News API Specification

|  |  |
| --- | --- |
| Version: | 1.2 |
| Date Modified: | 6 March 2016 |

[1. The context 3](#_Toc444879165)

[2. Function of the News API 3](#_Toc444879166)

[3. Module Specification 4](#_Toc444879167)

[Description of the input data involved 4](#_Toc444879168)

[What the module does 5](#_Toc444879169)

[An example 5](#_Toc444879170)

[4. Additional Information: 5](#_Toc444879171)

[5. References: 6](#_Toc444879172)

# The context

We are a company specialising in commodities trading. We are developing a complex system to provide an analytics platform for our traders. We wish to open our system to third parties software houses, offering them ability to “plug-in” independent software modules that implement particular functions. We therefore request all interested companies to provide an independent software module that implements an API as specified in the rest of this document.

# Function of the News API

The Thomson Reuters News Feed Direct (NDS) is an ultra-low latency feed of Thomson Reuters’ world-class news. This service enables the customers to receive news from all over the world almost instantly. The source of these news can be Reuters or other third parties.

You will be given files obtained through the NDS and containing all the news that were published during the fourth quarter of 2015. The files weight more than 3GB and represent millions of news, therefore it’s impossible to manually look for specific news. Thus the objective is to build an API that would easily enable a user to find all the news related to a specific topic and for a list of specific instruments over a given period of time.

The figure below illustrates how the API should work:

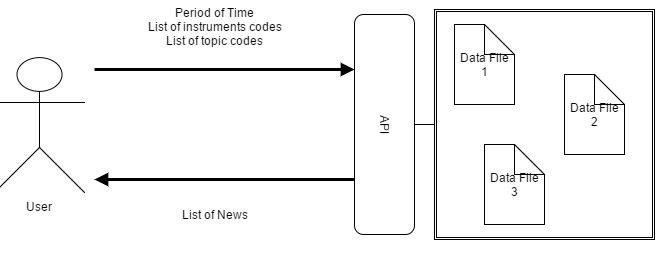


Figure 1 – API architecture

The language in which the module is to be written is not important as long as there is some way to invoke it from our system. Also, all exchanged data is in the form of text files to avoid any specific encoding. In other words, our system calls the module and supplies any data in the form of text files. After execution is finished, all results are also in text files. It is important that the module can be used without its source code being revealed.

# Module Specification

## Description of the input data involved

**The Period of Time:**

It’s composed of two inputs:

* *A start date*
* *An end date*

Both dates have the respect the following format: **“yyyy-MM-ddTHH:mm:ss.SSSZ”.**

These inputs can NOT be empty.

For example: if I want to retrieve all the news that were issued between the first of October 2015 at 08:45:10.295 (8:45am 10 seconds and 295 milliseconds) and the first of November 2015 at 19:37:12:193 (7:37pm 12 seconds and 193 milliseconds), my inputs will be:

* start\_date = “2015-10-01T08:45:10.295Z”
* end\_date = “2015-11-01T19:37:12.193Z”

**The list of instruments codes:**

This input contains a CSV list of all the instruments you want to get news about. Each instrument is identified by a unique RIC (Reuters Instrument Code). *See References for more details.*

This list can be empty in the case where the user doesn’t want to restrict his search.

For example: if you want to retrieve news related to ANZ bank, Woolworths Ltd, Apple Inc. and the Nasdaq Composite Index your input list should be :

* instr\_list = “ANZ.AX,WOW.AX,AAPL.O,.IXIC”

**The list of topic codes**

This input contains a CSV list of all the topics you want to get news about. As for the instruments each topic has a unique identifier. *See References for the list of topics identifiers.*

This list can be empty in the case where the user doesn’t want to restrict his search.

For example: if you are interested only in the news about the European Central Bank, crude oil or the aerospace sector, your input list should be:

* tpc\_list=”ECB,CRU,AER”

## What the module does

The module aims to isolate some specific news according to the user’s criteria.

The user has to input the 3 main information:

* period of interest
* instruments of interest
* topics of interest

The module will then filter the news first according to the period (start date + end date) entered by the user and only retain the news that were issued during that period. To do that for each news the module has to compare the value found in the <TimeStamp></TimeStamp> with the period entered.

After this first filter another one is applied on the instruments. The module will now select only the news related to the instruments inputted by the user if any. To accomplish that the module has to look in the tag <subject /> where the attributes qcode starts with an “R” (example : <subject qcode="R:KREXGR=ECI" />)and see if the RIC found after the “R” is in the lit entered by the user.

Finally, the module will filter once again the remaining news to only keep the ones with a topic that is part of the list of topics inputted by the user. For this part the module has to do the same as for the instruments codes but with the tags <subject /> where the attribute qcode starts with “N2”.

The module has to return in a text file the list of news that match the user criteria. Nevertheless, the choice of how the news are to be displayed within the file is completely yours.

## An example

Let’s consider a simplified example where the complete set of news is the following:

<ContentEnvelope majVers="1" minVers="0.7" pubStyle="Message" xmlns="http://data.schemas.tfn.thomson.com/Envelope/2008-05-01/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:news="http://news.schemas.tfn.thomson.com/2008-05-01/"><Header><Info><Id>20151001-000000000-nS6N0ZH01T-1-1</Id><TimeStamp>2015-10-01T00:00:00.092Z</TimeStamp></Info></Header><Body contentSet="News" majVers="1" minVers="3.0"><ContentItem action="Insert"><Entitlements><Product><Code>NP:M</Code><Code>NP:E</Code><Code>NP:D</Code><Code>NP:T</Code><Code>NP:C</Code><Code>NP:MTL</Code><Code>NP:GRO</Code><Code>NP:SOF</Code><Code>NP:O</Code></Product><Source><Group>NONE</Group></Source></Entitlements><Data xsi:type="news:NewsDataItem"><newsItem standard="NewsML-G2" standardversion="2.7" conformance="power" guid="20151001-000000000-nS6N0ZH01T-1-1" version="1" xml:lang="en" xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/ NAR\_1.8-spec-All-Power\_2.xsd <http://www.w3.org/1999/xhtml> xhtml1-strict.xsd" xmlns="http://iptc.org/std/nar/2006-10-01/" xmlns:rtr="http://www.reuters.com/ns/2003/08/content" xmlns:x="http://www.w3.org/1999/xhtml"><catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards\_14.xml" /><itemMeta><itemClass qcode="icls:text" /><provider qcode="NS:RTRS" /><versionCreated>2015-10-01T00:00:00.000Z</versionCreated><firstCreated>2015-10-01T00:00:00.000Z</firstCreated><pubStatus qcode="stat:usable" /><rtr:versionedId guid="20151001-000000000-nS6N0ZH01T-1-1" /><TakeSequence>1</TakeSequence></itemMeta><contentMeta><urgency>1</urgency><altId type="idType:USN" rtr:isOriginal="1">nS6N0ZH01T</altId><language tag="en" /><subject qcode="N2:KR" /><subject qcode="N2:EMRG" /><subject qcode="N2:ASIA" /><subject qcode="N2:MCE" /><subject qcode="N2:ECI" /><subject qcode="N2:NEWS1" /><subject qcode="N2:TRD" /><subject qcode="N2:LEN" /><subject qcode="N2:RTRS" /><subject qcode="R:KRW=" /><subject qcode="R:0#KRCOMP1=KQ" /><subject qcode="R:KREXGR=ECI" /><subject qcode="R:KRIMGR=ECI" /><subject qcode="R:KRTBAL=ECI" /><headline>S.KOREA SAYS SEPT EXPORTS -8.3 PCT VS YR EARLIER (REUTERS POLL -10.0 PCT)</headline></contentMeta><contentSet><inlineXML contenttype="application/xhtml+xml"><xhtml xmlns="http://www.w3.org/1999/xhtml"><body></body></xhtml></inlineXML></contentSet></newsItem></Data></ContentItem></Body></ContentEnvelope>

<ContentEnvelope majVers="1" minVers="0.7" pubStyle="Message" xmlns="http://data.schemas.tfn.thomson.com/Envelope/2008-05-01/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:news="http://news.schemas.tfn.thomson.com/2008-05-01/"><Header><Info><Id>20151001-000931000-nL1N11Z1QP-1-2</Id><TimeStamp>2015-10-01T00:09:31.242Z</TimeStamp></Info></Header><Body contentSet="News" majVers="1" minVers="3.0"><ContentItem action="Insert"><Entitlements><Product><Code>NP:G</Code><Code>NP:PSC</Code><Code>NP:RNP</Code><Code>NP:PGE</Code></Product><Source><Group>NONE</Group></Source></Entitlements><Data xsi:type="news:NewsDataItem"><newsItem standard="NewsML-G2" standardversion="2.7" conformance="power" guid="20151001-000931000-nL1N11Z1QP-1-2" version="1" xml:lang="en" xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/ NAR\_1.8-spec-All-Power\_2.xsd <http://www.w3.org/1999/xhtml> xhtml1-strict.xsd" xmlns="http://iptc.org/std/nar/2006-10-01/" xmlns:rtr="http://www.reuters.com/ns/2003/08/content" xmlns:x="http://www.w3.org/1999/xhtml"><catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards\_14.xml" /><itemMeta><itemClass qcode="icls:text" /><provider qcode="NS:RTRS" /><versionCreated>2015-10-01T00:09:31.000Z</versionCreated><firstCreated>2015-10-01T00:01:33.000Z</firstCreated><pubStatus qcode="stat:usable" /><rtr:versionedId guid="20151001-000931000-nL1N11Z1QP-1-2" /><TakeSequence>1</TakeSequence></itemMeta><contentMeta><urgency>3</urgency><altId type="idType:USN" rtr:isOriginal="1">nL1N11Z1QP</altId><language tag="en" /><subject qcode="N2:US" /><subject qcode="N2:NEWS1" /><subject qcode="N2:WASH" /><subject qcode="N2:VOTP" /><subject qcode="N2:AMERS" /><subject qcode="N2:GEN" /><subject qcode="N2:POL" /><subject qcode="N2:VOTE" /><subject qcode="N2:LEN" /><subject qcode="N2:RTRS" /><headline>Democrat Sanders raises more than $24 mln in last three months -NY Times</headline></contentMeta><contentSet><inlineXML contenttype="application/xhtml+xml"><xhtml xmlns="http://www.w3.org/1999/xhtml"><body> WASHINGTON, Sept 30 (Reuters) - U.S. Senator Bernie Sanders

of Vermont raised more than $24 million over the last three

months as he wages a populist challenge to become the Democratic

presidential nominee, the New York Times reported on Wednesday.

Sanders&apos; advisers, in announcing the fund-raising tally on

Wednesday night with several hours to go before the end of the

quarter, also said the senator had more than $25 million in cash

on hand, the Times said.

(Reporting by Eric Beech; Editing by Peter Cooney)

((eric.beech@thomsonreuters.com; 202-898-8319; Reuters

Messaging: eric.beech.reuters.com@reuters.net))

Keywords: USA ELECTON/SANDERS</body></xhtml></inlineXML></contentSet></newsItem></Data></ContentItem></Body></ContentEnvelope>

This set of news only contains two news (the news are delimited by the <ContentEnvelope></ContentEnvelope> tags). Now let’s apply different inputs to see what the module is supposed to return.

**Case 1:**

Inputs:

* start\_date = “2015-10-01T00:00:00.092Z”
* end\_date = “2015-10-01T00:05:00.000Z”
* instr\_list = “”
* tpc\_list=””

Return:

<ContentEnvelope majVers="1" minVers="0.7" pubStyle="Message" xmlns="http://data.schemas.tfn.thomson.com/Envelope/2008-05-01/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:news="http://news.schemas.tfn.thomson.com/2008-05-01/"><Header><Info><Id>20151001-000000000-nS6N0ZH01T-1-1</Id><TimeStamp>2015-10-01T00:00:00.092Z</TimeStamp></Info></Header><Body contentSet="News" majVers="1" minVers="3.0"><ContentItem action="Insert"><Entitlements><Product><Code>NP:M</Code><Code>NP:E</Code><Code>NP:D</Code><Code>NP:T</Code><Code>NP:C</Code><Code>NP:MTL</Code><Code>NP:GRO</Code><Code>NP:SOF</Code><Code>NP:O</Code></Product><Source><Group>NONE</Group></Source></Entitlements><Data xsi:type="news:NewsDataItem"><newsItem standard="NewsML-G2" standardversion="2.7" conformance="power" guid="20151001-000000000-nS6N0ZH01T-1-1" version="1" xml:lang="en" xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/ NAR\_1.8-spec-All-Power\_2.xsd <http://www.w3.org/1999/xhtml> xhtml1-strict.xsd" xmlns="http://iptc.org/std/nar/2006-10-01/" xmlns:rtr="http://www.reuters.com/ns/2003/08/content" xmlns:x="http://www.w3.org/1999/xhtml"><catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards\_14.xml" /><itemMeta><itemClass qcode="icls:text" /><provider qcode="NS:RTRS" /><versionCreated>2015-10-01T00:00:00.000Z</versionCreated><firstCreated>2015-10-01T00:00:00.000Z</firstCreated><pubStatus qcode="stat:usable" /><rtr:versionedId guid="20151001-000000000-nS6N0ZH01T-1-1" /><TakeSequence>1</TakeSequence></itemMeta><contentMeta><urgency>1</urgency><altId type="idType:USN" rtr:isOriginal="1">nS6N0ZH01T</altId><language tag="en" /><subject qcode="N2:KR" /><subject qcode="N2:EMRG" /><subject qcode="N2:ASIA" /><subject qcode="N2:MCE" /><subject qcode="N2:ECI" /><subject qcode="N2:NEWS1" /><subject qcode="N2:TRD" /><subject qcode="N2:LEN" /><subject qcode="N2:RTRS" /><subject qcode="R:KRW=" /><subject qcode="R:0#KRCOMP1=KQ" /><subject qcode="R:KREXGR=ECI" /><subject qcode="R:KRIMGR=ECI" /><subject qcode="R:KRTBAL=ECI" /><headline>S.KOREA SAYS SEPT EXPORTS -8.3 PCT VS YR EARLIER (REUTERS POLL -10.0 PCT)</headline></contentMeta><contentSet><inlineXML contenttype="application/xhtml+xml"><xhtml xmlns="http://www.w3.org/1999/xhtml"><body></body></xhtml></inlineXML></contentSet></newsItem></Data></ContentItem></Body></ContentEnvelope>

The return should only be the first news because it’s the only news that was created during the period inputted. If the end date would have been “2015-10-01T00:15:00.000Z” both news would have been returned.

**Case 2:**

Inputs:

* start\_date = “2015-10-01T00:00:00.092Z”
* end\_date = “2015-10-01T00:15:00.000Z”
* instr\_list = “KRW=,KREXGR=ECI”
* tpc\_list=””

Return:

<ContentEnvelope majVers="1" minVers="0.7" pubStyle="Message" xmlns="http://data.schemas.tfn.thomson.com/Envelope/2008-05-01/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:news="http://news.schemas.tfn.thomson.com/2008-05-01/"><Header><Info><Id>20151001-000000000-nS6N0ZH01T-1-1</Id><TimeStamp>2015-10-01T00:00:00.092Z</TimeStamp></Info></Header><Body contentSet="News" majVers="1" minVers="3.0"><ContentItem action="Insert"><Entitlements><Product><Code>NP:M</Code><Code>NP:E</Code><Code>NP:D</Code><Code>NP:T</Code><Code>NP:C</Code><Code>NP:MTL</Code><Code>NP:GRO</Code><Code>NP:SOF</Code><Code>NP:O</Code></Product><Source><Group>NONE</Group></Source></Entitlements><Data xsi:type="news:NewsDataItem"><newsItem standard="NewsML-G2" standardversion="2.7" conformance="power" guid="20151001-000000000-nS6N0ZH01T-1-1" version="1" xml:lang="en" xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/ NAR\_1.8-spec-All-Power\_2.xsd <http://www.w3.org/1999/xhtml> xhtml1-strict.xsd" xmlns="http://iptc.org/std/nar/2006-10-01/" xmlns:rtr="http://www.reuters.com/ns/2003/08/content" xmlns:x="http://www.w3.org/1999/xhtml"><catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards\_14.xml" /><itemMeta><itemClass qcode="icls:text" /><provider qcode="NS:RTRS" /><versionCreated>2015-10-01T00:00:00.000Z</versionCreated><firstCreated>2015-10-01T00:00:00.000Z</firstCreated><pubStatus qcode="stat:usable" /><rtr:versionedId guid="20151001-000000000-nS6N0ZH01T-1-1" /><TakeSequence>1</TakeSequence></itemMeta><contentMeta><urgency>1</urgency><altId type="idType:USN" rtr:isOriginal="1">nS6N0ZH01T</altId><language tag="en" /><subject qcode="N2:KR" /><subject qcode="N2:EMRG" /><subject qcode="N2:ASIA" /><subject qcode="N2:MCE" /><subject qcode="N2:ECI" /><subject qcode="N2:NEWS1" /><subject qcode="N2:TRD" /><subject qcode="N2:LEN" /><subject qcode="N2:RTRS" /><subject qcode="R:KRW=" /><subject qcode="R:0#KRCOMP1=KQ" /><subject qcode="R:KREXGR=ECI" /><subject qcode="R:KRIMGR=ECI" /><subject qcode="R:KRTBAL=ECI" /><headline>S.KOREA SAYS SEPT EXPORTS -8.3 PCT VS YR EARLIER (REUTERS POLL -10.0 PCT)</headline></contentMeta><contentSet><inlineXML contenttype="application/xhtml+xml"><xhtml xmlns="http://www.w3.org/1999/xhtml"><body></body></xhtml></inlineXML></contentSet></newsItem></Data></ContentItem></Body></ContentEnvelope>

The return in this case is still the first news because it’s the only one that satisfies the criteria on the instruments codes and the time period. If the input for instruments codes would have been “KRW=,KREXGR=ECI,AAPL” the return would still have been the news 1 because it matches one or more instrument codes in the input list. If the input list for instruments codes would have been empty both news would have been returned.

**Case 3:**

Inputs:

* start\_date = “2015-10-01T00:00:00.092Z”
* end\_date = “2015-10-01T00:15:00.000Z”
* instr\_list = “KRW=,KREXGR=ECI”
* tpc\_list=”WASH,VOTE”

Return:

NULL

The return is NULL because there are no news that both satisfy the instruments codes, topics codes and time period that were inputted. If the instruments codes list inputted by the user would have been empty, the News 2 would have been returned because it matches the topics codes inputted for that period but not the instruments codes.

# Additional Information:

Teams have the choice of running their system on two different platforms:

* + Standalone Program
    1. PC running Windows
    2. Unix/Linux platform
  + Web service (accessible via a REST interface)

Throughout the workshop, each team will need to have a Web page. As a minimum, the page is showing:

* The team name and members
* Consecutive releases of their module. Each release page must include a link to download the module and information about:
  + The date and version of the release
  + What has been implemented so far
  + Differences with previous version
  + Clear instructions on how to run the module in standalone mode
  + Guidelines on how to integrate the module with other systems
  + Any test software or data

# References:

What is a RIC:

<http://www.sirca.org.au/2011/01/what-is-a-ric/>

Topics Identifiers:

<https://customers.reuters.com/training/trainingCRMdata/promo_content/ReutersCodes.pdf>