

검색 색인

티키의 검색 엔진은 티키7 을 위하여 다시 제작되었으며 본 문서는 버전 7 및 그 이후에만 적용됩니다.

티키에서의 검색 능력은, **PluginList** 혹은 통합 검색 등에서 사용되며, 검색 색인에 의존합니다.

티키는 내부적으로 여러 개의 검색엔진들을 지원할 수 있습니다. 각각의 검색엔진은 다른 능력과 제한을 가지게 될 것입니다. 기본 엔진은 중소규모의 사이트에 대하여 충분한 능력을 제공하여야 합니다. 대형 사이트는 최대한의 성능을 얻기 위하여 추가 인프라스트럭처가 필요할 수도 있습니다.

Lucene (PHP 적용)

기본 구현은 Zend_Search_Lucene 에 기반하고 있으며, 이는 자바의 PHP 구현입니다 Lucene index engine. 엔진은 외부 의존도가 없으며 모든 호스트에서 실행될 수 있습니다. 하지만, 수용할만한 성능에 도달하기 위해서는 몇몇 구성설정이 요구될 수도 있습니다.

완벽한 색인을 빌드하기 위하여 필요한 시간은 사이트의 콘텐츠에 따라 다양할 것입니다. 참고로,

- doc.tiki.org (이 사이트), 는 1400개가 넘는 페이지가 있으며, 재인덱싱에 3분 정도 걸립니다
- themes.tiki.org, 는 몇몇 페이지와 수백개가 넘는 게시판 게시물이 있으며, 재인덱싱에 15초 정도 걸립니다
- 회사 인트라넷으로써 파일 갤러리에 수 기가바이트의 데이터는 한 시간 넘게 걸릴 수 있습니다

하지만, 증분적 색인이 사용될 수 있으며, 이 경우 완전 재인덱스는 특정 기간마다 혹은 업그레이드 시에만 필요합니다.

Alexander Veremy, 는 컴포넌트의 저자로서 변수를 조절에 대한 몇몇 이해 을 제공하고 있습니다.

필드

아래는 필드와 개체 유형간의 행렬입니다.
범례:

- 토큰화됨, 전체 텍스트 검색을 위한 단어들로 분해된 방식으로..
- 정렬가능

	토큰 화됨	정렬 가능	위키 페이 지	게시판 계 시물	블로그 계 시물	기사	파일	트래 커항 목	시트	댓글	사용 자	제공되는 터키 버전
Generic												
object_type		X	X	X	X	X	X	X	X	X	X	7
object_id		X	X	X	X	X	X	X	X	X	X	7
title	X	X	X	X	X	X	X	X	X	?	X	7
language		X	X	/	/	X	/	?	/	/		7
modification_date		X	X	X	X	X	X	X	X	X		7
contributors			X	X	X	X	X	X	X	X		7
description	X		X			X	X		X			7
contents	X		X	X	X	X	X	X	X	X	X	7
Specific												
wiki_content			X								X	7
wiki_uptodateness		X	X									7
wiki_approval_state			X									11
post_content				X						X		7
parent_thread_id (not to be confused with parent_object_id)				X								8
blog_id		X			X							7
blog_excerpt					X							7

blog_content				X						7
topic_id		X				X				7
article_content						X				7
article_topleftine						X				7
article_subtitle						X				7
article_author						X				9
article_type						X				9
gallery_id		X					X			7
filename_id		X					X			7
filetype	X	X					X			7
file_comment							X			7
file_content							X			7
tracker_id		X						X		7
tracker_status		X						X		7
tracker_field_ID	X	?						X		7
sheet_content								X		7
comment_content	X								X	7
user_country	X	X							X	10
internal										
view_permission			X			X		X		7
parent_object_type	X			X	X	X	X	X	X	7
parent_object_id	X			X	X	X	X	X	X	7
parent_view_permission				X	X	X	X	X	X	7
global_view_permission									X	7
hash			X							7
url			X						X	7
Global										
categories			X	X	X	X	X	X	X	7
deep_categories			X	X	X	X	X	X	X	7
allowed_groups			X	X	X	X	X	X	X	7

X 제공됨
/ 정적값
? 데이터에 의존적

트래커 필드에 대한 색인은 필드 종류에 따라 다양할 것입니다. 일반적인 규칙으로써, `tracker_field_ID` 가 필드로써 사용될 것이며 정렬 가능할 것입니다. 하지만, 몇몇 예외가 있습니다:

- 이미지와 파일 필드는 색인되지 않습니다
- 텍스트영역(TextArea) 은 정렬 가능하지 않습니다
- 다국어 필드는 복수 필드로 색인됩니다:
 - 보통의 경우 모든 언어를 포함합니다
 - `tracker_field_ID_lang` 은 하나의 언어만을 포함합니다 (예들 들면 `tracker_field_12_fr` 처럼)
- 등급과 관련 필드는 복수의 필드를 저장합니다
 - `tracker_field_ID` 는 평균을 포함합니다
 - `tracker_field_ID_sum` 은 투표 총합을 포함합니다
 - `tracker_field_ID_count` 는 투표수를 포함합니다
- 항목목록과 항목 링크 필드
 - `tracker_field_ID_text` 는 ID 대신에 텍스트를 포함합니다
- 만약 언어 필드가 트래커 항목에 대하여 설정되어있다면, 해당 언어는 아이템 언어로 색인됩니다.

buildQuery/tiki-searchindex.php 에 사용되는 일부(이들이 실제 필드인지 혹은 도우미인지에 대한 설명 필요):

type: 는 **object_type** 를 참조합니다

deep: 만약 이것이 설정되었다면, **categories** 나 **deep-categories** 로 고려될 것입니다

autocomplete: 은 해당내용으로 시작되는 제목을 가진 항목들을 찾을 것입니다

검색 색인 재빌드

색인은 temp/unified-index/ 에 저장됩니다. 재빌드가 진행되는 동안, 디렉터리 temp/unified-index-new/ 가 나타날 것입니다. (이는 새 것이 준비될 때까지 기존의 색인이 사용되는 것을 허용하기 위함입니다) 만약 temp/unified-index-new/ 가 색인작업의 종료 후 사라지지 않는다면, 무엇인가 잘못 된 것입니다. 이를 삭제하고 색인을 다시 시도해 보실 수 있습니다. 권한이 모두 OK 이라는 확신을 가지려면 "sh setup.sh" 를 실행하실 수도 있습니다

티키 인터페이스에서

다음 url 을 방문하실 수 있습니다: **tiki-admin.php?page=search&rebuild=now** 그리고 사이트가 소규모라면 검색 색인이 다시 빌드 될 것입니다. 로드가 중급에서 높은 사이트라면, 명령어에서 진행하실 수 있습니다.

명령어에서

검색 색인은 다음 명령어를 사용하여 명령입력에서 재빌드될 수 있습니다:

```
cd path_to_tiki/; php lib/search/shell.php rebuild;sh setup.sh -n
```

위 명령어를 cron job 에 위치하여 트키 표준 사이트에서 보다 적은 사용자가 활동적인 밤 시간에 진행되도록 할 수 있습니다. Cron 을 살펴
해보십시오.

문제가 발생하였고 기록을 남기고 싶으시다면,

```
php lib/search/shell.php rebuild log
```

이 기록(로그)은 temp/Search_Indexer.log 에 있을 것입니다.
관련:

- PluginList
- PluginCustomSearch
- Unified Search