

## Blog

# Alfresco – Indexing Document Metadata Only – Confirmation with Luke

by jhsu | Mar 5, 2014 | Alfresco, Input Management | 0 comments

[Enterprise Alfresco](#) 4.1.2 has two methods that control content and metadata indexing to facilitate search capabilities that apply to all content nodes. In this blog, we will explore the two methods, experiment with changing out-of-the-box capability, and verify those changes.

**NOTE:** There may be a way to apply indexing configuration only to certain content mimetypes. I am still verifying this and will update this blog in the future.

## Method 1

Enterprise Alfresco 4.1.2 automatically indexes document content and metadata out of the box. This is defined in the Alfresco Content Domain Model contentModel.xml by the following:

```
<aspect name="cm:indexControl">
  <title>Index Control</title>
  <properties>
    <property name="cm:isIndexed">
      <title>Is indexed</title>
      <type>d:boolean</type>
      <default>true</default>
    </property>
    <property name="cm:isContentIndexed">
```

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.

Ok

```
</aspect>
```

You can control indexing on only certain objects types (folder, content) by modifying the contentModel.xml data dictionary, by following [these instructions](#).

## Method 2

Solr is the default search / index for Alfresco 4.1.2, another way to alter index functionality is via configuration file change with each solr core in solrcore.properties. Using both the attributes below disables both content and the metadata, respectively:

```
alfresco.index.transformContent=false  
alfresco.ignore.datatype.1=d:content
```

## Use Case: Disable Content Indexing Only

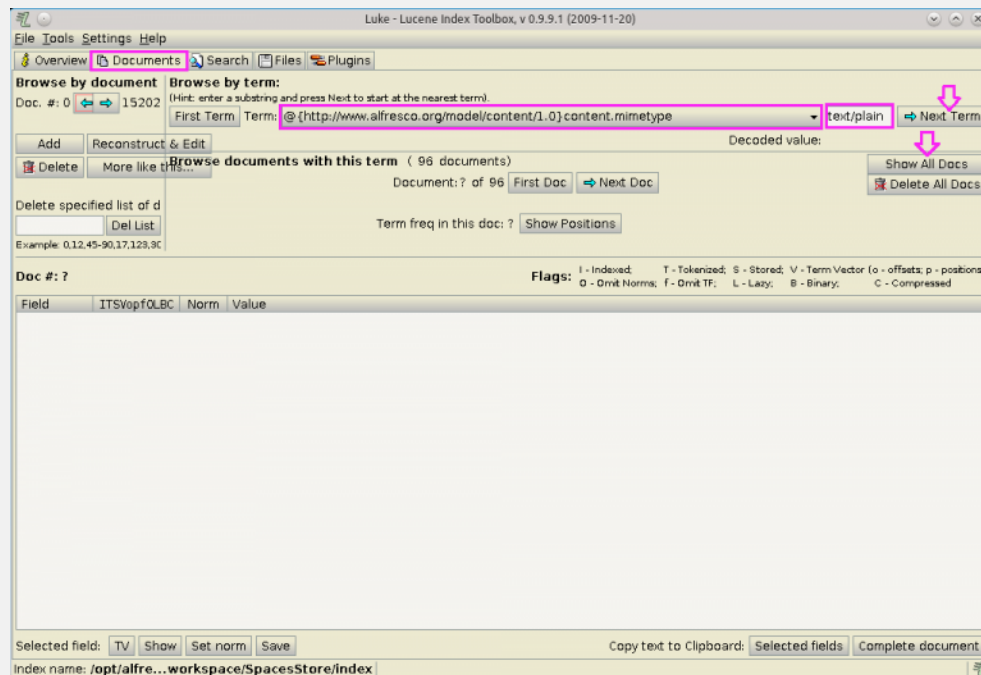
If your application does not need content full-text search capability, then turning off content indexing can increase performance. Method 2 controls indexing across all document types throughout the repository. Using only the first attribute (alfresco.index.transformContent=false) disables content indexing for all documents that are introduced into the system. This is described in [Alfresco's documentation](#); even though documentation state this only works in 4.1.3 and above, it also works in 4.1.2.

Setting transformContent to false disables content indexings because Alfresco performs context indexing by transforming a document into a plain/text document first, then indexing that text document. So if you turn off transformation of the document, then the document's content cannot be indexed.

## Verify Indexing (or Not Indexing) Using Luke

Since Solr uses the Lucene Java search library at its core for full-text indexing and search, let's verify that with [Luke](#). Luke is a handy development and diagnostic tool, which accesses already existing Lucene indexes and allows you to display and modify their content. We'll be using the

- First, find the system:node-dbid of the node index you wish to look at via Alfresco Explore node browser. In our case, that's 534.
- Start the Luke GUI by executing the luke jar:
  - `java -jar lukeall-0.9.9.1.jar`
- When the GUI opens, navigate to the index folder:
  - i.e.: `/opt/alfresco-4.1.2/alf_data/solr/workspace/SpacesStore/index`
- Navigate/find your document. For my example:
  - Navigated to 'Documents' tab. Chose '@{https://www.alfresco.org/model/content/1.0}content.mimetype' as the term. Clicked 'Next Term' until the 'text/plain' value was selected. Clicked 'Show All Docs'.



This brings you to the 'Search' tab. Click the 'right arrow' until you find the latest set of documents. Scroll to the right until you find the column 'dbid', and find row with 534.

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.

Ok

| ACTXCOMM                   | ASPECT | ASSOC TYPE | COBID | FTSSTATUS | ID        | INACTXID | INTXID | ISNODE | UID |
|----------------------------|--------|------------|-------|-----------|-----------|----------|--------|--------|-----|
| {http://www.alfresco.org/r |        |            | 537   |           | LEAF-537  | 6        | 6      | T      |     |
| {http://www.alfresco.org/r |        |            | 539   |           | LEAF-539  | 6        | 6      | T      |     |
| {http://www.alfresco.org/r |        |            | 540   |           | LEAF-540  | 6        | 6      | T      |     |
| {http://www.alfresco.org/r |        |            | 541   |           | LEAF-541  | 6        | 6      | T      |     |
| {http://www.alfresco.org/r |        |            | 542   |           | LEAF-542  | 6        | 6      | T      |     |
| {http://www.alfresco.org/r |        |            | 543   |           | LEAF-543  | 6        | 6      | T      |     |
| {http://www.alfresco.org/r |        |            | 544   |           | LEAF-544  | 6        | 6      | T      |     |
| {http://www.alfresco.org/r |        |            | 545   |           | LEAF-545  | 6        | 6      | T      |     |
| {http://www.alfresco.org/r |        |            | 546   |           | LEAF-546  | 6        | 6      | T      |     |
| {http://www.alfresco.org/r |        |            | 547   |           | LEAF-547  | 6        | 6      | T      |     |
| {http://www.alfresco.org/r |        |            | 548   |           | LEAF-548  | 6        | 6      | T      |     |
| {http://www.alfresco.org/r |        |            | 1062  |           | LEAF-1062 | 36       | 36     | T      |     |
| {http://www.alfresco.org/r |        |            | 6420  |           | LEAF-6420 | 4723     | 4723   | T      |     |

- Double click the row, this brings you to the 'Documents' tab again. Click on 'Reconstruct & Edit' to open up the document details window.

## Findings for Each Index Combination:

Using .txt plain/text files as examples.

1. Default out-of-box Alfresco, content and metadata indexing enabled

Both content and metadata are indexed. Notice the document details below that content, content.~, content.encoding, content.mimetype, content.size, and content.transformationStatus are all there. And you see the content in the tab on the right side (judy ed romano)

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.

Ok



## 2. Content and metadata indexing disabled

Add in the following to workspace-SpacesStore solrcore.properties (and any additional cores).

```
alfresco.index.transformContent=false  
alfresco.ignore.datatype.1=d:content
```

This causes both the content and the metadata to not be indexed, the content won't even show up in Luke since it wasn't indexed. You can double verify by going in Alfresco Explorer Node Browser and doing a lucene search for the mimetype in question and you won't find the document:  
@cm\:content.mimetype: "text/plain".

## 3. Metadata indexing only

Add in the following to workspace-SpacesStore solrcore.properties (and any additional cores).

```
alfresco.index.transformContent=false
```

This causes only the metadata to be indexed. Notice the document details below that only content.encoding and content.mimetype are available, the content itself was not indexed.

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.

Ok



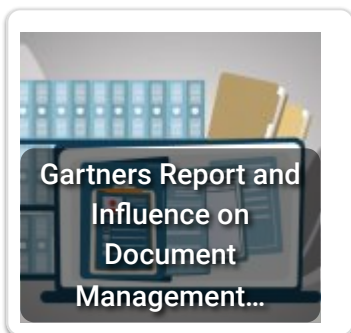
The post “[Alfresco – Indexing Document Metadata Only – Confirmation with Luke](#)” appeared first on [cherryshoe.blogspot.com](#).



Select Category



## Related Posts



Capabilities Brief



Contact Us

[Privacy](#) - [Terms](#)

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.

Ok



## Need a bit more info on how Armedia can help you?

Feel free to schedule a 30-minute no-obligations meeting.

Pick A Meeting Slot

# 0 Comments

We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.



Ok

Vienna, Virginia 22182  
+1 (571) 395-4800

5817 Stow Drive  
Tucker, Georgia 30084  
+1 (571) 395-4800

609 Balibay Road  
Apollo Beach, Florida 33572  
+1 (571) 395-4800

Vasil Gjorgov 20A  
1000 Skopje  
Republic of N. Macedonia

FOIA Alternatives

Gartner 2022 IT Modernization is here to stay

Gartner's Report and Influence on Document Management Procurement

FOIAOnline End of Life: How to Prepare Your Organization for a Smooth Transition

## **About Us**

Capabilities Brief

News

Events

Careers

Privacy Policy

## **Let's Connect**

(571) 395-4800

Contact Us

**Email us:** [info@armedia.com](mailto:info@armedia.com)





