

26 Content Update Checker

Function:	Logs message with event code when file or url content is updated
------------------	--

Module Name:	ContentUpdateChecker
Where to Use?	In a workflow
Why to Use?	To check file or url content is updated so a new task can be run
Description:	The ContentUpdateChecker is a module that can be used at the start of a workflow to see if new data is available so other tasks can be run.
Preconditions:	File or url should return plain text from which the first line dynamically changes
Outcome(s):	Log message with event code and content when new content is found, debug message when no new content is found
Scheendump (s):	
Remark(s):	
Available:	2013.01, 2014.01 and onwards

Contents

- [Contents](#)
- [Overview](#)
- [Configuration](#)
- [Optional elements](#)
- [Sample input and output](#)
- [Technical reference](#)

Overview

The ContentUpdateChecker is a module that can be used at the start of a workflow to see if new content is available from either a file or url. If so a log message with event code + the new content is logged, a debug message is logged when no new content is found. In case an empty file is found, the last modification date will be used as content in date format: 'yyyyMMddHHmmss'.

Configuration

A configuration example of the content update checker is given below:

Configuration for url

```
<?xml version="1.0" encoding="UTF-8"?>
<contentUpdateChecker xmlns="http://www.wldelft.nl/fews" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.wldelft.nl/fews https://fewsdocs.deltares.nl/schemas/version1.0
/contentUpdateChecker.xsd">
  <url>https://matroos.deltares.nl/direct/get_anal_times.php?database=maps&source=knmi_h11_v72&most_recent=1</url>
  <user>dummy_username</user>
  <password>dummy_password</password>
  <eventCode>HirLamMeteo.NewData</eventCode>
  <messagePrefix> New data for hirlam meteo for T0: </messagePrefix>
  <interval unit="second"/>
  <timeout unit="minute"/>
  <stopAfterNewContent>true</stopAfterNewContent>
  <contentIgnorePattern>Exact</contentIgnorePattern>
  <contentIgnorePattern>*ends</contentIgnorePattern>
  <contentIgnorePattern>Begins*</contentIgnorePattern>
  <contentIgnorePattern>*contains*</contentIgnorePattern>
</contentUpdateChecker>
```

The URL can be protected by authentication, the parser can supply username and password configured in the import together with the base URL (escape & in xml by using &#amp;):

```
<url>https://matroos.deltares.nl/direct/get_anal_times.php?database=maps&source=knmi_h11_v72&most_recent=1</serverUrl>
<user>dummy_username</user>
<password>dummy_password</password>
```

Configuration for file

```
<?xml version="1.0" encoding="UTF-8"?>
<contentUpdateChecker xmlns="http://www.wldelft.nl/fews" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.wldelft.nl/fews https://fewsdocs.deltares.nl/schemas/version1.0
/contentUpdateChecker.xsd">
  <file>$IMPORT_FOLDER$/KNMI_UKMO_MAPS/done</file>
  <eventCode>HirLamMeteo.NewData</eventCode>
  <messagePrefix> New data for hirlam meteo for T0: </messagePrefix>
  <interval unit="second"/>
  <timeout unit="minute"/>
  <stopAfterNewContent>true</stopAfterNewContent>
  <contentIgnorePattern>Exact</contentIgnorePattern>
  <contentIgnorePattern>*ends</contentIgnorePattern>
  <contentIgnorePattern>Begins*</contentIgnorePattern>
  <contentIgnorePattern>*contains*</contentIgnorePattern>
</contentUpdateChecker>
```

Optional elements

<interval unit="second"/>

Amount of time to wait between checks for new content, when absent content will be checked continuously.

<timeout unit="minute"/>

Total amount of time there must be checked and rechecked for new content, when absent it will do no rechecking which means only once.

<stopAfterNewContent>true</stopAfterNewContent>

This option makes sure there will not be checked for newer content after new content is found.

<contentIgnorePattern>Exact</contentIgnorePattern>

Pattern which determines whether content should be ignored as new. * can be used as wildcard.

*XXX means ends with XXX

XXX* means starts with XXX

XXX means contains XXX

when no * is present the content should match exactly in order to be ignored.

Sample input and output

A url returning a timestamp as plain text of most recent available data in format yyyyMMddHHmm, will result in the following logging.

When the event code has not been logged before:

```
11-07-2014 09:54:51 DEBUG - No logging of event code HirlamMeteo.NewContent because data still equals: 201407110600
11-07-2014 09:54:47 DEBUG - No logging of event code HirlamMeteo.NewContent because data still equals: 201407110600
11-07-2014 09:54:45 INFO - HirlamMeteo.NewContent: New data for hirlam meteo for T0: $$201407110600
11-07-2014 09:54:44 DEBUG - No content found in database for event code: HirlamMeteo.NewContent
```

When the event code has been logged before and content is still the same:

```
11-07-2014 09:51:50 DEBUG - No logging of event code HirlamMeteo.NewData because data still equals: 201407110600
11-07-2014 09:51:49 DEBUG - No logging of event code HirlamMeteo.NewData because data still equals: 201407110600
11-07-2014 09:51:48 DEBUG - Most recent content in database: 201407110600 for event code: HirlamMeteo.NewData
```

Technical reference

Link to schema:	https://fewsdocs.deltares.nl/schemas/version1.0/contentUpdateChecker.xsd
-----------------	---