values (one, two, three) the first number indicates the order, the second one the type (1 = list, 2 = input, 3 = select multiple, 4 = Text area)

```
okp\:technology.type=1,4,one,two,three
okp\:technology.language=2,3,java,c,c#,c++,python
okp\:technology.description=3,1
okp\:technology.comment=4,2
okp\:consulting.type=1,3,one,two,three
okp\:consulting.comment=2,2
```

- **PropertyGroupBundle.properties**: This file contains general translations for each value you can have more files for each language like PropertyGroupBundle_ES.properties etc.

```
okg\:technology=Tecnology
okp\:technology.type=Type
okp\:technology.type.one=Type One
okp\:technology.type.two=Type Two
okp\:technology.type.three=Type Three
okp\:technology.language=Language
okp\:technology.language.java=Java
okp\:technology.language.c=C
okp\:technology.language.c++=C++
okp\:technology.language.python=Python
okp\:tecnology.description=Description
okp\:technology.comment=Comment
okg\:consulting=Consulting
okp\:consulting.type=Type
okp\:consulting.type.one=Type One
okp\:consulting.type.two=Type Two
okp\:consulting.type.three=Type Three
okp\:consulting.comment=Comment
```

OpenKM tries to find the correct language in your property files. If it does not exist uses it uses *PropertyBundle.properties*.

Property Groups definition

Sometimes you want to add more information to a document, for example, you may want to add an internal code to the document for organization purposes and a description. OpenKM uses two files to describe this metadata: **PropertyGroups.cnd** and **PropertyGroups.xml**. The first one describes the properties to be added to the document node. This definition is used by Jackrabbit to retrieve and set these values. The second one is used by OpenKM to render the values described in the previous file.

> - For **OpenKM 5.1** the property groups DTD is **property-groups-2.0.dtd**.
> - For **OpenKM 5.0** the property groups DTD is **property-groups-1.1.dtd**.
> - For **OpenKM 4.1** is **property-groups-1.0.dtd**.

See this sample:

PropertyGroups.cnd

```
<okm = 'http://www.openkm.org/1.0'>
<okg = 'http://www.openkm.org/group/1.0'>
<okp = 'http://www.openkm.org/group/property/1.0'>

[okg:consulting] mixin
- okp:consulting.code (string) = '' autocreated
- okp:consulting.comment (string) = '' autocreated
```

PropertyGroups.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE property-groups PUBLIC "-//OpenKM//DTD Property Groups 2.0//EN"
```

```
"http://www.openkm.com/dtd/property-groups-2.0.dtd">
<property-groups>
  <property-group label="Consulting" name="okg:consulting">
    <input label="Code" type="text" name="okp:consulting.code" value="COD-0000" />
    <textarea label="Comment" name="okp:consulting.comment" value=""/>
  </property-group>
</property-groups>
```

Here we can see an XML which describes the "Consulting" property group with a couple of properties. In the description, we can see an INPUT where the user should type an internal code (with a default value of "COD-0000" and a comment which describes the document content.

As you can see in the XML DOCTYPE, there is a formal definition at http:/ / www. openkm. com/ dtd/ property-groups-2.0.dtd.

> If your OpenKM server can't access to Internet, you can download this DTD and copy to a convenient place at you server. Remember to update your PropertyGroups.xml with the TDT location.
>
> ```
> <!DOCTYPE property-groups PUBLIC "-//OpenKM//DTD Property Groups 2.0//EN"
>
> "file:///path/to/property-groups-2.0.dtd">
> ```

But let's see which components it can handle:

- input
- suggestbox
- checkbox
- textarea
- select
- separator
- text

See Form Element description for a detailed definition of every form element. These form elements have also an additional property called "readonly" with can be used to make a property not modifiable by the user.

Debugging OpenKM

You can debug your OpenKM installation using the JBoss logging facility. This is an useful thing when you have problems with your configuration. Default OpenKM installation tries to log important events like errors and warnings. Is possible to change this configuration editing the file *$JBOSS_HOME/server/default/conf/jboss-log4j.xml*.

Default JBoss log configuration can generate a lot of messages. These files are stored at *$JBOSS_HOME/server/default/log*. It is configured to use the DailyRollingFileAppender. This appender create a new log file for every day. This is better than have a unique huge log file, os course. The rollover is performed at midnight each day, but you can configure it to make the rollover every hour (uncomment the proper line).

```xml
<!-- A time/date based rolling appender -->
<appender name="FILE"
          class="org.jboss.logging.appender.DailyRollingFileAppender">
   <errorHandler class="org.jboss.logging.util.OnlyOnceErrorHandler"/>
   <param name="File" value="${jboss.server.log.dir}/server.log"/>
   <param name="Append" value="false"/>
   <!-- Rollover at midnight each day -->
   <param name="DatePattern" value="'.'yyyy-MM-dd"/>
   <!-- Rollover at the top of each hour
   <param name="DatePattern" value="'.'yyyy-MM-dd-HH"/>
   -->
   <layout class="org.apache.log4j.PatternLayout">
      <!-- The default pattern: Date Priority [Category] Message\n -->
      <param name="ConversionPattern" value="%d %-5p [%c] %m%n"/>
      <!-- The full pattern: Date MS Priority [Category] (Thread:NDC) Message\n
      <param name="ConversionPattern" value="%d %-5r %-5p [%c] (%t:%x) %m%n"/>
      -->
   </layout>
</appender>
```

You can reduce the amount of log messages produced by OpenKM, or can increase them. In this example we limit the log messages produced by the class OKMAccessManager for those of type ERROR.

```xml
<category name="com.openkm.core.OKMAccessManager">
   <priority value="ERROR" />
</category>
```

If you create this configuration:

```xml
<category name="com.openkm">
   <priority value="DEBUG" />
</category>
```

All the log messages generated by OpenKM will be shown. As you can see, you can increase debug messages in some parts of OpenKM to check a determinate behavior.

Debugging JAAS configuration

If you are trying to setup another authentication source different from the default provided by OpenKM, you can afford some problems. The JBoss login-config.xml is supposed to have the right configuration, but you can't log into the application. The most common case is a bad or missing JAAS configuration. So if you need to debug the JAAS, you can add to the log4j.xml file the following:

```xml
<category name="org.jboss.security">
   <priority value="TRACE" class="org.jboss.logging.XLevel"/>
   <appender-ref ref="SECURITY_F"/>
</category>
<appender name="SECURITY_F" class='org.jboss.logging.appender.DailyRollingFileAppender'>
   <param name="Append" value="true"/>
   <param name="DatePattern" value="'.'yyyy-MM-dd"/>
   <param name="File" value="${jboss.server.home.dir}/log/jboss.security.log"/>
   <layout class="org.apache.log4j.PatternLayout">
      <param name="ConversionPattern" value="%d{ABSOLUTE} %-5p [%c] %m%n"/>
   </layout>
</appender>
```

This is more or less in the middle of the file, just where the <category-name> section begins. And you should look at this new log file:

```
$ tailf -f $JBOSS_HOME/server/default/log/jboss.security.log
```

Email error notification

Always is good idea to be notified when things goes wrong. There are some log appenders that can help you. The SMTPAppender will mail you log messages with threshold ERROR by default. You can lower this threshold, but you will got lots of useless mail messages. Here you must configure some properties:

- **To**: The mail account where the messages will arrive.
- **From**: You can set it simply as noreply@your-domain.com.
- **Subject**: Here you can specify the subject of the mail. If you have several OpenKM installations, you can create a filter in your mail client using this value.
- **SMTPHost**: The mail server server. Can be localhost if there is a mail server installed in this computer.

```xml
<!-- EMail events to an administrator -->
<appender name="SMTP" class="org.apache.log4j.net.SMTPAppender">
  <errorHandler class="org.jboss.logging.util.OnlyOnceErrorHandler"/>
  <param name="Threshold" value="ERROR"/>
  <param name="To" value="admin@myhost.domain.com"/>
  <param name="From" value="nobody@myhost.domain.com"/>
  <param name="Subject" value="JBoss Sever Errors"/>
  <param name="SMTPHost" value="localhost"/>
  <param name="BufferSize" value="10"/>
  <layout class="org.apache.log4j.PatternLayout">
      <param name="ConversionPattern" value="[%d{ABSOLUTE},%c{1}] %m%n"/>
  </layout>
</appender>
```

In Unix / Linux systems there is a centralized log manager called syslog. You can configure Log4J to use this system using the SyslogAppender:

```xml
<!-- Syslog events -->
<appender name="SYSLOG" class="org.apache.log4j.net.SyslogAppender">
  <errorHandler class="org.jboss.logging.util.OnlyOnceErrorHandler"/>
  <param name="Facility" value="LOCAL7"/>
  <param name="FacilityPrinting" value="true"/>
  <param name="SyslogHost" value="localhost"/>
  <layout class="org.apache.log4j.PatternLayout">
      <param name="ConversionPattern" value="[%d{ABSOLUTE},%c{1}] %m%n"/>
  </layout>
</appender>
```

• **SyslogHost**: This configuration parameters allows you to specify the local syslog or a remote syslog server used to centralize the network log management.

For more info, visit:

• http://jboss.org/community/docs/DOC-11280
• http://jboss.org/community/docs/DOC-9372
• http://jakarta.apache.org/log4j
• http://primalcortex.wordpress.com/2007/11/28/jboss-and-jaas-debug
• can I get log4j to delete old rotating log files? [1]

References

[1] http://stackoverflow.com/questions/1050256/how-can-i-get-log4j-to-delete-old-rotating-log-files|How

Troubleshooting

This is a compilation of most common OpenKM user problems with its solution. If you have any other problem not listed here, please take a look at http://forum.openkm.com/to be helped by our community. Also you can request professional support at http://www.openkm.com/Contact.

Can't access OpenKM from other computers

By default, JBoss only listen for connections on localhost. If you want to access from other computer, you need to start JBoss passing the parameter -b 0.0.0.0

GNU/Linux

```
./bin/run.sh -b 0.0.0.0
```

Windows

```
\bin\run.bat -b 0.0.0.0
```

Other options

```
./bin/run.sh -Djboss.bind.address=0.0.0.0
```

Error downloading document as PDF

```
Code: OKM-004006
Message: conversion failed: could not save output document; OOo
errorCode: 2074
```

This error occurs when you want to download a document as PDF and there is some error with OpenOffice.org. Perhaps the OpenOffice.org installation has some missing components. Try to install these packages:

```
$ aptitude install openoffice.org2-l10n-es openoffice.org-java openoffice.org-java-common openoffice.org
```

```
Code: OKM-004013
Message: connection failed:
socket,host=localhost,port=8100,tcpNoDelay=1:java.net.ConnectException:
 Connection refused'
```

This error occurs when you want to download a document as PDF and there is some error with OpenOffice.org connection. Perhaps there is no OpenOffice.org running in server mode (headless). See Third-party software integration: OpenOffice.org for more info.

OpenKM not deployed due to a webservice error

Perhaps your computer have no Internet access? JBoss tries to validate WSDL descriptors on application deployment, can't access to DTD and throws an exception. This is a know JBoss issue [1]. More on JBoss webservices at http://jboss.org/community/docs/DOC-10823.

A workaround is disabling webservices in OpenKM:

- Stop JBoss.
- Delete the folder
  $JBOSS_HOME/server/default/deploy/OpenKM.ear/OpenKM.war/WEB-INF/classes/es/git/openkm/ws.
- Start JBoss again.

It should solve the deployment problem. If the EAR is not exploded you can browse it with Midnight Commander (from Linux). or WinRAR (from Windows).

Error when click on "Copy to clipboard"

To enable JavaScript "Copy to clipboard" buttons on Firefox and Mozilla, enter the following line into browser address bar: "about:config"

The list of preferences should open in browser window. Click right mouse button somewhere in active part of preference list window and choose New > Boolean option from context menu. Paste the following Preference name into the New boolean value window input box: "signed.applets.codebase_principal_support"

Error opening contextual menu - right click on firefox"

I right-click on a document, but OpenKM contextual menu is hidden by the browser. If you use Firefox, go to Edit → Preferences → Content, check "Active JavaScript", click on "Advanced" and check "Deactivate or replace contextual menus" option.

The Administration tab is empty

This is a known bug of IExplorer. Switch to Firefox: its JavaScript engine also is faster, so your OpenKM experience will be better. Or access it from Tools → Show administration. Or point your browser to http://localhost/OpenKM/es.git.openkm.backend.Main/.

IMAP SSL certificate error

First obtained my certificate with the openssl command and after this import the certificate with:

```
keytool -import -alias mail.yourcompany.com -keystore /etc/java-6-sun/security/cacerts -file imapd.pem
```

You have to restart JBoss to refresh the SSL certifications. See also: http://confluence.atlassian.com/display/JIRA/Connecting+to+SSL+services

Unable to run Scanner appler under Linux

If you enable the Java Console and see this error:

```
java.lang.UnsatisfiedLinkError: uk.co.mmscomputing.util.JarLib.loadX: Could not find library [libjsane.so]
```

You have to install the **libsane-dev** package:

```
$ sudo aptitude install libsane-dev
```

Some PDF file not shown in preview

Perhaps the PDF is copy protected. Hence, you're not allowed to convert it.

```
FATAL    PDF disallows copying
```

If you created this PDF using Acrobat Distiller, try switching the copy protection checkbox off. If you can't disable the protection you can modify the *'pdf2swf* source code to disable this check (see http://www.foolabs.com/xpdf/cracking.html).

Other problem may be due to complex PDF documents.

```
ERROR This file is too complex to render- SWF only supports 65536 shapes at once
```

You can invoke the *pdf2swf* utility with the *-s poly2bitmap* option. The generated document quality is worse, but you can see it at least. This can be simplified to be used by OpenKM in the preview generation. Create a script called *pdf2swf_wrapper.sh*:

```bash
#!/bin/bash

/usr/bin/pdf2swf -s poly2bitmap $*
```

Make it executable:

```
$ chmod +x pdf2swf_wrapper.sh
```

And don't forget to change the OpenKM.cfg file:

```
system.pdf2swf=/path/to/pdf2swf_wrapper.sh
```

See SWFTools for more info.

Random errors in preview

We have detected errors in special with Firefox under Linux, because the Flash plugin is quite old or you have installed the GNASH flash plugin and not the original from Adobe. To fix this, install the "Flash Aid" plugin under Firefox and run the wizard. It will install the right Flash Plugin version into your preferred browser.

Export to zip file gives wrong charset

This issue is corrected since OpenKM 4.1-RC2 and exported zip archives should be opened without problems from WinZip 14 at least. But you may have an issue in Ubuntu Linux. The problem is the *unzip* package, which should be compiled with a patch to enable UTF-8 chars in file names. You can find a patched *unzip* for Ubuntu/Debian at http://packages.ubuntu.com/jaunty/unzip.To keep informed about the progress in this issue, go to Bug #10979: unzip does not support UTF-8 filenames [2].

Also you can try to install the port of the 7-Zip compression tool for Linux:

```
$ sudo aptitude install p7zip-full
```

Jackrabbit: Bad check digit

If you've got some problem like this when starting OpenKM

```
2010-06-03 17:56:14,640 ERROR
[org.apache.jackrabbit.webdav.simple.DavResourceImpl] Error while
importing resource: java.io.IOException: cannot read item
dd85e9b7-6398-4720-a00d-87374467e4f9
2010-06-03 17:56:14,640 WARN
[org.apache.jackrabbit.core.lock.LockManagerImpl] Bad lock token: Bad
check digit.
```

Probably you got a problem with the crappy LVM setup that was responsible for this problem. Fresh installation could solve the problem!

Take a look at forum post at http://forum.openkm.com/viewtopic.php?f=3&t=3806&start=0

Errors caused by proxy configuration / Error OKM-009006

If you've configured OpenKM under a proxy, sometimes is needed to set jboss params to indicate it, because could be some connection errors in rpc calls.

```
-Dhttp.proxyHost=x.x.x.x -Dhttp.proxyPort=yyyy
```

Microsoft Office Addin disappears after restarting Office

Could be a problem with default template (normal.dot) to solve it:

- Open normal.dot
- Go to menu tools / personalize / Tool bars
- Open OpenKM toolbar
- Save normal.dot

OpenKM server has no direct Internet connection

If you need a proxy to access Internet, edit your **/etc/bash.bashrc** file as root and put these line at the end of your /etc/bash.bashrc file :

```
export http_proxy=http://username:password@proxyserver.net:port/
export ftp_proxy=http://username:password@proxyserver.net:port/
```

You can omit the username:password, if your proxy server has no password.

But may be the server is can't access Internet at all. In this case you will have problems when validating *$JBOSS_HOME/PropertyGroups.xml*. Read Property Groups definition to see there is a easy solution. Also edit OpenKM.cfg and set:

```
update.info=off
```

Access denied installing OpenKM as service on Windows 7

The command shell works slightly differently in Vista. By default when you run cmd.exe, it does not give administrator rights. To get around it, create a shortcut to cmd.exe on desktop, open Properties, Compatibility tab and check box for Administrator rights. Then you can do admin stuff on cmd line. More information here Configure JBoss service

Problems with character encodings

If your language has non-ASCII chars and they are displayed as garbage you have a problem with encodings. To solve it, you have to edit the JBoss start script.

> This process, as described here will destroy OpenKM database and will create a new one. Be careful! If you want a non-destructive way of achieve this, you need to delete actual inserted translations and insert them again.

In case of Linux, edit the file $JBOSS_HOME/bin/run.sh and change the line:

```
JAVA_OPTS="-Xms256m -Xmx2048m -XX:PermSize=128m -XX:MaxPermSize=256m"
```

By this others:

```
JAVA_OPTS="-Xms256m -Xmx2048m -XX:PermSize=128m -XX:MaxPermSize=256m -Dfile.encoding=utf-8"
```

In case of Windows, edit the file $JBOSS_HOME/bin/run.bat and change the line:

```
set JAVA_OPTS=%JAVA_OPTS% -Dprogram.name=%PROGNAME%
```

By this others:

```
set JAVA_OPTS=%JAVA_OPTS% -Dprogram.name=%PROGNAME% -Dfile.encoding=utf-8
```

You need to stop JBoss, delete the $JBOSS_HOME/server/default/data/hypersonic folder and start JBoss again. Now the OpenKM database will be created again and the language translation will be imported right.

Checking error "Document name is different"

This is a known error in Internet Explorer family. To disable name mismatch check, take a look at Disable document name mismatch check [3].

Error dowloading files under IE with https / SSL

There's a very older bug from version 5.5 until our days that seems Microsoft has not decided to solve, that causes a problem downloading files with IE behind SSL protocol (the force may be with you, hope some day will decide to solve it).

The problem is caused by "**Cache-control: no-cache HTTP header to the response message**", obviously that's valid for any browser except IE family ... other Microsoft IE hidden feature ?

Can see some supporting and forum information about it:

- IE SSL download problem [4]
- http://support.microsoft.com/kb/815313
- http://support.microsoft.com/kb/323308
- http://support.microsoft.com/kb/316431

The solution is change a couple of entries in the Windows Registry:

- Create a file called ss_download_patch.reg

- Copy the contents

```
REGEDIT4
[HKEY_LOCAL_MACHINE\Software\Microsoft\Windows\CurrentVersion\Internet Settings]
"DisableCachingOfSSLPages"=dword:00000000
"BypassSSLNoCacheCheck"=dword:00000001
"BypassHTTPNoCacheCheck"=dword:00000001
"SSLPragmaNoCache"=dword:00000001
```

- Execute it (will be imported into Windows Registry)
- Then you need to restart Windows to take effect.

	This IE bug has been fixed in OpenKM 5.1.8 (nighbuild from 6 Oct 2011).

Error in uploader / scanner applet

To provide a more complete information in order to determine the issue, please activate the Java Console. Read How do I enable and view the Java Console? [5] for more info.

PermGen errors

If you see this error when executing OpenKM:

```
java.lang.OutOfMemoryError: PermGen space
```

Edit your $JBOSS_HOME/bin/run.sh (or $JBOSS_HOME/bin/run.bat if using Windows) and set the **JAVA_OPTS** variable to this value:

```
JAVA_OPTS="-Xms256m -Xmx2048m -XX:PermSize=128m -XX:MaxPermSize=256m"
```

See also [Presenting the Permanent Generation [6]].

References

[1] http://www.jboss.com/index.html?module=bb&op=viewtopic&p=4125074
[2] https://bugs.launchpad.net/ubuntu/+source/unzip/+bug/10979
[3] http://wiki.openkm.com/index.php/OpenKM.cfg#Disable_document_name_mismatch_check
[4] http://www.sibsoft.net/forum/post1670.html
[5] http://www.java.com/en/download/help/javaconsole.xml
[6] http://blogs.oracle.com/jonthecollector/entry/presenting_the_permanent_generation

Error codes

OpenKM error codes are composed by OKM-XXX-YYY codification error. First 3 digits indicates the origin and second tree digits indicates the cause.

First tree digits - Origin

OKM-001 Some problem getting folders to server.

OKM-002 Some problem getting documents to server.

OKM-003 Some problem in remote server session.

OKM-004 Some problem downloading document form server.

OKM-005 Some problem uploading document to server.

OKM-006 Some problem on server session.

OKM-007 Some problem on authentication.

OKM-008 Some problem making a search.

OKM-009 Some problem with property groups (metadata).

OKM-010 Some problem with notification service.

OKM-011 Some problem with bookmark.

OKM-012 Some problem with repository (internal).

OKM-013 Some problem into general utils service.

OKM-014 Some problem with authetication in admin mode.

OKM-015 Some problem with search in admin mode.

OKM-016 Some problem in remote session in admin mode.

OKM-017 Some problem with property groups (metadata) in admin mode.

OKM-018 Some problem in repository (internal) in admin mode.

OKM-019 Some problem getting folder in admin mode.

OKM-020 Some problem getting dashboard

OKM-021 Some problem getting workspace.

OKM-022 Some problem executing a workflow.

OKM-023 Some problem uploading a workflow in admin mode.

OKM-024 Some problem with mail service.

OKM-025 Some problem getting properties.

OKM-026 Some browser problem.

OKM-027 Some problem in contact.

OKM-028 Some problem in activity log.

Second tree digits - Cause by

001 Caused by repository.

002 Caused by item not found.

003 Caused by item not exist.

004 Caused by lock.

005 Caused by unlock.

006 Caused by general cause.

007 Caused by general cause.

008 Caused by GWT shell environment.

009 Caused by access denied.

010 Caused by unsupported mime type.

011 Caused by file size exceed.

012 Caused by no such group.

013 Caused by IO exception.

014 Caused by no such property.

015 Caused by path not found.

016 Caused by version.

017 Caused by session lost.

018 Caused by file not found.

019 Caused by parse exception.

020 Caused by invalid node type definition.

021 Caused by sql exception.

022 Caused by configuration problem.

023 Caused by quota exceed.

024 Caused by database problem.

025 Caused by principal adapter problem.

026 Caused by workflow problem.

027 Caused by document name problem.

028 Caused by number format problem.

029 Caused by login problem.

030 Caused by document problem.

031 Caused by eval problem.

032 Caused by virus problem.

033 Caused by conversion problem.

034 Caused by authentication problem.

035 Caused by google problem.

Using OpenKM with other databases

By default, OpenKM uses an embedded database called HSQLDB [1]. This database is integrated into JBoss and offers a good performance and low hardware requirements. But sometimes it not enough for some users and need OpenKM to be deployed within their IT structure, which is based on any other database vendor.

Startign with OpenKM 5, you can create the databases automatically configuring the **hibernate.dialect** and **hibernate.hbm2ddl** properties in OpenKM.cfg.

```
hibernate.dialect=org.hibernate.dialect.HSQLDialect
hibernate.hbm2ddl=create
```

Once the tables are created, change the **hibernate.hbm2ddl** property from *create* to *none*. Also are interesting a couple of new configuration property:

```
hibernate.show_sql=false
hibernate.statistics=true
```

With these properties you can configure if you want to see the SQL sentences generated by Hibernate, and enable Hibernate statistics. Only meaningful for development and fine tuning.

Generating database script

If you want to see the table creation script, create, compile and run this piece of code:

```java
import java.io.File;

import org.hibernate.cfg.Configuration;
import org.hibernate.tool.hbm2ddl.SchemaExport;

public class Test {
  public static void main(String[] args) throws Exception {
    Configuration cfg = new Configuration();
    cfg.configure(new File("/path/to/hibernate.cfg.xml"));
    SchemaExport se = new SchemaExport(cfg);
    se.setOutputFile("/path/to/schema.sql");
    se.setFormat(true);
    se.create(false, false);
  }
}
```

> This configuration property should be set before the database creation. Once the database has been initialized don't modify it because can damage your installation. If your OpenKM installation has been already configured with another database (default one is an embedded one called HSQL) you can't switch to another database simply changing this property.

Here we will discuss the changes needed to run OpenKM with these databases:

- MySQL 🔒
- PostgreSQL 🔒
- MS SQLServer 🔒
- Oracle 🔒

See also:

- How To Generate DDL Scripts from Hibernate [2]

References

[1] http://hsqldb.org/

[2] http://jandrewthompson.blogspot.com/2009/10/how-to-generate-ddl-scripts-from.html

MySQL

Download MySQL JDBC driver from MySQL Home Page [1] and move it to *$JBOSS_HOME/server/default/lib*.

note	If you get an error like this: ``` Packet for query is too large (1708726 > 1048576). You can change this value on the server by setting the 'max_allowed_packet' variable. ``` You need to modify your MySQL server configuration file and increase the value of the **max_allowed_packet** property. Don't forget to restart the MySQL after any change in this file. In Debian based distros like Ubuntu, this configuration file is located at */etc/mysql/my.cnf*.

💡	InnoDB [2] seems to work better than MyISAM [3].

- OpenKM 5.0 🔒 (also valid for OpenKM 5.1)
- OpenKM 4.1 🔒

References

[1] http://www.mysql.com/products/connector/

[2] http://en.wikipedia.org/wiki/InnoDB

[3] http://en.wikipedia.org/wiki/MyISAM

PostgreSQL

Download PostgreSQL JDBC driver from PostgreSQL Driver Page [1] and move it to *$JBOSS_HOME/server/default/lib*.

> **OpenKM 5.1.8** has been verified to run with **PostgreSQL 9.1** using the JDBC driver **postgresql-9.1-901.jdbc4.jar** under JBoss 4.2.3.

- OpenKM 5.0 🔒 (also valid for OpenKM 5.1)
- OpenKM 4.1 🔒

If you want to obtain the maximun performance from PostgreSQL, pgFouine [2] will help in your task.

References

[1] http://jdbc.postgresql.org/

[2] http://pgfouine.projects.postgresql.org/index.html

PostgreSQL-OpenKM 5.0

> These instructions are also valid for OpenKM 5.1.

In this OpenKM release you can create the databases automatically configuring the **hibernate.dialect** and **hibernate.hbm2ddl** properties in OpenKM.cfg.

```
hibernate.dialect=org.hibernate.dialect.PostgreSQLDialect
hibernate.hbm2ddl=create
```

Once the tables are created, change the **hibernate.hbm2ddl** property from *create* to *none*.

> This configuration property should be set before the database creation. Once the database has been initialized don't modify it because can damage your installation. If your OpenKM installation has been already configured with another database (default one is an embedded one called HSQL) you can't switch to another database simply changing this property.

More info about this at:

- Hibernate: SQL Dialects [1]

Database creation

Starting with OpenKM 5.0, only two databases are needed:

```sql
DROP DATABASE IF EXISTS okm_repo;
DROP DATABASE IF EXISTS okm_app;

CREATE USER openkm WITH PASSWORD '*secret*';

CREATE DATABASE okm_repo WITH OWNER openkm ENCODING 'UTF8';
CREATE DATABASE okm_app WITH OWNER openkm ENCODING 'UTF8';
```

More info at:

- PostgreSQL 8.4.5 Documentation [2]

JBoss datasources

```
$ vim $JBOSS_HOME/server/default/deploy/openkm-ds.xml
```

```xml
<local-tx-datasource>
    <jndi-name>OpenKMDS</jndi-name>
    <connection-url>jdbc:postgresql://localhost:5432/okm_app</connection-url>
    <driver-class>org.postgresql.Driver</driver-class>
    <user-name>openkm</user-name>
    <password>*****</password>
    <min-pool-size>5</min-pool-size>
    <max-pool-size>20</max-pool-size>
    <idle-timeout-minutes>28680</idle-timeout-minutes>
    <metadata>
        <type-mapping>PostgreSQL</type-mapping>
        <!-- <type-mapping>PostgreSQL 8.0</type-mapping> -->
        <!-- <type-mapping>PostgreSQL 7.2</type-mapping> -->
    </metadata>
</local-tx-datasource>
```

The type mapping should match a type-mapping/name element from
$JBOSS_HOME/server/default/conf/standardjbosscmp-jdbc.xml. Example configurations for many third-party JDBC drivers are included in the *$JBOSS_HOME/docs/examples/jca* directory.

You may be interested in Encrypting DataSource Passwords [3].

Login configuration

```
$ vim $JBOSS_HOME/server/default/conf/login-config.xml
```

This is for 5.0 version:

```xml
<!-- OpenKM -->
<application-policy name = "OpenKM">
    <authentication>
        <login-module code="org.jboss.security.auth.spi.DatabaseServerLoginModule" flag = "required">
            <module-option name="dsJndiName">java:/OpenKMDS</module-option>
            <module-option name="principalsQuery">select usr_password as PASSWD from OKM_USER where usr_id=?
and usr_active=true</module-option>
            <module-option name="rolesQuery">select ur_role as ROLEID, 'Roles' from OKM_USER_ROLE where
ur_user=?</module-option>
            <module-option name="hashAlgorithm">md5</module-option>
            <module-option name="hashEncoding">hex</module-option>
        </login-module>
    </authentication>
</application-policy>
```

This is for 5.1 version:

```xml
<!-- OpenKM -->

<application-policy name = "OpenKM">

    <authentication>

        <login-module code="org.jboss.security.auth.spi.DatabaseServerLoginModule" flag = "required">

            <module-option name="dsJndiName">java:/OpenKMDS</module-option>

            <module-option name="principalsQuery">select usr_password as PASSWD from OKM_USER where usr_id=?

and usr_active='T'</module-option>

            <module-option name="rolesQuery">select ur_role as ROLEID, 'Roles' from OKM_USER_ROLE where

ur_user=?</module-option>

            <module-option name="hashAlgorithm">md5</module-option>

            <module-option name="hashEncoding">hex</module-option>

        </login-module>

    </authentication>

</application-policy>
```

Read also: Dynamic login-config.xml [4].

Repository configuration

```
$ vim $JBOSS_HOME/repository.xml
```

```xml
<?xml version="1.0"?>
<!DOCTYPE Repository PUBLIC "-//The Apache Software Foundation//DTD Jackrabbit 1.6//EN"


"http://jackrabbit.apache.org/dtd/repository-1.6.dtd">

<Repository>
    <!-- virtual file system where the repository stores global state
        (e.g. registered namespaces, custom node types, etc.) -->
    <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">
        <param name="path" value="${rep.home}/repository"/>
    </FileSystem>


    <!-- Security configuration -->
    <Security appName="OpenKM">
        <!-- Access manager: FQN of class implementing the AccessManager interface -->
        <AccessManager class="com.openkm.core.OKMAccessManager"/>
        <!-- <AccessManager class="org.apache.jackrabbit.core.security.SimpleAccessManager"/> -->
        <!-- <AccessManager class="org.apache.jackrabbit.core.security.DefaultAccessManager"> -->
            <!-- <param name="config" value="${rep.home}/access.xml"/> -->
        <!-- </AccessManager> -->
    </Security>


    <!-- Location of workspaces root directory and name of default workspace -->
    <Workspaces rootPath="${rep.home}/workspaces" defaultWorkspace="default"/>


    <!-- Workspace configuration template:
        used to create the initial workspace if there's no workspace
yet -->
    <Workspace name="${wsp.name}">
```

```xml
      <!-- Virtual file system of the workspace:
            class: FQN of class implementing the FileSystem interface
-->
      <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">
          <param name="path" value="${wsp.home}"/>
      </FileSystem>


      <!-- Persistence manager of the workspace:
            class: FQN of class implementing the PersistenceManager
interface -->
      <PersistenceManager class="org.apache.jackrabbit.core.persistence.bundle.PostgreSQLPersistenceManager">
        <param name="driver" value="org.postgresql.Driver"/>
        <param name="url" value="jdbc:postgresql://localhost:5432/okm_repo"/>
        <param name="schema" value="postgresql"/>
        <param name="user" value="openkm"/>
        <param name="password" value="*****"/>
        <param name="schemaObjectPrefix" value="${wsp.name}_"/>
        <param name="externalBLOBs" value="false"/>
      </PersistenceManager>


      <!-- Search index and the file system it uses.
            class: FQN of class implementing the QueryHandler
interface -->
      <SearchIndex class="org.apache.jackrabbit.core.query.lucene.SearchIndex">
          <param name="path" value="${wsp.home}/index"/>
          <param name="textFilterClasses" value="
          org.apache.jackrabbit.extractor.PlainTextExtractor,
          org.apache.jackrabbit.extractor.MsWordTextExtractor,
          org.apache.jackrabbit.extractor.MsExcelTextExtractor,
          org.apache.jackrabbit.extractor.MsPowerPointTextExtractor,
          org.apache.jackrabbit.extractor.OpenOfficeTextExtractor,
          org.apache.jackrabbit.extractor.RTFTextExtractor,
          org.apache.jackrabbit.extractor.HTMLTextExtractor,
          org.apache.jackrabbit.extractor.XMLTextExtractor,
          org.apache.jackrabbit.extractor.PngTextExtractor,
          org.apache.jackrabbit.extractor.MsOutlookTextExtractor,
          com.openkm.extractor.PdfTextExtractor,
          com.openkm.extractor.AudioTextExtractor,
          com.openkm.extractor.ExifTextExtractor,
          com.openkm.extractor.TiffTextExtractor,
          com.openkm.extractor.SourceCodeTextExtractor,
          com.openkm.extractor.MsOffice2007TextExtractor"/>
          <param name="extractorPoolSize" value="2"/>
          <param name="supportHighlighting" value="false"/>
          <param name="indexingConfiguration" value="${wsp.home}/../../../indexing_configuration.xml"/>
      </SearchIndex>
   </Workspace>
```

```xml
    <!-- Configures the versioning -->
    <Versioning rootPath="${rep.home}/version">
        <!-- Configures the filesystem to use for versioning for the respective
             persistence manager -->
        <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">
            <param name="path" value="${rep.home}/version" />
        </FileSystem>


        <!-- Configures the persistence manager to be used for persisting version state.
             Please note that the current versioning implementation is
based on
             a 'normal' persistence manager, but this could change in
future
             implementations. -->
        <PersistenceManager class="org.apache.jackrabbit.core.persistence.bundle.PostgreSQLPersistenceManager">
          <param name="driver" value="org.postgresql.Driver"/>
          <param name="url" value="jdbc:postgresql://localhost:5432/okm_repo"/>
          <param name="schema" value="postgresql"/>
          <param name="user" value="openkm"/>
          <param name="password" value="*****"/>
          <param name="schemaObjectPrefix" value="version_"/>
          <param name="externalBLOBs" value="false"/>
        </PersistenceManager>
    </Versioning>


    <!-- Search index for content that is shared repository wide
         (/jcr:system tree, contains mainly versions) -->
    <SearchIndex class="org.apache.jackrabbit.core.query.lucene.SearchIndex">
        <param name="path" value="${rep.home}/repository/index"/>
        <param name="textFilterClasses" value=""/>
        <param name="extractorPoolSize" value="2"/>
        <param name="supportHighlighting" value="false"/>
    </SearchIndex>


    <!-- DataStore improve file handling performance -->
    <DataStore class="org.apache.jackrabbit.core.data.FileDataStore">
        <param name="path" value="${rep.home}/repository/datastore"/>
        <param name="minRecordLength" value="100"/>
    </DataStore>
</Repository>
```

More info about this at Jackrabbit Configuration [5].

References

[1] http://docs.jboss.org/hibernate/core/3.3/reference/en/html/session-configuration.html#configuration-optional-dialects

[2] http://www.postgresql.org/docs/8.4/static/index.html

[3] http://community.jboss.org/wiki/encryptingdatasourcepasswords

[4] http://community.jboss.org/wiki/DynamicLoginConfig

[5] http://jackrabbit.apache.org/jackrabbit-configuration.html

PostgreSQL-OpenKM 4.1

Table creation

```
#
# Generic activity database definition
#
CREATE TABLE activity(act_date TIMESTAMP, act_user VARCHAR(32),
act_token VARCHAR(48), act_action VARCHAR(48), act_item VARCHAR(256),
act_params VARCHAR(256));
```

```
#
# Generic auth database definition
#
CREATE TABLE users(usr_id VARCHAR(32), usr_name VARCHAR(64), usr_pass
VARCHAR(32) NOT NULL, usr_email VARCHAR(32) NOT NULL, usr_active
BOOLEAN, PRIMARY KEY(usr_id));
CREATE TABLE roles(rol_id VARCHAR(32), PRIMARY KEY(rol_id));
CREATE TABLE user_role(ur_user VARCHAR(32), ur_role VARCHAR(32),
PRIMARY KEY(ur_user, ur_role));
CREATE TABLE mail_accounts(ma_id SERIAL, ma_user VARCHAR(32), ma_mhost
VARCHAR(32), ma_muser VARCHAR(32), ma_mpass VARCHAR(32), ma_mfolder
VARCHAR(32), ma_active BOOLEAN, PRIMARY KEY(ma_id));
CREATE TABLE twitter_accounts(ta_id SERIAL, ta_user VARCHAR(32),
ta_tuser VARCHAR(32), ta_active BOOLEAN, PRIMARY KEY(ta_id));

# INSERT DEFAULT USER / ROLES
INSERT INTO users (usr_id, usr_name, usr_pass, usr_email, usr_active)
VALUES ('okmAdmin', 'Administrator',
'21232f297a57a5a743894a0e4a801fc3', '', true);
INSERT INTO roles (rol_id) VALUES ('AdminRole');
INSERT INTO roles (rol_id) VALUES ('UserRole');
INSERT INTO user_role (ur_user, ur_role) VALUES ('okmAdmin',
'AdminRole');
```

```
#
# Generic dashboard stats database definition
#
CREATE TABLE dashboard_stats(ds_user VARCHAR(32), ds_source
VARCHAR(64), ds_node VARCHAR(256), ds_date TIMESTAMP);
```

Workflow engine

You have to modify the file *WEB-INF/classes/hibernate.cfg.xml* located inside the OpenKM.war archive. These are the entries to change:

```xml
<!-- hibernate dialect -->
<property name="hibernate.dialect">org.hibernate.dialect.HSQLDialect</property>
```

to

```xml
<!-- hibernate dialect -->
<property name="hibernate.dialect">org.hibernate.dialect.PostgreSQLDialect</property>
```

And in are the sententeces needed to create the tables. More info about this at:

- The jBPM Database [1]
- Hibernate: SQL Dialects [1]

JBoss datasources

```xml
<?xml version="1.0" encoding="UTF-8"?>
<datasources>
  <!-- OpenKM User Activity -->
  <local-tx-datasource>
    <jndi-name>OKMActivityDS</jndi-name>
    <connection-url>jdbc:postgresql://localhost:5432/openkm</connection-url>
    <driver-class>org.postgresql.Driver</driver-class>
    <user-name>db_user</user-name>
    <password>db_pass</password>
    <min-pool-size>5</min-pool-size>
    <max-pool-size>20</max-pool-size>
    <idle-timeout-minutes>0</idle-timeout-minutes>
    <track-statements/>
    <prepared-statement-cache-size>32</prepared-statement-cache-size>
    <metadata>
      <type-mapping>PostgreSQL</type-mapping>
    </metadata>
  </local-tx-datasource>

  <!-- OpenKM User Auth -->
  <local-tx-datasource>
    <jndi-name>OKMAuthDS</jndi-name>
    <connection-url>jdbc:postgresql://localhost:5432/openkm</connection-url>
    <driver-class>org.postgresql.Driver</driver-class>
    <user-name>db_user</user-name>
    <password>db_pass</password>
    <min-pool-size>5</min-pool-size>
```

```xml
      <max-pool-size>20</max-pool-size>
      <idle-timeout-minutes>0</idle-timeout-minutes>
      <track-statements/>
      <prepared-statement-cache-size>32</prepared-statement-cache-size>
      <metadata>
        <type-mapping>PostgreSQL</type-mapping>
      </metadata>
  </local-tx-datasource>

  <!-- OpenKM Dashboard Stats -->
  <local-tx-datasource>
    <jndi-name>OKMDashboardStatsDS</jndi-name>
    <connection-url>jdbc:postgresql://localhost:5432/openkm</connection-url>
    <driver-class>org.postgresql.Driver</driver-class>
    <user-name>db_user</user-name>
    <password>db_pass</password>
    <min-pool-size>5</min-pool-size>
    <max-pool-size>20</max-pool-size>
    <idle-timeout-minutes>0</idle-timeout-minutes>
    <track-statements/>
    <prepared-statement-cache-size>32</prepared-statement-cache-size>
    <metadata>
      <type-mapping>PostgreSQL</type-mapping>
    </metadata>
  </local-tx-datasource>

  <!-- OpenKM Workflow -->
  <local-tx-datasource>
    <jndi-name>OKMWorkflowDS</jndi-name>
    <connection-url>jdbc:postgresql://localhost:5432/openkm</connection-url>
    <driver-class>org.postgresql.Driver</driver-class>
    <user-name>db_user</user-name>
    <password>db_pass</password>
    <min-pool-size>5</min-pool-size>
    <max-pool-size>20</max-pool-size>
    <idle-timeout-minutes>0</idle-timeout-minutes>
    <track-statements/>
    <prepared-statement-cache-size>32</prepared-statement-cache-size>
    <metadata>
      <type-mapping>PostgreSQL</type-mapping>
    </metadata>
  </local-tx-datasource>
</datasources>
```

More info about this at Configuring JDBC DataSources [2].

Repository configuration

```xml
<?xml version="1.0" encoding="UTF-8"?>

<Repository>

  <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">

    <param name="path" value="${rep.home}/repository"/>

  </FileSystem>

  <Security appName="OpenKM">

    <AccessManager class="com.openkm.core.OKMAccessManager"/>

  </Security>

  <Workspaces rootPath="${rep.home}/workspaces" defaultWorkspace="default"/>

  <Workspace name="${wsp.name}">

    <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">

      <param name="path" value="${wsp.home}"/>

    </FileSystem>

    <PersistenceManager class="org.apache.jackrabbit.core.persistence.bundle.PostgreSQLPersistenceManager">

      <param name="driver" value="org.postgresql.Driver"/>

      <param name="url" value="jdbc:postgresql://localhost:5432/openkm?autoReconnect=true"/>

      <param name="schema" value="postgresql"/>

      <param name="user" value="db_user"/>

      <param name="password" value="db_pass"/>

      <param name="schemaObjectPrefix" value="${wsp.name}_"/>

      <param name="externalBLOBs" value="false"/>

    </PersistenceManager>

    <SearchIndex class="org.apache.jackrabbit.core.query.lucene.SearchIndex">

      <param name="path" value="${wsp.home}/index"/>

      <param name="useCompoundFile" value="true"/>

      <param name="minMergeDocs" value="100"/>

      <param name="volatileIdleTime" value="3"/>

      <param name="maxMergeDocs" value="100000"/>

      <param name="mergeFactor" value="10"/>

      <param name="bufferSize" value="10"/>

      <param name="cacheSize" value="1000"/>

      <param name="forceConsistencyCheck" value="false"/>

      <param name="autoRepair" value="true"/>

      <param name="analyzer" value="org.apache.lucene.analysis.standard.StandardAnalyzer"/>

      <param name="respectDocumentOrder" value="false"/>

      <param name="indexingConfiguration" value="$ {wsp.home}/../../../indexing_configuration.xml"/>

      <param name="textFilterClasses" value="
org.apache.jackrabbit.core.query.lucene.TextPlainTextFilter,
org.apache.jackrabbit.extractor.PdfTextExtractor,
org.apache.jackrabbit.extractor.HTMLTextExtractor,
org.apache.jackrabbit.extractor.XMLTextExtractor,
org.apache.jackrabbit.extractor.RTFTextExtractor,
org.apache.jackrabbit.extractor.OpenOfficeTextExtractor,
com.openkm.extractor.MsExcelTextExtractor,
com.openkm.extractor.MsPowerPointTextExtractor,
com.openkm.extractor.MsWordTextExtractor,
```

```
com.openkm.extractor.MsOffice2007TextExtractor,

com.openkm.extractor.ExifTextExtractor,

com.openkm.extractor.TiffTextExtractor,

com.openkm.extractor.AudioTextExtractor" />

    </SearchIndex>

  </Workspace>

  <Versioning rootPath="${rep.home}/version">

    <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">

      <param name="path" value="${rep.home}/version"/>

    </FileSystem>

    <PersistenceManager class="org.apache.jackrabbit.core.persistence.bundle.PostgreSQLPersistenceManager">

      <param name="driver" value="org.postgresql.Driver"/>

      <param name="url" value="jdbc:postgresql://localhost:5432/openkm?autoReconnect=true"/>

      <param name="schema" value="postgresql"/>

      <param name="user" value="db_user"/>

      <param name="password" value="db_pass"/>

      <param name="schemaObjectPrefix" value="version_"/>

      <param name="externalBLOBs" value="false"/>

    </PersistenceManager>

  </Versioning>

  <!-- Also see DatabaseDataStore-->

  <DataStore class="org.apache.jackrabbit.core.data.FileDataStore"/>

</Repository>
```

More info about this at Jackrabbit Configuration [5].

References

[1] http://docs.jboss.org/jbpm/v3/userguide/thejbpmdatabase.html

[2] http://www.jboss.org/file-access/default/members/jbossas/freezone/docs/Server_Configuration_Guide/4/html/
     Connectors_on_JBoss-Configuring_JDBC_DataSources.html

MS SQLServer

> This configuration has been tested with Microsoft SQL Server 2005 [1] and Microsoft SQL Server 2005 express edition [2].

Download MS SQLServer JDBC driver from MS SQLServer Home Page [3] and move it to *$JBOSS_HOME/server/default/lib*.

> Other JDBC driver for MS SQLServer we've tested is jdts [4].

In this OpenKM release you can create the databases automatically configuring the **hibernate.dialect** and **hibernate.hbm2ddl** properties in OpenKM.cfg.

```
hibernate.dialect=org.hibernate.dialect.SQLServerDialect
hibernate.hbm2ddl=create
```

Once the tables are created, change the **hibernate.hbm2ddl** property from *create* to *none*.

> This configuration property should be set before the database creation. Once the database has been initialized don't modify it because can damage your installation. If your OpenKM installation has been already configured with another database (default one is an embedded one called HSQL) you can't switch to another database simply changing this property.

More info about this at:

- Hibernate: SQL Dialects [5]

Database creation

Starting with OpenKM 5.0, only two databases are needed:

```
okm_repo
okm_app
```

Start SQL Server Management Studio Express [6] and create the two databases.
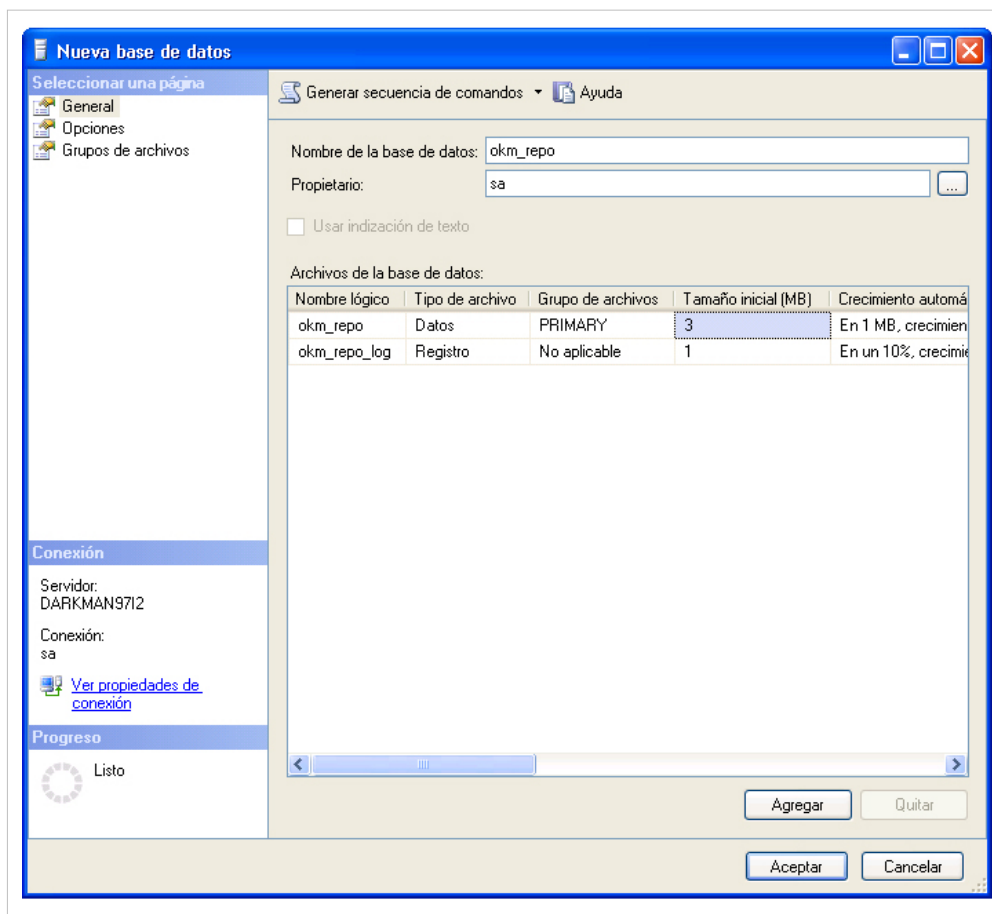
Select your appropiate database collate.
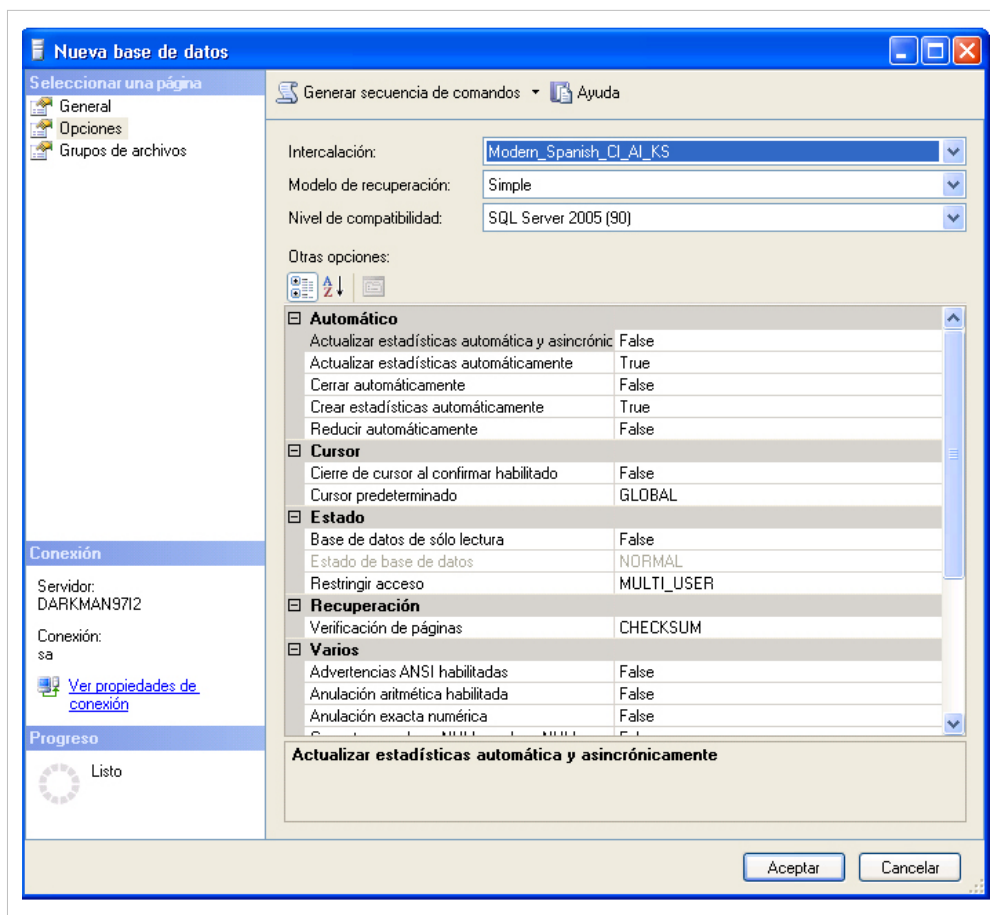
Troubleshooting, error you can not connect to MS SQL Server

Ensure you've got tcp protocol enabled (by default is disabled)

Solving hibernate schema creation problems

> Starting with OpenKM 5.1.6 the tables and initial database import has no problem, so you don't need all this work.

We've found some minimal problems in default hibernate creation database script in **okm_app** database. It'll be needed to connect to database and make minimal changes and execute some script.

Connect to **okm_app** database and alter **OKM_ACTIVITY** the column **ACT_ITEM** from type **text** to **varchar(MAX)**.

Connect to **okm_app** database and alter **OKM_EXTENSION** the column **EX_DESCRIPTION** from type **text** to **varchar(MAX)**.

Connect to **okm_app** database and alter **OKM_MIME_TYPE** the column **MT_IMAGE_CONTENT** from type **text** to **varchar(MAX)**.

Then open a sql console and execute the script to create OKM_PROFILE table and insert some fields that can not been executed by default in hibernate creation process.


```
create table OKM_PROFILE (
        PRF_ID int identity not null,
        PRF_NAME varchar(255) not null,
        PRF_ACTIVE tinyint not null,
        PRF_MSC_ADVANCED_FILTERS tinyint default 0 null,
        PRF_MSC_USER_QUOTA numeric(19,0) default 0 null,
        PRF_MSC_WEB_SKIN varchar(255) default 'default' null,
        PRF_MSC_PRINT_PREVIEW tinyint default 1 null,
```

```
        PRF_CHT_CHAT_EN tinyint default 0 null,
        PRF_CHT_AUTO_LOGIN_EN tinyint default 0 null,
        PRF_WZRD_KEYWORDS_EN tinyint default 0 null,
        PRF_WZRD_CATEGORIES_EN tinyint default 0 null,
        PRF_WZRD_PROPERTY_GROUPS  tinyint DEFAULT 0 NULL,
        PRF_WZRD_WORKFLOWS   tinyint DEFAULT 0 NULL,
        PRF_STCK_TAXONOMY_VIS tinyint default 1 null,
        PRF_STCK_CATEGORIES_VIS tinyint default 1 null,
        PRF_STCK_THESAURUS_VIS tinyint default 1 null,
        PRF_STCK_TEMPLATES_VIS tinyint default 1 null,
        PRF_STCK_PERSONAL_VIS tinyint default 1 null,
        PRF_STCK_MAIL_VIS tinyint default 1 null,
        PRF_STCK_TRASH_VIS tinyint default 1 null,
        PRF_TB_DESKTOP_VIS tinyint default 1 null,
        PRF_TB_SEARCH_VIS tinyint default 1 null,
        PRF_TB_DASHBOARD_VIS tinyint default 1 null,
        PRF_TB_ADMIN_VIS tinyint default 1 null,
        PRF_TB_DOC_PROPS_VIS tinyint default 1 null,
        PRF_TB_DOC_SECURITY_VIS tinyint default 1 null,
        PRF_TB_DOC_NOTES_VIS tinyint default 1 null,
        PRF_TB_DOC_VERSIONS_VIS tinyint default 1 null,
        PRF_TB_DOC_PREVIEW_VIS tinyint default 1 null,
        PRF_TB_DOC_PROP_GRPS_VIS tinyint default 1 null,
        PRF_TB_FLD_PROPS_VIS tinyint default 1 null,
        PRF_TB_FLD_SECURITY_VIS tinyint default 1 null,
        PRF_TB_FLD_NOTES_VIS tinyint default 1 null,
        PRF_TB_ML_PROPS_VIS tinyint default 1 null,
        PRF_TB_ML_SECURITY_VIS tinyint default 1 null,
        PRF_DB_USER_VIS tinyint default 1 null,
        PRF_DB_MAIL_VIS tinyint default 1 null,
        PRF_DB_NEWS_VIS tinyint default 1 null,
        PRF_DB_GENERAL_VIS tinyint default 1 null,
        PRF_DB_WORKFLOW_VIS tinyint default 1 null,
        PRF_DB_KEYWORDS_VIS tinyint default 1 null,
        PRF_MN_FILE_VIS tinyint default 1 null,
        PRF_MN_EDIT_VIS tinyint default 1 null,
        PRF_MN_TOOLS_VIS tinyint default 1 null,
        PRF_MN_BOOKMARKS_VIS tinyint default 1 null,
        PRF_MN_HELP_VIS tinyint default 1 null,
        PRF_MN_FI_CREATE_FLD_VIS tinyint default 1 null,
        PRF_MN_FI_FIND_FLD_VIS tinyint default 1 null,
        PRF_MN_FI_GO_FLD_VIS tinyint default 1 null,
        PRF_MN_FI_DOWNLOAD_VIS tinyint default 1 null,
        PRF_MN_FI_DOWNLOAD_PDF_VIS tinyint default 1 null,
        PRF_MN_FI_ADD_DOC_VIS tinyint default 1 null,
        PRF_MN_FI_START_WORKFLOW_VIS tinyint default 1 null,
        PRF_MN_FI_REFRESH_VIS tinyint default 1 null,
```

```
            PRF_MN_FI_SCANNER_VIS tinyint default 1 null,
            PRF_MN_FI_UPLOADER_VIS tinyint default 1 null,
            PRF_MN_FI_PURGE_VIS tinyint default 1 null,
            PRF_MN_FI_PURGE_TRASH_VIS tinyint default 1 null,
            PRF_MN_FI_RESTORE_VIS tinyint default 1 null,
            PRF_MN_FI_EXPORT_VIS tinyint default 1 null,
            PRF_MN_FI_CREATE_FROM_TPL_VIS tinyint default 1 null,
            PRF_MN_FI_SEND_DOC_LINK_VIS tinyint default 1 null,
            PRF_MN_FI_SEND_DOC_ATTACH_VIS tinyint default 1 null,
            PRF_MN_ED_LOCK_VIS tinyint default 1 null,
            PRF_MN_ED_UNLOCK_VIS tinyint default 1 null,
            PRF_MN_ED_CIN_VIS tinyint default 1 null,
            PRF_MN_ED_COUT_VIS tinyint default 1 null,
            PRF_MN_ED_CANCEL_COUT_VIS tinyint default 1 null,
            PRF_MN_ED_DELETE_VIS tinyint default 1 null,
            PRF_MN_ED_RENAME_VIS tinyint default 1 null,
            PRF_MN_ED_COPY_VIS tinyint default 1 null,
            PRF_MN_ED_MOVE_VIS tinyint default 1 null,
            PRF_MN_ED_ADD_PROP_GRP_VIS tinyint default 1 null,
            PRF_MN_ED_REM_PROP_GRP_VIS tinyint default 1 null,
            PRF_MN_ED_ADD_SUBS_VIS tinyint default 1 null,
            PRF_MN_ED_REM_SUBS_VIS tinyint default 1 null,
            PRF_MN_BM_MNG_BOOKMARKS_VIS tinyint default 1 null,
            PRF_MN_BM_ADD_BOOKMARK_VIS tinyint default 1 null,
            PRF_MN_BM_SET_HOME_VIS tinyint default 1 null,
            PRF_MN_BM_GO_HOME_VIS tinyint default 1 null,
            PRF_MN_TL_LANGS_VIS tinyint default 1 null,
            PRF_MN_TL_SKIN_VIS tinyint default 1 null,
            PRF_MN_TL_DEBUG_VIS tinyint default 1 null,
            PRF_MN_TL_ADMIN_VIS tinyint default 1 null,
            PRF_MN_TL_PREFS_VIS tinyint default 1 null,
            PRF_MN_HLP_DOC_VIS tinyint default 1 null,
            PRF_MN_HLP_BUG_TRACKING_VIS tinyint default 1 null,
            PRF_MN_HLP_SUPPORT_VIS tinyint default 1 null,
            PRF_MN_HLP_FORUM_VIS tinyint default 1 null,
            PRF_MN_HLP_CHANGELOG_VIS tinyint default 1 null,
            PRF_MN_HLP_WEB_SITE_VIS tinyint default 1 null,
            PRF_MNU_HLP_ABOUT_VIS tinyint default 1 null,
            primary key (PRF_ID)
);


alter table OKM_USER_CONFIG add constraint FK7798F4E88FDAFE34 foreign
key (UC_PROFILE) references OKM_PROFILE;


insert into OKM_USER (USR_ID, USR_NAME, USR_PASSWORD, USR_EMAIL,
USR_ACTIVE)
values ('okmAdmin', 'Administrator',
```

```
'21232f297a57a5a743894a0e4a801fc3', 'admin@noreply.com', 1);


insert into OKM_ROLE (ROL_ID, ROL_ACTIVE) values ('AdminRole', 1);
insert into OKM_ROLE (ROL_ID, ROL_ACTIVE) values ('UserRole', 1);


insert into OKM_USER_ROLE (UR_USER, UR_ROLE) values ('okmAdmin',
'AdminRole');


insert into OKM_PROFILE (PRF_NAME, PRF_ACTIVE) values ('Default', 1);
```

JBoss datasources

```xml
<local-tx-datasource>

    <jndi-name>OpenKMDS</jndi-name>

    <connection-url>jdbc:sqlserver://localhost:1433;databaseName=okm_app</connection-url>

    <driver-class>com.microsoft.sqlserver.jdbc.SQLServerDriver</driver-class>

    <user-name>sa</user-name>

    <password>****</password>

    <connection-property name="autoReconnect">true</connection-property>

    <min-pool-size>5</min-pool-size>

    <max-pool-size>100</max-pool-size>

    <blocking-timeout-millis>5000</blocking-timeout-millis>

    <idle-timeout-minutes>28680</idle-timeout-minutes>

    <metadata>

      <type-mapping>MS SQLSERVER2000</type-mapping>

    </metadata>

</local-tx-datasource>
```

The type mapping should match a type-mapping/name element from *$JBOSS_HOME/server/default/conf/standardjbosscmp-jdbc.xml*. Example configurations for many third-party JDBC drivers are included in the *$JBOSS_HOME/docs/examples/jca* directory.

You may be interested in Encrypting DataSource Passwords [3].

Login configuration

```xml
<!-- OpenKM -->

<application-policy name = "OpenKM">

    <authentication>

        <login-module code="org.jboss.security.auth.spi.DatabaseServerLoginModule" flag = "required">

            <module-option name="dsJndiName">java:/OpenKMDS</module-option>

            <module-option name="principalsQuery">select usr_password as PASSWD from OKM_USER where usr_id=?
and usr_active=1</module-option>

            <module-option name="rolesQuery">select ur_role as ROLEID, 'Roles' from OKM_USER_ROLE where
ur_user=?</module-option>

            <module-option name="hashAlgorithm">md5</module-option>

            <module-option name="hashEncoding">hex</module-option>

        </login-module>

    </authentication>
```

```
</application-policy>
```

Read also: Dynamic login-config.xml [4].

Repository configuration

```xml
<?xml version="1.0"?>
<!DOCTYPE Repository PUBLIC "-//The Apache Software Foundation//DTD Jackrabbit 1.6//EN"


"http://jackrabbit.apache.org/dtd/repository-1.6.dtd">

<Repository>
    <!-- virtual file system where the repository stores global state

        (e.g. registered namespaces, custom node types, etc.) -->

    <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">

        <param name="path" value="${rep.home}/repository"/>

    </FileSystem>



    <!-- Security configuration -->

    <Security appName="OpenKM">

        <!-- Security manager: FQN of class implementing the JackrabbitSecurityManager interface -->

        <!--<SecurityManager class="org.apache.jackrabbit.core.DefaultSecurityManager" workspaceName="security">-->

            <!-- workspace access: FQN of class implementing the WorkspaceAccessManager interface -->

            <!-- <WorkspaceAccessManager class="..."/> -->

            <!-- <param name="config" value="${rep.home}/security.xml"/> -->

        <!--</SecurityManager>-->



        <!-- Access manager: FQN of class implementing the AccessManager interface -->

        <AccessManager class="com.openkm.core.OKMAccessManager"/>

        <!-- <AccessManager class="org.apache.jackrabbit.core.security.SimpleAccessManager"/> -->

        <!-- <AccessManager class="org.apache.jackrabbit.core.security.DefaultAccessManager"> -->

            <!-- <param name="config" value="${rep.home}/access.xml"/> -->

        <!-- </AccessManager> -->



        <!-- <LoginModule class="org.apache.jackrabbit.core.security.simple.SimpleLoginModule"> -->

        <!-- <LoginModule class="org.apache.jackrabbit.core.security.authentication.DefaultLoginModule"> -->

            <!-- Anonymous user name ('anonymous' is the default value) -->

            <!-- <param name="anonymousId" value="anonymous"/> -->

            <!-- Administrator user id (default value if param is missing is 'admin') -->

            <!-- <param name="adminId" value="admin"/> -->

            <!-- <param name="principalProvider" value="..."/> -->

        <!--</LoginModule>-->

    </Security>



    <!-- Location of workspaces root directory and name of default workspace -->

    <Workspaces rootPath="${rep.home}/workspaces" defaultWorkspace="default"/>



    <!-- Workspace configuration template:

        used to create the initial workspace if there's no workspace
```

```
yet -->

    <Workspace name="${wsp.name}">

        <!-- Virtual file system of the workspace:

             class: FQN of class implementing the FileSystem interface

-->

        <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">

            <param name="path" value="${wsp.home}"/>

        </FileSystem>


        <!-- Persistence manager of the workspace:

             class: FQN of class implementing the PersistenceManager

interface -->

        <PersistenceManager class="org.apache.jackrabbit.core.persistence.bundle.MSSqlPersistenceManager">

          <param name="driver" value="com.microsoft.sqlserver.jdbc.SQLServerDriver"/>

          <param name="url" value="jdbc:sqlserver://darkman97i2:1433;databaseName=okm_repo;autoReconnect=true"/>

          <param name="schema" value="mssql"/>

          <param name="user" value="sa"/>

          <param name="password" value="****"/>

          <param name="schemaObjectPrefix" value="${wsp.name}_"/>

          <param name="externalBLOBs" value="false"/>

        </PersistenceManager>


        <!-- Search index and the file system it uses.

             class: FQN of class implementing the QueryHandler

interface -->

        <SearchIndex class="org.apache.jackrabbit.core.query.lucene.SearchIndex">

            <param name="path" value="${wsp.home}/index"/>

            <param name="textFilterClasses" value="

            org.apache.jackrabbit.extractor.PlainTextExtractor,

            org.apache.jackrabbit.extractor.MsWordTextExtractor,

            org.apache.jackrabbit.extractor.MsExcelTextExtractor,

            org.apache.jackrabbit.extractor.MsPowerPointTextExtractor,

            org.apache.jackrabbit.extractor.OpenOfficeTextExtractor,

            org.apache.jackrabbit.extractor.RTFTextExtractor,

            org.apache.jackrabbit.extractor.HTMLTextExtractor,

            org.apache.jackrabbit.extractor.XMLTextExtractor,

            org.apache.jackrabbit.extractor.PngTextExtractor,

            org.apache.jackrabbit.extractor.MsOutlookTextExtractor,

            com.openkm.extractor.PdfTextExtractor,

            com.openkm.extractor.AudioTextExtractor,

            com.openkm.extractor.ExifTextExtractor,

            com.openkm.extractor.TiffTextExtractor,

            com.openkm.extractor.SourceCodeTextExtractor,

            com.openkm.extractor.MsOffice2007TextExtractor"/>

            <param name="extractorPoolSize" value="2"/>

            <param name="supportHighlighting" value="false"/>

            <param name="indexingConfiguration" value="${wsp.home}/../../../indexing_configuration.xml"/>
```

```xml
        </SearchIndex>

    </Workspace>


    <!-- Configures the versioning -->

    <Versioning rootPath="${rep.home}/version">

        <!-- Configures the filesystem to use for versioning for the respective

            persistence manager -->

        <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">

            <param name="path" value="${rep.home}/version" />

        </FileSystem>


        <!-- Configures the persistence manager to be used for persisting version state.

            Please note that the current versioning implementation is

based on

            a 'normal' persistence manager, but this could change in

future

            implementations. -->

        <PersistenceManager class="org.apache.jackrabbit.core.persistence.bundle.MSSqlPersistenceManager">

          <param name="driver" value="com.microsoft.sqlserver.jdbc.SQLServerDriver"/>

          <param name="url" value="jdbc:sqlserver://darkman97i2:1433;databaseName=okm_repo;autoReconnect=true"/>

          <param name="schema" value="mssql"/>

          <param name="user" value="sa"/>

          <param name="password" value="****"/>

          <param name="schemaObjectPrefix" value="version_"/>

          <param name="externalBLOBs" value="false"/>

        </PersistenceManager>

    </Versioning>


    <!-- Search index for content that is shared repository wide

        (/jcr:system tree, contains mainly versions) -->

    <SearchIndex class="org.apache.jackrabbit.core.query.lucene.SearchIndex">

        <param name="path" value="${rep.home}/repository/index"/>

        <param name="textFilterClasses" value=""/>

        <param name="extractorPoolSize" value="2"/>

        <param name="supportHighlighting" value="false"/>

    </SearchIndex>


    <!-- DataStore improve file handling performance -->

    <DataStore class="org.apache.jackrabbit.core.data.FileDataStore">

        <param name="path" value="${rep.home}/repository/datastore"/>

        <param name="minRecordLength" value="100"/>

    </DataStore>

</Repository>
```

More info about this at Jackrabbit Configuration [5].

References

[1]   http://www.microsoft.com/sqlserver/2005/en/us/default.aspx

[2]   http://www.microsoft.com/sqlserver/2005/en/us/express.aspx

[3]   http://www.microsoft.com/spain/sql/technologies/jdbc/default.mspx

[4]   http://jtds.sourceforge.net/

[5]   http://docs.jboss.org/hibernate/core/3.5/reference/en/html/session-configuration.html#configuration-optional-dialects

[6]   http://www.microsoft.com/downloads/details.aspx?FamilyID=c243a5ae-4bd1-4e3d-94b8-5a0f62bf7796&DisplayLang=en

Oracle

In this OpenKM release you can create the databases automatically configuring the **hibernate.dialect** and **hibernate.hbm2ddl** properties in OpenKM.cfg. This configuration is optimized for OpenKM 5.1.5 and newer. For older releases you may nee to set the dialect to "org.hibernate.dialect.OracleDialect".

```
hibernate.dialect=org.hibernate.dialect.Oracle10gDialect
hibernate.hbm2ddl=none
```

> In others databases Hibernate can create the tables automagically, but due to some Oracle mapping issues you should create the tables yourself. See Solving schema table creation problems for more info. We will fix this behavior in the next major OpenKM release.

> This configuration property should be set before the database creation. Once the database has been initialized don't modify it because can damage your installation. If your OpenKM installation has been already configured with another database (default one is an embedded one called HSQL) you can't switch to another database simply changing this property.

More info about this at:

• Hibernate: SQL Dialects [5]

Database creation

Starting with OpenKM 5.0, only one database is needed because there is no table name collision.

The right sequence is:

• Edit OpenKM.cfg and set:

```
hibernate.dialect=org.hibernate.dialect.Oracle10gDialect
hibernate.hbm2ddl=none
```

• Edit and properly modify JBoss datasources.
• Edit and properly modify repository.xml (See Repository configuration)
• Execute the table creation script.
• Execute the inserts to create a minimal environment (see next point)
• Start JBoss.

The application will create some tables related to Jackrabbit persistence, so the Oracle user should have the right permissions. If not possible, this creation script need to be execute to create the Jackrabbit related tables:

```
DROP TABLE DEFAULT_BUNDLE cascade constraints;
DROP TABLE DEFAULT_REFS cascade constraints;
DROP TABLE DEFAULT_BINVAL cascade constraints;
DROP TABLE DEFAULT_NAMES cascade constraints;
DROP SEQUENCE DEFAULT_seq_names_id;
```

```
DROP TABLE VERSION_BUNDLE cascade constraints;
DROP TABLE VERSION_REFS cascade constraints;
DROP TABLE VERSION_BINVAL cascade constraints;
DROP TABLE VERSION_NAMES cascade constraints;
DROP SEQUENCE VERSION_seq_names_id;
CREATE TABLE DEFAULT_BUNDLE (NODE_ID raw(16) NOT NULL, BUNDLE_DATA blob
 NOT NULL);
CREATE UNIQUE INDEX DEFAULT_BUNDLE_IDX ON DEFAULT_BUNDLE (NODE_ID);
CREATE TABLE DEFAULT_REFS (NODE_ID raw(16) NOT NULL, REFS_DATA blob NOT
 NULL);
CREATE UNIQUE INDEX DEFAULT_REFS_IDX ON DEFAULT_REFS (NODE_ID);
CREATE TABLE DEFAULT_BINVAL (BINVAL_ID varchar2(64) NOT NULL,
BINVAL_DATA blob NULL);
CREATE UNIQUE INDEX DEFAULT_BINVAL_IDX ON DEFAULT_BINVAL (BINVAL_ID);
CREATE TABLE DEFAULT_NAMES (ID INTEGER PRIMARY KEY, NAME varchar2(255)
NOT NULL);
CREATE UNIQUE INDEX DEFAULT_NAMES_IDX ON DEFAULT_NAMES (NAME);
CREATE sequence DEFAULT_seq_names_id;
CREATE TRIGGER DEFAULT_t1 before INSERT ON DEFAULT_NAMES FOR each row
begin SELECT DEFAULT_seq_names_id.NEXTVAL INTO :new.id FROM dual; end;
/
CREATE TABLE VERSION_BUNDLE (NODE_ID raw(16) NOT NULL, BUNDLE_DATA blob
 NOT NULL);
CREATE UNIQUE INDEX VERSION_BUNDLE_IDX ON VERSION_BUNDLE (NODE_ID);
CREATE TABLE VERSION_REFS (NODE_ID raw(16) NOT NULL, REFS_DATA blob NOT
 NULL);
CREATE UNIQUE INDEX VERSION_REFS_IDX ON VERSION_REFS (NODE_ID);
CREATE TABLE VERSION_BINVAL (BINVAL_ID varchar2(64) NOT NULL,
BINVAL_DATA blob NULL);
CREATE UNIQUE INDEX VERSION_BINVAL_IDX ON VERSION_BINVAL (BINVAL_ID);
CREATE TABLE VERSION_NAMES (ID INTEGER PRIMARY KEY, NAME varchar2(255)
NOT NULL);
CREATE UNIQUE INDEX VERSION_NAMES_IDX ON VERSION_NAMES (NAME);
CREATE sequence VERSION_seq_names_id;
CREATE TRIGGER VERSION_t1 before INSERT ON VERSION_NAMES FOR each row
begin SELECT VERSION_seq_names_id.NEXTVAL INTO :new.id FROM dual; end;
/
```

In addition if you want the datastore stored in database:

```
create table DATASTORE (ID VARCHAR(255) PRIMARY KEY, LENGTH NUMBER,
LAST_MODIFIED NUMBER, DATA BLOB);
```

Solving hibernate schema creation problems

> Starting with OpenKM 5.1.6 the tables and initial database import has no problem, so you don't need all this work.

We've found some problems in default hibernate creation database script in okm_app database. It'll be needed to

connect to database and make minimal changes and execute this script at .

Also run these inserts to create a minimal environment:

```sql
insert into OKM_USER (USR_ID, USR_NAME, USR_PASSWORD, USR_EMAIL,
USR_ACTIVE) values ('okmAdmin', 'Administrator',
'21232f297a57a5a743894a0e4a801fc3', 'admin@noreply.com', 1);
insert into OKM_ROLE (ROL_ID, ROL_ACTIVE) values ('AdminRole', 1);
insert into OKM_ROLE (ROL_ID, ROL_ACTIVE) values ('UserRole', 1);
insert into OKM_USER_ROLE (UR_USER, UR_ROLE) values ('okmAdmin',
'AdminRole');
insert into OKM_PROFILE (PRF_ID, PRF_NAME, PRF_ACTIVE) values (1,
'Default', 1);
```

JBoss datasources

```xml
<local-tx-datasource>

    <jndi-name>OpenKMDS</jndi-name>

    <connection-url>jdbc:oracle:thin:@localhost:1521:sid</connection-url>

    <driver-class>oracle.jdbc.driver.OracleDriver</driver-class>

    <user-name>openkm</user-name>

    <password>*****</password>

    <min-pool-size>5</min-pool-size>

    <max-pool-size>20</max-pool-size>

    <idle-timeout-minutes>28680</idle-timeout-minutes>

    <exception-sorter-class-name>org.jboss.resource.adapter.jdbc.vendor.OracleExceptionSorter</exception-sorter-class-name>

    <valid-connection-checker-class-name>org.jboss.resource.adapter.jdbc.vendor.OracleValidConnectionChecker</valid-connection-checker-class-name>

    <metadata>

        <type-mapping>Oracle9i</type-mapping>

    </metadata>

</local-tx-datasource>
```

> This type-mapping applies both to Oracle 9i and Oracle 10g. Make sure that you have the latest Oracle JDBC driver [1]. OpenKM needs JDK6 to run, so you will need **ojdbc6.jar** which is for Oracle 11g but also support Oracle 10g and Oracle 9i. See also JDBC - Oracle FAQ [2]. Read more about datasource configuration at Configuring JDBC DataSources [3].

The type mapping should match a type-mapping/name element from *$JBOSS_HOME/server/default/conf/standardjbosscmp-jdbc.xml*. Example configurations for many third-party JDBC drivers are included in the *$JBOSS_HOME/docs/examples/jca* directory.

You may be interested in Encrypting DataSource Passwords [3].

Login configuration

Read also: Dynamic login-config.xml [4].

OpenKM 5.0

```xml
<!-- OpenKM -->
<application-policy name = "OpenKM">
    <authentication>
        <login-module code="org.jboss.security.auth.spi.DatabaseServerLoginModule" flag = "required">
            <module-option name="dsJndiName">java:/OpenKMDS</module-option>
            <module-option name="principalsQuery">select usr_password as PASSWD from OKM_USER where usr_id=?
and usr_active=1</module-option>
            <module-option name="rolesQuery">select ur_role as ROLEID, 'Roles' from OKM_USER_ROLE where
ur_user=?</module-option>
            <module-option name="hashAlgorithm">md5</module-option>
            <module-option name="hashEncoding">hex</module-option>
        </login-module>
    </authentication>
</application-policy>
```

OpenKM 5.1

```xml
<!-- OpenKM -->
<application-policy name = "OpenKM">
    <authentication>
        <login-module code="org.jboss.security.auth.spi.DatabaseServerLoginModule" flag = "required">
            <module-option name="dsJndiName">java:/OpenKMDS</module-option>
            <module-option name="principalsQuery">select usr_password as PASSWD from OKM_USER where usr_id=?
and usr_active='T'</module-option>
            <module-option name="rolesQuery">select ur_role as ROLEID, 'Roles' from OKM_USER_ROLE where
ur_user=?</module-option>
            <module-option name="hashAlgorithm">md5</module-option>
            <module-option name="hashEncoding">hex</module-option>
        </login-module>
    </authentication>
</application-policy>
```

Repository configuration

```xml
<?xml version="1.0"?>
<!DOCTYPE Repository PUBLIC "-//The Apache Software Foundation//DTD Jackrabbit 1.6//EN"

"http://jackrabbit.apache.org/dtd/repository-1.6.dtd">
<Repository>
    <!-- virtual file system where the repository stores global state
        (e.g. registered namespaces, custom node types, etc.) -->
    <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">
```

```xml
            <param name="path" value="${rep.home}/repository"/>
    </FileSystem>


    <!-- Security configuration -->
    <Security appName="OpenKM">
        <!-- Access manager: FQN of class implementing the AccessManager interface -->
        <AccessManager class="com.openkm.core.OKMAccessManager"/>
        <!-- <AccessManager class="org.apache.jackrabbit.core.security.SimpleAccessManager"/> -->
        <!-- <AccessManager class="org.apache.jackrabbit.core.security.DefaultAccessManager"> -->
            <!-- <param name="config" value="${rep.home}/access.xml"/> -->
        <!-- </AccessManager> -->
    </Security>


    <!-- Location of workspaces root directory and name of default workspace -->
    <Workspaces rootPath="${rep.home}/workspaces" defaultWorkspace="default"/>


    <!-- Workspace configuration template:
         used to create the initial workspace if there's no workspace
yet -->
    <Workspace name="${wsp.name}">
        <!-- Virtual file system of the workspace:
             class: FQN of class implementing the FileSystem interface
-->
        <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">
            <param name="path" value="${wsp.home}"/>
        </FileSystem>


        <!-- Persistence manager of the workspace:
             class: FQN of class implementing the PersistenceManager
interface -->
        <PersistenceManager class="org.apache.jackrabbit.core.persistence.bundle.OraclePersistenceManager">
          <param name="driver" value="oracle.jdbc.driver.OracleDriver"/>
          <param name="url" value="jdbc:oracle:thin:@localhost:1521:XE"/>
          <param name="schema" value="oracle"/>
          <param name="user" value="openkm"/>
          <param name="password" value="*****"/>
          <param name="schemaObjectPrefix" value="${wsp.name}_"/>
          <param name="externalBLOBs" value="false"/>
        </PersistenceManager>


        <!-- Search index and the file system it uses.
             class: FQN of class implementing the QueryHandler
interface -->
        <SearchIndex class="org.apache.jackrabbit.core.query.lucene.SearchIndex">
            <param name="path" value="${wsp.home}/index"/>
            <param name="textFilterClasses" value="
            org.apache.jackrabbit.extractor.PlainTextExtractor,
```
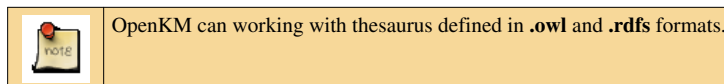
```xml
            org.apache.jackrabbit.extractor.MsWordTextExtractor,

            org.apache.jackrabbit.extractor.MsExcelTextExtractor,

            org.apache.jackrabbit.extractor.MsPowerPointTextExtractor,

            org.apache.jackrabbit.extractor.OpenOfficeTextExtractor,

            org.apache.jackrabbit.extractor.RTFTextExtractor,

            org.apache.jackrabbit.extractor.HTMLTextExtractor,

            org.apache.jackrabbit.extractor.XMLTextExtractor,

            org.apache.jackrabbit.extractor.PngTextExtractor,

            org.apache.jackrabbit.extractor.MsOutlookTextExtractor,

            com.openkm.extractor.PdfTextExtractor,

            com.openkm.extractor.AudioTextExtractor,

            com.openkm.extractor.ExifTextExtractor,

            com.openkm.extractor.TiffTextExtractor,

            com.openkm.extractor.SourceCodeTextExtractor,

            com.openkm.extractor.MsOffice2007TextExtractor"/>

            <param name="extractorPoolSize" value="2"/>

            <param name="supportHighlighting" value="false"/>

            <param name="indexingConfiguration" value="${wsp.home}/../../../indexing_configuration.xml"/>

        </SearchIndex>

    </Workspace>


    <!-- Configures the versioning -->

    <Versioning rootPath="${rep.home}/version">

        <!-- Configures the filesystem to use for versioning for the respective
             persistence manager -->

        <FileSystem class="org.apache.jackrabbit.core.fs.local.LocalFileSystem">

            <param name="path" value="${rep.home}/version" />

        </FileSystem>


        <!-- Configures the persistence manager to be used for persisting version state.
             Please note that the current versioning implementation is
based on
             a 'normal' persistence manager, but this could change in
future
             implementations. -->

        <PersistenceManager class="org.apache.jackrabbit.core.persistence.bundle.OraclePersistenceManager">

          <param name="driver" value="oracle.jdbc.driver.OracleDriver"/>

          <param name="url" value="jdbc:oracle:thin:@localhost:1521:XE"/>

          <param name="schema" value="oracle"/>

          <param name="user" value="openkm"/>

          <param name="password" value="*****"/>

          <param name="schemaObjectPrefix" value="version_"/>

          <param name="externalBLOBs" value="false"/>

        </PersistenceManager>

    </Versioning>


    <!-- Search index for content that is shared repository wide
```

```
            (/jcr:system tree, contains mainly versions) -->
    <SearchIndex class="org.apache.jackrabbit.core.query.lucene.SearchIndex">

        <param name="path" value="${rep.home}/repository/index"/>

        <param name="textFilterClasses" value=""/>

        <param name="extractorPoolSize" value="2"/>

        <param name="supportHighlighting" value="false"/>

    </SearchIndex>


    <!-- DataStore improve file handling performance -->
    <DataStore class="org.apache.jackrabbit.core.data.FileDataStore">

        <param name="path" value="${rep.home}/repository/datastore"/>

        <param name="minRecordLength" value="100"/>

    </DataStore>
</Repository>
```

If you want all your information stored in the database, you can replace the FileDataStore configuration by DbDatastore:

```
<!-- DataStore improve file handling performance -->
<DataStore class="org.apache.jackrabbit.core.data.db.DbDataStore">
    <param name="driver" value="oracle.jdbc.driver.OracleDriver"/>
    <param name="url" value="jdbc:oracle:thin:@localhost:1521:XE"/>
    <param name="databaseType" value="oracle"/>
    <param name="user" value="openkm"/>
    <param name="password" value="*****"/>
    <param name="minRecordLength" value="100"/>
</DataStore>
```

More info about this at Jackrabbit Configuration [5].

References

[1] http://www.oracle.com/technetwork/database/features/jdbc
[2] http://www.orafaq.com/wiki/JDBC
[3] http://docs.jboss.org/jbossas/docs/Server_Configuration_Guide/4/html/Connectors_on_JBoss-Configuring_JDBC_DataSources.html

Thesaurus

A **thesaurus** is a book that lists **words grouped together according to similarity of meaning** (containing synonyms and sometimes antonyms), in contrast to a dictionary, which contains definitions and pronunciations.

> Thesaurus is only available starting from OpenKM 5.0.
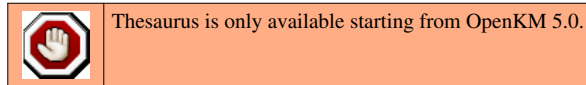
In Information Science, Library Science, and **Information Technology, specialized thesauri are designed for information retrieval**. They are a type of controlled vocabulary, for indexing or tagging purposes. If you want more information about the term you can take a look at http://en.wikipedia.org/wiki/Thesaurus.

> OpenKM can working with thesaurus defined in **.owl** and **.rdfs** formats.

There's a great utility http://protege.stanford.edu/ for creating and maintaining your own thesaurus. Specific thesauri can be found on the Internet. For example, Agrovoc [1] (FAO), NASA Thesuarus [2], UNESCO Thesaurus [3], Public administration Thesaurus [4] among others. Some of them are free and some have usage restrictions.

In order to correctly configure OpenKM to use a thesaurus the following OpenKM.cfg entries must be set:

```
kea.thesaurus.owl.file
kea.thesaurus.base.url
kea.thesaurus.tree.root
kea.thesaurus.tree.childs
```

Set OWL file

```
kea.thesaurus.owl.file=/vocabulary/file.owl
```

Set base url

```
kea.thesaurus.base.url=http://www.someweb.org
```

Query to get root node

```
kea.thesaurus.tree.root=SELECT DISTINCT UID, TEXT FROM {UID} Y
{OBJECT}, {UID} rdfs:label {TEXT} ; [rdfs:subClassOf {CLAZZ}] where not
bound(CLAZZ)
and lang(TEXT)="en" USING NAMESPACE foaf=<http://xmlns.com/foaf/0.1/>, dcterms=<http://purl.org/dc/terms/>,
rdf=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>, owl=<http://www.w3.org/2002/07/owl#>, rdfs=<http://www.w3.org/2000/01/rdf-schema#>,
skos=<http://www.w3.org/2004/02/skos/core#>, dc=<http://purl.org/dc/elements/1.1/>
```

Query to get childs node

```
kea.thesaurus.tree.childs=SELECT DISTINCT UID, TEXT FROM {UID}
rdfs:subClassOf {CLAZZ}, {UID} rdfs:label {TEXT} where
xsd:string(CLAZZ) =
"RDFparentID" and lang(TEXT)="en" USING NAMESPACE foaf=<http://xmlns.com/foaf/0.1/>, dcterms=<http://purl.org/dc/terms/>,
rdf=<http://www.w3.org /1999/02/22-rdf-syntax-ns#>, owl=<http://www.w3.org/2002/07/owl#>, rdfs=<http://www.w3.org/2000/01/rdf-schema#>,
skos=<http://www.w3.org/2004/02/skos/core#>, dc=<http://purl.org/dc/elements/1.1/>
```

You could be interested in:

- A Method for Converting Thesauri to RDF/OWL [5]

- OWL to SKOS online conversor [6]
- SKOS to OWL online conversor [7]
- SKOS Simple Knowledge Organization System [8]

Also check Thesaurus full example 🔒 for a step-by-step guide.

References

[1] http://aims.fao.org/website/AGROVOC-Thesaurus/sub

[2] http://www.sti.nasa.gov/thesfrm1.htm

[3] http://www2.ulcc.ac.uk/unesco/

[4] http://redined.r020.com.ar/es/index.php?tema=55

[5] http://thesauri.cs.vu.nl/

[6] http://owl.cs.manchester.ac.uk/owltoskos/

[7] http://www.ebusiness-unibw.org/tools/skos2owl/

[8] http://www.w3.org/2004/02/skos/

Thesaurus full example

OpenKM.cfg configuration

```
kea.thesaurus.owl.file=~/vocabulary/inroute.owl

kea.thesaurus.base.url=http://www.inroutenetwork.org

kea.thesaurus.tree.root=SELECT DISTINCT UID, TEXT FROM {UID} ?

{OBJECT}, {UID} rdfs:label {TEXT} ; [ rdfs:subClassOf {CLASS}] where not

bound(CLASS) and lang(TEXT)="es" USING NAMESPACE foaf=<http://xmlns.com/foaf/0.1/>, dcterms=<http://purl.org/dc/terms/>,

rdf=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>, owl=<http://www.w3.org/2002/07/owl#>, rdfs=<http://www.w3.org/2000/01/rdf-schema#>, skos=<http://www.w3.org/2004/02/skos/core#>, dc=<http://purl.org/dc/elements/1.1/>

kea.thesaurus.tree.childs=SELECT DISTINCT UID, TEXT FROM {UID}

rdfs:subClassOf {CLASS}, {UID} rdfs:label {TEXT} where

sesame:directSubClassOf(CLASS) = "%CMparentUUID" and lang(TEXT)="es" USING NAMESPACE

foaf=<http://xmlns.com/foaf/0.1/>, dcterms=<http://purl.org/dc/terms/>, rdf=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>, owl=<http://www.w3.org/2002/07/owl#>, rdfs=<http://www.w3.org/2000/01/rdf-schema#>, skos=<http://www.w3.org/2004/02/skos/core#>, dc=<http://purl.org/dc/elements/1.1/>
```

inroute.owl file

```xml
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:dcterms="http://purl.org/dc/terms/"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:skos="http://www.w3.org/2004/02/skos/core#"
    xmlns:dc="http://purl.org/dc/elements/1.1/">


<rdf:Description rdf:about="http://www.inroutenetwork.org">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Ontology"/>
    <rdfs:comment>INROUTE</rdfs:comment>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_1">
```

```
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
    <rdfs:label xml:lang="en">Visitors trips and activities</rdfs:label>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_2">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
    <rdfs:label xml:lang="en">Supply and employment</rdfs:label>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_3">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
    <rdfs:label xml:lang="en">Tourism and territory</rdfs:label>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_4">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
    <rdfs:label xml:lang="en">Tourism destination management</rdfs:label>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_5">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
    <rdfs:label xml:lang="en">Research and tools</rdfs:label>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_1_1">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
    <rdfs:label xml:lang="en">Visitor</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_1"/>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_1_2">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
    <rdfs:label xml:lang="en">Trip</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_1"/>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_1_3">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
    <rdfs:label xml:lang="en">Activities</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_1"/>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_1_4">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
    <rdfs:label xml:lang="en">Measurement and analysis tools (A.5)</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_1"/>
</rdf:Description>
```

```xml
<rdf:Description rdf:about="http://www.inroutenetwork.org#c_1_5">
     <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
     <rdfs:label xml:lang="en">Forecast modellinsg (A.6)</rdfs:label>
     <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_1"/>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_1_6">
     <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
     <rdfs:label xml:lang="en">Other accounting tools (A7)</rdfs:label>
     <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_1"/>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_1_7">
     <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
     <rdfs:label xml:lang="en">Oter terms</rdfs:label>
     <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_1"/>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_1_8">
     <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
     <rdfs:label xml:lang="en">Forms of toruism</rdfs:label>
     <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_1"/>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_2_1">
     <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
     <rdfs:label xml:lang="en">Establishments ( productive )</rdfs:label>
     <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_2"/>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_2_2">
     <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
     <rdfs:label xml:lang="en">Companies</rdfs:label>
     <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_2"/>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_2_3">
     <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
     <rdfs:label xml:lang="en">Tourism industries</rdfs:label>
     <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_2"/>
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_2_4">
     <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
     <rdfs:label xml:lang="en">Job</rdfs:label>
     <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_2"/>
```

```
</rdf:Description>


<rdf:Description rdf:about="http://www.inroutenetwork.org#c_2_5">
      <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
      <rdfs:label xml:lang="en">Other terms</rdfs:label>
      <rdfs:subClassOf rdf:resource="http://www.inroutenetwork.org#c_2"/>
</rdf:Description>


</rdf:RDF>
```

Additional steps

- Create a folder under your %JBOSS_HOME%/vocabulary
- Copy the file inroute.owl into vocabulary folder
- Login with some user with adminitrator grants
- Select administrator tab
- Select generate thesaurus option

- Select the show level and execute send

```
Founded 5 terms in level 0
2010/09/20 01:36:34 Creating term > [Research and tools] - with uid:http://www.inroutenetwork.org#c_5
Founded 0 terms in level 1
2010/09/20 01:36:34 Creating term > [Supply and employment] - with uid:http://www.inroutenetwork.org#c_2
Founded 5 terms in level 1
2010/09/20 01:36:34 Creating term -> [Companies] - with uid:http://www.inroutenetwork.org#c_2_2
Founded 0 terms in level 2
2010/09/20 01:36:34 Creating term -> [Establishments ( productive )] - with uid:http://www.inroutenetwork.org#c_2_1
Founded 0 terms in level 2
2010/09/20 01:36:35 Creating term -> [Job] - with uid:http://www.inroutenetwork.org#c_2_4
Founded 0 terms in level 2
2010/09/20 01:36:35 Creating term -> [Other terms] - with uid:http://www.inroutenetwork.org#c_2_5
Founded 0 terms in level 2
2010/09/20 01:36:35 Creating term -> [Tourism industries] - with uid:http://www.inroutenetwork.org#c_2_3
Founded 0 terms in level 2
2010/09/20 01:36:35 Creating term > [Tourism and territory] - with uid:http://www.inroutenetwork.org#c_3
Founded 0 terms in level 1
2010/09/20 01:36:35 Creating term > [Tourism destination management] - with uid:http://www.inroutenetwork.org#c_4
Founded 0 terms in level 1
2010/09/20 01:36:35 Creating term > [Visitors trips and activities] - with uid:http://www.inroutenetwork.org#c_1
Founded 8 terms in level 1
2010/09/20 01:36:35 Creating term -> [Activities] - with uid:http://www.inroutenetwork.org#c_1_3
Founded 0 terms in level 2
2010/09/20 01:36:35 Creating term -> [Forecast modellinsg (A.6)] - with uid:http://www.inroutenetwork.org#c_1_5
Founded 0 terms in level 2
2010/09/20 01:36:35 Creating term -> [Forms of toruism] - with uid:http://www.inroutenetwork.org#c_1_8
Founded 0 terms in level 2
2010/09/20 01:36:35 Creating term -> [Measurement and analysis tools (A.5)] - with uid:http://www.inroutenetwork.org#c_1_4
Founded 0 terms in level 2
2010/09/20 01:36:35 Creating term -> [Oter terms] - with uid:http://www.inroutenetwork.org#c_1_7
Founded 0 terms in level 2
2010/09/20 01:36:35 Creating term -> [Other accounting tools (A7)] - with uid:http://www.inroutenetwork.org#c_1_6
Founded 0 terms in level 2
2010/09/20 01:36:35 Creating term -> [Trip] - with uid:http://www.inroutenetwork.org#c_1_2
Founded 0 terms in level 2
2010/09/20 01:36:35 Creating term -> [Visitor] - with uid:http://www.inroutenetwork.org#c_1_1
Founded 0 terms in level 2
```

Level '2'

Now in thesaurus view you'll see
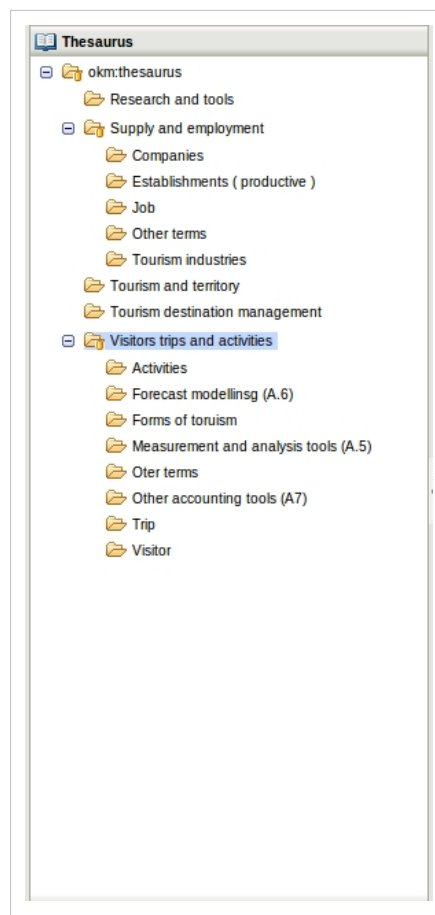
Automatic key extraction

> Automatic key extraction is only available from OpenKM 5.0 and upper.

OpenKM uses **KEA** for extracting keyphrases from text documents. **KEA** it by default can be either used for free indexing or for indexing with a controlled vocabulary, but with OpenKM is **mandatory having a controled vocabulary**. OpenKM automatic extractrion keyphrases is based in KEA 5.0.

If order having KEA running in OpenKM must be a well done configured vocabulary (Thesaurus).

KEA is a training module that uses a Thesaurus as the controller vocabulary. In order how to configure OpenKM Thesaurus take a look at Thesaurus in installation guide.

To creating KEA model must checkout openkm and thesaurus modules:

Select the svn type and type the url https://openkm.svn.sourceforge.net/svnroot/openkm/trunk/openkm to refer openkm:

Select the svn type and type the url https://openkm.svn.sourceforge.net/svnroot/openkm/trunk/thesausus to refer thesaurus:

In KEA web page could downloading file that comes with some example how to creating KEA model. In similar way using class ModelBuilder in thesaurus modules must be created the KEA model based in some vocabulary controller (Thesaurus).

For training KEA module is needed execute ModelBuilder class with that params:

```
sourceFolder
trainingFolder
vocabularyFile
vocabularyType
stopwordFile
modelFileName
porterStemmerClass
stopwordClass
language
documentEncoding
```

In order to correctly configure OpenKM thesaurus you must set this OpenKM.cfg entries:

```
kea.thesaurus.skos.file
kea.thesaurus.vocabulary.serql
kea.model.file
kea.stopwords.file
kea.automatic.keyword.extraction.number
kea.automatic.keyword.extraction.restriction
```

Setting the SKOS file

```
kea.thesaurus.skos.file=file.rdf
```

Setting vocabulary query

kea.thesaurus.vocabulary.serql=SELECT X,UID FROM {X} skos:prefLabel {UID} WHERE lang(UID) ="en" USING NAMESPACE rdf=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>, skos=<http://www.w3.org/2004/02/skos/core#>,rdfs=<http://www.w3.org/2000/01/rdf-schema#>, dc=<http://purl.org/dc/elements/1.1/>, dcterms=<http://purl.org/dc/terms/>, foaf=<http://xmlns.com/foaf/0.1/>

Setting model file

```
kea.model.file=file.model
```

Setting stop words

```
kea.stopwords.file=stopwords.txt
```

Setting max keywords extraction

```
kea.automatic.keyword.extraction.number=10
```

Setting dictionary restriction

Only dictionary words are allowed

```
kea.automatic.keyword.extraction.restriction=on
```

Automatic key extraction full example 🔒

You could be interested in:

- KEA [1]
- WEKA - Data mining with Open Source machine learning in Java [2]
- Aperture framework [3]
- RDF2GO [4]
- OpenRDF [5]

References

[1] http://www.nzdl.org/Kea/index.html

[2] http://www.cs.waikato.ac.nz/~ml/weka/

[3] http://aperture.sourceforge.net/

[4] http://mavenrepo.fzi.de/semweb4j.org/site/rdf2go/

[5] http://www.openrdf.org/

Automatic key extraction full example

SVN checkout modules

To creating KEA model must checkout openkm and thesaurus modules:

Select the svn type and type the url https://openkm.svn.sourceforge.net/svnroot/openkm/trunk/openkm to refer openkm:

Select the svn type and type the url https://openkm.svn.sourceforge.net/svnroot/openkm/trunk/thesaurus to refer thesaurus:

Installing openkm classes into maven repository

Ensure you've intalled openkm into your local maven repository, to ensure it you can execute the command:

```
mvn clean package install –Dmaven.test.skip=true
```

Donwloading AGROVOC thesaurus

We'll use agrovoc for testing purposes, you can downloading from http://oaei.ontologymatching.org/2007/environment/please read terms of use.

Copy into **thesaurus/src/test/resources/vocabulary** folder the file **ag_skos_20070219.rdf** and agrovoc_oaei2007.owl

Into **vocabulary** folder there's **testdocs** folders are some agrovoc training docs to creating KEA module.

Create runtime configuration

Now we can create runtime configuration, it must be executed the ModelBuilder class with some params
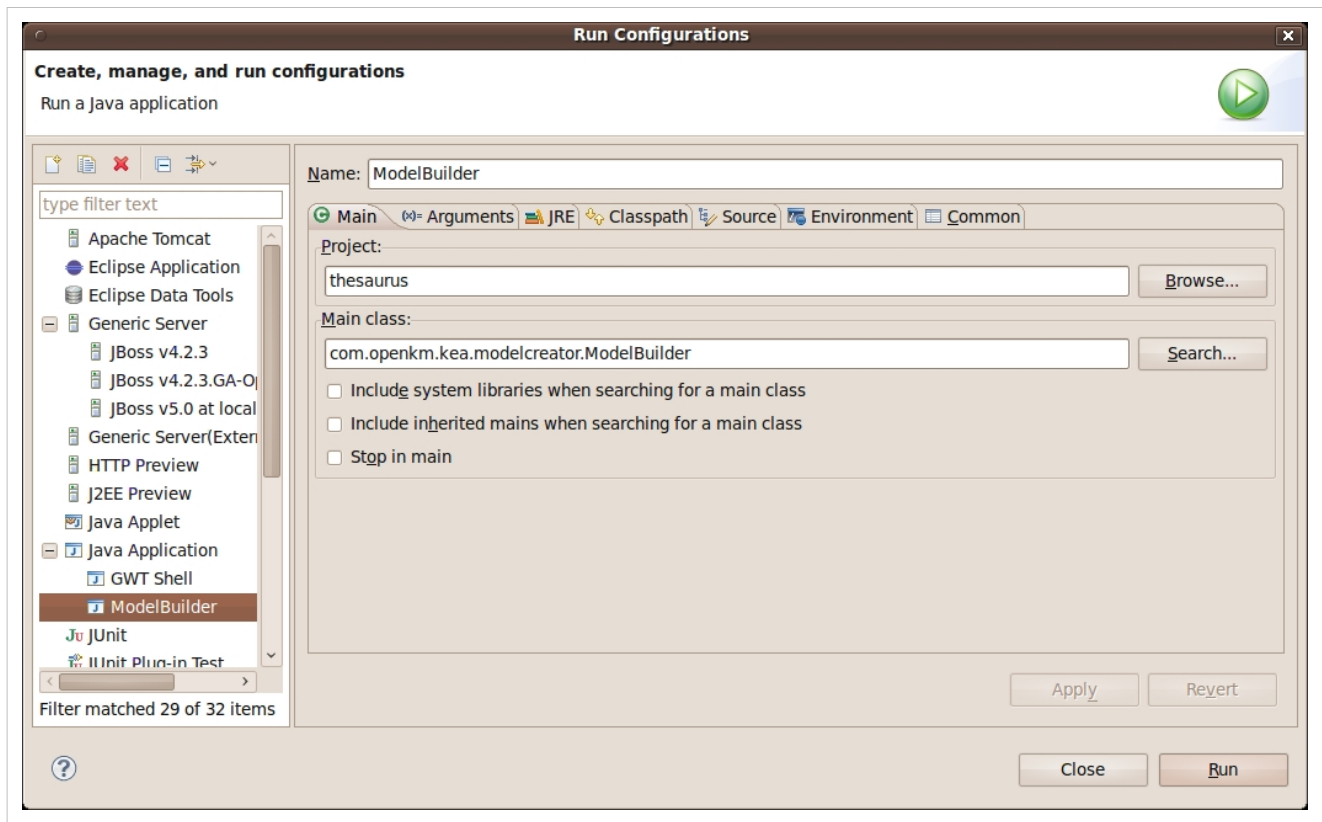
For training KEA module is needed execute ModelBuilder class with that params:

```
sourceFolder
trainingFolder
vocabularyFile
vocabularyType
stopwordFile
modelFileName
porterStemmerClass
stopwordClass
language
documentEncoding
testDocs ( optional )
```

In my case

sourceFolder=/home/jllort/softwareFactoryGalileo/thesaurus/vocabulary (all path are relative to sourceFolder)

trainingFolder=testdocs/en/train

vocabularyFile=ag_skos_20070219.rdf

vocabularyType=skos

stopwordFile=stopwords_en.txt

modelFileName=ag_skos_20070219.model

porterStemmerClass=com.openkm.kea.stemmers.PorterStemmer

stopwordClass=com.openkm.kea.stopwords.StopwordsEnglish

language=en

documentEncoding=UTF-8

testDocs=testdocs/en/test

The params to execute ModelBuilder class are "**/home/jllort/softwareFactoryGalileo/thesaurus/vocabulary testdocs/en/train ag_skos_20070219.rdf skos stopwords_en.txt ag_skos_20070219.model com.openkm.kea.stemmers.PorterStemmer com.openkm.kea.stopwords.StopwordsEnglish en UTF-8 testdocs/en/test**" and VM argument "**-Xmx526M**" as you can see in next screenshot
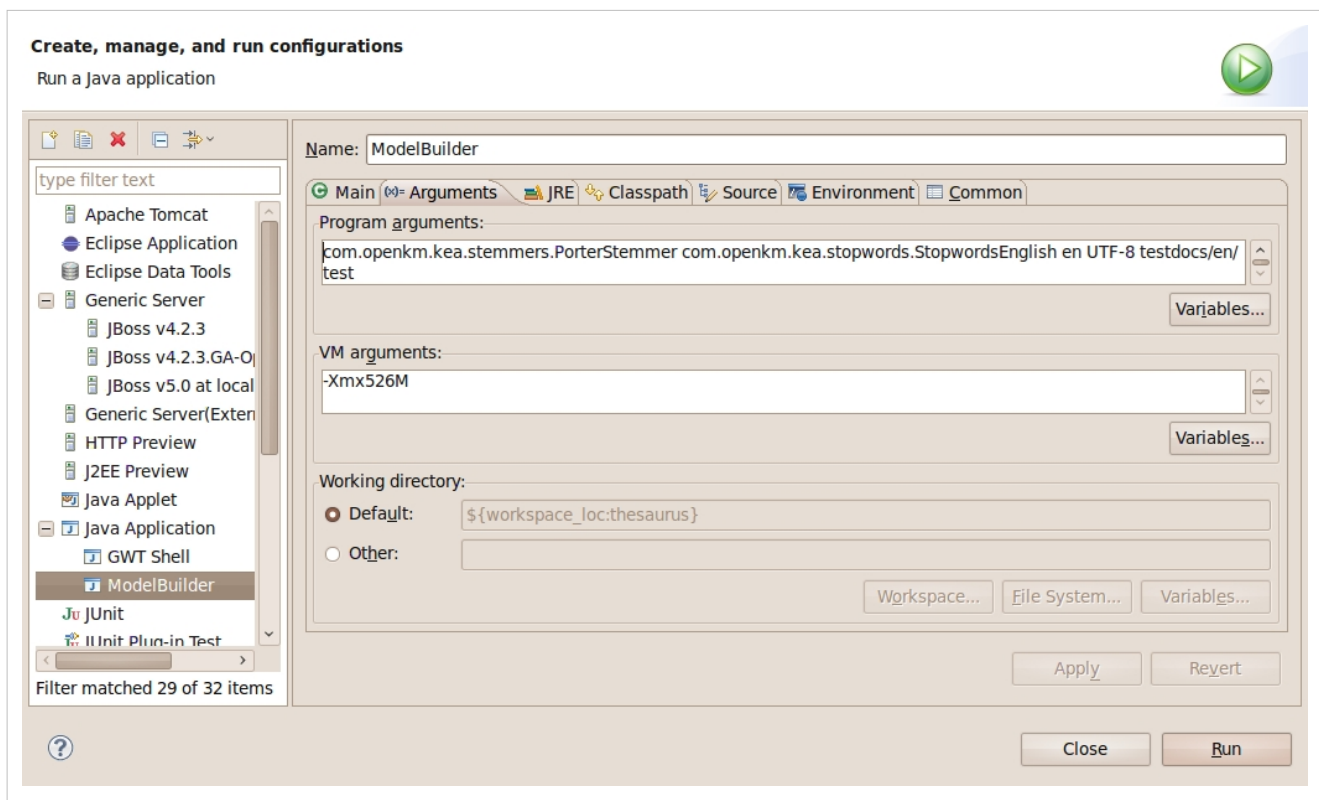
Classpath must be shown as
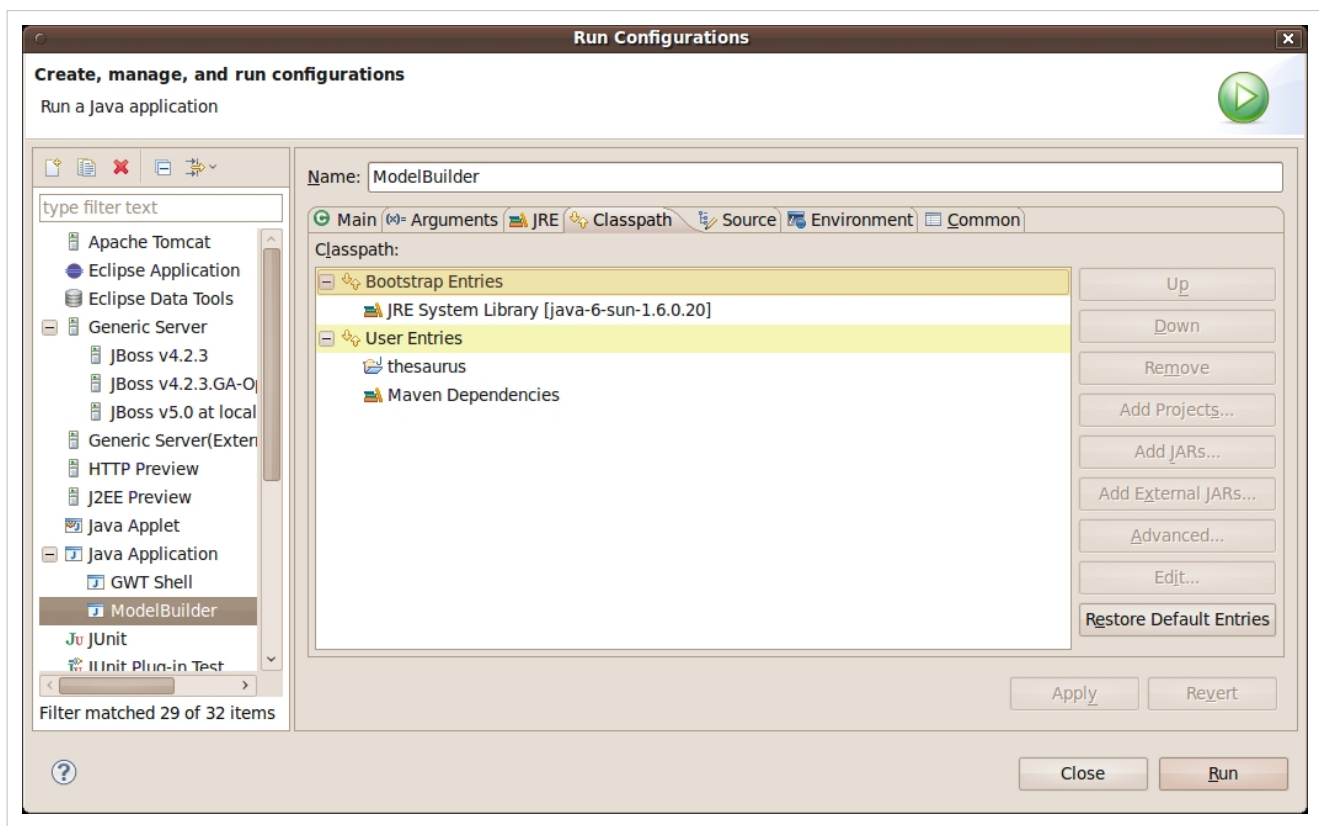
It all goes fine it has been generated into vocabulary folder a file called agrovoc_oaei2007.model

```
INFO   com.openkm.core.Config - Using default dir: /home/jllort/softwareFactoryGalileo/thesaurus/src/test
INFO   com.openkm.kea.modelcreator.ModelBuilder - * Source folder: /home/jllort/softwareFactoryGalileo/th
INFO   com.openkm.kea.modelcreator.ModelBuilder - * Training folder: /home/jllort/softwareFactoryGalileo/
INFO   com.openkm.kea.modelcreator.ModelBuilder - * Vocabulary file: /home/jllort/softwareFactoryGalileo/
INFO   com.openkm.kea.modelcreator.ModelBuilder - * Vocabulary type: skos
INFO   com.openkm.kea.modelcreator.ModelBuilder - * Stopwords file: stopwords_en.txt
INFO   com.openkm.kea.modelcreator.ModelBuilder - * Model file: /home/jllort/softwareFactoryGalileo/thesa
INFO   com.openkm.kea.modelcreator.ModelBuilder - * Stemmer class: com.openkm.kea.stemmers.PorterStemmer
INFO   com.openkm.kea.modelcreator.ModelBuilder - * Stopword class: com.openkm.kea.stopwords.StopwordsEng
INFO   com.openkm.kea.modelcreator.ModelBuilder - * Language: en
INFO   com.openkm.kea.modelcreator.ModelBuilder - * Document encoding: UTF-8
INFO   com.openkm.kea.modelcreator.ModelBuilder - * Testing folder: /home/jllort/softwareFactoryGalileo/t
INFO   com.openkm.kea.modelcreator.ModelBuilder - Creating the model...
WARN   com.hp.hpl.jena.rdf.model.impl.RDFDefaultErrorHandler - file:///home/jllort/softwareFactoryGalile
WARN   com.hp.hpl.jena.rdf.model.impl.RDFDefaultErrorHandler - file:///home/jllort/softwareFactoryGalile
INFO   com.openkm.kea.modelcreator.KEAModelBuilder - -- Reading the Documents...
INFO   com.openkm.kea.modelcreator.ModelBuilder - Extracting keyphrases from test documents...
WARN   com.hp.hpl.jena.rdf.model.impl.RDFDefaultErrorHandler - file:///home/jllort/softwareFactoryGalile
WARN   com.hp.hpl.jena.rdf.model.impl.RDFDefaultErrorHandler - file:///home/jllort/softwareFactoryGalile
INFO   com.openkm.kea.modelcreator.KEAKeyphraseExtractor - -- Extracting Keyphrases...
INFO   com.openkm.kea.modelcreator.KEAKeyphraseExtractor - Avg. number of matching keyphrases compared tc
INFO   com.openkm.kea.modelcreator.KEAKeyphraseExtractor - Based on 5 documents
INFO   com.openkm.kea.modelcreator.ModelBuilder - Look into /home/jllort/softwareFactoryGalileo/thesaurus
INFO   com.openkm.kea.modelcreator.ModelBuilder - and compare them to /home/jllort/softwareFactoryGalilec
```

Copying vocabulary files into OpenKM

Create a folder called vocabulary int %JBOSS_HOME%, copy into files called **ag_skos_20070219.rdf**, **agrovoc_oaei2007.owl**, **ag_skos_20070219.model**, and **stopwords_en.txt**

Configuring OpenKM.cfg

Thesaurus configuration values

```
kea.thesaurus.owl.file=/vocabulary/agrovoc_oaei2007.owl
kea.thesaurus.base.url=http://www.fao.org/aos/agrovoc
kea.thesaurus.tree.root=SELECT DISTINCT UID, TEXT FROM {UID} Y
{OBJECT}, {UID} rdfs:label {TEXT} ; [rdfs:subClassOf {CLAZZ}] where not
 bound(CLAZZ) and lang(TEXT)="en" USING NAMESPACE
foaf=<http://xmlns.com/foaf/0.1/>,
dcterms=<http://purl.org/dc/terms/>,
rdf=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>,
owl=<http://www.w3.org/2002/07/owl#>,
rdfs=<http://www.w3.org/2000/01/rdf-schema#>,
skos=<http://www.w3.org/2004/02/skos/core#>,
dc=<http://purl.org/dc/elements/1.1/>
kea.thesaurus.tree.childs=SELECT DISTINCT UID, TEXT FROM {UID}
rdfs:subClassOf {CLAZZ}, {UID} rdfs:label {TEXT} where
xsd:string(CLAZZ) = "RDFparentID" and lang(TEXT)="en" USING NAMESPACE
```

```
foaf=<http://xmlns.com/foaf/0.1/>,
dcterms=<http://purl.org/dc/terms/>,
rdf=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>,
owl=<http://www.w3.org/2002/07/owl#>,
rdfs=<http://www.w3.org/2000/01/rdf-schema#>,
skos=<http://www.w3.org/2004/02/skos/core#>,
dc=<http://purl.org/dc/elements/1.1/>
```

KEA model configuration values

```
kea.thesaurus.skos.file=/vocabulary/ag_skos_20070219.rdf
kea.thesaurus.vocabulary.serql=SELECT X,UID FROM {X} skos:prefLabel
{UID} WHERE lang(UID) ="en" USING NAMESPACE
rdf=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>,
skos=<http://www.w3.org/2004/02/skos/core#>,rdfs=<http://www.w3.org/2000/01/rdf-schema#>,
 dcterms=<http://purl.org/dc/terms/>,
foaf=<http://xmlns.com/foaf/0.1/>
kea.model.file=/vocabulary/ag_skos_20070219.model
kea.stopwords.file=/vocabulary/stopwords_en.txt
kea.automatic.keyword.extraction.number=10
kea.automatic.keyword.extraction.restriction=on
```

kea.automatic.keyword.extraction.restriction is an optional paramater to indicate that only words in thesaurus are enabled to be extracted.

Creating thesaurus

Login into OpenKM with some user with administrator grants, go to Administration tab and select Generate Thesaurus option. Then select the "show level" and execute the "send" option.

Please be patient it's needed some time to building all thesaurus. Depending your hardware configuration (RAM) could take some hours before process it'll be finishing.

```
Founded 1954 terms in level 0
2010/09/20 06:12:57 Creating term > [ability] - with uid:http://www.fao.org/aos/agrovoc#c_49830
Founded 1 terms in level 1
2010/09/20 06:13:02 Creating term -> [interoperability] - with uid:http://www.fao.org/aos/agrovoc#c_49831
Founded 0 terms in level 2
2010/09/20 06:13:03 Creating term > [Abiotic injuries] - with uid:http://www.fao.org/aos/agrovoc#c_26768
Founded 16 terms in level 1
2010/09/20 06:13:03 Creating term -> [Browsing damage] - with uid:http://www.fao.org/aos/agrovoc#c_1119
Founded 0 terms in level 2
2010/09/20 06:13:03 Creating term -> [Bruises of products] - with uid:http://www.fao.org/aos/agrovoc#c_16138
Founded 1 terms in level 2
2010/09/20 06:13:03 Creating term --> [Trampling] - with uid:http://www.fao.org/aos/agrovoc#c_24093
2010/09/20 06:13:07 Creating term -> [Cold injury] - with uid:http://www.fao.org/aos/agrovoc#c_3114
Founded 0 terms in level 2
2010/09/20 06:13:07 Creating term -> [Fouling] - with uid:http://www.fao.org/aos/agrovoc#c_25400
Founded 0 terms in level 2
2010/09/20 06:13:07 Creating term -> [Freeze damage (environment)] - with uid:http://www.fao.org/aos/agrovoc#c_3114
Founded 0 terms in level 2
2010/09/20 06:13:08 Creating term -> [Frost damage] - with uid:http://www.fao.org/aos/agrovoc#c_3114
Founded 0 terms in level 2
2010/09/20 06:13:08 Creating term -> [Frost injury] - with uid:http://www.fao.org/aos/agrovoc#c_3114
Founded 0 terms in level 2
2010/09/20 06:13:11 Creating term -> [Frostbite] - with uid:http://www.fao.org/aos/agrovoc#c_3114
Founded 0 terms in level 2
2010/09/20 06:13:12 Creating term -> [Ice damage] - with uid:http://www.fao.org/aos/agrovoc#c_3114
Founded 0 terms in level 2
```

After finishing Thesarus creation in your desktop could see the thesaurus folders representation as is shown:
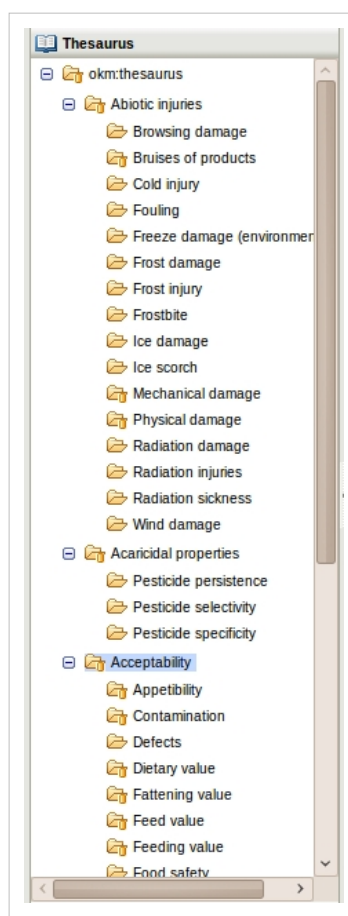
Automatic key extraction in new uploaded document

Upload a new document, for example some document from testdocs/en/train

In your jboss console it'll appears something like this:

```
18:43:34,098 INFO  [SubjectExtractor] Subject extraction completed in 880ms
18:43:34,099 INFO  [DirectDocumentModule] Creator:
18:43:34,099 INFO  [DirectDocumentModule] Title: Calliandra calothyrsus - Production and use: A Field Manual
18:43:34,099 INFO  [DirectDocumentModule] Mime type: text/html
18:43:34,099 INFO  [DirectDocumentModule] Filename: null
18:43:34,099 INFO  [DirectDocumentModule] Content created: null
18:43:34,099 INFO  [DirectDocumentModule] Content last modified: null
18:43:34,099 INFO  [DirectDocumentModule] Term:Seed
18:43:34,102 INFO  [DirectDocumentModule] Term:Calliandra
18:43:34,103 INFO  [DirectDocumentModule] Term:Production
18:43:34,105 INFO  [DirectDocumentModule] Term:Forests
18:43:34,106 INFO  [DirectDocumentModule] Term:Calliandra calothyrsus
18:43:34,106 INFO  [DirectDocumentModule] Term:Planting
18:43:34,108 INFO  [DirectDocumentModule] Term:Water
18:43:34,111 INFO  [DirectDocumentModule] Term:Self management
18:43:34,116 INFO  [DirectDocumentModule] Term:Trees
18:43:34,119 INFO  [DirectDocumentModule] Term:Arkansas
```

And in your OpenKM UI the extracted keywords as shown:

Propiedades	Notas	Seguridad	Historial	Previsualizar

UUID 152084ce-f5d8-4576-b545-ab0892ad677c

Nombre winrock_wi10ce.txt

Carpeta /okm:root

Tamaño 104.1 KB

Creado 20/09/2010 18:43:34 por okmAdmin

Modificado 20/09/2010 18:43:35 por okmAdmin

Tipo MIME text/plain

Palabras clave Escribe la palabra

Calliandra   Self_management
Trees   Production
Planting   Seed   Forests
Calliandra_calothyrsus   Water
Arkansas

Usuarios suscritos

Nube de palabras clave
Calliandra Self_management Trees Production Planting Seed
Forests Calliandra_calothyrsus Water Arkansas

Categorías

Creating automatic key extraction training files

Creating training files is so easy you simply must create a couple of files that KEA will use for creating KEA model extractor.

The main file to be analyzed by kea must be a foo.txt file (if you've got pdf, doc, rtf or other type of file, that must be converted to txt). Each file foo.txt must have a foo.key file. The foo.key file contains the keys which you identifies the document, that keys must be present into your thesaurus.

Example of foo.key

```
AMARANTHUS
PLANT PRODUCTION
GEOGRAPHICAL DISTRIBUTION
NUTRITIVE VALUE
SEEDS
MERCHANTS
```

Both files among other pair of couples must be under some directory. That directoy path is what it'll be used by KEA to create the model. Take a look at Automatic_key_extraction_full_example and the use of the trainingFolder param used by application to creation the KEA model.

You need a significative couples of documents in order making a good key extraction model. Upper 100 or more files (depending how large is your thesaurus, etc...) it's good size to starting.

We suggest you take a look at KEA project in order to see how that files are defined in training folder [1]

How optimize model

The KEA model is something alive. The idea is that users tunning the KEA model in OpenKM. For doing it we suggest creation of some metadata (property group) to indicating that user has validated some documents key (flag to indicate that are documents that can be used to creating a new model). After passed some time you can create a minimal application to extract relevant documents (using openoffice conversion can created easilly txt files) and key files too (assigned keywords to documents).

While your repository is growing your KEA model it'll become more efficient.

References

[1] http://www.nzdl.org/Kea

Article Sources and Contributors

Installation Guide *Source*: http://wiki.openkm.com/index.php?oldid=5072 *Contributors*: Jllort, Pavila, Wangmj

Software and hardware requirements *Source*: http://wiki.openkm.com/index.php?oldid=4734 *Contributors*: Pavila

OpenKM installation *Source*: http://wiki.openkm.com/index.php?oldid=691 *Contributors*: Pavila

Java installation *Source*: http://wiki.openkm.com/index.php?oldid=2370 *Contributors*: Pavila

Download OpenKM *Source*: http://wiki.openkm.com/index.php?oldid=693 *Contributors*: Pavila

OpenKM configuration *Source*: http://wiki.openkm.com/index.php?oldid=2049 *Contributors*: Pavila

JBoss mail configuration *Source*: http://wiki.openkm.com/index.php?oldid=4519 *Contributors*: Jllort, Kahtan.it.tep, Pavila

JBPM configuration *Source*: http://wiki.openkm.com/index.php?oldid=861 *Contributors*: Pavila

Application configuration *Source*: http://wiki.openkm.com/index.php?oldid=5044 *Contributors*: Jllort, Pavila, Sean.fmq

Several OpenKM instances in the same JBoss *Source*: http://wiki.openkm.com/index.php?oldid=4341 *Contributors*: Liska.aj, Pavila

Configure JBoss service *Source*: http://wiki.openkm.com/index.php?oldid=5199 *Contributors*: Javier.rodriguez.dgd, Jllort, Pavila, Wangmj

Several JBoss in the same server *Source*: http://wiki.openkm.com/index.php?oldid=3554 *Contributors*: Pavila

Several JBoss instances in the same server *Source*: http://wiki.openkm.com/index.php?oldid=1893 *Contributors*: Pavila, Wangmj

Securing JBoss *Source*: http://wiki.openkm.com/index.php?oldid=1592 *Contributors*: Pavila

Adding APR to JBoss *Source*: http://wiki.openkm.com/index.php?oldid=3767 *Contributors*: Pavila

Changing OpenKM timeout session *Source*: http://wiki.openkm.com/index.php?oldid=1216 *Contributors*: Jllort, Pavila

Test OpenKM installation *Source*: http://wiki.openkm.com/index.php?oldid=1023 *Contributors*: Pavila

Change default okmAdmin password *Source*: http://wiki.openkm.com/index.php?oldid=1019 *Contributors*: Pavila

Create users and roles *Source*: http://wiki.openkm.com/index.php?oldid=1020 *Contributors*: Pavila

Create password validator *Source*: http://wiki.openkm.com/index.php?oldid=3228 *Contributors*: Jllort, Pavila

Notification and subscription messages *Source*: http://wiki.openkm.com/index.php?oldid=3748 *Contributors*: Pavila

OpenKM authentication *Source*: http://wiki.openkm.com/index.php?oldid=5009 *Contributors*: Jllort, Pavila

Active Directory *Source*: http://wiki.openkm.com/index.php?oldid=5012 *Contributors*: Jllort, Pavila

Open Directory *Source*: http://wiki.openkm.com/index.php?oldid=5013 *Contributors*: Pavila

Central Authentication Service *Source*: http://wiki.openkm.com/index.php?oldid=5077 *Contributors*: Pavila

Testing LDAP configuration *Source*: http://wiki.openkm.com/index.php?oldid=5015 *Contributors*: Pavila

Repository configuration *Source*: http://wiki.openkm.com/index.php?oldid=758 *Contributors*: Pavila

Third-party software integration *Source*: http://wiki.openkm.com/index.php?oldid=4920 *Contributors*: Pavila

Third-party software integration: Apache *Source*: http://wiki.openkm.com/index.php?oldid=4732 *Contributors*: Jllort, Pavila, Sean.fmq

Third-party software integration: OCR *Source*: http://wiki.openkm.com/index.php?oldid=5038 *Contributors*: Pavila

Third-party software integration: OpenOffice.org *Source*: http://wiki.openkm.com/index.php?oldid=4539 *Contributors*: Jllort, Pavila

Third-party software integration: SWFTools *Source*: http://wiki.openkm.com/index.php?oldid=4278 *Contributors*: Jllort, Pavila

Third-party software integration: Antivirus *Source*: http://wiki.openkm.com/index.php?oldid=3608 *Contributors*: Pavila

Repository backup *Source*: http://wiki.openkm.com/index.php?oldid=3604 *Contributors*: Jllort, Pavila

Backup scripts *Source*: http://wiki.openkm.com/index.php?oldid=5193 *Contributors*: Jllort, Pavila

Backup restoring *Source*: http://wiki.openkm.com/index.php?oldid=5036 *Contributors*: Pavila

Disaster recovering *Source*: http://wiki.openkm.com/index.php?oldid=5050 *Contributors*: Pavila

Repository mirroring *Source*: http://wiki.openkm.com/index.php?oldid=1489 *Contributors*: Pavila, Wangmj

WebDAV access *Source*: http://wiki.openkm.com/index.php?oldid=4974 *Contributors*: Jllort, Pavila

Property Groups *Source*: http://wiki.openkm.com/index.php?oldid=1945 *Contributors*: Jllort, Pavila

Property Groups definition *Source*: http://wiki.openkm.com/index.php?oldid=4671 *Contributors*: Kahtan.it.tep, Liska.aj, Pavila

Debugging OpenKM *Source*: http://wiki.openkm.com/index.php?oldid=3795 *Contributors*: Pavila

Troubleshooting *Source*: http://wiki.openkm.com/index.php?oldid=5120 *Contributors*: Jllort, Pavila

Error codes *Source*: http://wiki.openkm.com/index.php?oldid=4210 *Contributors*: Jllort, Pavila

Using OpenKM with other databases *Source*: http://wiki.openkm.com/index.php?oldid=4296 *Contributors*: Jllort, Pavila

MySQL *Source*: http://wiki.openkm.com/index.php?oldid=5058 *Contributors*: Pavila

PostgreSQL *Source*: http://wiki.openkm.com/index.php?oldid=5002 *Contributors*: Pavila

PostgreSQL-OpenKM 5.0 *Source*: http://wiki.openkm.com/index.php?oldid=4713 *Contributors*: Mediagraphics.qatar.ijq, Pavila

PostgreSQL-OpenKM 4.1 *Source*: http://wiki.openkm.com/index.php?oldid=2302 *Contributors*: Pavila

MS SQLServer *Source*: http://wiki.openkm.com/index.php?oldid=4518 *Contributors*: Fjsalazar.c, Jllort, Pavila

Oracle *Source*: http://wiki.openkm.com/index.php?oldid=5057 *Contributors*: Bcheath.svo, Jllort, Pavila

Thesaurus *Source*: http://wiki.openkm.com/index.php?oldid=4729 *Contributors*: Jllort, Megamansgo.ok, Pavila

Thesaurus full example *Source*: http://wiki.openkm.com/index.php?oldid=2789 *Contributors*: Jllort, Pavila

Automatic key extraction *Source*: http://wiki.openkm.com/index.php?oldid=3634 *Contributors*: Jllort, Pavila

Automatic key extraction full example *Source*: http://wiki.openkm.com/index.php?oldid=2413 *Contributors*: Jllort

Creating automatic key extraction training files *Source*: http://wiki.openkm.com/index.php?oldid=2457 *Contributors*: Jllort

MS SQLServer *Source*: http://wiki.openkm.com/index.php?oldid=4518 *Contributors*: Fjsalazar.c, Jllort, Pavila

Oracle *Source*: http://wiki.openkm.com/index.php?oldid=5057 *Contributors*: Bcheath.svo, Jllort, Pavila

Thesaurus *Source*: http://wiki.openkm.com/index.php?oldid=4729 *Contributors*: Jllort, Megamansgo.ok, Pavila

Thesaurus full example *Source*: http://wiki.openkm.com/index.php?oldid=2789 *Contributors*: Jllort, Pavila

Image Sources, Licenses and Contributors