

## Infrastructure - Bug #3514

### CN solr search returns the 'text' field

2013-01-23 17:53 - Rob Nahf

<b>Status:</b>	Closed	<b>Start date:</b>	2013-01-23
<b>Priority:</b>	Normal	<b>Due date:</b>	2013-04-13
<b>Assignee:</b>	Skye Roseboom	<b>% Done:</b>	100%
<b>Category:</b>	d1_cn_solr_extensions	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	2013.14-Block.2.3	<b>Story Points:</b>	
<b>Milestone:</b>	CCI-1.2		
<b>Product Version:</b>	*		

#### Description

The 'text' field in CN solr is used to facilitate searching, but is an unstructured amalgam of science metadata, is usually very long, and probably not very useful.

for example:

[https://cn-dev.test.dataone.org/cn/v1/query/solr/?q=text:\\*&rows=1](https://cn-dev.test.dataone.org/cn/v1/query/solr/?q=text:*&rows=1)

With opening up Solr search as a public method, it would be better to make the 'text' field non-returnable by default. (Many clients will fail to specify return fields, and so get a lot of content.)

#### History

##### #1 - 2013-01-23 17:58 - Dave Vieglais

the config for specifying the default field list is in d1-index-solrconfig.xml:

```
...  
*,score
```

The \* returns all fields and should be replaced with the list of fields that should be returned.

I think this might be preferable over making the field non-returnable as there *may* be a use case for taking advantage of the full text returned for fine tuning the indexing process.

##### #2 - 2013-02-21 01:09 - Skye Roseboom

Currently it is necessary to return the text field - due to our strategy for modeling relationships in the index.

When the ORE document is indexed, each references document index record must be updated with the ORE information (resourceMap, documents, documentedBy) fields. If the full text field is not returned as part of the index record - then the original science metadata documents must be re-indexed off disk - rather than just writing the records back (with updated ORE data) - since the full text field cannot be regenerated from the index data - must be created using the original document text.

Given the above - Dave's solution looks preferable and should still allow indexing strategy to continue to work with updated query strings.

##### #3 - 2013-03-01 18:29 - Skye Roseboom

- Due date changed from 2013-01-23 to 2013-03-02

- Target version changed from 2013.2-Block.1.1 to 2013.8-Block.1.4

**#4 - 2013-03-01 18:33 - Skye Roseboom**

- Due date changed from 2013-03-02 to 2013-03-16

- Target version changed from 2013.8-Block.1.4 to 2013.10-Block.2.1

**#5 - 2013-03-04 15:17 - Skye Roseboom**

- Milestone changed from None to CCI-1.2

**#6 - 2013-04-02 16:10 - Skye Roseboom**

- Target version changed from 2013.10-Block.2.1 to 2013.14-Block.2.3

- Due date changed from 2013-03-16 to 2013-04-13

**#7 - 2013-04-03 23:19 - Skye Roseboom**

- Status changed from New to In Progress

**#8 - 2013-04-04 17:58 - Skye Roseboom**

- Status changed from In Progress to Testing

Updated deployment and test resources for the CN search index so that the search handler now contains a default 'fl' (field list) parameter. It lists all fields that are returnable except the (full) 'text' field.

Installed in the cn-dev environment.

**#9 - 2013-05-13 23:37 - Skye Roseboom**

- Status changed from Testing to Closed

Deployed to all three prod CN. text field only returned if listed in the fl parameter.