

Kepler - Bug #3179

Distinguish "download" and "information" attribute values in eml distribution online url

03/18/2008 03:40 PM - Jing Tao

Status:	Resolved	Start date:	03/18/2008
Priority:	Normal	Due date:	
Assignee:	Jing Tao	% Done:	0%
Category:	data access	Estimated time:	0.00 hour
Target version:	1.0.0	Spent time:	0.00 hour
Bugzilla-Id:	3179		

Description

I agree, I don't think it does handle this, but this is a bug in my opinion. It should distinguish these URL types. The intention of the "function" attribute in EML was to handle exactly what Wade is trying to do, so Kepler should look for it and only really try to parse and download data from 'download' URLs. If a "function" attribute has not been provided on the URL, then maybe it should try to download it as well, but that is open to discussion. I've been looking for the query specification in Kepler -- but to no avail.

History

#1 - 03/18/2008 03:45 PM - ben leinfelder

from Jing:

I am not exactly sure how hard the query will be implemented, but I don't think it will be too hard. However, i have a concern - we add more conditions into query, the performance will be get worse.

Alternative way maybe is: query will get back packages with urls having both "download" and "information" attributes (which is same as we do now). However, when user drag a package from search resultset panel to canvas, an alert window will show up if the data distribution url in this eml package has the attribute value "information".

Any comment will be appreciated.

Thanks,

Jing

#2 - 03/18/2008 03:50 PM - ben leinfelder

Whether or not the query is altered, I think Kepler will still need to handle both types of distribution urls. I'm thinking about the case where there are two datatables in an EML document - one is available as a "download" and the other is protected as "information". Kepler should still be able to download the first data table and ignore the second (maybe with a warning).

I'm not sure if the second datatable should even show up in the "selected entity" dropdown list for the eml actor (nor am I sure what kind of data it might be reasonable for it to output if it were indeed selected).

#3 - 03/18/2008 04:00 PM - Jing Tao

Yeah. So here is my approach:

I am not exactly sure how hard the query will be implemented, but I don't think it will be too hard. However, i have a concern - we add more conditions into query, the performance will be get worse.

Alternative way maybe is: query will get back packages with urls having both "download" and "information" attributes (which is same as we do now). However, when user

drag a package from search resultset panel to canvas, an alert window will show up if the data distribution url in this eml package has the attribute value

"information".

Any comment will be appreciated

#4 - 04/01/2008 03:39 PM - Jing Tao

Here are solutions after the 04/01/2008 telephone conference:

1. Modify the query to exclude the eml packages which only have function=information in url element during the search (we should keep packages which is with function=download and without function at all.)
2. If an eml package has more than one data tables, it only shows the data tables with the url which has download value.

#5 - 04/14/2008 06:31 PM - Jing Tao

Modified the Entity and EML2Parser class. The value of attribute "function" in distribution url will be stored. In EML2DataSource actor, only the entity with the "download" function value will be shown, the entity with "information" value will be skipped.

EML declare "download" value as the default value for attribute "function" in online url element. When we insert an eml document into metacat, xerces parser will automatically added default value - "download" into attribute "function" if the eml document didn't specify function value. So there are only two options in Metacat, either "download" or "information". This make things easier.

So we added a condition with "AND" operator to eml query, it will return the package only contain "download" value:
<condition concept="dataset/dataTable/physical/distribution/online/url/@function" operator="EQUALS">download</condition>

Note: we use "EQUALS" as operator. Metacat wouldn't add "%" symbol into the query. Our db index will work for this part.

#6 - 04/15/2008 06:29 PM - Jing Tao

Tested it against server and it worked.

#7 - 03/27/2013 02:22 PM - Redmine Admin

Original Bugzilla ID was 3179