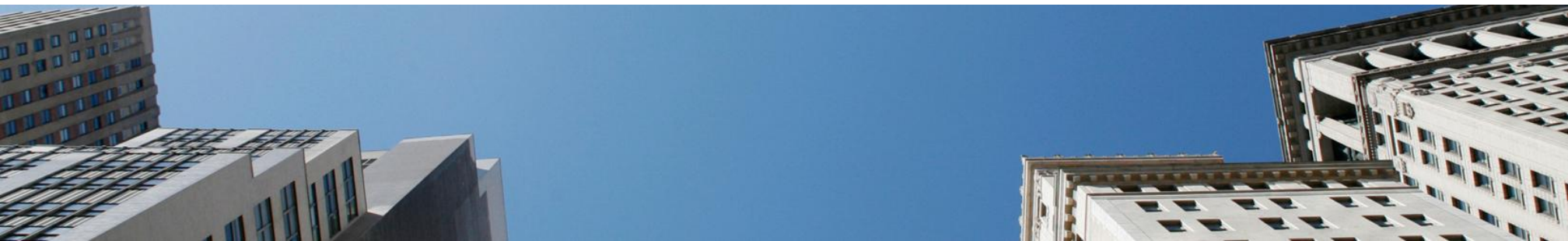


BUILDING A WEBSITE FOR THE NUMBER ONE CHILDREN'S HOSPITAL IN THE U.S.

May 10, 2011



Introduction

About me and NorthPoint...

- NorthPoint is a USA-based organization
- Specializing in Open Source technologies for web applications
 - Of course, we are experts in OpenCms
- 100+ on-staff technologist focusing on Content Management and Enterprise Digital Architectures
- Areas of focus in Media, Publishing, Healthcare, Education, and Not-for-Profit markets

The Site : The Children's Hospital of Philadelphia

Search

[Health Information](#)

[For Patients and Visitors](#)

[Specialties and Services](#)

[Locations](#)

[Find a Doctor](#)

[About Us](#)

A World Leader in Children's Health

The Children's Hospital of Philadelphia, the first pediatric hospital in the United States, is ranked the #1 pediatric hospital by Parents magazine and is one of only eight on the prestigious U.S. News & World Report Honor Roll for 2010. Our physicians, nurses, researchers and other professionals are recognized regionally and nationally for their expertise. CHOP's exceptional patient care and major research initiatives benefit children worldwide.



Important information for patients who have received family planning services. [Read More »](#)

My Child Has:

Search

Search by Disease, Condition, Test or Procedure

Find Specialties and Services by first letter

A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z

Contact Us

1-800-879-2467

(1-800-TRY-CHOP)

- » [Contact Us Online](#)
- » [Request an Appointment](#)

Welcome Kit

A hospital can be a confusing place. Whether this is the first time you are visiting or you've been here before, these links can help prepare you and your family before you arrive.



International Visitors

Learn about the services we

Building a Website and Intranet for CHOP

What we will be covering today

- The situation – Re-platform of existing CMS
- The problem – Scaling of OpenCms
- The solution – A new OpenCms module
- How it all works – A Step by Step review
- How can this benefit you

The Situation at CHOP

- Running on Legacy CMS (Broadvision CMS) with old servers and old technology
- Needed a new CMS to support their growth initiatives
- Rebranding of the site, new site design
- Wanted a new site architecture to focus on the patients and children
- New Intranet site to run on the same platform as the internet
- New Solution needs to handle 10gb of binary assets for the intranet
- New Solution needs to be easy to use and have great search to find these assets
 - Assets mostly PDF's, Word Doc's and Presentations

Re-platform of CMS to OpenCms at CHOP

- NorthPoint was part of the early adoption process of OpenCms
- CHOP reviewed and prototyped 3 other CMS platforms
- NorthPoint proto-typed OpenCms and it was chosen to be CHOP's CMS platform for the future
- Re-platform included re-branding of the site, new I/A
- Migration of existing content, major content cleanup
- 3rd Party integrations of content from outside content providers
- Integrated Search, Lucene
- New Architecture and hosting provider to improve performance and save \$
- Architecture to support large-scale Intranet and additional sites

The Challenges - Scaling the Solution

- The Internet implementation was successful
- Developed a custom Portal 'MyOpenCms' for non-admin editors
- Challenges for the Intranet
- Do we use the same instance of OpenCms as the Internet?
- Do we need to share content between these systems?
- What is the hardware configuration?
- How do we support 10 gb of binary assets that need to be served through OpenCms
- Needed a way to allow the editors to reference the binary asset in OpenCms just like they would with any other content type.
- Need to be able to search the content and index the content using built-in Lucence support
- No Known solution to house binary assets outside of OpenCms DB

So What did We Do?

- As you all know putting 10gb of assets in OpenCms means that all these assets have to be in the database.
- This causes issues with performance of the site and issues with database backups and replication (taking 4 hours without the additional 10gb of binary assets)
- First we looked at the Architecture of OpenCms
- As any good developers – we prototyped a few different solutions.
- Could we upload the files from a web form?
- Could we parse the data to capture the information needed to keep it referenced in OpenCms – find the right tools?
- Make sure we could keep the OpenCms search working thru Lucene
- Built a Content type to support the solution
- Once it worked decided to create a module for OpenCms community

Creating The Solution

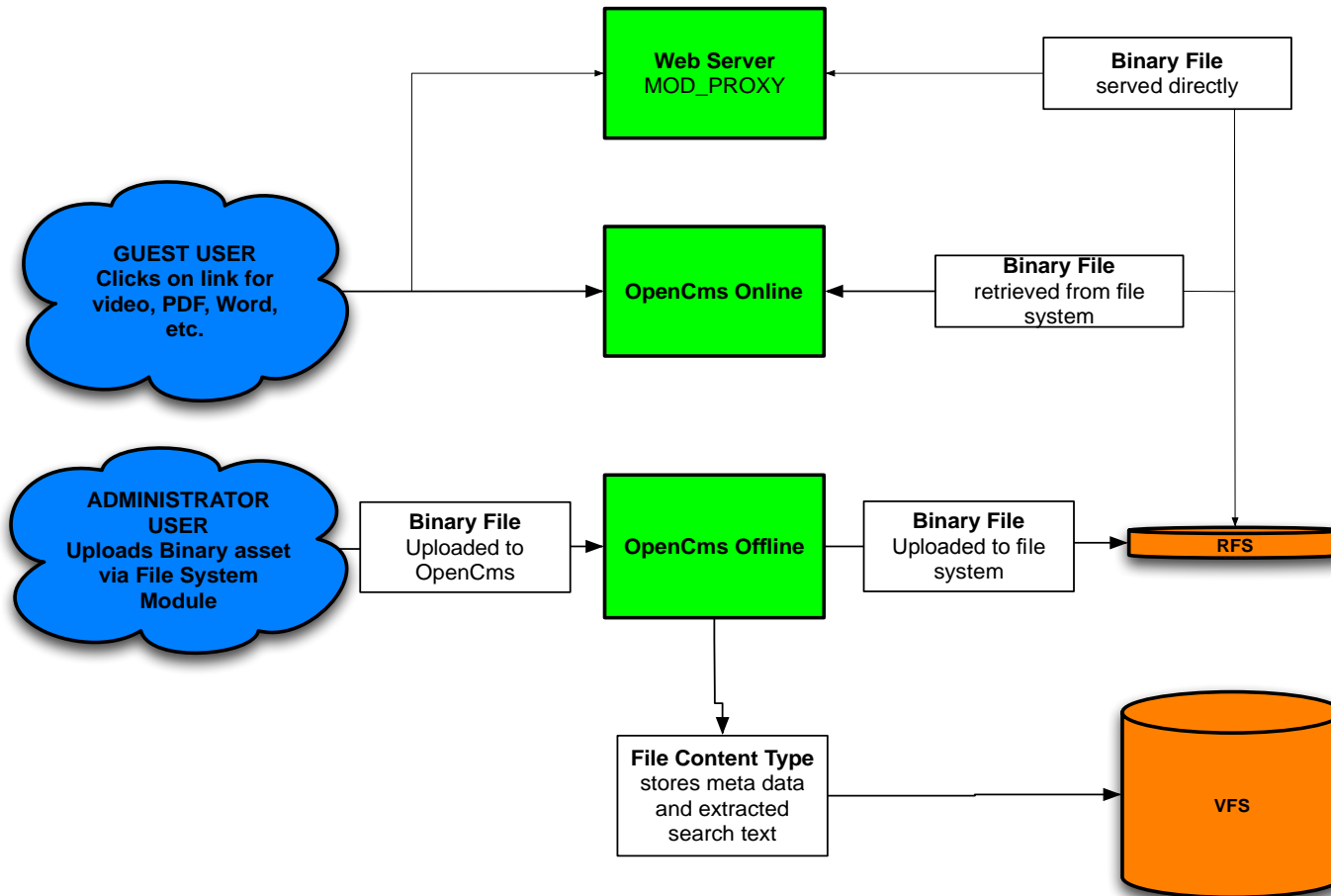
What Would the Module need to Do?

- Store Binary files outside the VFS
- Store the actual binary File in an RFS
- Parse the Content from the Binary files to OpenCms via Apache Tika
- Leverage native OpenCms APIs to create the content type to hold all the data
- Needed to come up with an interface – so we created a new view ‘File System Module’
- Configure Search so it can leverage the content type

What tools did we use

- Tika, JQuery, OpenCms, Java, JSP

So How Does it All Work



How Do You Make it All Work?

- There are a number of Steps that you need to go through to make this new module work for you
- Let's take a look at how easy it is to solve this problem for your sites
- Steps we will cover
 - Installing the Module
 - Configuring the Module
 - Configuring Search
 - Start using the Module to Upload Binary Files

Installation Overview

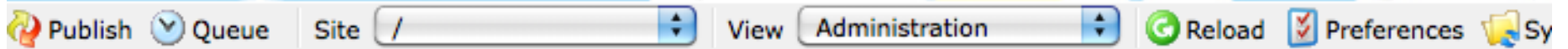
- Install the Module (NorthPoint Solutions FileSystem Module)

www.northps.com/sites/default/NorthPoint_Solutions_OpenCms_FileSystem_1.0.zip

- Configure the Module, Application Server & Search
- Navigate to the File System Manager View (like explorer or administration view)
- Select File System Module View
- Load up the HTML Interface – Ready to Use and Upload Assets
- Next – Upload your files
 - Uploads file to RFS
 - Parses the file (using Apache Tika)
 - Uses meta data from Tika to create a new content type (file type)
 - This file type contains the meta data and the reference to the RFS

Let's Walk Through the Process

Install Module : Click on Module Management



Administration View

Administration



Project Management



Publish Queue



Account Management



Database Management



Module Management



Gallery Overview



Search Management



Scheduled Jobs



Cache Administration



Content Tools

Install Module : Import Module

Administration View > Module Management

Module Management

Module actions



New Module



Import Module from
Server






Import Module with
HTTP



Install Module : Import The Module


Module Management













Module actions

 New Module  Import Module from Server  Import Module with HTTP

Installed Modules

Installed Modules (34)  Author info  Relevance

 Search

E	X	D	Module ▲	Title
			com.alkacon.opencms.calendar	Alkacon OpenCms Calendar Module
			com.alkacon.opencms.common	Alkacon OpenCms Common Utilities
			com.alkacon.opencms.formgenerator	Alkacon OpenCms Add-On Module Package Webform
			com.northps.filesystem	NorthPoint Solutions File System Module

Configure the Module

Administration View > Module Management > Edit Module

Edit Module

Module actions



Edit Module



Module Resources



Module Parameters



Module exportpoints



Module dependencies



Export Module

Module information

Package name: com.northps.filesystem
Module name: NorthPoint Solutions File System Module
Module description: This module allows for the upload of files to the real filesystem and to the virtual filesystem
Module version: 1.0
Module group: NorthPoint Solutions
Action class:

Author information

Author name: NorthPoint Solutions
Author email: infoUS@northps.com

Configure the Module : Configuration Parameters

Parameter 1 : RFS Location


Parameter 2 : 0 is unlimited default is 10mb

Parameter 3 : location on VFS where content will be stored

Administration View > Module Management > Edit Module > Module Parameters

Module Parameters

Module parameters

Parameter [1]:  directoryPath=/usr/local/repo/

Parameter [2]: fileSizeLimitBytes=0

Parameter [3]: openCmsPath=/system/galleries/download/filesystem/

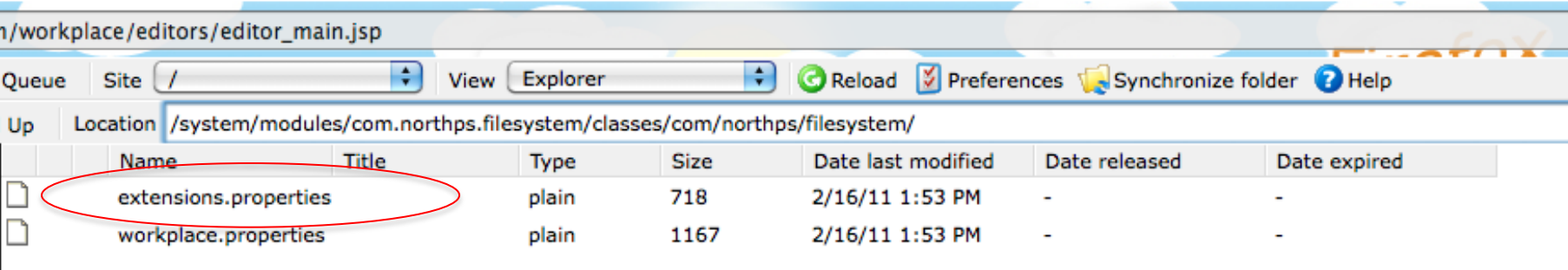
Ok

Cancel



Configure : Allowable File Extensions

- We keep a list of the allowed file extensions in the module by using a properties file called extensions.properties
- You can modify this list if you choose.
- The “parsableExtension=\\.pdf..” is the extensions that Tika can actually parse
- You can modify this list within the VFS in the classes directory of the module
- When you publish the file it will be pushed to the tomcat server and will require a restart of the server

VFS LOCATION



The screenshot shows a web-based file explorer interface. The address bar displays the path: /workplace/editors/editor_main.jsp. Below the address bar, there are navigation buttons: Queue, Site (/), View (Explorer), Reload, Preferences, Synchronize folder, and Help. The current location is shown as /system/modules/com.northps.filesystem/classes/com/northps/filesystem/. A table lists the files in this directory:

	Name	Title	Type	Size	Date last modified	Date released	Date expired
	extensions.properties		plain	718	2/16/11 1:53 PM	-	-
	workplace.properties		plain	1167	2/16/11 1:53 PM	-	-

The file 'extensions.properties' is circled in red in the original image.

Configure : Extensions.properties

CONTENT OF FILE extensions.properties

For example if you wanted to allow .zip files you would add under .rtf. You would not add it to the parableExtension list because tika will not parse it.

.rtf=text/rtf

.zip= application/*zip*

Extensions List

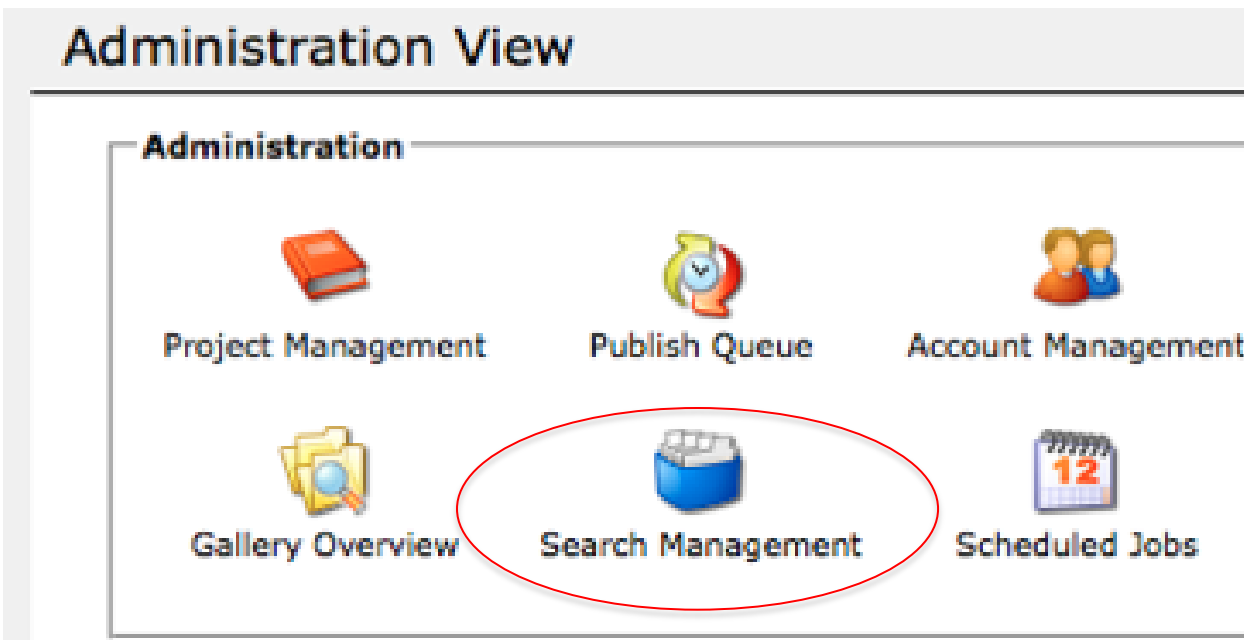
```
parseableExtension=\\.pdf|\\.doc|\\.flv|\\.docx|\\.mp3|\\.rtf|\\.ppt|\\.pptx|\\.xl
```

```
#ALL Supported extensions including the ones that are not parseable
```

```
.swf=application/x-shockwave-flash  
.mov=video/quicktime  
.avi=video/x-msvideo  
.mpg=video/mpeg  
.mpeg=video/mpeg  
.pdf=application/pdf  
.docx=application/vnd.openxmlformats-officedocument.wordprocessingml.document  
.flv=video/x-flv  
.doc=application/msword  
.mp3=audio/mpeg  
.pptx=application/vnd.openxmlformats-officedocument.presentationml.presentation  
.ppt=application/vnd.ms-powerpoint  
.xlsx=application/vnd.openxmlformats-officedocument.spreadsheetml.sheet  
.xls=application/vnd.ms-excel  
.txt=application/text  
.rtf=text/rtf
```

Configure Search Configuration (Optional)

- Search setup is an advanced function and is optional. Setting up search properly is somewhat time consuming if you are unfamiliar with set-up
- Highly recommend going through the readme file provided with the new module
- Basis search implementation has been provided and will work based on the steps provided in the readme file
- Modifications to Search would require customization of this module






Search Configuration : Step 1 (high level)

- This screenshot is the result of going through the steps of clicking on the new index button
- Creates a new search index for the File System Module
- Search index will update when content is created, updated or deleted in the VFS

Administration View > Search Management


























Search Management Up

Index actions

 New Index  View index sources  View field configurations

Search Indices

Search Indices (5) Show index sources Show field configuration Print Delete Rebuild

E	S	D	R	S	Name	Field config	Rebuild	Project	Locale	
					filetype	filetype	auto Offline		en	<input type="checkbox"/>
					German online help	standard	auto Online		de	<input type="checkbox"/>
					Offline project EN (VFS)	standard	offline Offline		en	<input type="checkbox"/>
					Online project DE (VFS)	standard	auto Online		de	<input type="checkbox"/>
					Online project EN (VFS)	standard	auto Online		en	<input type="checkbox"/>

Search Configuration : Step 2 New Field Configuration

- This is where you configure the fields on your content types that you want to be indexed.
 - For example the author...

There are two sub steps to Step 2...so lets continue

Administration View > Search Management > View field configurations

View field configurations

Field configuration actions



New field configuration

Field configurations

Field configurations (2)

Show fields

E	F	D	Name*	Description
			filetype	filetype search
			standard	The standard OpenCms 7.0 search index field configuration.

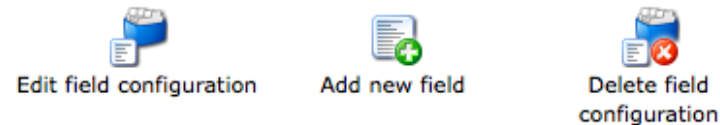
Search Configuration : Step 2A

- This adds a new field configuration which has to be assigned to the index in order for it to work.
- In this configuration, you will specify mappings

Administration View > Search Management > View field configurations > Field configuration overview: filetype

Field configuration overview: filetype

Field configuration actions



Field configuration

Name : filetype
Description: filetype search

Fields

Fields (1)

Show mappings

E	M	Name	Display name	Stored	Excerpt	Indexed	Boost	D
		author	author			untokenized	1.0	

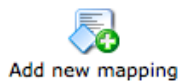
Search Configuration : Step 2B

- This adds a mapping to the field configuration
- Basically this will tell the index which fields in the content you want to search through the index

Administration View > Search Management > View field configurations > Field configuration overview: filetype > Field overview: author

Field overview: author

Field actions




Field

Name: author
Display: author
Store: True
Index: untokenized
Excerpt: False
boost: 1.0
Default:

Mappings

Mappings (1)

M	Parameter	Type [▲]	Default
	meta_author[1]	item	-

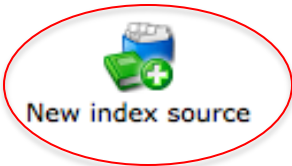
Search Configuration : Step 3

- This configuration is so you can point the Lucene index at a folder in your VFS and it will index the content
- For our module, we will be indexing “/system/galleries/download/filesystem/”
 - This is our default location for filetype content

Administration View > Search Management > View index sources

View index sources













Index source actions



Index sources

Index sources (3)

Show document types

E	R	T	D	Name	Indexer
				filetype	org.opencms.search.CmsVfsIndexer
				source1	org.opencms.search.CmsVfsIndexer
				source2	org.opencms.search.CmsVfsIndexer

Additional Configuration (Mandatory)

- This step is not part of OpenCms.
- This is a sever configuration of an xml file on tomcat and is normally performed by the system administrator or developer
- The file is web.xml. Filters declarations and mappings are declared above servlet declaration and mappings
- In order for the VFS to serve content from the RFS, a servlet filter needs to intercept the http request
- The Servlet filter mapping needs to be added to web.xml
 - This filter provides URL resolution from content type to binary asset when requested in OpenCms
 - Without this set up the content type can not serve the binary asset from the RFS

Snap Shot of web.xml file

```
<listener-class>org.opencms.main.OpenCmsListener</listener-class>
</listener>

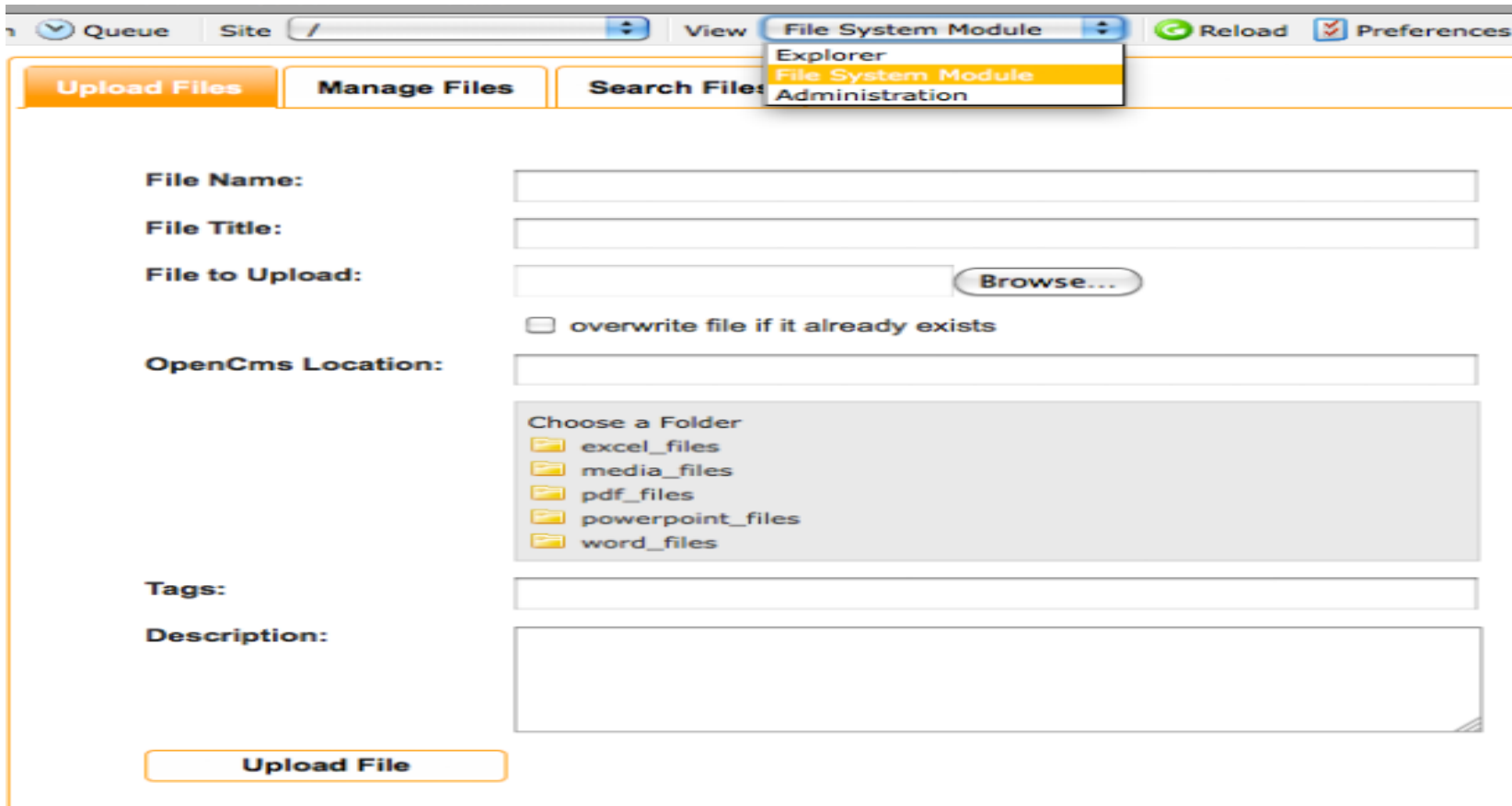
<filter>
  <filter-name>fileupload</filter-name>
  <filter-class>com.northps.opencms.rfs.Rfsuploadfilter</filter-class>
  <description>upload filter</description>
</filter>

<filter-mapping>
  <filter-name>fileupload</filter-name>
  <url-pattern>/opencms/*</url-pattern>
</filter-mapping>

<servlet>
  <description>
    The main servlet that handles all requests to the OpenCms VFS.
  </description>
  <servlet-name>OpenCmsServlet</servlet-name>
  <servlet-class>org.opencms.main.OpenCmsServlet</servlet-class>
```

Let's Get Started

Now that we have things configured let's upload some files



The screenshot shows the OpenCms File System Module interface. At the top, there is a navigation bar with a 'Queue' icon, a 'Site' dropdown, a 'View' dropdown set to 'File System Module', and 'Reload' and 'Preferences' buttons. Below the navigation bar are three tabs: 'Upload Files' (highlighted in orange), 'Manage Files', and 'Search Files'. The main content area contains a form for uploading files. The form includes the following fields and controls:

- File Name:** A text input field.
- File Title:** A text input field.
- File to Upload:** A text input field with a 'Browse...' button to its right.
- overwrite file if it already exists**
- OpenCms Location:** A text input field.
- Choose a Folder:** A list of folders with folder icons:
 - excel_files
 - media_files
 - pdf_files
 - powerpoint_files
 - word_files
- Tags:** A text input field.
- Description:** A large text area.
- Upload File:** A button at the bottom of the form.

Upload your First File

It's Easy –

Just Click & Upload

Note:

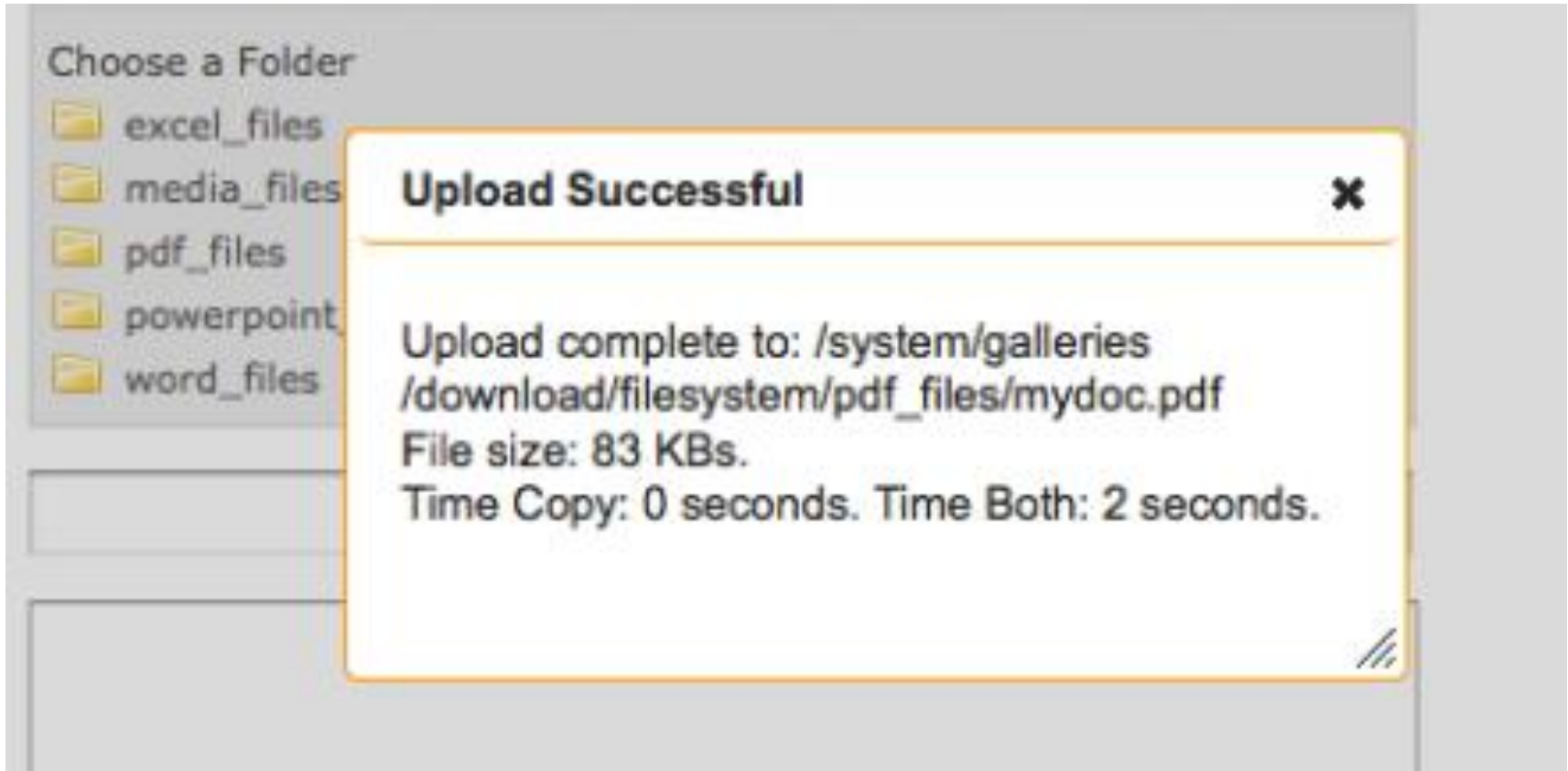
File name location

The screenshot shows the OpenCms file upload interface. At the top, there are three tabs: "Upload Files" (highlighted in orange), "Manage Files", and "Search Files". Below the tabs, the form contains the following fields and options:

- File Name:** A text input field containing "my document".
- File Title:** A text input field containing "my title".
- File to Upload:** A text input field containing ":uments/ebooks/mydoc.pdf" and a "Browse..." button.
- overwrite file if it already exists
- OpenCms Location:** A text input field containing "/system/galleries/download/filesystem/pdf_files/".
- Choose a Folder:** A list of folders: excel_files, media_files, pdf_files (highlighted in blue), powerpoint_files, and word_files.
- Tags:** A text input field containing "Java, OpenCms".
- Description:** A text area containing "My document description.".

At the bottom of the form is an "Upload File" button.

Erfolg



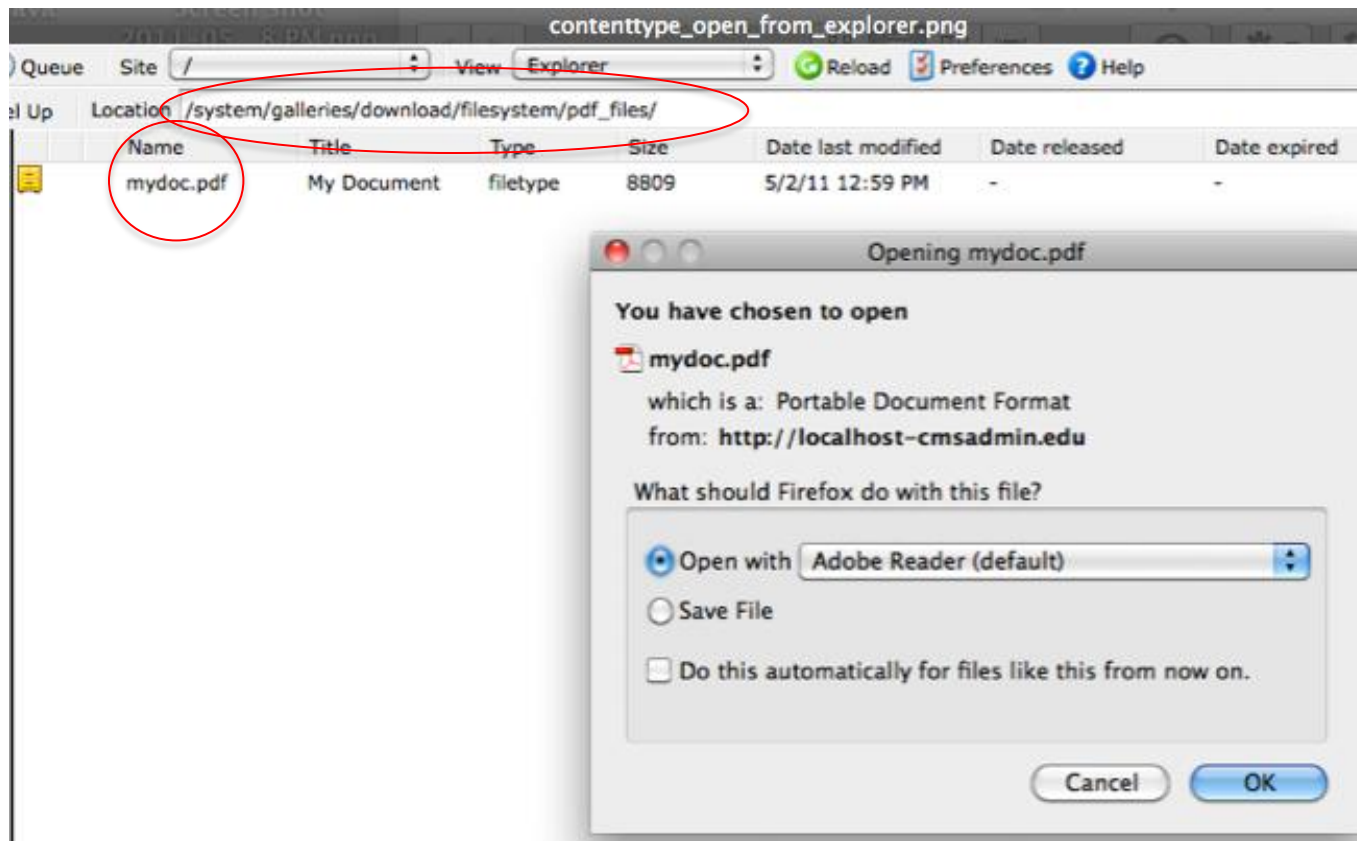
Let's See the Results

- What occurred behind the scenes
 - The physical binary file is stored on the RFS
 - The NP module calls Apache Tika to extract the meta data and store it on the VFS (your file system content type)
 - Apache Tika will extract content from PDF-s, Word Docs, Video and a host of other formats
 - <http://tika.apache.org/0.9/formats.html>
 - Search will work since all the information needed has been extracted and stored in the content type

Let's look at the new Content Type

Reminder :

The content type has been saved in the OpenCms path during configuration



Let's Open from the Icon

Note:

RFS Path

Content is the extracted meta data & Mime Type is captured from Apache Tika

The screenshot shows a document metadata form with the following fields:

- Link Name: My Document
- RFS Path: /usr/local/share/pdfshare/system/galleries/download/filesystem/pdf_files/mydoc.pdf
- Title: My Document
- Author: jcampedelli
- Keywords:
- Description: (empty)
- Content: C Abstract Factory S Facade S Proxy
Memento Memento
S Adapter C Factory Method B Observer Caretaker -state
Type: Behavioral
S Bridge S Flyweight C Singleton
- Tags: Java, OpenCms
- Mime Type: application/pdf
- Date Created: 2008-01-23T15:30:03Z
- Date Modified:
- Release Date:
- Expire Date:

Red circles highlight the RFS Path and Content fields.

Basic Administrative Features

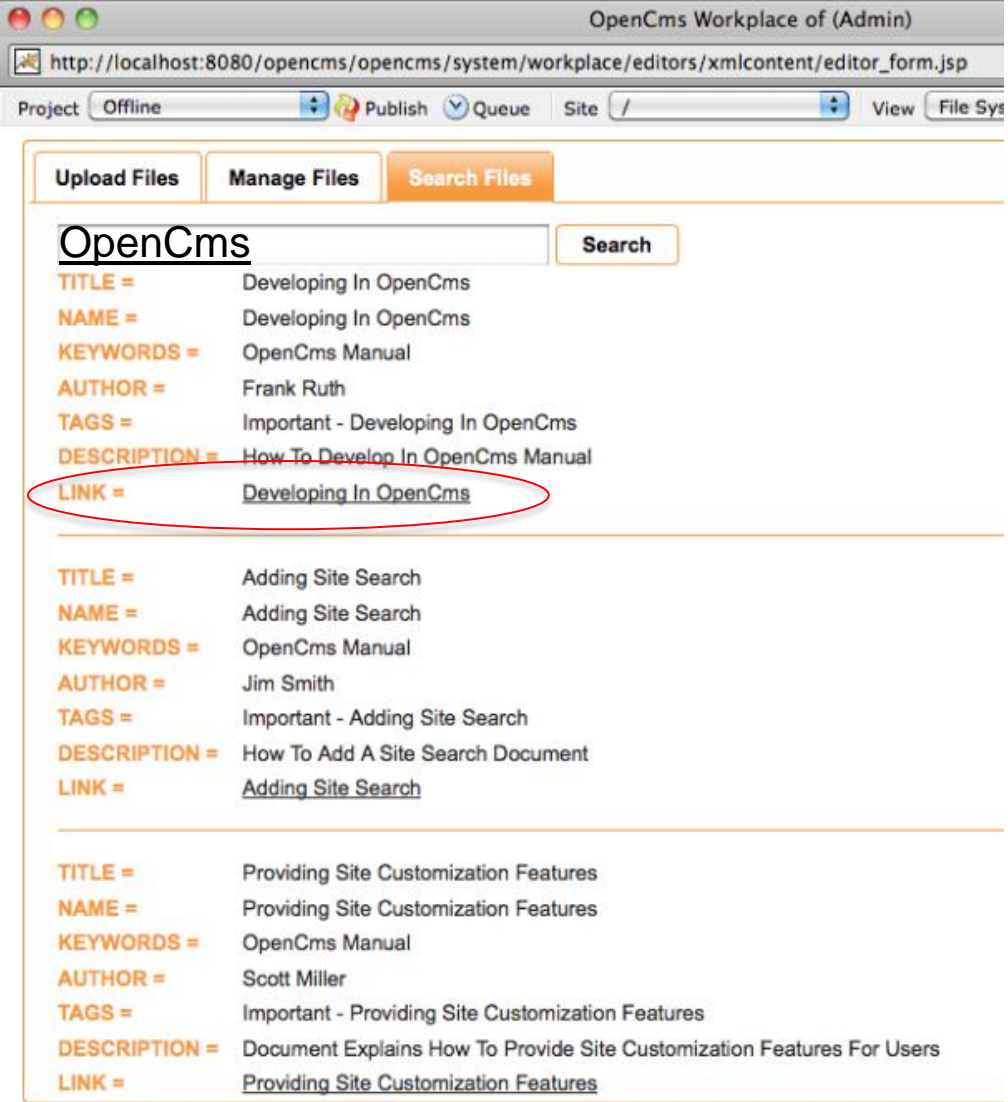
- As an Administrator, you can delete the uploaded files from the system
 - Both RFS and VFS content are deleted

The screenshot displays a web-based file management interface. At the top, there are three navigation buttons: 'Upload Files', 'Manage Files' (which is highlighted in orange), and 'Search Files'. Below these buttons, the 'File Name:' field contains the path '/system/galleries/download/filesystem/pdf_files/'. A dropdown menu is open, showing a list of folders and files: 'excel_files', 'media_files', 'pdf_files', 'mydoc.pdf' (which is highlighted with a blue bar), 'powerpoint_files', and 'word_files'. Below the dropdown is a 'Delete File' button.

Let's do some Searching : Find OpenCms

Note:

Link will open up the RFS
Binary asset!



The screenshot shows the OpenCms Workplace of (Admin) interface. The browser address bar displays the URL: `http://localhost:8080/opencms/opencms/system/workplace/editors/xmlcontent/editor_form.jsp`. The interface includes navigation tabs for 'Upload Files', 'Manage Files', and 'Search Files'. A search input field contains the text 'OpenCms', and a 'Search' button is visible. The search results are displayed in a list format, with the first result circled in red. The first result is:

TITLE =	Developing In OpenCms
NAME =	Developing In OpenCms
KEYWORDS =	OpenCms Manual
AUTHOR =	Frank Ruth
TAGS =	Important - Developing In OpenCms
DESCRIPTION =	How To Develop In OpenCms Manual
LINK =	Developing In OpenCms

The second result is:

TITLE =	Adding Site Search
NAME =	Adding Site Search
KEYWORDS =	OpenCms Manual
AUTHOR =	Jim Smith
TAGS =	Important - Adding Site Search
DESCRIPTION =	How To Add A Site Search Document
LINK =	Adding Site Search

The third result is:

TITLE =	Providing Site Customization Features
NAME =	Providing Site Customization Features
KEYWORDS =	OpenCms Manual
AUTHOR =	Scott Miller
TAGS =	Important - Providing Site Customization Features
DESCRIPTION =	Document Explains How To Provide Site Customization Features For Users
LINK =	Providing Site Customization Features

Customize : Custom Implementation for CHOP

My IntranetCms

Home Search **File Upload**

Upload Form

File Name:

File Title:

File to Upload:

Overwrite file if it already exists

OpenCms Location:

Choose a Folder

- excel_files
- media_files
- pdf_files
- word_files

and your "Recently Published" files for easy access

Unpublished Recently Published

The following files have been edited and have not been published

Alerts

Jennifer J Jenn



News

Schizophrenia Genes Involved in Brain Signaling



Refined Tools Help Pinpoint Disease-Causing Genes



Remember : Follow the Readme....

- Install the new module : `com.northps.filesystem`
- Configure
- Configure Search
- Customize if you like
- Enjoy having a separate File system for you Binary Assets

RICH COOLEY

rich_cooley@northps.com

TALK TO US. CHALLENGE US.

CALL US at 212.819.1700

VISIT US at www.northps.com

FOLLOW US @northps

© NorthPoint and NorthPoint Solutions LLC