Search Indexing tab

Overview

Use this tab to configure specific MIME types in order to search within files (such as PDF files) that have been uploaded to the Tiki File Gallery.

Related Topics

Search within files

To Access

From the File Gallery Admin page, click the **Search Indexing** tab.

Note

In order to search within uploaded files, your server may require additional applications, such as **strings** or **pdftotext**.

		,
Option	Description	Default
Automatic indexing of file content	Uses command line tools to extract the information from the files based on their MIME types.	Disabled
Automatic indexing of emails stored as files	Parses message/rfc822 types of files (aka eml files) and stores individual email headers and content in search index.	Disabled
Asynchronous indexing		Enabled
OCR Files	Extract and index text from supported file types.	Disabled
OCR Every File	Attempt to OCR every supported file.	Disabled
Allow file level OCR languages	Allow users to change the default languages that will be used to OCR a file.	Enabled
DCR limit languages	Limit the number of languages one can select from this list. E Auto detect languages Afrikaans (Afrikaans) Albanian (Shqip) Amharic (አማáˆáŠ›) Arabic Arabic (العربية) Armenian Armenian (Õ€ÕiÕμÕ¥Ö€Õ¥Õ¶) Assamese (à¦à¦¸à¦®à§€à¦¯à¦¼à¦¾) Azerbaijani (azÉ™rbaycan dili) Azerbaijani (azÉ™rbaycan dili) (cyrl) Basque (euskara, euskera) Belarusian (бĐμĐ»Đ°ÑƒÑĐ⁰аN Đ¼Đ¾Đ²Đ°) Bengali Bengali (বাà¦,লা) Bosnian (bos	None
tesseract path	Path to the location of the binary. Defaults to the \$PATH location. If blank, the \$PATH will be used, but will likely fail with scheduler.	/usr/bin/tesserac
		,

Path to the location of the binary. Defaults to the \$PATH location. If blank, the \$PATH will be used, but will likely fail with scheduler. Description	Pdfimages	
Description		
	Default	
Uses command line tools to extract the information from the files based on their MIME types.	Disabled	
Parses message/rfc822 types of files (aka eml files) and stores individual email headers and content in search index.	Disabled	
	Enabled	
Extract and index text from supported file types.	Disabled	
Attempt to OCR every supported file.	Disabled	
Allow users to change the default languages that will be used to OCR a file.	Enabled	
Limit the number of languages one can select from this list. \equiv Auto detect languages Afrikaans (Afrikaans) Albanian (Shqip) Amharic (አá^ \hat{a}) Arabic Arabic (اÙ,,عربية) Armenian Armenian (Õ€ÕiÕµÕ¥Ö€Õ¥Õ¶) Assamese (à $ $ à $.$ à $.$ ®à§€à $.$ ā $.$ ³¼à $.$ 3¾) Azerbaijani (azÉ Tbaycan dili) Azerbaijani (azÉ tbaycan dili) (cyrl) Basque (euskara, euskera) Belarusian (белаÑ $.$ Ñ $.$ D°Đ°Ñ $.$ D° $.$ 0% D $.$ 4 $.$ 2°D°) Bengali Bengali (à $.$ 7à $.$ 3¼à $.$ 34 $.$ 3¾) Bosnian (bos	None	
Path to the location of the binary. Defaults to the \$PATH location. If blank, the \$PATH will be used, but will likely fail with scheduler.	/usr/bin/tesseract	
Path to the location of the binary. Defaults to the \$PATH location. If blank, the \$PATH will be used, but will likely fail with scheduler.	Pdfimages	
Description	Default	
Uses command line tools to extract the information from the files based on their MIME types.	Disabled	
Parses message/rfc822 types of files (aka eml files) and stores individual email headers and content in search index.	Disabled	
	Enabled	
Extract and index text from supported file types.	Disabled	
Attempt to OCR every supported file.	Disabled	
Allow users to change the default languages that will be used to OCR a file.	Enabled	
	Uses command line tools to extract the information from the files based on their MIME types. Parses message/rfc822 types of files (aka eml files) and stores individual email headers and content in search index. Extract and index text from supported file types. Attempt to OCR every supported file. Allow users to change the default languages that will be used to OCR a file. Limit the number of languages one can select from this list. Extract and index text from supported file. Allow detect languages Afrikaans (Afrikaans) Albanian (Shqip) Amharic (ጠá⁻ゝá⁻ጛ) Arabic Arabic (Øśū,ø¹ð⊕ð°0 g) Armenian Armenian (ô€ôì∂µðvô€ôvôð) Assamese (a¡a¡, aḥ @àĕçaì⁻ai,¼aḥa) Azerbaijani (azē¹™rbaycan dili) (cyrl) Basque (euskara, euskera) Belarusian (ð⊕ðþæðþ°Ñeðþ°Ñ ðþæðþ²ðþ°) Bengali Bengali (a¡¬aḥ¼aḥa)²a)²a Bosnian (bos Path to the location of the binary. Defaults to the \$PATH location. If blank, the \$PATH will be used, but will likely fail with scheduler. Path to the location of the binary. Defaults to the \$PATH location. If blank, the \$PATH will be used, but will likely fail with scheduler. Description Uses command line tools to extract the information from the files based on their MIME types. Parses message/rfc822 types of files (aka eml files) and stores individual email headers and content in search index. Extract and index text from supported file types. Attempt to OCR every supported file.	

Option	Descript	Default	
OCR limit languages	Limit the	None	
esseract path	Path to the state of the state	/usr/bin/tesserac	
odfimages path	Path to the state of the state	Pdfimages	
Option		Description	Default
Automatic indexing of file content		Uses command line tools to extract the information from the files based on their MIME types.	Disabled
Automatic indexing of emails stored as files		Parses message/rfc822 types of files (aka eml files) and stores individual email headers and content in search index.	Disabled
Asynchronous indexing			Enabled
OCR Files		Extract and index text from supported file types.	Disabled
OCR Every File		Attempt to OCR every supported file.	Disabled
Allow file level OCR languages		Allow users to change the default languages that will be used to OCR a file.	Enabled
OCR limit languages		Limit the number of languages one can select from this list. Auto detect languages	None
esseract path		Path to the location of the binary. Defaults to the \$PATH location. If blank, the \$PATH will be used, but will likely fail with scheduler.	Tesseract
odfimages path		Path to the location of the binary. Defaults to the \$PATH location. If blank, the \$PATH will be used, but will likely fail with scheduler.	/usr/bin/pdfimages
Option	Description		efault
Automatic indexing of file content	Uses command line tools to extract the information from the files based on their MIME types.		sabled
Asynchronous indexing		Er	nabled
OCR Files	Extract and index text from supported file types.		sabled
OCR Every File	Attempt to OCR every supported file.		sabled
Allow file level OCR languages	Allow users to change the default languages that will be used to OCR a file. Enabl		nabled

DCR limit languages	Limit the number of languages one can select from this list.	None
tesseract path	Path to the location of the binary. Defaults to the \$PATH location. If blank, the \$PATH will be used, but will likely fail with scheduler.	sh: 1: where: not found
odfimages path	Path to the location of the binary. Defaults to the \$PATH location. If blank, the \$PATH will be used, but will likely fail with scheduler.	sh: 1: where: not found

Default

Description

Option