

OAIFetch

This is the detailed design document for the OAI ([Open Archives Initiative](#)) fetch portion of the VIVO harvester. This document is neither feature-complete nor comprehensive and will evolve as new requirements or optimizations are recognized.

The OAIFetch method used to ingest data from OAI Data sources. Brings in data as XML selected by date range and returns raw XML which is then stored in a file as determined by the configuration file.

Command Line Parameters

Short Option	Long Option	Parameter Value Map	Description	Required
u	url	URL	repository url without http://	true
s	start	DATE	beginning date of date range (YYYY-MM-DD)	true
e	end	DATE	ending date of date range (YYYY-MM-DD)	true
o	output	CONFIG_FILE	RecordHandler config file path	true
O	outputOverride	VALUE	override the RH_PARAM of output record handler using VALUE	false

Flow

1. Executed by Fetch.java
2. Read in configuration parameters.
3. Run the OAI Harvest as determined by the configuration parameters
4. Fetch records until there are none remaining
5. Return XML stream of record data.
6. Write XML stream to file.

Inputs

Parameters defined in delimited text file.

1. Address of OAI Repository
2. Start Date
3. End Date
4. Output Filename

Outputs

XML Stream of records written to file.

Class Variables

Log - Static, logfactory.

arrRequiredParamaters - An array of which parameters are required to run the OAI Fetch. They are: "address", "startDate", "endDate", and "filename".

strAddress - The website address of the OAI repository, without the protocol prefix. (No http://)

strStartDate - The start date for the range of records to pull, format is YYYY-MM-DD. If time is required, format is YYYY-MM-DDTHH:MM:SS:MSZ. Some repositories do not support millisecond resolution. Example 2010-01-15T13:45:12:50Z

strEndDate - The end date for the range of records to pull, format is YYYY-MM-DD. If time is required, format is YYYY-MM-DDTHH:MM:SS:MSZ. Some repositories do not support millisecond resolution. Example 2010-01-15T13:45:12:50Z

strFileName - The filename to write the XML data, should be XMLVault/OAI/outputfilenamegoeshere.xml

Functions

Execute

Executes the OAI Fetch using the parameters defined in the configuration file.

Inputs

strAddress - The website address of the OAI repository, without the protocol prefix. (No http://)
strStartDate - The start date for the range of records to pull, format is YYYY-MM-DD. If time is required, format is YYYY-MM-DDTHH:MM:SS:MSZ. Some repositories do not support millisecond resolution. Example 2010-01-15T13:45:12:50Z
strEndDate - The end date for the range of records to pull, format is YYYY-MM-DD. If time is required, format is YYYY-MM-DDTHH:MM:SS:MSZ. Some repositories do not support millisecond resolution. Example 2010-01-15T13:45:12:50Z
strFileName - The filename to write the XML data, should be XMLVault/OAI/outputfilenamegoeshere.xml

Outputs

Nothing, it writes to the output stream during execution and returns nothing.

acceptParams

Documentation needed.

runTask

Documentation needed.

Configuration file example

```
address:www.twmuseums.org.uk/pnds/memorynet/  
startDate:1000-01-01  
endDate:2010-01-01  
filename:XMLVault/OAI/MemoryNet.xml
```

OAI Data Sources

This is a listing of OAI data sources that conform to the OAI specification and will work with OAIFetch. In no way is this list comprehensive and if other repositories are found they should be added to this list.

CiteSeer

Address

<http://cs1.ist.psu.edu/cgi-bin/oai.cgi>

Example Configuration File

```
address:cs1.ist.psu.edu/cgi-bin/oai.cgi  
startDate:2005-01-01  
endDate:2010-01-01  
filename:XMLVault/OAI/CiteSeer.xml
```