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.

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**. It also relies on utilities on the server, more information on searching within files at Search-within-files. To check support on your server, use the admin tool, Tiki Check.*

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.	None
	☰ Auto detect languages | Afrikaans (Afrikaans) | Albanian (Shqip) | Amharic (አማርኛ) | Arabic | Arabic (العربية) | Armenian | Armenian (Հայերէն) | Assamese (অসমীয়া) | Azerbaijani (azərbaycan dili) | Azerbaijani (azərbaycan dili) (cyrl) | Basque (euskara, euskera) | Belarusian (беларуская мова) | Bengali | Bengali (বাংলা) | Bosnian (bos...	

Option	Description	Default
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/tesseract
pdfimages 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	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 | Afrikaans (Afrikaans) | Albanian (Shqip) | Amharic (አማርኛ) | Arabic | Arabic (العربية) | Armenian | Armenian (Հայերէն) | Assamese (অসমীয়া) | Azerbaijani (azərbaycan dili) | Azerbaijani (azərbaycan dili) (cyrl) | Basque (euskara, euskera) | Belarusian (беларуская мова) | Bengali | Bengali (বাংলা) | 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/tesseract
pdfimages 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