Merge

Harvester . Utilities

Merge

Combines multiple related rdf records into a single rdf record

This tool Merge takes a set of records and a regular expression to find related records and combines them in another set of records

Merge Parameters

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. Merge 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 ('Building List of Primary Records' and 'Beginning Merging into Primary Records') and how many primary records it found.
- <Param name="wordiness">DEBUG</Param> - Results in all messages above and including DEBUG level messages to be displayed. DEBUG level messages detail each primary record name into which it will merge the matching records. Additionally, it will display stacktrace information if an error occurs.
- <Param name="wordiness">ALL</Param> or 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 record as it is added to the primary record's merge set.

baseRegex - Regex for finding primary records (with a grouping for the subsection to use to find sub-records)

Example:
<Param name="baseRegex">tableName_(id_-_.?)</Param> - A regular expression to match record IDs and grouping capture to isolate the id.

input - (optional - at least one of this and/or inputOverride) the configuration file that describes the input record set. The parameters for this config file are described in the Record Sets section below.

Example:

- <Param name="input">/absolute/path/to/file.conf.xml</Param> - An absolute path to a recordhandler config file on linux/unix/macosx based systems.
- <Param name="input">C:/absolute/path/to/file.conf.xml</Param> - An absolute path to a recordhandler config file on a windows operating system.
- <Param name="input">relative/path/to/file.conf.xml</Param> - A path to a recordhandler config file that is relative to the folder the shell was in when this command was executed.

inputOverride - (optional - at least one of this and/or input) specify the parameters for the record set without a config file and/or override specific parameters from the given config file. The parameters that can be set/overridden are described in the Record Sets section below.

Example:

- <Param name="inputOverride">paramName=valueToUse</Param>

output - (optional - at least one of this and/or outputOverride) the configuration file that describes the output record set. The parameters for this config file are described in the Record Sets section below.

Example:

- <Param name="output">/absolute/path/to/file.conf.xml</Param> - An absolute path to a recordhandler config file on linux/unix/macosx operating systems.
- <Param name="output">C:/absolute/path/to/file.conf.xml</Param> - An absolute path to a recordhandler config file on a windows operating system.
- <Param name="output">relative/path/to/file.conf.xml</Param> - A path to a recordhandler config file that is relative to the folder the shell was in when this command was executed.

outputOverride - (optional - at least one of this and/or output) specify the parameters for the record set without a config file and/or override specific parameters from the given config file. The parameters that can be set/overridden are described in the Record Sets section below.

Example:

- <Param name="outputOverride">paramName=valueToUse</Param>

Configuration Example

```
<Config>
        <Param name="wordiness">INFO</Param>
        <Param name="baseRegex">tableName_(id_-_.*?)</Param>
        <Param name="input">record-set.conf.xml</Param>
        <Param name="outputOverride">rhClass=org.vivoweb.harvester.util.repo.JenaRecordHandler</Param>
        <Param name="outputOverride">dataFieldType=http://yourDomain.com/propbase#myProp</Param>
        <Param name="outputOverride">jenaConfig=jena-model.conf.xml</Param>
</Config>
```