

Currently customers do analysis of various market reports on their own or hire experts to make procurement decision. The expert analyses the reports captured from various data source. This can be time consuming and creates dependency on experts, every time a procurement decision should be made. There is also probability of human error if expert misses important factor. e.g. Assume experts needs to analyse various report, different suppliers and their location. If he/she missed to consider some of these factor like plant shutdown and he/she might request a material from the supplier. Now this supplier who has plant down may not able serve the order immediately and that might impact production. By using our intelligent procurement system customer can get such expert analysis faster and more accurate. The customer should initially train the model/system with various use cases(reports). The target end user of this system is person who is working for a procurement section in a company or any other stakeholder who has authority to make procurement decisions.

Watson Knowledge Studio (WKS)

- We build Type System specific to business domain/use case
- Upload the document set (corpus documents).
- Bootstrap the annotation process: The bootstrap process helps to pre-annotate the document. This will reduce human efforts and speeds up human annotation process.
The following are the options to pre-annotate the corpus documents.
 - a. Pre-annotate using Dictionary
 - b. Pre-annotate using Rule
 - c. Pre-annotate using Alchemy Language
 - d. Pre-annotate using Machine-learning (existing wks model)

I will be focussing on a and b options in this article. Before we start pre-annotations, let us understand how to build dictionary and rules.

What is dictionary: Dictionary groups together words and phrases that share something in common. An entry in the dictionary does not mean that all words in the entry mean the same thing, but that the words are to be treated equivalently by an annotator component.

Build Dictionary

Navigate to Pre-annotators menu under Assets and tools. Under Dictionary tab choose Manage Dictionaries → Create Dictionary

a. Create new Dictionary Supplier

The screenshot shows the 'IBM Watson Knowledge Studio' interface. On the left, there's a navigation sidebar with sections like 'Model Management', 'Assets & Tools', 'Documents', 'Entity Types', 'Relation Types', 'Pre-annotators' (which is currently selected), and 'Document Annotation'. The main area is titled 'Manage Dictionaries'. A modal window is open with the title 'Create Dictionary'. Inside the modal, there's a 'Name:' field containing 'Supplier', a 'Cancel' button, and a 'Save' button. Below the modal, there's a list of dictionaries: 'Product_Others' with 9 entries and 'Plant' with 5 entries.

b. Map Entity type and Rule Class

The screenshot shows the 'Manage Dictionaries' interface for the 'Supplier' dictionary. The 'Entity type' dropdown is set to 'Supplier' and the 'Rule Class' dropdown is set to 'Supplier_Dict'. There are 'Add Entry' and 'Upload' buttons. A 'Download' button is also present. A table below shows 'No entries'. The bottom of the screen displays a URL: <https://oatewsv-fra.watsonplatform.net/knowledge-studio/tools/aop/rzhz10/tbxo05wd5du72o56/u/r>.

c. Create new entry Lucite

IBM Watson Knowledge Studio

Manage Dictionaries

Supplier
Language: English | 1 entries

Entity type: Supplier
Rule Class: Supplier_Dict

	Lemma	Surface Forms	Part of Speech	Action
<input checked="" type="checkbox"/>	Lucite	Lucite	Noun	Edit Delete

Add Entry Delete Entries Upload Download

d. Below two screenshots for Supplier and PLantProduction_Down entry list

IBM Watson Knowledge Studio

Manage Dictionaries

Supplier
Language: English | 20 entries

Entity type: Supplier
Rule Class: Supplier_Dict

	Lemma	Surface Forms	Part of Speech	Action
<input type="checkbox"/>	local supplier	local supplier	Noun	Edit Delete
<input type="checkbox"/>	Sinopec Qilu	Sinopec Qilu	Noun	Edit Delete
<input type="checkbox"/>	Sumitomo Chemical	Sumitomo Chemical	Noun	Edit Delete
<input type="checkbox"/>	Fushun	Fushun	Noun	Edit Delete
<input type="checkbox"/>	Jilin Chemical	Jilin Chemical, Jilin	Noun	Edit Delete
<input type="checkbox"/>	Sailboat	Sailboat	Noun	Edit Delete
First 1 Last				

Add Entry Upload Download

We are done with creating dictionaries as required. Let us see how to evaluate these dictionaries.

Navigate to Dictionaries under Document Annotation

- On right side there is class panel, create classes mapping to dictionaries.

- Add sample document

- c. Any references to terms in the dictionaries are annotated with corresponding classes in sample document.

The screenshot shows the workspace interface with the following panels:

- Dictionaries Panel:** Shows a list of dictionaries:
 - Product_Others (highlighted in orange)
 - Plant
 - Supplier
 - Market_Region
 - Price (highlighted in yellow)
 - Supply
 - PlantProductionDown
 - Month
 - Location
- Sample Document Panel:** Displays a text snippet with annotations:

MMA supply continues to be tightened and August prices firme up by US \$50-100 pmt MMA prices regardless of customers' resistance. As several MMA plants unexpected and scheduled shut down during the period between June and August, like Thailand's PTT Asahi Chemicals, South Korea's Lotte Chemical, Taiwan Formosa Plastics Corporation and Japan's Mitsubishi Gas Chemical, MMA producers struggle with low inventories, and are compelled to raise the price along with allocating MMA quantity to regular customer for August shipment. In USA and Europe, it's heard that a major MMA plants suffered from production outage in America and the tight MMA supply did n't release in European. In addition, previously tight supply scenario hasn't eased in China, and the domestic MMA prices were spiraling up during July and August. Thus, the surge MMA demand from Asia boosted Asia MMA
- Class Panel:** Shows a list of classes:
 - Capacity_RegEx (highlighted in yellow)
 - DayofMonth_Dict
 - Location_Dict
 - Market_Dict (highlighted in green)
 - Month_Dict
 - Plant_Dict
 - PlantProductionDown_Dict (highlighted in cyan)
 - Price_Dict (highlighted in orange)
 - Price_RegEx (highlighted in pink)
 - PriceRange_Rule
 - Product_Alternative...
 - Product_Other_Dict (highlighted in orange)
 - ProductofInterest_Dict

- d. Search annotations for class PlantProductionDown_Dict using search options in right class panel (click search lens icon for the class you want to search).

The screenshot shows the workspace interface with the following panels:

- Dictionaries Panel:** Same as the previous screenshot.
- Find Annotations Panel:** Shows results for the class **PlantProductionDown_Dict**:

Results: PlantProductionDown_Dict 20 annotations

Text	Document
... As several MMA plants unexpected and scheduled shut down during the period between June and August, like ...	borica-Aug-2017
... it's heard that a major MMA plants suffered from production outage in America and the tight MMA supply didn't ...	borica-Aug-2017
... LGP) grade remains stable but most PMMA producers cut production rate on the back of short MMA supply ...	borica-Aug-2017
... 's heard that a major 100kpta MMA plant in Shanghai shut down for two week on 14th of July due to ...	borica-Aug-2017
... 100kpta MMA plant in Yeosu, South Korea, unexpected shut down 7days due to technical issues, resulting to loss ...	borica-Aug-2017
... Chemical (MGC) 40kpta MMA plant in Niigata will shut down for maintenance from 12 August to 3 October, ...	borica-Aug-2017
... It's heard a major 70kpta MMA plant cut production rate in July on the back of upstream acrylonitrile (...	borica-Aug-2017
- Class Panel:** Same as the previous screenshot.

Verified with dictionary created earlier.

Let us move to regular expressions. Navigate to Regex under Document Annotations.

- a. Just like mapping class to dictionary using class panel present at right side, we will create class for regular expression. The below screenshot shows we have created class for capacity (Capacity_RegEx), price (Price_RegEx) and Year_RegEx.

IBM Watson Knowledge Studio

Regular Expressions

- Capacity_RegEx
- Capacity_RegEx** (highlighted)
- Price_RegEx
- Year_RegEx_1
- Year_RegEx

Documents

- borica-Aug-2017**

2017/8/2 MARKET OVERVIEW MMA supply continues to be tightened and August prices firme up by US\$50...
- borica_Sept_2017**

2017/9/1 MARKET OVERVIEW MMA market demand is relatively weak though MMA producers have implemented ...

Class

- Capacity_RegEx
- DayofMonth_Dict
- Location_Dict
- Market_Dict
- Month_Dict
- Plant_Dict
- PlantProductionDown...
- Price_Dict
- Price_RegEx
- PriceRange_Rule
- Product_Alternative...
- Product_Other_Dict
- ProductofInterest_Dict

- b. From the sidebar, click Regex, and then click the Create a regular expression icon next to Regular Expressions.

IBM Watson Knowledge Studio

Regular Expressions

Name: Capacity_RegEx

Class:

- Capacity_RegEx
- DayofMonth_Dict
- Location_Dict
- Market_Dict
- Month_Dict
- Plant_Dict
- PlantProductionDown...
- Price_Dict
- Price_RegEx
- PriceRange_Rule
- Product_Alternative...
- Product_Other_Dict
- ProductofInterest_Dict

Regular Expressions

Regular Expression	Minimum Word Tokens	Maximum Word Tokens	Action

Class

- Capacity_RegEx
- DayofMonth_Dict
- Location_Dict
- Market_Dict
- Month_Dict
- Plant_Dict
- PlantProductionDown...
- Price_Dict
- Price_RegEx
- PriceRange_Rule
- Product_Alternative...
- Product_Other_Dict
- ProductofInterest_Dict

c. Click the Add Entry button.

The screenshot shows the 'Regular Expressions' section of the IBM Watson Knowledge Studio. On the left, there's a sidebar with navigation links like 'Back to Workspaces', 'Model Management', 'Assets & Tools', 'Document Annotation' (selected), 'Mentions', 'Relations', 'Coreferences', 'Rules', 'Regex' (selected), 'Dictionaries', 'Settings', and 'Help'. The main area has a search bar 'Enter text to filter expressions' and a table for managing regular expressions. The table columns are 'Regular Expression', 'Minimum Word Tokens', 'Maximum Word Tokens', and 'Action'. Two entries are listed: '([0-9][1,4])(kpta\$)' with tokens 1 and 1, and '([0-9][1,4])(\s+(kpta\$))' with tokens 1 and 2. A modal window titled 'Add Entry' is open, showing the same regular expression and token counts. To the right, a 'Class' sidebar lists various document classes with checkboxes, including 'Capacity_RegEx' (checked and highlighted in yellow), 'Price_RegEx' (highlighted in pink), and 'Year_RegEx' (highlighted in purple).

d. Now regular expression is applied to capacity term 250kkpta marked in yellow background in sample document. Similarly, regular expression is applied to price range in pink background.

This screenshot shows the 'Regular Expressions' interface with a sample document loaded. The sidebar and regular expression management interface are similar to the previous screenshot. The sample document text discusses feedstock prices, mentioning 'MMA' and 'Naphtha' prices stabilizing around US \$400-\$420 pmt, and acetone and methanol prices around US \$690-\$720 pmt. It also mentions propylene and MTBE prices. The document side highlights specific terms: 'Capacity_RegEx' (yellow), 'Price_RegEx' (pink), and 'Year_RegEx' (purple). The right-hand 'Class' sidebar shows the same list of document classes as the previous screenshot, with the same checked status for 'Capacity_RegEx' and 'Price_RegEx'.

Done with regular expression.

Let us navigate to Rules under Document Annotation. Here we will use dictionary and regular expression classes created to build rules.

- a. Select tokens highlighted with blue rectangles, required to build rule.

The screenshot shows the IBM Watson Knowledge Studio interface. On the left, there's a sidebar with navigation links like 'Back to Workspaces', 'Model Management', 'Assets & Tools', 'Document Annotation', 'Rules' (which is selected), 'Regex', 'Dictionaries', 'Settings', and 'Help'. The main area displays a document titled 'borica-Aug-2017' with several words highlighted with blue rectangles, such as 'MMA', 'supply', 'August', 'prices', 'June', 'August', 'Thailand', 'PTT', 'Asahi Chemicals', 'South Korea', 'Lotte Chemical', 'Taiwan', 'Formosa', 'Plastics Corporation', 'Japan', 'Mitsubishi Gas Chemical', 'MMA', 'price', 'plants', 'outage', 'America', 'tight', 'Europe', 'Chinese', and 'China'. To the right of the document is a sidebar titled 'Class' with a list of checked classes: Capacity_RegEx, DayofMonth_Dict, Location_Dict, Market_Dict, Month_Dict, Plant_Dict, PlantProductionDown_Dict, Price_Dict, Price_RegEx, PriceRange_Rule, Product_Alternative_Dict, Product_Other_Dict, and ProductofInterest_Dict.

- b. Select all text boxes using shift key and map the rule to relevant class.

This screenshot shows the 'Create a rule' dialog box. In the center, there's a large text area containing the document text with highlighted tokens. A 'Assign class' dropdown menu is open over the word 'Capacity_RegEx'. The menu lists various classes: Capacity_RegEx (selected), DayofMonth_Dict, Location_Dict, Market_Dict, Month_Dict, Plant_Dict, PlantProductionDown_Dict, Price_Dict, Price_RegEx, PriceRange_Rule, Product_Alternative_Dict, Product_Other_Dict, and ProductofInterest_Dict. To the right of the dialog is a sidebar titled 'Class' with a list of checked classes: Capacity_RegEx, DayofMonth_Dict, Location_Dict, Market_Dict, Month_Dict, Plant_Dict, PlantProductionDown_Dict, Price_Dict, Price_RegEx, PriceRange_Rule, Product_Alternative_Dict, Product_Other_Dict, and ProductofInterest_Dict.

c. Map the class name

The screenshot shows the 'Create a rule' interface in IBM Watson Knowledge Studio. On the left, a sidebar lists various workspace categories like Model Management, Assets & Tools, Document Annotation, and Rules. Under the 'Rules' category, a list of existing rules is shown, including 'PriceRange.Rule1', 'PriceRange.Rule2', 'PriceRange.Rule3', 'TimePeriod.Rule1', 'TimePeriod.Rule2', 'TimePeriod.Rule3', 'TimePeriod.Rule4', and 'PriceRange.Rule5'. In the center, the 'Create a rule' panel displays a visual pattern editor. A pink box labeled 'Price_RegEx' is connected to several other tokens: 'US' (with a checkmark and edit icon), a dollar sign '\$', a gray circle '100', and 'pmt' (with a checkmark and edit icon). Below the editor, a legend defines symbols: a circle with a 'C' for 'Repeat Setting', a checkmark for 'Required (Exactly 1)', a plus sign for 'Repeating 1 or more times', an asterisk for 'Repeating 0 or more times', a question mark for 'Occurring 0 or 1 time', and a checkmark with a minus sign for 'Advanced: Exactly 1'. On the right, a 'Class' sidebar lists 17 available classes with checkboxes, including Capacity_RegEx, DayofMonth_Dict, Location_Dict, Market_Dict, Month_Dict, Plant_Dict, PlantProductionDow..., Price_Dict, Price_RegEx, PriceRange.Rule, Product_Alternative..., Product_Other.Dict, and ProductofInterest_Dict.

d. Provide Rule name

This screenshot is identical to the one above, showing the 'Create a rule' interface in IBM Watson Knowledge Studio. It displays the same workspace sidebar, rule list, visual pattern editor for 'PriceRange.Rule', and class sidebar. The only difference is that the 'PriceRange.Rule' pattern has been partially expanded to show its internal structure with tokens like 'Price_RegEx', 'US', '\$', '100', and 'pmt'.

e. You find rule is applied to US\$50~100 pmt.

borica-Aug-2017

2017/8/2 MARKET OVERVIEW

MMA supply continues to be tightened and August prices firm up by US \$50-100 pmt MMA prices regardless of customers' resistance.

As several MMA plants unexpected and scheduled shut down during the period between June and August, like Thailand's PTT Asahi Chemicals, South Korea's Lotte Chemical, Taiwan Formosa Plastics Corporation and Japan's Mitsubishi Gas Chemical, MMA producers struggle with low inventories, and are compelled to raise the price along with allocating MMA quantity to regular customer for August shipment.

In USA and Europe, it's heard that a major MMA plants suffered from production outage in America and the tight MMA supply didn't release in European.

In addition, previously tight supply scenario hasn't eased in China.

Class

- Capacity_RegEx
- DayofMonth_Dict
- Location_Dict
- Market_Dict
- Month_Dict
- Plant_Dict
- PlantProductionDow...
- Price_Dict
- Price_RegEx
- PriceRange_Rule
- Product_Alternative...
- Product_Other_Dict
- ProductofInterest_Dict

f. From the sidebar, select Model Management > Versions and click the Rule-based model type mapping tab. Map the entity to relevant class as shown below.

Entity Type Name	Class Name	Actions
Facility_Production_Down	PlantProductionDown_Dict	Edit
Capacity	Capacity_RegEx	Edit
Facility	Plant_Dict	Edit
Supplier	Supplier_Dict	Edit
Market	Market_Dict	Edit
Demand		Edit
Supply	Supply_Dict	Edit
Time_Period		Edit

Now let us back to pre-annotation. We will see how to use dictionary, regular expression and rules in bootstrapping the annotation process.

Let us consider a new document and follow the pre-annotation steps.

1. Upload new document

The screenshot shows the IBM Watson Knowledge Studio interface. On the left, there's a sidebar with options like 'Back to Workspaces', 'Model Management', 'Assets & Tools' (selected), 'Documents', 'Entity Types', 'Relation Types', 'Pre-annotators', 'Document Annotation' (selected), 'Settings', and 'Help'. The main area has tabs for 'Documents' and 'Document Sets (8)'. A central modal window titled 'Add a Document Set' contains a file upload area with a placeholder 'Drag files here, or click to browse for the files that you want to add.' Below it, instructions say 'For best performance, individual documents should not contain more than 2000 words.' and provide links to 'Hide details' and 'CSV file', 'Text files', and 'ZIP file' formats. To the right of the modal is a table of 'Last Modified' document sets, with columns for 'Name', 'Last Modified', 'Action' (Rename, Delete), and a 'Download Document Sets' button.

2. Create annotation set

The screenshot shows the 'Create Annotation Sets' dialog. It includes fields for 'Base set' (set to 'borica_Dec_2017.txt_set'), 'Overlap' (set to '100 % of 1 documents'), and 'New Sets' (with a '+ Add another set and human annotator' button). Below these, a note says: 'To view inter-annotator agreement scores, assign at least two human annotators and ensure that a percentage of documents overlap between the sets.' Further down are fields for 'Annotator' (set to 'Rajesh Gudikoti') and 'Set name' (set to 'Dec-DocSet'). At the bottom are 'Cancel' and 'Generate' buttons.

3. Navigate to Pre-annotators under Assets and Tool and apply Dictionary Pre-Annotator for annotation set we created above.

IBM Watson Knowledge Studio

Pre-annotators

Dictionaries Natural Language Understanding AlchemyLanguage

A dictionary annotator creates an annotation for each mention in the text that matches a term in a dictionary. Dictionaries must be mapped to a type or class in order to apply an annotation.

[Apply This Pre-annotator](#)

Dictionaries and Mapping

Manage Dictionaries Upload Dictionary Download Dictionaries

Dictionary Name	Entity Type	Class
Product_Others	OtherProduct	Product_Other_Dict
Plant	Facility	Plant_Dict
Supplier	Supplier	Supplier_Dict
Market_Region	Market	Market_Dict
Price	Price	Price_Dict

4. Select the annotation set

IBM Watson Knowledge Studio

Run Annotator

Close Run

Documents

Select the document sets or annotation sets that you want to pre-annotate.

- borica_Jul-Aug_2017.txt_set
- borica_Sep-Oct_2017.txt_set
- borica_Jun_2017.txt_set
- dem.txt_set
- Jul-Aug-DocSet
- Sep-Oct-DocSet
- Jun-DocSet
- borica_Dec_2017.txt_set
- Dec-DocSet

5. Success message post running the pre-annotator

The screenshot shows the 'Pre-annotators' section of the IBM Watson Knowledge Studio interface. On the left, there's a sidebar with navigation links like 'Back to Workspaces', 'Model Management', 'Assets & Tools', 'Document Annotation', 'Settings', and 'Help'. The main area is titled 'Pre-annotators' and has tabs for 'Dictionaries', 'Natural Language Understanding', and 'AlchemyLanguage'. A success message 'Success: Dictionary annotator application complete.' is displayed at the top right. Below it, there's a note about dictionary annotators and an 'Apply This Pre-annotator' button. The 'Dictionaries and Mapping' section contains buttons for 'Manage Dictionaries', 'Upload Dictionary', and 'Download Dictionaries'. A table lists dictionaries with their entity types and classes:

Dictionary Name	Entity Type	Class
Product_Others	OtherProduct	Product_Other_Dict
Plant	Facility	Plant_Dict
Supplier	Supplier	Supplier_Dict
Market_Region	Market	Market_Dict
Price	Price	Price_Dict

6. Navigate to Versions menu under Model Management

7. Select Rule-based tab

The screenshot shows the 'Versions' section of the IBM Watson Knowledge Studio interface. The sidebar includes 'Back to Workspaces', 'Model Management' (selected), 'Performance', 'Versions' (selected), 'Assets & Tools', 'Document Annotation', 'Settings', and 'Help'. The main area has tabs for 'Machine Learning', 'Rule-based' (selected), and 'Rule-based model type mapping'. There are two boxes: one for running a rule-based model and another for exporting the current version. Below is a 'Version History and Deployment' table:

Model	Creation Date	Description	Action	Status
Current	-		Save for Deployment	N/A

8. Select the annotation set

The screenshot shows the 'Run Annotator' dialog. At the top is a title 'Run Annotator' with 'Close' and 'Run' buttons. Below is a section titled 'Documents' with the sub-instruction 'Select the document sets or annotation sets that you want to pre-annotate.' A list of annotation sets is shown with checkboxes:

- borica_Jul-Aug_2017.txt_set
- borica_Sep-Oct_2017.txt_set
- borica_Jun_2017.txt_set
- dem.txt_set
- Jul-Aug-DocSet
- Sep-Oct-DocSet
- Jun-DocSet
- borica_Dec_2017.txt_set
- Dec-DocSet

9. Success message post running the pre-annotator

The screenshot shows the 'Versions' section of the IBM Watson Knowledge Studio interface. At the top right, there is a green success message box that says 'Success: Jan 17 4:32 Rule-based annotator application complete.' On the left, the navigation menu includes 'Back to Workspaces', 'Model Management', 'Assets & Tools', 'Document Annotation', 'Settings', and 'Help'. Under 'Model Management', 'Rule-based' is selected. In the center, there are two main sections: one for 'Machine Learning' and one for 'Rule-based'. The 'Rule-based' section contains a note about rule-based models being used independently or as part of training a machine learning model, with a 'Run this model' button. Another section allows exporting the current version of the model to other Watson applications, with an 'Export current model' button. Below these is a table for 'Version History and Deployment'.

Model	Creation Date	Description	Action	Status
Current	-		Save for Deployment	N/A

10. Create task for the annotation created above

The screenshot shows the 'Tasks' tab in the 'Documents' section of the IBM Watson Knowledge Studio. The left sidebar has 'Documents' selected under 'Assets & Tools'. The main area shows three tasks: 'Jun-Jul-Aug-Task' and 'Sep-Oct-Task', both with 100% progress. A new task creation dialog is open on the left, titled 'Add Task', with 'Dec-Task' entered in the 'Task name' field and '01/17/2018' in the 'Deadline' field. Buttons for 'Cancel' and 'Create' are at the bottom of the dialog.

11. Add annotation set to task.

The screenshot shows the 'Add Annotation Sets to the Task' dialog. It lists four annotation sets: 'Jul-Aug-DocSet', 'Sep-Oct-DocSet', 'Jun-DocSet', and 'Dec-DocSet'. The 'Dec-DocSet' row is highlighted with a blue checkmark in the first column. Buttons for 'Cancel' and 'Create Task' are at the top right.

Annotation Set Name	Annotator Name	Task Name
Jul-Aug-DocSet	Rajesh Gudikoti	Jun-Jul-Aug-Task
Sep-Oct-DocSet	Rajesh Gudikoti	Sep-Oct-Task
Jun-DocSet	Rajesh Gudikoti	Jun-Jul-Aug-Task
Dec-DocSet	Rajesh Gudikoti	

12. Start annotation clicking annotate.

IBM Watson Knowledge Studio

Back to Workspaces

Model Management

Assets & Tools

Documents

Entity Types

Relation Types

Pre-annotators

Document Annotation

Settings

Help

Documents

Document Sets (9) Annotation Sets (4) Documents (All, 8) Tasks

Go Back to Task List

Dec-Task

Deadline: 01/17/2018

Annotations added to annotation sets are not considered ground truth until the annotation sets are submitted and accepted. When an annotation set is accepted, documents that are annotated in only one annotation set immediately become ground truth. Documents that are annotated in two or more annotation sets become overlapping documents that will become ground truth after conflicts are resolved.

Edit Annotation Sets Refresh Calculate Inter-Annotator Agreement Check Overlapping Documents for Conflicts Apply Type System Updates

Annotation Set Name	Annotator Name	Status	Action
Dec-DocSet	Rajesh Gudikoti	IN PROGRESS	Annotate

13. click the document from the list.

Change Task

Dec-Task

Annotation Set

Dec-DocSet

Showing 1-1 of 1

Document Name	Status	Last Modified
borica_Dec_2017.txt	Ready	2018-01-17 12:46:06 +0530

Submit All Documents

14. Below screen shows most of occurrences of mentions are identified within the document. Now a human annotator can begin work on documents that are pre-annotated and complete rest of annotation process.

IBM Watson Knowledge Studio

Back to Workspaces

View Details Replace Concordance Attribute View Ready Save

Alphab... 14pt 1

Mentions

Relations

Coreferences

Rules

Regex

Dictionaries

Settings

Help

borica_Dec_2017.txt

1 2017/12/29 MARKET OVERVIEW

2 In January 2018, MMA supply continues to be tightened and MMA producers announced US\$30~50 pmt MMA prices hike regardless of customers' resistance and relative weakness of MMA market demand for upcoming year-end holidays.

3 Several MMA plants scheduled shut down during the period between January and February, for example, Evonik 100kpta MMA plant in Shanghai, China will shut down for 10 days from 30th January, Formosa Plastics Corp (FPC) 98kpta MMA plant in Mailiao, Taiwan will shut down from 12th January to 6th February, and Mitsui Chemicals 40kpta MMA plant in Sakai, Japan will shut down from 10th to 20th January.

4 Meanwhile, it's heard that Saudi Methacrylate SAMAC (JV between Mitsubishi Chemical and market newcomer SABIC) 250kpta MMA plant at Al Jubail, Saudi Arabia can't fully operate till December and runs at low production rate.

5 Thus, there is no bulk vessel to Asia till January.

6 In USA and Europe, it's heard that a major MMA plant suffered from the shortage of HCN supply and kept low production rate during December and January.

Type	Subtype	Role
a	Alternate_Product	
n	Attribute_Number	
c	Capacity	
-	Currency	
-	Date	
-	DayofMonth	
D	Demand	
-	Demand_Attribute_Negative	
-	Demand_Attribute_Neutral	
-	Demand_Attribute_Positive	
-	Downstream_products	
f	Facility	
7	Facility_Production_Down	
8	Facility_Production_Up	

15. Once human annotation process is complete, all documents should be submitted and accepted. You can refer below to know more about human annotation process.

<https://console.bluemix.net/docs/services/watson-knowledge-studio/annotate-documents.html#annotate-documents>

16. We create machine learning model and train the model till we are satisfied with model.

a. Click Train and evaluate.

procurement-borica-2

Language of documents: English

3 Training Set
1 Test Set
1 Blind Set

Training Set: View Ground Truth | Test Set: View Ground Truth | View Decoding Results

Last trained on: Jan 12, 2018 5:48 PM | Last evaluated on: Jan 12, 2018 5:50 PM

Train and evaluate | Low performance? Click here to train.

b. Click the edit settings.

Training / Test / Blind Sets

Last trained on: Jan 12, 2018 5:48 PM
Last evaluated on: Jan 12, 2018 5:50 PM

Train | Evaluate | Train & Evaluate | Edit Settings

Set Name	Number of Documents	Action
Training Set	3	View Ground Truth
Test Set	1	View Ground Truth View Decoding Results
Blind Set	1	

c. Choose the annotation set to train/evaluate/train-evaluate model.

IBM Watson Knowledge Studio

Last trained on: Jan 12, 2018 5:48 PM
Last evaluated on: Jan 12, 2018 5:50 PM

[Train](#) [Evaluate](#) [Train & Evaluate](#)

[Go Back to View Mode](#)

⚠ Some tasks aren't completed yet. Machine learning uses current ground truth. The ground truth can change when a task completes.

Document Set	Task Status
All	
borica_Jul-Aug_2017.txt_set	
borica_Sep-Oct_2017.txt_set	
borica_Jun_2017.txt_set	
dem.txt_set	
Jul-Aug-DocSet	COMPLETED
Sep-Oct-DocSet	COMPLETED
Jun-DocSet	COMPLETED

Create new sets by splitting the selected document sets

Ratio
Enter the percentage of documents to include in each set.

70	Training Set (70% Recommended)
23	Test Set (23% Recommended)
7	Blind Set (7% Recommended)

Add documents in the selected sets to:

Set Name	Number of Documents
Training Set	3

d. Training and evaluation process in progress.

IBM Watson Knowledge Studio

Performance

procurement-borica-2

Language of documents
English

3 Training Set
1 Test Set
1 Blind Set

Training Set [View Ground Truth](#) [View Decoding Results](#)

Test Set [View Ground Truth](#) [View Decoding Results](#)

Last trained on:
Jan 12, 2018 5:48 PM

Last evaluated on:
Jan 12, 2018 5:50 PM

[Train and evaluate](#) Low performance? Click here to train.

e. Training and evaluation process complete

IBM Watson Knowledge Studio

Performance

procurement-borica-2

Language of documents
English

3 Training Set
1 Test Set
1 Blind Set

Training Set [View Ground Truth](#) [View Decoding Results](#)

Test Set [View Ground Truth](#) [View Decoding Results](#)

Success: Jan 17 5:31
Machine learning model evaluation complete.

Last trained on:
Jan 17, 2018 5:29 PM

Last evaluated on:
Jan 17, 2018 5:31 PM

[Train and evaluate](#) Low performance? Click here to train.

Document set evaluation [i](#)

Model over time

[View Log](#)

Mention: .84 Precision: .88 Recall: .80

Relation: .51 Precision: .60 Recall: .43

17. Deploy the machine learning model

We can deploy the model created to Alchemy Discovery or NLU services. Below screenshots shows deployment to Discovery service.

- Take a snapshot

The screenshot shows the IBM Watson Knowledge Studio interface. The top navigation bar includes 'IBM Watson Knowledge Studio', a user icon, and the email 'ragudiko@in.ibm.com'. On the left, there's a sidebar with 'Back to Workspaces', 'Model Management' (selected), 'Performance', 'Versions' (selected), 'Assets & Tools', 'Document Annotation', 'Settings', and 'Help'. The main content area is titled 'Versions' and has tabs for 'Machine Learning' (selected), 'Rule-based', and 'Rule-based model type mapping'. A central panel contains text about machine learning models and buttons for 'Run this model' and 'Export current model'. Below this is a table titled 'Version History and Deployment' with columns: Version, Base, Creation Date, Entity Types, Relation Types, Description, Action, and Status. It lists three versions: 1.2 (Current Version), 1.1 (01/17/2018, Sample), and 1.0 (01/12/2018, Model used for article). Actions like 'Take Snapshot', 'Promote', 'Delete', 'Deploy', 'Undeploy', and 'WDS' are available for each version.

- Choose Deploy the model option

Deploy Model v1.1

Select a service to deploy to.

All services require a subscription. [Learn more](#)

- Natural Language Understanding
- Discovery
- AlchemyLanguage

- c. Choose space and Discovery service instance

Deploy Model v1.1

Deploying to Discovery

You must have a subscription to the IBM Watson™ Discovery service, and know the names of your IBM Cloud space and service instance. [Learn more](#)

IBM Cloud Information

IBM Cloud is the IBM cloud platform. Click [here](#) to open IBM Cloud and create an account or look up details for an existing service.

Region

Germany

Space

devge

Service instance

procurement

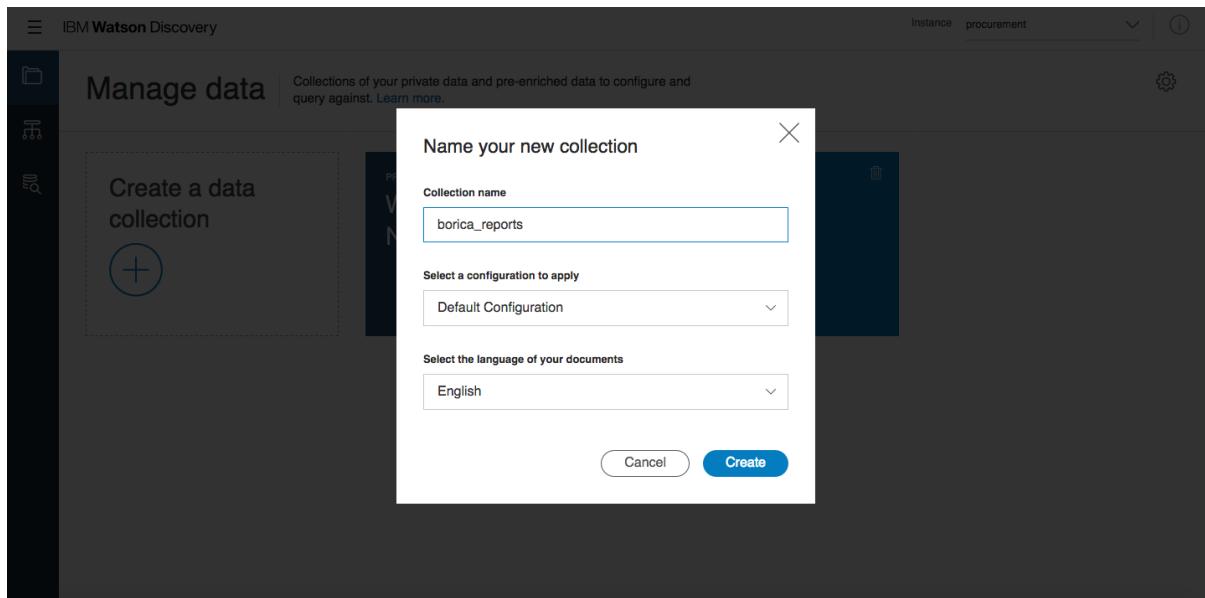
[Cancel](#)

[Deploy](#)

Discovery Service

- Login into IBM cloud and create discover instance.
- We create collection with customized configuration which points to wks model id.

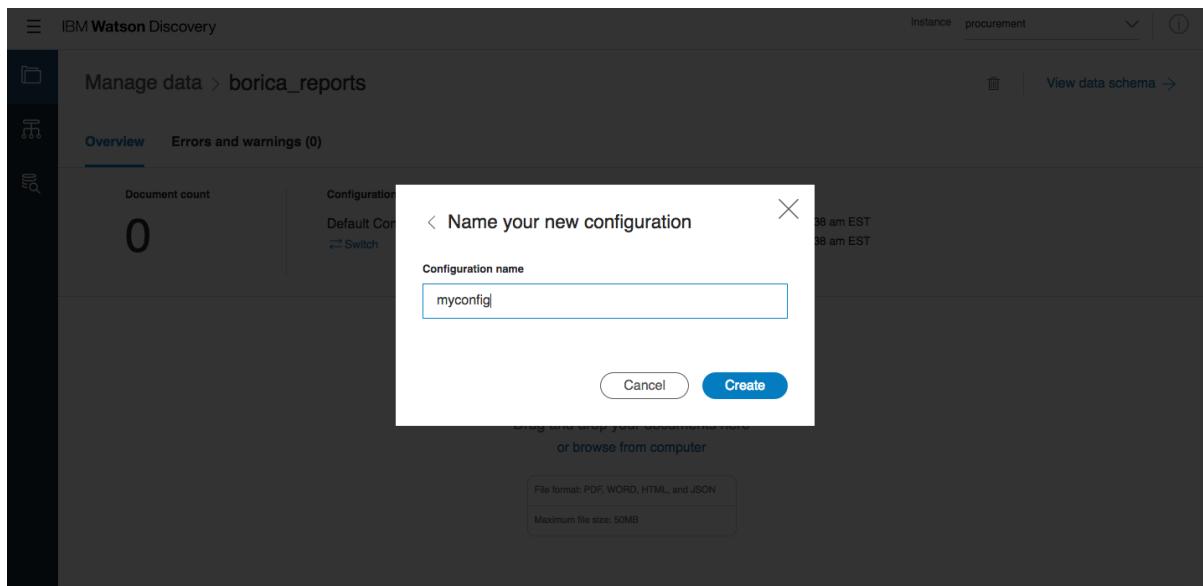
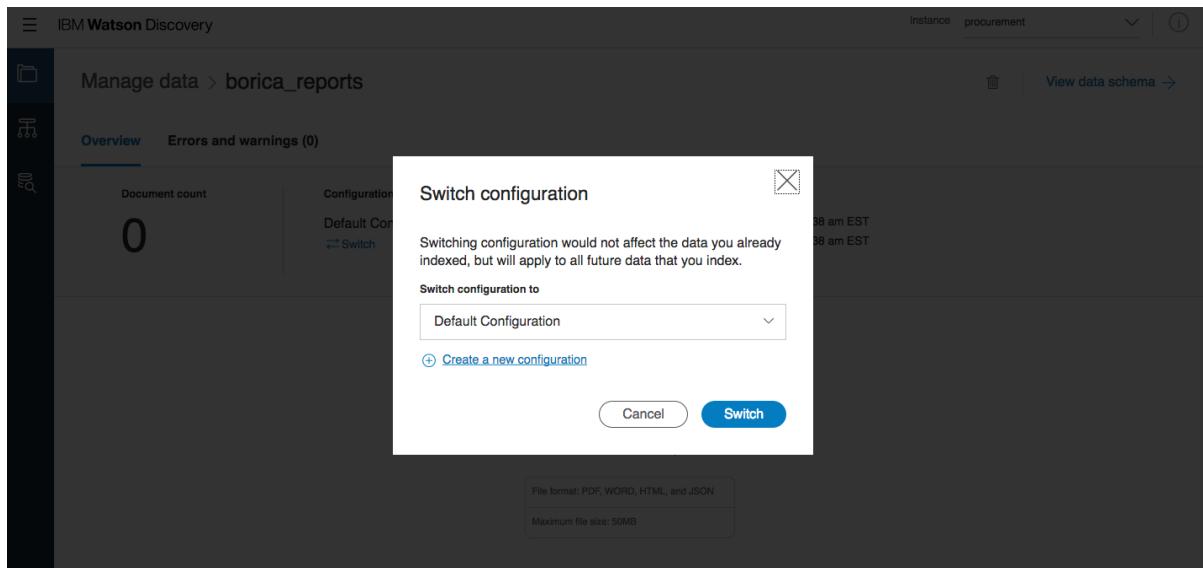
a. Create new collection for



b. Click switch link

The screenshot shows the 'Overview' tab for the 'borica_reports' collection. The 'Document count' is 0. The 'Configuration' section shows 'Default Configuration' with a 'Switch' link. The 'Collection info' section shows the collection was created on 1/21/2018 10:50:38 am EST and last updated on 1/21/2018 10:50:38 am EST. Below this, there is a cloud-up arrow icon and a placeholder text 'Drag and drop your documents here or browse from computer'. A note at the bottom specifies file formats (PDF, WORD, HTML, and JSON) and maximum file size (50MB).

c. Create new configuration



d. Add enrichments

IBM Watson Discovery

myconfig

Convert Enrich Normalize

Enrich your data with additional Watson insights
Set up rules for which fields you want to apply enrichments to. [Learn more.](#)

Add a field to enrich Choose a field

Fields to be enriched	Enrichments
text	entities X sentiment X categories X concepts X

+ Add enrichments

Last saved on: 1/21/2018 10:51:44 am EST

Apply & Save Close

Upload sample documents
You can preview your configuration by using sample documents that are representative of your data collection. [Learn more.](#)

Drag and drop up to 10 documents or browse from computer

Sample documents expire after 1 month
File format: PDF, WORD, HTML, and JSON
Maximum file size: 5MB

e. Configure wks model id for entity and relation extraction

IBM Watson Discovery

myconfig

Convert Enrich Normalize

Enrich your data with additional Watson insights
Set up rules for which fields you want to apply enrichments to.

Add a field to enrich Choose a field

Fields to be enriched
text

Concept Tagging
Identifies general concepts that aren't necessarily directly referenced in this field.
[Learn more](#) Added!

Category Classification
Classifies this field into a hierarchy of categories that's five levels deep.
[Learn more](#) Added!

Semantic Role Extraction
Parses sentences into subject, action, and object form and returns additional semantic information.
[Learn more](#) Add

Emotion Analysis
Analyzes the emotions (anger, disgust, fear, joy, and sadness) in this field.
[Learn more](#) Add

Entity Extraction
Extracts people, companies, organizations, cities, geographic features, and more from this field. You can also create and add custom entity models with Watson Knowledge Studio.
[Learn more](#) Added!
Custom Model ID jfcb-4548-abf9-9a6951cd2d5d Apply

Relation Extraction
Recognizes when two entities are related and identifies the type of relation. You can also create and add custom relation models with Watson Knowledge Studio.
[Learn more](#) Added!
Custom Model ID jfcb-4548-abf9-9a6951cd2d5d Apply

Last saved on: 1/21/2018 10:51:44 am EST

Upload sample documents
You can preview your configuration by using sample documents that are representative of your data collection. [Learn more.](#)

Drag and drop up to 10 documents or browse from computer

Sample documents expire after 1 month
File format: PDF, WORD, HTML, and JSON
Maximum file size: 5MB



View data schema | Collection: borica

Build queries →

Collection view Document view

Documents

- borica-dec.html >
- borica-nov.html >
- borica-aug.html** >

enriched_text.relations.type	isaSupplierOf
enriched_text.relations.sentence	It's heard that Lucite 170kpta MMA plant in Shanghai cut 20~30% production rate in July owing to upstream Secco acrylo-nitrile (AN) plant reduced running rate.
enriched_text.relations.score	0.60073
enriched_text.relations.arguments.0.text	Lucite
enriched_text.relations.arguments.0.location.0	2170
enriched_text.relations.arguments.0.location.1	2176
enriched_text.relations.arguments.0.entities.0.type	Supplier
enriched_text.relations.arguments.0.entities.0.text	Lucite
enriched_text.relations.arguments.0.entities.0.disambiguation.subtype.0	NONE
enriched_text.relations.arguments.1.text	MMA