TYPO3

IndexedSearch Crawler for tt_news

The content of this document is related to TYPO3
- a GNU/GPL CMS/Framework available from www.typo3.org

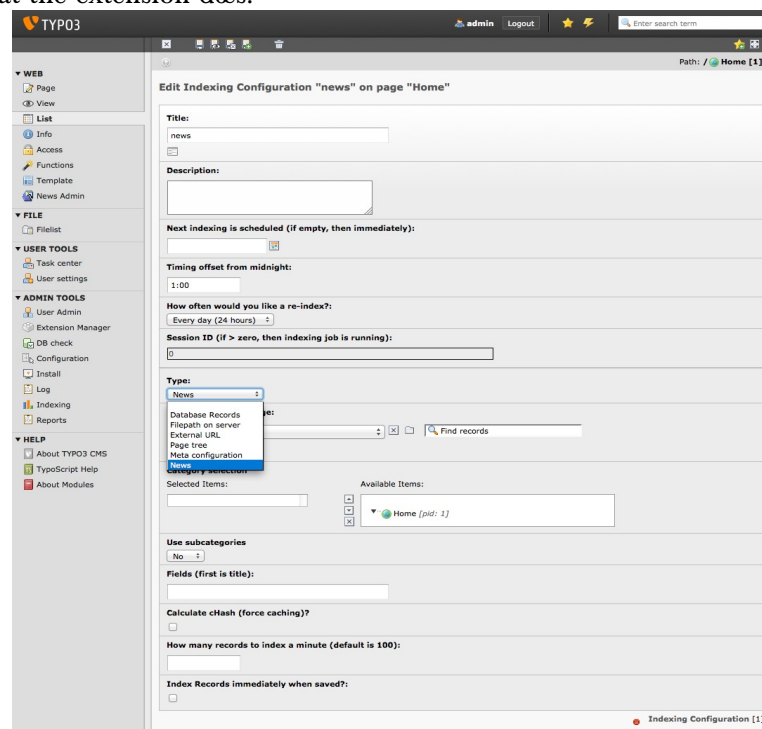# Table of Contents

Introduction

What dœs it do?

It gives you the possibility to index tt_news entries using crawler and indexed_search.

The benefit of this extension, over choosing the build-in option of indexed_search (choosing "Database Records" as type and "News" as table to index) is, that you can filter the news, to be indexed, by categories.

For example if you have two main-categories, that are stored in one storage-folder and you have a single-view for each of them, this extension will fit your needs.

Screenshots

Here you see what the extension dœs:

It adds a new type to the Indexing Configuration called "News" which let's you set several settings specially for news. The settings you don't see here are called "Alternative Source Page" and "Category selection". All functionalities are described in the Administration section of this document.

Users manual

No user manual since you have to configure this extension once and let it do it's work.

Administration

You probably already read the documentation of indexed_search. Following that, you can create Indexing Configuration at pages where you have extended content that should be indexed.

In this case, we're indexing news. Please create an Indexing Configuration on every page you show the news on as SingleView. The pid of the Indexing Configuration Record will be used as id in the generated url for the news.

If you choose "News" in the type-field of an Indexing Configuration, the following additional fields will appear:

- Alternative Source Page
  The pool where your news-entries are saved in. If you have multiple you should create one for each source page on this page.

- Category selection
  Choose the categories you want to index here.

- Use subcategories
  This feature has been added that you don't have to add every category to the record, you add in the system. You can just add it's parent-category and all child-categories can be taken as well.

- Fields
  A comma-separated list of database-columns that will be pushed to the crawler. The content of the first column will be used as title. Prefered value: title, bodytext

The other fields (Calculate cHash, How many records to index in a minute, Index Records immediately when saved) have the same functionalities as the one at the type Database Records.

Configuration

Don't know what to write here ... drop me a mail if you're missing something here and I will update the manual. I don't have special configuration or any TypoScript.
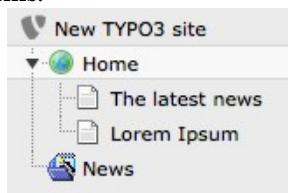
There's just one hook you can use:

```
$GLOBALS['TYPO3_CONF_VARS']['EXTCONF']['indexedsearch_ttnews_crawler']
['indexSingleNewsRecord']
```

It is thought for manipulating a document before it get's send to the indexer. This is for example used by the extension indexedsearch_rendered_ttnews, which replaces the column "bodytext" by what is shown if you'd show the news-entry on your website.

Tutorial

First you have to think about where your news are stored (Alternative Source Pages) and where your news are displayed on the page.

I, for example, have a structure like this:

In addition, I also have three news-categories.

Two called "latest" and "TYPO3". All news in here are displayed on the page "The latest news", the LIST and the SINGLE view. But just the fact, that the SINGLE view is on this page, is of importance here.

The third category is called "lorem ipsum". The news linked to this category are visible on the page "Lorem Ipsum".

This said, I create an Indexing Configuration at the page "The latest news" and fill it with the following data:

Property:	Value:	Reason:
Title	News – Latest and TYPO3	Because I like the name :) This is not used by any logic.
Type	News	To use this extension :)
Alternative Source Page	3	Id of the page where my news are stored (storage-folder called "News")
Category selection	latest, TYPO3	Selected the categories who'se single-view is on "The latest news". No selection will not index any news!
Use subcategories	Yes	Include all news, linked to a subcategory of one of the categories chosen in "category selection".
Fields	title,short,bodytext	A list of database-columns I want to have indexed. The first value is the title. In this case I want to have the news-title as title and the short- and bodytext as content of the element that can be found in the search.
Index Records immediately when saved?	Yes	I want my records to be added to the index when I save them. This lets new news-entries turn up in the search as soon as they're published. Same appeals when I update the content. (Functionality provided by crawler)

After that, I create another Indexing Configuration at the page "Lorem Ipsum" having nearly the same configuration (attention to the category-selection):

Property:	Value:	Reason:
Title	News – Lorem Ipsum	See above ...
Type	News	See above ...
Alternative Source Page	3	See above ...
Category selection	lorem ipsum	Selected the categories who'se single-view is on "Lorem Ipsum". No selection will not index any news!
Use subcategories	Yes	See above ...
Fields	title,short,bodytext	See above ...
Index Records immediately when saved?	Yes	See above ...

Known problems

Not that I know of...

To-Do list

- The crawler should get an information when a news-entry is deleted or hidden. Currently all news will be added, but never removed from the index.

ChangeLog

Version:	Changes:
2.0.1	Added documentation and icon
2.0.0	Complete code-rewrite.
1.0.0	Wrote the code of the extension.