

## Metacat - Bug #3174

### Metacat performance issue in Sanparks skin

03/13/2008 03:37 PM - Jing Tao

<b>Status:</b>	Resolved	<b>Start date:</b>	03/13/2008
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Jing Tao	<b>% Done:</b>	0%
<b>Category:</b>	metacat	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	1.8.1	<b>Spent time:</b>	0.00 hour
<b>Bugzilla-Id:</b>	3174		
<b>Description</b>			
Matt and Mike reported it would like about 4 or 5 minutes to do a search in sanparks skin of production server. We should fixed before 1.8.1 release.			
<b>Related issues:</b>			
Is duplicate of Metacat - Bug #3032: Sanparks skin search is slower than othe...		<b>Resolved</b>	<b>12/14/2007</b>
Blocked by Metacat - Bug #3146: Include FGDC metadata in KNB and NCEAS skin s...		<b>Resolved</b>	<b>02/13/2008</b>

### History

#### #1 - 03/13/2008 04:22 PM - Jing Tao

- Bug 3032 has been marked as a duplicate of this bug. \*\*\*

#### #2 - 03/18/2008 10:30 AM - ben leinfelder

duane's comments about the indexPath in LTER pointed me toward the search pathquery in the sanparks/saeon skins. Perhaps the fix is as simple as adding the additional returnfields that are requested (from FGDC documents). We had previously only added the FGDC field "placekey" because it was used as part of the queryterm/pathexpr element.

The additional returnfields are:

```
<returnfield>idinfo/citation/citeinfo/title</returnfield>
<returnfield>idinfo/citation/citeinfo/origin</returnfield>
<returnfield>idinfo/keywords/theme/themekey</returnfield>
```

#### #3 - 03/18/2008 02:56 PM - ben leinfelder

adding those extra returnfields to the index paths did not seem to make a difference with searches on dev. The sanparks skin ends up generating SQL that uses an "intersect" so as to handle organization filters for both EML and FGDC document types- i believe this is where the performance hit is being introduced.

It might be nice to include a search option so that you could search for only one document type at a time. If you wanted all types it would just take longer...

\*dev.nceas.ucsb.edu is giving about 85 seconds for a 'kruger' search vs. the default skin's 23 seconds for the same search term.

#### #4 - 03/18/2008 03:39 PM - ben leinfelder

oh! one more thing:

looks like sanparks is always searching for the search term in *any* nodedata (not limited to title, abstract, etc....)

that means it does not use the xml\_patch\_index table to find some of the valid docids....

It is usually an option in the other skins' search interface to specify that searches only look in certain [predefined] fields. Perhaps we should add this to the sanparks and saeon skins

#### #5 - 03/24/2008 01:53 PM - Jing Tao

Here is the selection query:

```
SELECT docid,docname,doctype,date_created, date_updated, rev FROM xml_documents WHERE docid IN (((SELECT DISTINCT docid FROM xml_path_index WHERE (UPPER LIKE '%PLANT%' AND path IN ('abstract/para','surName','givenName','organizationName','title','keyword','para','geographicDescription','literalLayout','@packageId','abstract','idinfo/citation/citeinfo/title','idinfo/citation/citeinfo/origin','idinfo/keywords/theme/themekey')) UNION (SELECT DISTINCT docid FROM xml_nodes WHERE UPPER LIKE '%PLANT%' AND parentnodeid IN (SELECT nodeid FROM xml_index WHERE path LIKE 'idinfo/keywords/theme/placekey')) ) ) INTERSECT (SELECT DISTINCT docid FROM xml_path_index WHERE (UPPER LIKE '%SANPARKS, SOUTH AFRICA%' AND path IN ('placekey','keyword')) OR (UPPER LIKE '%SAEON, SOUTH AFRICA%' AND path IN ('placekey','keyword')))) AND (docid IN (SELECT docid from xml_access WHERE = 'public') AND perm_type = 'allow' AND permission > 3)) AND docid NOT IN (SELECT docid from xml_access WHERE = 'public') AND perm_type = 'deny' AND perm_order = 'allowFirst' AND permission > 3))
```

**#6 - 03/25/2008 02:34 PM - ben leinfelder**

Added optional "Search All fields" checkbox for SANParks and SAEON skins. Default search (unchecked) searches only a handful of indexed document elements (as specified by the indexPath metacat property).

**#7 - 04/02/2008 11:10 AM - Jing Tao**

Modified the search query - change the search field from "idinfo/keywords/theme/placekey", which is not in the path index, to "placekey" which is in the path index. The search now only takes about 10 seconds in the first time.

**#8 - 04/10/2008 11:05 AM - Jing Tao**

callie did some test too and it is fine to her to close the bug.

**#9 - 03/27/2013 02:22 PM - Redmine Admin**

Original Bugzilla ID was 3174