

Lucene

Lucene Jumping Off Point

This page will provide links to various Lucene pages in this wiki. More information about Lucene can be found at their website <http://lucene.apache.org>

It should be noted that rather than using Lucene in-process the preferred solution nowadays is to use a separate SolR server.

Nutch used to be a Lucene sub project but became a top level project in 2010.

Lucene dynamic attributes

Torsten Krah asks:

```
> pre 1.0 Days, it was possible to have dynamic attributes in lucene, because
> the API let you do such things (Lucene document access).
>
> How to do the same in 1.0> - using 1.1 the API the NutchDocument does only
> know name and value, but if i don't know the name (dynamic attribute via
> HtmlParser, meta tags indexing) - how can i still index them? Or is this
> impossible with the lucene backend now?
```

Andrzej Bialecki replies:

It's still possible to do this, but it's undocumented...

Here's a quick howto: in your [IndexingFilter](#), whenever you want to add a previously undeclared field you need to declare its Lucene options on a per-document level like this:

```
String fieldName = "myMetaField";
String value = "undeclared meta value";
Metadata meta = nutchDocument.getDocumentMeta();
meta.add(LuceneConstants.FIELD_PREFIX + fieldName,
LuceneConstants.STORE_YES);
meta.add(LuceneConstants.FIELD_PREFIX + fieldName,
LuceneConstants.INDEX_TOKENIZED);
//... etc, add those field options that you want
// and add the field value
nutchDocument.add(fieldName, value);
```