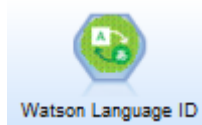


Watson Language ID

Use Watson Service through Bluemix to identify the language of a given text



Product: IBM® SPSS® Modeler

Extension type: Utility

Table of Contents

Description.....	3
Requirements.....	3
Installation.....	3
R Packages used.....	3
User Interface.....	4
Example.....	5
Load file and identify languages.....	5
Results.....	6-7
Important links.....	7
Learn.....	7
Discuss.....	7

Description:

This node allows you to use the Watson Language Translation service on the Bluemix platform from within SPSS Modeler. This extension identifies the language of the input text and provides a two letter language ID for the source language. The output for this node is a data table that contains the original input data with an additional new column for the language ID.

Requirements:

- SPSS Modeler v16.0 or later
- R: <http://www.r-project.org/>
- 'R Essentials for SPSS Modeler' plugin: <https://developer.ibm.com/predictiveanalytics/downloads/>
- A Bluemix account with Watson Language Translation Service activated

Installation:

Close SPSS Modeler. Save the .cfe file in the CDB folder of the IBM SPSS Modeler installation directory for Windows and Linux. The copy should reside in that same folder and not in a sub-folder.

For example, for Windows 7 the default location is "C:\ProgramData\IBM\SPSS\Modeler\16\CDB". If the ProgramData folder is hidden type the path manually.

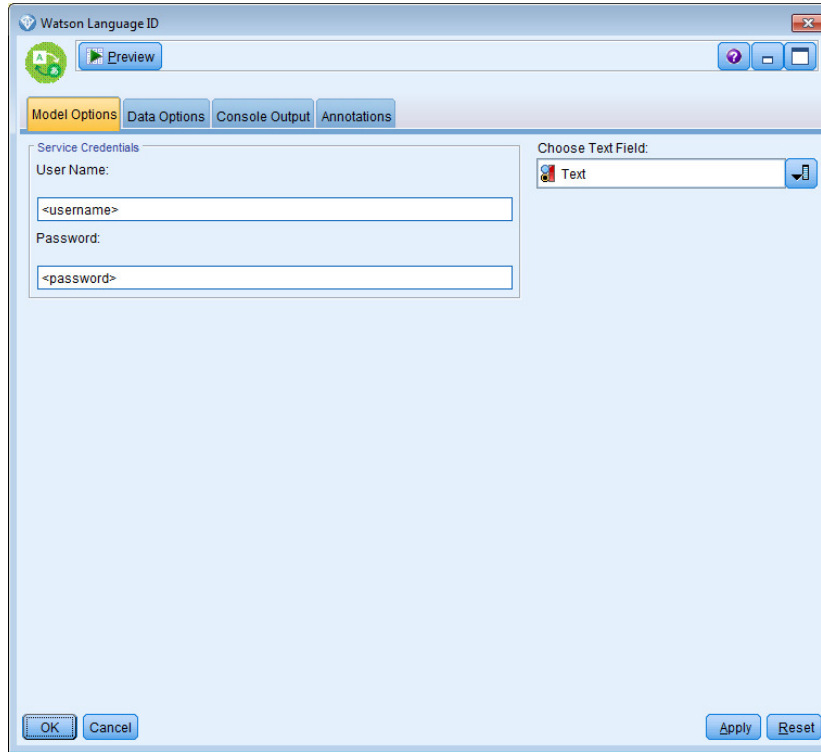
Restart SPSS Modeler: the node will now appear in the Field Ops palette.

R Packages used:

- 'httr' package created by Hadley Wickham Lang <https://cran.r-project.org/web/packages/httr/>
- 'XML' package created by Duncan Temple Lang <https://cran.r-project.org/web/packages/XML/>

User Interface

- Double click on the node to get to the options. There are the following fields:
 - User Name and Password: These are the credentials obtained in Bluemix when Watson Language Translation is added to your account (This is not your Bluemix user name and password).
 - Choose a Text Field: click the drop down to select the column containing the text that needs to be identified



Watson Language ID

Preview

Model Options | Data Options | Console Output | Annotations

Service Credentials

User Name:

Password:

Choose Text Field:

OK Cancel Apply Reset

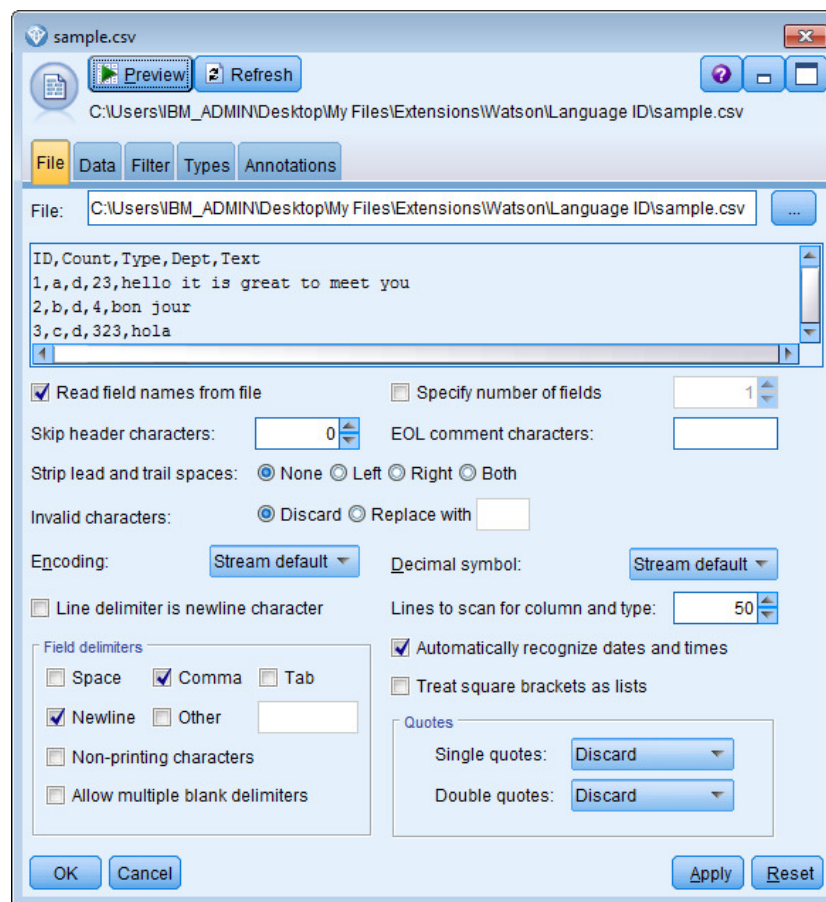


Example

This example will demonstrate how to identify language from text in a CSV file. Text input can be used from any source node, a CSV file is used in this example for demonstration. For this example, the source text will be English, French, and Spanish.

Load file and identify languages

1. In the sources palette, add a Var. File node to the stream
2. Save the “sample.csv” file in the Example directory of this repository and load the file in the Var. File node. This file has sample text from 3 different languages in the Text column.



3. Add the Watson Language ID node to the stream and connect it to the ‘sample.csv’ node

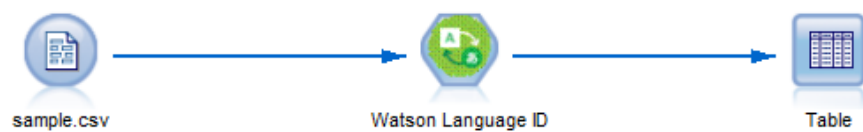


- Now fill the fields:

Choose Text Field: Click the drop down and pick “Text”

User Name and Password: Enter the credentials obtained by adding Watson Translation Service to your Bluemix Dashboard

- From the Output palette, connect a Table to the translation node
- Click on the Table and run the stream



Results

If you used the sample CSV file, your results should match the table below:

Preview from Watson Language ID Node (6 fields, 3 records)

FileEditGenerate

TableAnnotations

	ID	Count	Type	Dept	Text	Language Code
1	1 a	d	23	hello it is great to meet you	en	
2	2 b	d	4	bon jour	fr	
3	3 c	d	323	hola	es	

OK



Notice that one column was added to the original 'sample.csv' data. The Language Code is the two letter identifier for the language of text in the Text column. The table below can be used as a reference for the language codes:

Language	Code
English	En
Spanish	Es
French	Fr
Portugese	Pt
Arabic	Ar

Important Links

Learn

- Learn more about [SPSS software](#).
- Visit [developerWorks Business analytics](#) for more technical analytics resources for developers.
- Create your Bluemix account here: <https://console.ng.bluemix.net/solutions/watson>
- The [Comprehensive R Archive Network](#) is the main site for the R project and each R package. The help pages and manuals that are associated with optimx, nlrmr, and Rcgmin are detailed. Numerous references are provided.
- Read "[Do I need to learn R?](#)" (Catherine Dalzell, developerWorks, September 2013) to learn why R is a valuable tool for data analytics that was expressly designed to reflect the way that statisticians think and work.
- "[Calling R from SPSS](#)" describes how to use R code inside IBM SPSS Modeler 16.
- Read "[Create new nodes for IBM SPSS Modeler 16 using R](#)" to learn how to create new extensions easily.

Discuss

- Visit the [IBM SPSS Community](#) to share tips and experiences with other IBM SPSS developers.
- Follow [developerWorks on Twitter](#) to be among the first to hear about new resources.