

EML - Bug #2507

Data Manager Library: Create a EML parser lib to digest eml document

08/01/2006 10:52 AM - Jing Tao

Status:	Resolved	Start date:	08/01/2006
Priority:	Normal	Due date:	
Assignee:	Jing Tao	% Done:	0%
Category:	datamanager	Estimated time:	0.00 hour
Target version:	DataManager 1.0.0	Spent time:	0.00 hour
Bugzilla-Id:	2507		
Description			
Currently, the EML actor in Kepler can download eml document and parse it. After parsing, the entity information in eml document will be stored in java object and data file will be download into local file system and also be stored in relation db too. We want to separete this process from Kepler and make it as lib in eml module. So this library can be used in Kepler, Metacat and some other projects.			
Related issues:			
Is duplicate of EML - Bug #2504: Create a EML parser lib to digest eml document		Resolved	07/31/2006

History

#1 - 08/01/2006 11:12 AM - Jing Tao

Here is our plan:

Creating 3 packages in eml src dir:

1. org.ecoinformatics.eml.digester package and main class is EML200Parser. The main class can be copied from kepler module.
2. org.ecoinformatics.eml.download package and main class is DataDistributionHandler. The this class will implement Runnable interface and API is
DataDistributionHandler(Entity entity);
run();

3. org.ecoinformatics.eml.db package and main class is is TableGenerator. The API of the class is:
TableGenerator(Enity entity, File localFile);
generateTable();
getTableName();
loadDataToTable();

The function of download package is very similar to cache system of Kepler. I am thinking how to reuse those code in kepler.

#2 - 08/02/2006 11:04 AM - Jing Tao

In order to make download package more configurable, I would like to change to constructor to:

DataDistributionHandler(Entity entity, File cacheDir, File fileName);

#3 - 08/04/2006 04:49 PM - Jing Tao

Here is the change in org.ecoinformatics.eml.db package:

Main class is SQLCommandHandler and API is:

SQLCommandHandler(DBConnection conn, String plugInName)
generateTable(Entity entity, File fileName) and it will return the generated table name as string;
dropTable(String tableName);
excuteSelectionSQLComman(String sqlCommand) and it return a Resultset object;

The org.ecoinformatics.eml.degestor package API is:

EML200Parser(InputStream stream);
EML200Parser(InputSource source);
getEntityList() and it will return a vector;
parse();

#4 - 08/08/2006 01:35 PM - Jing Tao

New package name are suggested:

org.ecoinformatics.eml.digester.parser
org.ecoinformatics.eml.digester.download
org.ecoinformatics.eml.digester.db

#5 - 08/08/2006 01:58 PM - Matt Jones

digester is a bit of a crude name. How about "loader"?

org.ecoinformatics.eml.loader.parser
org.ecoinformatics.eml.loader.download
org.ecoinformatics.eml.loader.database

This is an improvement but still not totally great. Suggestions welcome.

#6 - 08/08/2006 02:15 PM - James Brunt

I like loader but it's not a perfect fit for the way the work is divided which is more like parse (eml) -> create (table) -> source (data). Correct?

#7 - 10/26/2006 09:45 AM - Duane Costa

We named the top-level package "org.ecoinformatics.datamanager". The complete set of packages is:

org.ecoinformatics.datamanager
org.ecoinformatics.datamanager.database
org.ecoinformatics.datamanager.download
org.ecoinformatics.datamanager.parser
org.ecoinformatics.datamanager.parser.eml

#8 - 10/26/2006 09:58 AM - Duane Costa

- Bug 2504 has been marked as a duplicate of this bug. ***

#9 - 01/11/2010 11:50 AM - ben leinfelder

this has been completed. Moreover, it has been extended to support any XML schema that makes use of the EML dataSet module.

#10 - 03/27/2013 02:20 PM - Redmine Admin

Original Bugzilla ID was 2507