



JiFile



Filesystem



Index



Config



Support

[FAQ - Quick Tutorial - Video Tutorial - extensions.joomla.org](#)

**Problems with JiFile? [Click here](#) for download your configuration and attach it on the [help forums](#) or by email [info@isapp.it](mailto:info@isapp.it).**

*For security you can use the forum code [confidential] to view the file only to administrators. Example:*

[confidential]

[attachment]sysinfo\_0.833778001330285188\_localhost\_505.dat[/attachment]

[confidential]

# JiFile

Documentation

and

PDF Optimization

Manual

By Matthias Kleespies, Germany

[germanclimateblog.com](http://germanclimateblog.com)

[principia-scientific.org](http://principia-scientific.org)

## Preface:

JiFile is a very powerful and versatile **indexing** and **search** software enabling you to render many of your documents searchable that otherwise could not be searched by Joomla's search engines.

This, of course, will bring more value to your site.

However, it contains some "cliffs" which are not easy to circumnavigate if you are not familiar with its *modus operandi*.

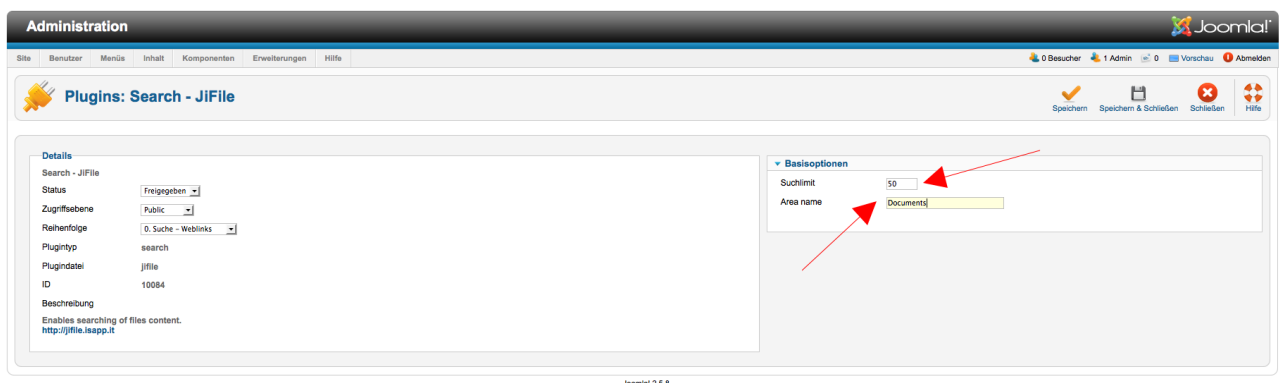
I spent several hours together with JiFile's developers, whose support is tremendous, to have it operate on two sites now exactly the way I wanted to.

Therefore, to make it easier for you to set up and optimize JiFile, I decided to write this little documentation and manual.

## Let's Get Started – Setting Up JiFile After Your First Installation

Installation is a breeze. I especially like that the installation routine reminds you of enabling the SEARCH plugin which is **absolutely necessary** to search the index JiFile creates out of your documents.

However, already here you may encounter the first "cliff".



*JiFile's search plugin*

What does "Area name" mean? How does any entry in this field effect anything?

And what does "Search Limit" determine?

The answer, fortunately, is easy: whatever you enter in the *Area name* will be the exact designation of JiFile's search plugin displayed at the frontend as additional search filter:

# SEARCH

Search

Search Keyword:

All words  Any words  Exact Phrase

Ordering:  ▾

Search Only:  Categories  Contacts  Articles  Newsfeeds  Weblinks  Publications  Kunena Forum

Total: 55 results found.

*Designation of JiFile's search plugin. In this case it was named "Publications" in the "Area name" of the plugin.*

Wait, this is MOST IMPORTANT!

"Search Limit" is NOT the maximum number of characters accepted by the search field (at the frontend).

This value sets the maximum number of *search results displayed* by JiFile's own search plugin. Therefore, if you have many documents in your database, make sure to adjust this value accordingly. Otherwise all other documents exceeding the "Search Limit" will simply be truncated from being displayed at the frontend.

## *Basic Configuration – This is CRUCIAL*

When you open JiFile for the first time after installation you will see this screen and message:

No path configured for the index

Report JiFile				
Type	Directory	State	Info	
Filesystem	Configure	Not exists		
Index	Configure	Not exists		

Report check system [Warning check here](#)

Check update

FAQ - Quick Tutorial - Video Tutorial - extensions.joomla.org

Problems with JiFile? Click here for download your configuration and attach it on the help forums or by email [info@sapp.it](mailto:info@sapp.it).

For security you can use the forum code [confidential] to view the file only to administrators. Example:  
[confidential]  
[attachmentid=999999\_0\_833778001230285188\_localhost\_505.dcf(attachment)]  
[confidential]

*Initial JiFile screen*

The message tells you "No path configured for the index". To fix this, please click on "Filesystem/Configure" below "Report JiFile" OR "Index/Configure". This opens another screen:



*Configuration Screen*

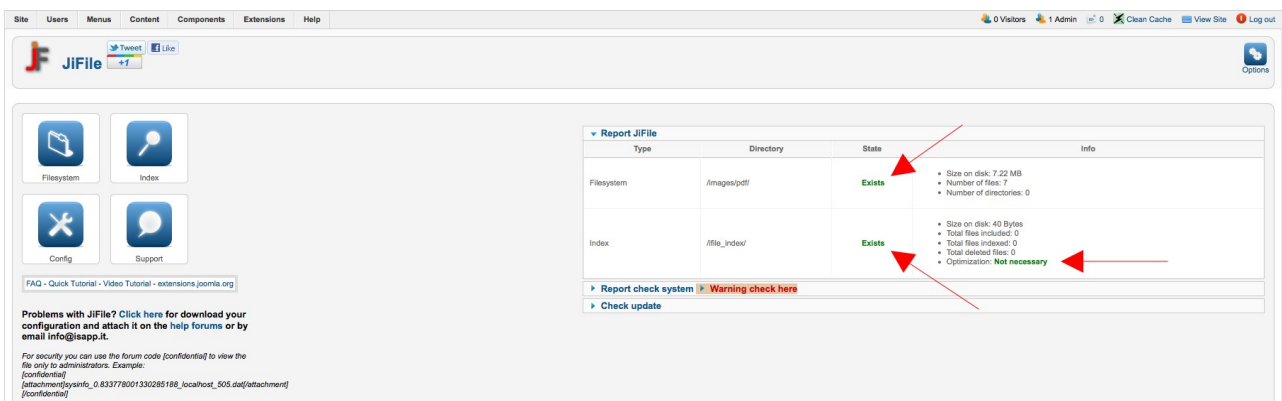
**Filesystem path:** here you can enter any path you like. NOTE: the path can be nested! I. e., subfolders are possible.

NOTE: JiFile allows only ONE Folder as "Filesystem path". However, inside this folder, you can create (nested) subfolders.

**Index path:** You can either enter your own preferred folder or leave this entry as it is. Upon clicking "Save" or "Save & Close" this folder will be created.

NOTE: you can change these settings at any time by clicking "Options" in the top right corner of JiFile's control panel.

After clicking "Save & Close" JiFile will show you if your settings were correct:



*Confirmation Screen*

When everything is GREEN, your settings were correct. If not, the wrong settings will be marked and you have to change them.

Okay, this was the easy part. Now let's turn to the – somewhat – harder part.

Perhaps you already noticed that "Report Check System" says: "Warning check here". Please click on "Report Check System", which will open this screen:

Report JiFile				
Report check system <span style="color: red;">Warning check here</span>				
Label	Check	Requirements	Info	Use & WebSite
<b>IFile Framework</b>				
IFile Framework	Exists	Version 1.1.3 or later	IFile Framework is installed Version 1.1.5	Used by JiFile WebSite: <a href="http://ifile.isapp.it">http://ifile.isapp.it</a>
<b>Zend Framework</b>				
Zend Lucene	Exists	Version 1.10.1 or later	Zend Lucene is installed in /home2/nonamesc/public_html/germanclimateblog/libraries/ifile - Version 1.11.3	Used by Lucene and MySQL Interface WebSite: <a href="http://www.zend.com">http://www.zend.com</a>
Zend Cache	Exists	Version 1.10 or later	Zend Cache is installed in /home2/nonamesc/public_html/germanclimateblog/libraries/ifile Version 1.11.3	Used by JiFile WebSite: <a href="http://www.zend.com">http://www.zend.com</a>
<b>Encoding</b>				
PCRE	OK	Not defined	PCRE unicode support is enabled in PHP	Used by Zend Search Lucene Framework WebSite: <a href="http://www.php.net/manual/en/book.pcre.php">http://www.php.net/manual/en/book.pcre.php</a>
<b>SERVER</b>				
Server	64bit	Not defined	Note: If the OS is 64bit but running a 32 bit version of php, the check will return (32 bit)	Only for linux/windows. Copy adapter/helpers/binaries/[linux windows]/bin64/pdftotext in adapter/helpers/binaries/[linux windows] WebSite: <a href="http://www.php.net/manual/en/install.php">http://www.php.net/manual/en/install.php</a>
<b>XPDF</b>				
XPDF Binaries File	Unexecutable	CHMOD 0755	Permission XPDF Binaries File (adapter/helpers/binaries/linux/pdftotext): 0644 - Please set to 0755 for binaries XPDF in LIN	Used only for PDF file parser WebSite: <a href="http://www.foolabs.com">http://www.foolabs.com</a>
<b>ANTIWORD</b>				
ANTIWORD Binaries File	Unexecutable	CHMOD 0755	Permission ANTIWORD Binaries File (adapter/helpers/binaries/linux/antiword): 0644 - Please set to 0755 for binaries ANTIWORD in LIN	Used only for DOC file parser WebSite: <a href="http://www.winfield.demon.nl/">http://www.winfield.demon.nl/</a>
<b>PHP</b>				
PHP Version	OK	Version 5.1.0 or later	Version installed is 5.3.20	Not defined WebSite: <a href="http://www.php.net">http://www.php.net</a>

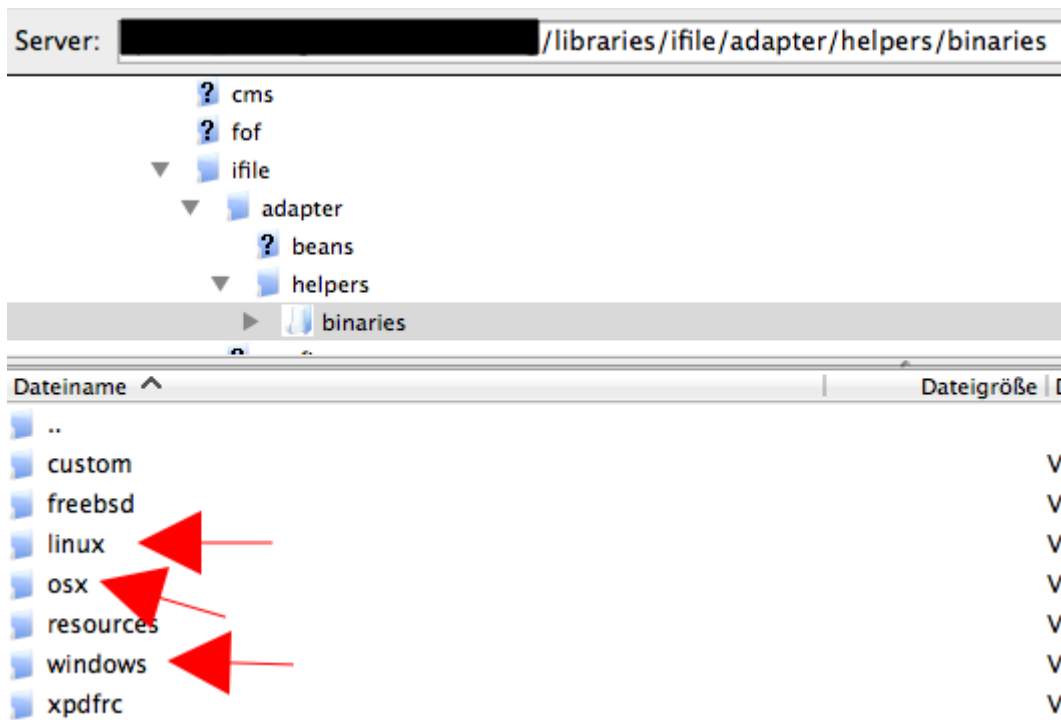
If anything else apart from "Unexecutable" is RED within your system configuration (besides "id3 and "com\_dotnet" saying "KO"), you will most likely have to ask your ISP for help.

These two "unexecutable" files are the most important ones for *pdf* and *.doc* (MS Word) indexing. Therefore, we'll have to fix this. To fix it you will need FTP access to your server. The message "0644 -Please set to 0755 for binaries XPDF in LIN" already – somewhat – tells us what to do.

Unfortunately, it does not tell the COMPLETE path to the files.

The *complete path* is: ROOT(your root folder on the server in which Joomla is hosted)/**libraries/ifile/adapter/helpers/binaries**.

Inside the "binaries" folder you will find the following subfolders:



*Folders inside JiFile's "binaries" folder*

THIS IS IMPORTANT: In most cases you will have to open the "linux" folder – JiFile said " XPDF in LIN".

If it says "OSX" you are on an Apple Server and you'll have to open the "osx" folder; if on a Windows server, please open the "windows" folder.

Inside the appropriate folder, highlight both "antiword" and "pdftotext" and choose "Change permissions" in your FTP client.

Then set the permissions of these two files to "755" and click "OK", or anything suitable in your FTP client, to alter file permissions.



After that, check your changes by SELECTING JIFILE FROM THE "COMPONENTS" MENU AGAIN – you have to **reload it** in order for the changes to take effect.

Then click "Report Check System" again and see if both formerly "Unexecutable" red messages now are green and "Executable".

Server	64bit	Not defined	Note: If the OS is 64bit but running a 32 bit version of php, the check will return (32 bit)	Only for linux/windows. Copy adapter/helpers/binaries/[linux windows]/bin64/pdfotext in adapter/helpers/binaries/[linux windows] WebSite: <a href="http://www.php.net/manual/en/install.php">http://www.php.net/manual/en/install.php</a>
<b>XPDF</b>				
XPDF Binaries File	Executable	CHMOD 0755	Permission XPDF Binaries File (adapter/helpers/binaries/linux/pdfotext): 0755	Used only for PDF file parser WebSite: <a href="http://www.foolabs.com">http://www.foolabs.com</a>
<b>ANTIWORD</b>				
ANTIWORD Binaries File	Executable	CHMOD 0755	Permission ANTIWORD Binaries File (adapter/helpers/binaries/linux/antiword): 0755	Used only for DOC file parser WebSite: <a href="http://www.winfield.demon.nl/">http://www.winfield.demon.nl/</a>


*XPDF and ANTIWORD are both executable*

The following step is **MOST IMPORTANT!** If you skip it, you will not be able to index anything!

Please check your "Server" settings. They can either be 32 or 64 bit. In our case the system is set to (capable of running) 64 bit.

So, please click on "Config" in JiFile's control panel and adjust the "Server bit" to the appropriate value. If you forget or skip this, you may not be able to index any file and may rather get an error message saying "Error recovery index: empty body".

Site	Users	Menus	Content	Components	Extensions	Help
------	-------	-------	---------	------------	------------	------



## JiFile [iFile configuration]

### Configuration

Root-application \*

Timelimit

Memorylimit(Mb)

Duplicate

Encoding  (Recommended: leave blank)

Server bit  64bit  32bit (Important for the indexing of PDF)

**Microsoft Word To Text**

Type  ANTIWORD works only with Windows, OSX and Linux to 32bit

At this step you do not need to change anything else. However, PLEASE READ THE SECTION "OPTIMIZATION OF PDF INDEXING" later in this manual.

Save the configuration.

Congratulations! Your basic settings are done and we may now turn to the indexing of files.

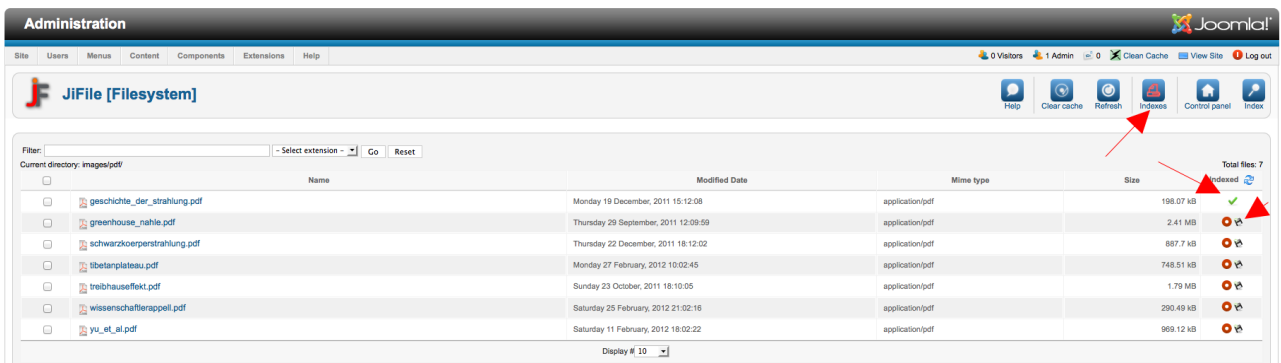
## *Indexing of Files*

In JiFile's control panel, please click on "Filesystem". This will open the screen showing you all the files inside the folder you defined under "Filesystem path" when you started to configure JiFile.

In this screen you have several options:

- Browse through/Go to subfolders
- Manual indexing of files
- Automatic file indexing





*The Filesystem screen*

- A green checkmark below "Indexed" tells you the file has been successfully indexed.
- Clicking on "Indexes" starts the automatic indexing of CHECKED files. **Note:** you can check all files at once. You can even check *subfolders* which will be completely and automatically indexed as well.
- Clicking on the little disk symbol next to a red dot starts manual indexing.

Please just "play around" with these options and see what fits your needs best.

## Checking the Indexes

After file indexing you may want to check the indexes. To do this, please click on "Index", either in the control panel or in the "Filesystem" screen.



*Index screen*

In this screen you have the following options:

- Click on the little "+" icon below "Detail". This will open another screen showing you the content of the indexed file.

- Delete one or several index files by CHECKING them.
- Delete ALL index files without checking them by clicking "Delete All".

**NOTE:** After deleting one or several index files the message next to "Optimization" will turn red, saying "Necessary". In this case it is **mandatory** to click on "Optimize" in order to optimize the index.

## *Optimization of PDF Indexing*

While in many cases the standard settings we applied before may be sufficient to fulfill your needs, in other cases you may not be satisfied with JiFile's *search results on the frontend*.

### **NOTE – MOST IMPORTANT:**

**If you do not find ANY of your documents (*PDFs*) on the frontend even though they were successfully indexed, please try the following before investigating anything else or turning to JiFile's support:**

- Create a test pdf containing ONLY ONE WORD. Ideally this is a completely *artificial* word not present anywhere else on your site – neither inside the "normal" text, nor in any document.
- Upload this test file (pdf)
- Index it
- Search for it at the frontend
- If THIS document is found, you can be sure that JiFile is WORKING on your site but you will have to optimize JiFile's settings and your indexes

Okay, let's do the optimization now. This, again, is *a bit* hard.

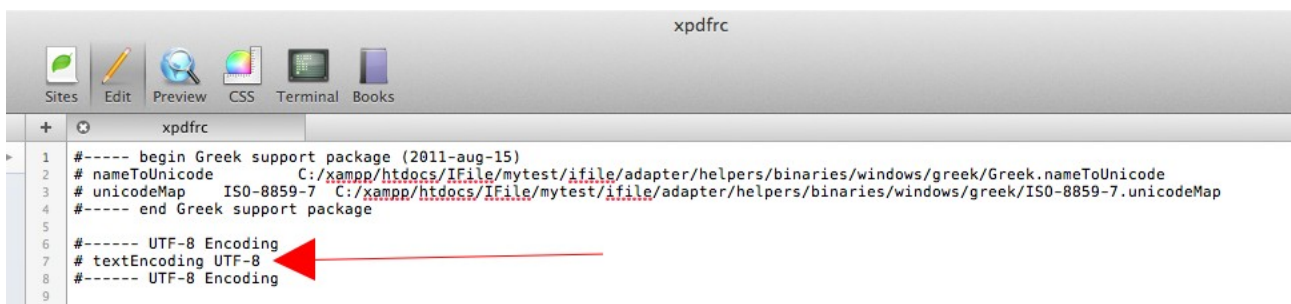
We will have to change the CONTENT of another file which is also found in ***/libraries/ifile/adapter/helpers/binaries***

It's found inside the folder named "xpdfrc". Open it and DOWNLOAD the file named "xpdfrc" to your local machine.

Now you will need an html (or php) editor. On a Mac, this would be something like "Coda" – please go and search (google) for it if you don't already have it.

Windows users: sorry, but I'm a Mac user ;-).

After opening "xpdfrc" in the appropriate editor, we will have to "uncomment" one single line, line #7:

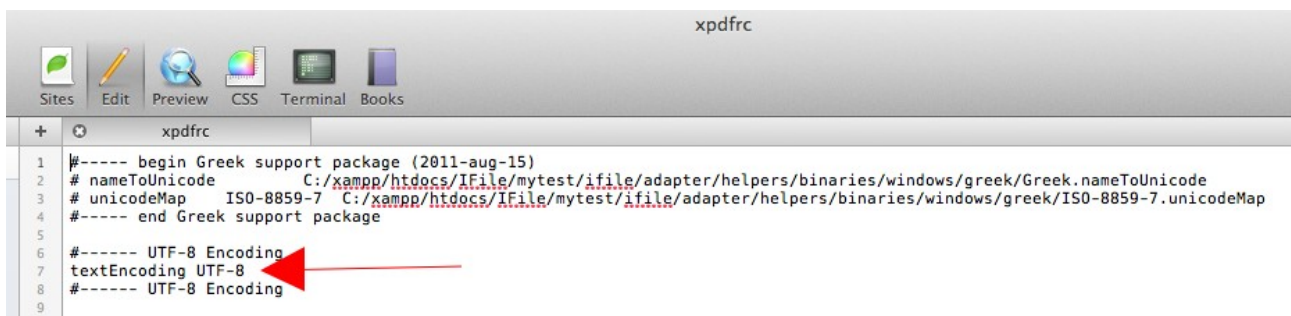


```
xpdfrc
1 #----- begin Greek support package (2011-aug-15)
2 # nameToUnicode C:/xampp/htdocs/IFile/mytest/iframe/adaptor/helpers/binaries/windows/greek/Greek.nameToUnicode
3 # unicodeMap ISO-8859-7 C:/xampp/htdocs/IFile/mytest/iframe/adaptor/helpers/binaries/windows/greek/ISO-8859-7.unicodeMap
4 #----- end Greek support package
5
6 #----- UTF-8 Encoding
7 # textEncoding UTF-8
8 #----- UTF-8 Encoding
9
```

*xpdfrc opened in editor, line #7 "commented"*

The "#" symbol next to any line in php tells you this is only a comment and thus NOT executable.

Okay, let's simply delete the "#" next to line #7 and xpdfrc is ready for optimization:



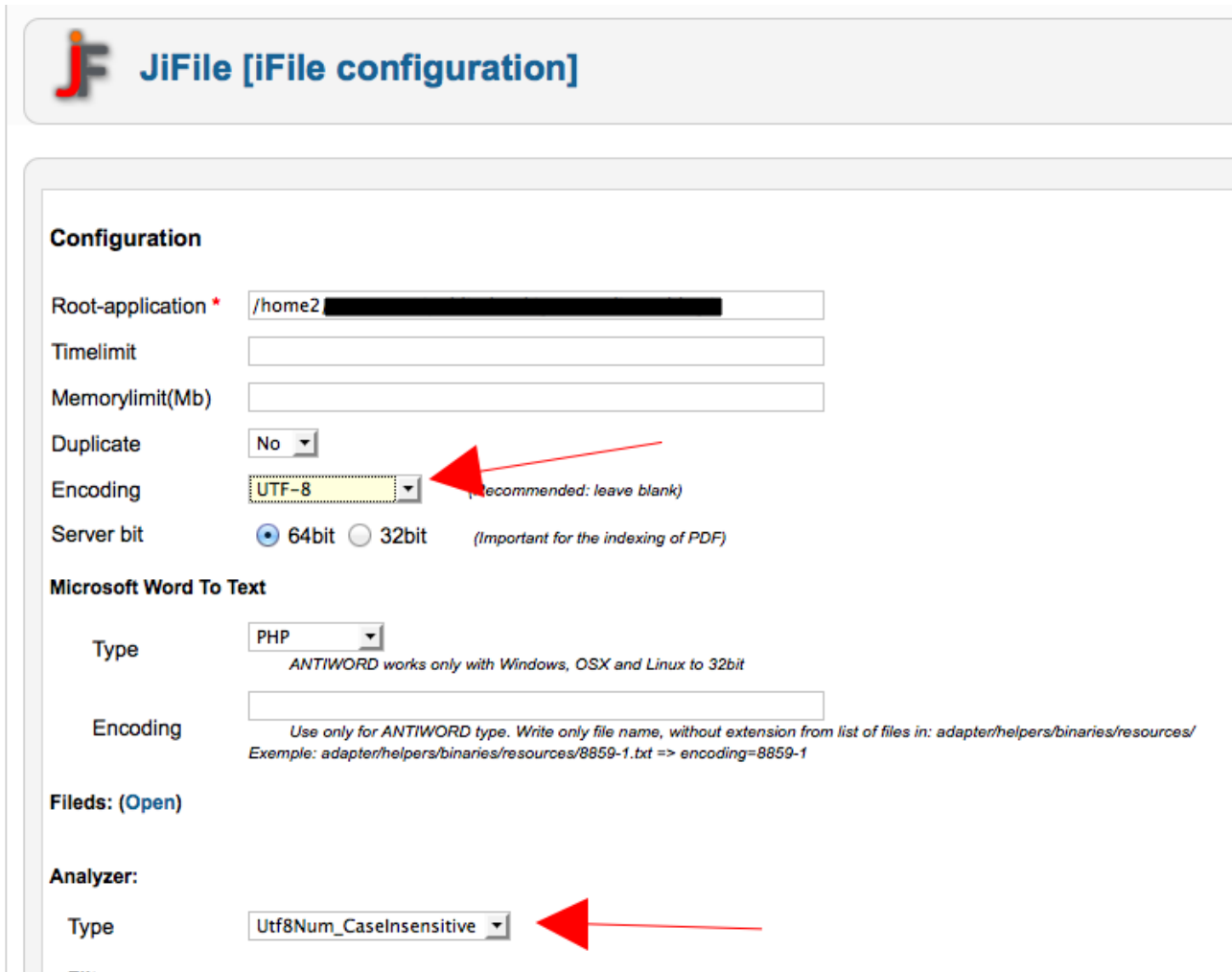
```
xpdfrc
1 #----- begin Greek support package (2011-aug-15)
2 # nameToUnicode C:/xampp/htdocs/IFile/mytest/iframe/adaptor/helpers/binaries/windows/greek/Greek.nameToUnicode
3 # unicodeMap ISO-8859-7 C:/xampp/htdocs/IFile/mytest/iframe/adaptor/helpers/binaries/windows/greek/ISO-8859-7.unicodeMap
4 #----- end Greek support package
5
6 #----- UTF-8 Encoding
7 textEncoding UTF-8
8 #----- UTF-8 Encoding
9
```

*xpdfrc opened in editor, line #7 "Uncommented" and thus executable*

**SAVE** and **UPLOAD** this modified file to the "xpdfrc" folder on your server.

NOW let's turn to JiFile's control panel again. Open "Config" and change "Encoding" to "UTF8" and "Analyzer Type" to "UTF8Num\_CaseInsensitive" or any other "UTF8" setting you like. But it

MUST be "UTF8" because encoding the index with "UTF8" and reading from such a "UTF8" encoded index is KEY to the whole optimization process.



The screenshot shows the 'JiFile [iFile configuration]' interface. Under the 'Configuration' section, the 'Encoding' dropdown is set to 'UTF-8', highlighted with a red arrow. Below it, the 'Server bit' section has '64bit' selected. In the 'Microsoft Word To Text' section, the 'Type' is set to 'PHP'. Under 'Files: (Open)', the 'Analyzer:' section has 'Type' set to 'Utf8Num\_CaseInsensitive', also highlighted with a red arrow.

*JiFile Config set to optimized PDF indexing and searching*

NOW, if you already indexed your files and the indexes didn't work for you, you will have to DELETE ALL indexes and REBUILD them with the new settings.

After that, your frontend search should find all files (pdfs) indexed this way.

### *Known Issues, Drawbacks, and Solutions*

Currently JiFile has one drawback: it breaches any "anti-leaching" technology for file downloading, such as provided by e. g. "Akeeba Release System", or "Docman". Normally, files managed by

these systems cannot be accessed directly and may thus be hidden from the public, i. e., they may only be visible and accessible by a restricted group of users like "registered", or alike.

Since JiFile, via index search, points directly to any file present in its index, each file found by such a search is completely accessible and thus downloadable by anyone, i. e., by the public.

Presently, there is NO direct solution to this. The JiFile developers promise to address this issue in a future update.

The only solution I could think of therefore is an indirect protection of files via password-protection: protect the files you want to exclude from public access and deliver the password to the exclusive group of users you want to address, hidden from the public.

However, since JiFile cannot index password-protected pdfs, we have to find a way to circumvent this. This is rather easy:

- Upload and index the unprotected file
- Then simply replace the file with the protected file of IDENTICAL file name

JiFile will then show that the file is not indexed which is correct, of course, since this specific password-protected file never was indexed.

However, since the index has already been created from the unprotected file it will find and point to the *protected* file.

During my work on 500+ different files with JiFile it sometimes said "File already in the index" although the file NAMES were different, as shown in the following image.

AR5/IPCC_GreyLiterature-GN 2.pdf	(251.08 kB)	File already in the index <a href="#">Update!</a>	00:00:01	✘
AR5/IPCC_Meeting_Report_Noordwijkerhout.pdf	(2.06 MB)	Indexing completed	00:04:03	✔
AR5/IPCC_p33_Decision_error_protocol.pdf	(344.5 kB)	Indexing completed	00:00:06	✔
AR5/IPCC_p34_Decision_COI.pdf	(66.09 kB)	Indexing completed	00:00:04	✔
AR5/IPCC_p34_Decision_COI 2.pdf	(66.09 kB)	File already in the index <a href="#">Update!</a>	00:00:01	✘
AR5/IPCC_RoleofReviewEditors.pdf	(216.1 kB)	Indexing completed	00:00:02	✔
AR5/IPCC_RoleofReviewEditors_Rev.pdf	(214.51 kB)	Indexing completed	00:00:02	✔
AR5/IPCC_Rules&Procedures.pdf	(75.83 kB)	Indexing completed	00:00:06	✔
AR5/IPCC_Rules&Procedures 2.pdf	(75.6 kB)	Indexing completed	00:00:04	✔
AR5/IPCC_Uncertainties-GN_brochure.pdf	(6.17 MB)	Indexing completed	00:00:03	✔
AR5/IPCC_Uncertainties-GN_brochure 2.pdf	(6.72 MB)	Indexing completed	00:00:03	✔
AR5/IPCC_Uncertainties-GN_brochure 3.pdf	(6.72 MB)	File already in the index <a href="#">Update!</a>	00:00:01	✘
AR5/IPCC_WoSES_Report_final.pdf	(583.57 kB)	Indexing completed	00:00:30	✔

First I thought this was an error. But then I checked the files via direct content comparison and I noticed they were in fact identical. So it seems that JiFile is very smart and actually compares the files' contents rather than only their file names. Great feature!!

I also noticed that JiFile would not index *Excel* files when set to UTF8 encoding. Also, some MS Word files were not indexed with "Antiword".

Solution: convert those files to pdfs and index them. There are several batch converters on the net doing this job for you with ease.

For Mac users here's the easiest and free solution: <https://bitbucket.org/codepoet/cups-pdf-for-mac-os-x/wiki/Home>

Depending on which OS you are using, setting up of cups pdf may be a little difficult. With Mountain Lion it's a breeze: run the installer and follow the instructions given in the wiki. Then, from System Preferences – Print and Scan – simply drag the cups pdf printer icon to your dock. Now you have a "desktop printer". Drag and drop any amount of files you like onto that printer icon and the Mac will do the rest for you.

Another error message you may encounter is: "Text of body not indexing. Check the type of encoding".

In this case the pdf actually does not contain any text but only images; at least this is what I found out.

If, however, the error message says: "Error recovery document. Empty body" you can be sure that your settings are wrong (most likely you didn't set the "Server bit" to the correct value). In this case, please see above.

That's currently all, "folks".

I hope this little manual is of some value to you and you do enjoy JiFile now.

It's well worth every effort!