

Медиа>Индексация

Ключ расширения: **dam_index**

Авторские права 2003-2005, René Fritz, <r.fritz@colorcube.de>

Этот документ опубликован под Open Content License
доступной на <http://www.opencontent.org/opl.shtml>

Содержимое этого документа относится к TYPO3
- GNU/GPL CMS/Framework доступны на www.typo3.com

Содержание

Медиа>Индексация	1	Управление	2
Введение	1	Список To-Do	3
Руководство пользователя	1	Список изменений	3

Введение

Этот модуль — часть системы доступа к цифровому содержимому (DAM) в TYPO3.

Модуль индексации может индексировать файлы и директории с поддиректориями (в отличие от модуля Файл, который индексирует лишь единственный файл после его загрузки).

Возможности:

- массовое индексирование файлов и папок
- применение правил при индексации (например, назначение категории по названию папки расположения файла)
- предустановка дополнительных данных перед индексацией
- установка сохраняемых дополнительных данных, которые не должны быть перезаписаны в процессе индексации
- сохранение настроек индексации

Особая возможность: правила индексации могут быть расширены своими расширениями. Скажем, названия ваших файлов имеют специальный формат, который позволяет легко их категоризировать. Это можно оформить в собственное правило индексации.

Важно

Для получения дополнительных данных из файлов в DAM, необходимо установить сервисы, типа metaExtract, textExtract и textLang.

Руководство пользователя

Модуль индексации находится внутри основного модуля Медиа или в его подмодуле Файл, это зависит от настроек.

Индексация файлов проходит в 4 этапа, которые интуитивно понятны.

Indexing

fileadmin/: dam/ Indexing ▾

OPTIONS

1 **2** 3 4 Back Next

Options:

Index sub folders

Categorize files from folder names
 Categories will be attached to files if a category has the same name as the files folder.
 Use fuzzy folder/category comparison.

Reindexing
 Reindex files and update meta data.
 Only get meta data which is still missing.
 Overwrite current meta data with data from file. Will overwrite the meta data but preserve those where's no data available from the file itself.

Indexing

fileadmin/: dam/ Indexing ▾

INDEXING FIELD PREDEFINITION

1 2 **3** 4 Back Next

Preset meta descriptions for files.
 The content may be replaced by meta content from the files themselves.

Check the fields you want to set **fixed**.
 The input will then not be replaced.

Title:

Keywords (,):

Description:

Source / Original location:

Source / Original location description:

Ident key (sku):

Управление

В процессе установки расширения можно настроить, где во внутреннем интерфейсе будет расположен модуль.

CONFIGURATION:

(Notice: You may need to clear the cache after configuration of the extension. This is required if the extension adds TypoScript depending on these settings.)

Enable features

Media>File>Indexing [add_media_file_indexing]
 Add submodule function "Indexing" to Media>File.

 Default: 1

Media>Indexing [add_media_indexing]
 Add module "Indexing".

 Default: 0

Update

Однако, доступ к модулю можно получить через Медиа>Индексация или Медиа>Файл>Индексация. Это должно быть сделано в записи внутреннего пользователя или их группы, членом которой он является.

Для получения дополнительных данных из файлов в DAM, необходимо установить сервисы, типа metaExtract, textExtract и textLang.

Indexing				
Info				
AVAILABLE SERVICES FOR INDEXING:				
Indexing needs the help of some services to extract meta data or read text content from the files.				
Used service types are: metaExtract - get meta data from files. textExtract - get text content out of files. textLang - detect the language of text content.				
Services:	Types:	OS:	External:	Avail.:
 metaExtract (Service Type) Read meta data from files.				
EXIF extraction [tx_ccmetaexec_sv1] Extract EXIF data from images using external program "exiftags".	image:exif		exiftags	✓
PDF meta extraction [tx_ccmetaexec_sv2] Extract meta data from PDF files using external program "pdffinfo".	pdf		pdffinfo	✓
Services:	Types:	OS:	External:	Avail.:
 textExtract (Service Type) Extract pure text out of files like doc, pdf, rtf, ...				
Text extraction for txt, html, sgml, xml [tx_cctxtextphp_sv2] This service depends on PHP only.	txt, html, sgml, xml			✓
Text extraction for pdf [tx_cctxtextexec_sv1] This service depends on pdftotext	pdf		pdftotext	✓
Text extraction for Word documents (doc) [tx_cctxtextexec_sv2a] This service depends on catdoc	doc, dot		catdoc	✓
Text extraction for Excel documents (xls) [tx_cctxtextexec_sv2b] This service depends on xls2csv	xls, xlt, xlw		xls2csv	✓
Services:	Types:	OS:	External:	Avail.:
 textLang (Service Type) Detects languages by example text.				
Text-Cat [tx_cctxtextcat_sv1] TextCat is an implementation of the text categorization algorithm "N-Gram-Based Text Categorization" in Perl.			perl	✓

Список To-Do

Основное

- Возможность: авто/фоновая индексация новых файлов
- Возможность: руководство (пользователей и разработчиков)

Список изменений

Текущая версия

- улучшен процесс индексации: индексация отображается внутри невидимого фрейма и перезагружается каждые 20 секунд, дополнительно отображается индикатор выполнения. Это предоставляет обратную связь и процесс не прерывается из-за ограничения времени выполнения сценария PHP.
- Надлежащая поддержка exim и iptc