The lucene Module

Table of contents

1. Introduction

The `lucene` module integrates the Lucene search engine.

2. Configuration

Each publication has a configuration at `$PUB_HOME/config/search/lucene_index.xml` which is being defined by `org.pache.cocoon.components.search.components.impl.IndexManagerImpl`.

3. Indexing

Indexing is typically triggered by the `org.apache.lenya.cms.lucene.IndexUpdaterImpl`, which attaches itself as a listener to the repository on startup.

Indexing can also be executed explicitly by calling the `lucene.index` usecase which is defined by `config/cocoon-xconf/usecase-lucene.index.xconf` and uses as main entry point `org.apache.lenya.cms.lucene.IndexDocument`.

In order to make a resource type indexable one needs to add the format luceneIndex to the resource type configuration (e.g. `src/modules/xhtml/config/cocoon-xconf/resource-type-xhtml.xconf`) One needs to create or reuse a pipeline for this format within the specified sitemap.

4. Indexing Issues

When the system is under high load, it can occur that a document is not indexed after a change because the indexer is busy. In this case, a notification message is sent to a user which can be configured in `modules/lucene/sitemap.xmap`:

```
<map:transformer name="index2"
    logger="sitemap.transformer.luceneindextransformer2"
    src="org.apache.cocoon.transformation.LuceneIndexTransformer2">
  <notify user="lenya"/>
</map:transformer>
```

5. Searching

To be documented.