

WOSFetch

WOSFetch

The Web of Science (WOS) is a data resource available from [Thomson Reuters](#). This harvester tool pulls records from the WOS and places them into a record handler. It uses the [SOAPMessenger](#) to communicate through two forms of queries available for getting information from the WOS. For more information on the use of the WOSFetch, look at the [Web Of Science Example Script](#)

Reason for use

If the WOS has data about your chosen publications and access is authorized, then this fetch can be used to pull the data into a [RecordHandler](#) in order to put it into a vivo instance.

Parameters

The parameters for the tool can be placed in the specified config file.

wordiness

wordiness - (optional) sets the lowest level of log messages to be displayed to the console. The lower the log level, the more detailed the messages. Possible Values:

- `<Param name="wordiness">OFF</Param>` - Results in no messages being displayed.
- `<Param name="wordiness">ERROR</Param>` - Results in only messages from the ERROR level to be displayed.
 - Error messages detail when the tool has experienced an error preventing it from completing its task
- `<Param name="wordiness">WARN</Param>` - Results in only messages above and including WARN level messages to be displayed. Match does not produce any WARN level messages.
- `<Param name="wordiness">INFO</Param>` - (Default) Results in all messages above and including INFO level messages to be displayed. INFO level messages detail when the tool has started and ended and when it begins/ends a phase ('Finding matches' and 'Beginning Rename of matches') and how many matches have been found.
- `<Param name="wordiness">DEBUG</Param>` - Results in all messages above and including DEBUG level messages to be displayed. DEBUG level messages detail each matching input URI to its VIVO URI as they are processed. Additionally, it will display stacktrace information if an error occurs.
- `<Param name="wordiness">ALL</Param>` or `<Param name="wordiness">TRACE</Param>` - Results in all messages above and including TRACE level messages to be displayed, since trace is the lowest level it is the same as ALL in practice. TRACE level messages details every matching set as it is processed in each phase along with SPARQL queries and start and stop for their execution.

Authorization

`<Param name="authurl">` needs the web address to provide the encoded session identifier which is then used to allow the search query to be authorized to get a response. If this address is incorrect then the entire harvest run will fail.

Default value:

`<Param name="authurl">****/WOKMWSAuthenticate</Param>`

`<Param name="authmessage">` contains the soap message which would provide a username and password authorizing use of the Web of Science. If the parameter is not present then WOSFetch will construct a message which relies on the IP authorization service provided by Thomson Reuters.

Default value:

The parameter is left out to use the IP authorization

Search Query

`<Param name="searchconnection">` contains the web address to send the search message requesting data related to the contents of that query. If it is not specified or incorrect the search will not function and the entire harvest run will fail.

Default value:

`<Param name="searchconnection">****/WokSearchLite</Param>`

`<Param name="searchmessage">` contains the soap message which will specify the query terms and other search properties. WOSFetch will automatically attempt to harvest all of the "records found" in batches the size of the value in the "count" tag.

Default value:

`<Param name="searchmessage">wos-query-message.xml</Param>`

Llinks Article Match Retrieval Service (LAMR) Query

`<Param name="lamrconnection">` this parameter contains the address to send the LAMR query message. If this site is incorrect then the soap messenger will not return the correct information and thus the harvest will either fail, be incorrect, or incomplete.

<Param name="lamrmessage"> contains the soap message which specifies the query return items which determines the fields returned by the harvest calls. The map tag named lookup will be populated with the UT values found from the search call. If this message is faulty then the harvest will be incorrect or incomplete.

Overview

Short Option	Long Option	Parameter Value Map	Description	Required
u	authurl	URL	The URL used to authorize and close the session	true
a	authmessage	AUTHMESSAGE	File path to authorization message	false
c	searchconnection	URL	The URL that will receive the search request and will return the results	true
s	searchmessage	SEARCHMESSAGE	File path for search query message	true
l	lamrconnection	URL	The URL that will receive the LAMR request and will return the results	true
m	lamrmessage	LAMRMESSAGE	File path for the LAMR search query message	false
p	usernamepassword	USERNAMEPASSWORD	The combined user name and password string to be encoded for authorization.	false
o	output	OUTPUT_FILE	The XML config file for the RecordHandler for the output	true
O	outputOverride	Override of the RH_PARAM of the output RecordHandler using VALUE	false	

Usage

WOSFetch is often the first part of a harvest of data from The Web of science service. It pulls the information into a local [RecordHandler](#) which can then be [Translated](#) before being [transferred](#) into a jena model.

For some detail on how WOSFetch is used see [Web Of Science Example Script](#).

Messages

There are a couple provided messages within the WOSFetch. They are the default messages to acquire a session with an ID and then close it.

This is the default authentication request message. the response of which contains the session ID.

```
<?xml version="1.0" encoding="UTF-8"?>
<soap:Envelope xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/"
  xmlns:ns2="http://auth.cxf.wokmws.thomsonreuters.com">
  <soap:Body>
    <ns2:authenticate/>
  </soap:Body>
</soap:Envelope>
```

This is the session closing message, after performing the various queries the session gets closed by sending this message with the authorized session ID.

```
<?xml version="1.0" encoding="UTF-8"?>
<soap:Envelope xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/"
  xmlns:ns2="http://auth.cxf.wokmws.thomsonreuters.com">
  <soap:Body>
    <ns2:closeSession/>
  </soap:Body>
</soap:Envelope>
```