

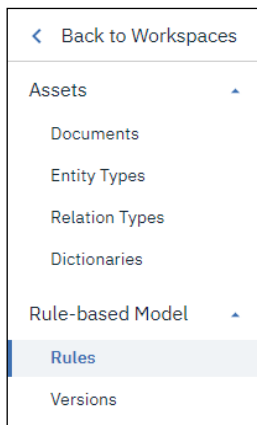
# Watson Knowledge Studio Rule Lab

**Step 1: Provision your Knowledge Studio service instance and create a workspace**

**[Follow the instructions in Appendix A](#)**

**Step 2: Overview of the core UI components**

Select [Rule-based Model] – [Rules] in left navigation pane to open the Rules page.



This view serves to give an overview of the core UI components:



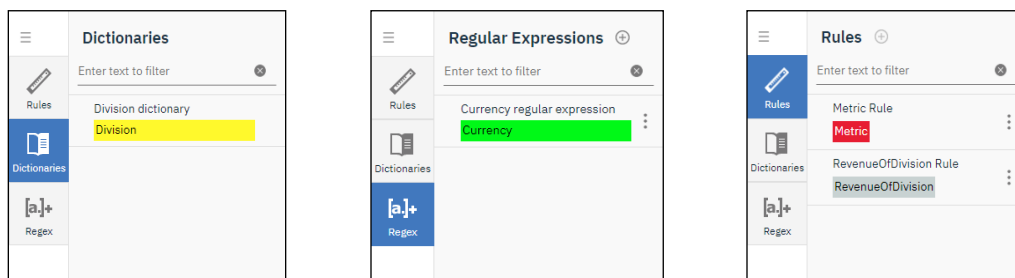
There are three tools that can be used to create a search pattern.

- **Dictionaries** - Any word that is found from the dictionary is annotated with the assigned class.

**Regular expressions (Regex)** - A regular expression is a sequence of characters that defines a search pattern. A basic example is `[A-Z][a-z]*` which finds capitalized words. Any sequence of characters that is found by the regular expression is annotated with the assigned class.

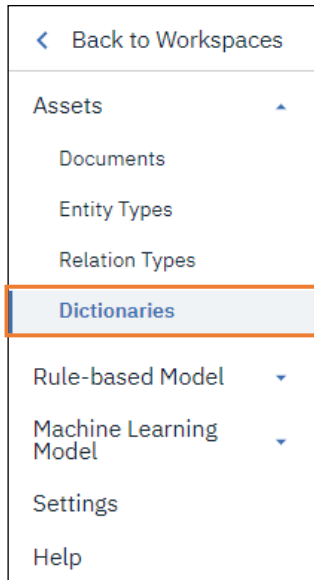
- **Rules** - A rule is complex search pattern which is defined by a particular sequence of one or more annotations that are generated by dictionaries, regular expressions, tokens (such as words and their corresponding parts of speech), punctuation, numbers, or other rules. Any sequence of annotations that is found by the rule is annotated with the assigned class.

Each tool can be edited by selecting it from the tab container on the leftmost side of the tool pane:

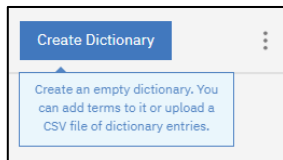


### Step 3: Upload the “Division” dictionary

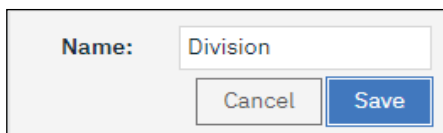
1. Select [Assets] – [Dictionaries] in left navigation pane to open Dictionaries page.



2. Click the **Create Dictionary** button placed at the top of the left pane.



3. Enter “**Division**” in the Name field and then click **Save**.

A screenshot of a form for creating a dictionary. It has a label 'Name:' followed by a text input field containing the word 'Division'. Below the input field are two buttons: 'Cancel' and 'Save'.

4. Click the **Upload** button in the main pane, and upload the “**dictionary\_division.csv**” file from the lab zip.

Dictionary

Create Dictionary

Division 0

Language: English | 0 entry

Entity type: None Rule class: None

Add Entry Upload Add a CSV file that contains dictionary entries. Download

Add a dictionary term.

☐ Lemma Surface Forms Part of Speech Action

No entries

Upload Dictionary Entries

Drag a file here, or click to browse for the file that you want to upload.  
You can upload any CSV file that contains dictionary terms.  
Maximum file size: 1MB

Close Upload

5. Make sure that the following 5 entries were listed in the table after uploading is completed.

Division

Language: English | 5 entries

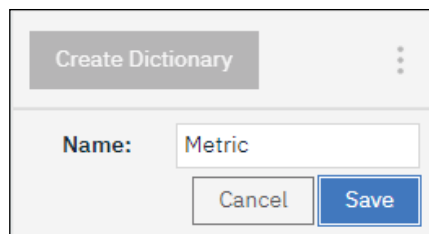
Entity type: None Rule class: None

Add Entry Upload Download

<input type="checkbox"/>	Lemma	Surface Forms	Part of Speech	Action
<input type="checkbox"/>	GTS	GTS, Global Technology Services	Noun	<a href="#">Edit</a> <a href="#">Delete</a>
<input type="checkbox"/>	GBS	GBS, Global Business Services	Noun	<a href="#">Edit</a> <a href="#">Delete</a>
<input type="checkbox"/>	Hardware	Hardware	Noun	<a href="#">Edit</a> <a href="#">Delete</a>
<input type="checkbox"/>	Software	Software	Noun	<a href="#">Edit</a> <a href="#">Delete</a>
<input type="checkbox"/>	S&TG	S&TG, Systems and Technology	Noun	<a href="#">Edit</a> <a href="#">Delete</a>

## Step 4: Manually create the “Metric” dictionary

1. Click **Create Dictionary** button placed at the top of left side pane again, and enter “**Metric**” in Name field, then click **Save** button.

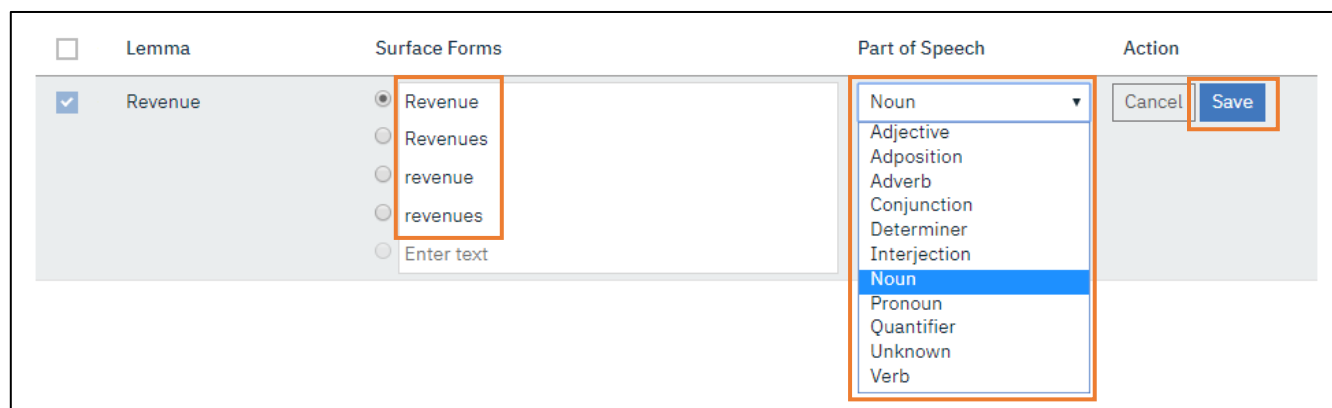


Create Dictionary

Name: Metric

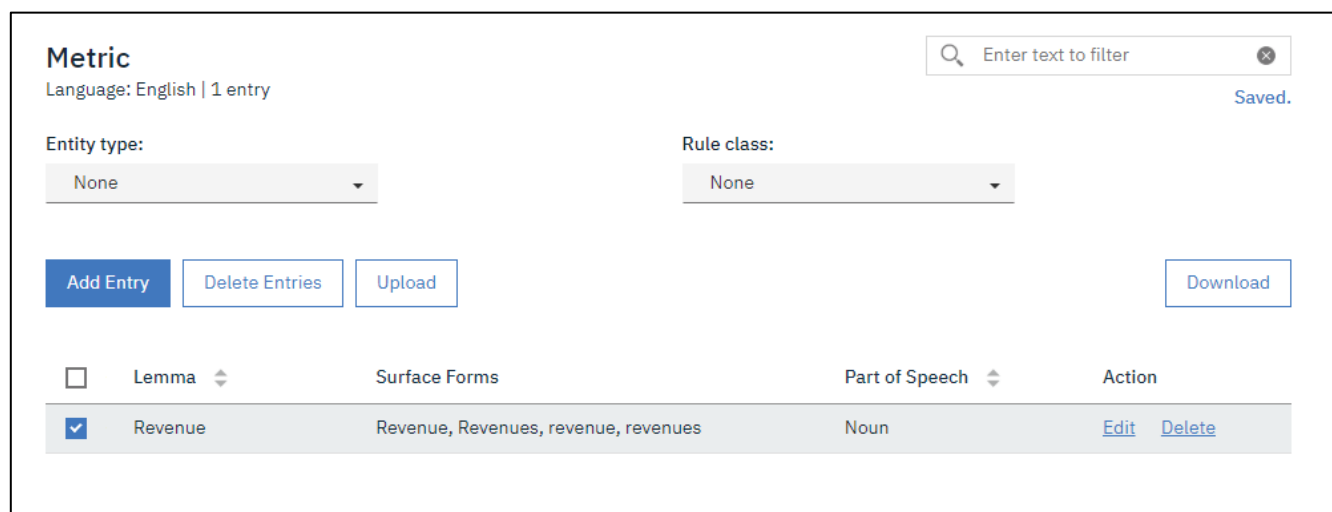
Cancel Save

2. In right side pane, click **Add Entry** button, then enter “**Revenue**”, “**Revenues**”, “**revenue**” and “**revenues**” in Surface Forms field and select “**Noun**” from dropdown menu in Part of Speech column. After that, click **Save** button.



<input type="checkbox"/> Lemma	Surface Forms	Part of Speech	Action
<input checked="" type="checkbox"/> Revenue	<input checked="" type="radio"/> Revenue <input type="radio"/> Revenues <input type="radio"/> revenue <input type="radio"/> revenues <input type="radio"/> Enter text	Noun Adjective Adposition Adverb Conjunction Determiner Interjection Noun Pronoun Quantifier Unknown Verb	Cancel Save

3. Make sure that “Revenue”, “Revenues”, “revenue” and “revenues” are added into Metric dictionary correctly.



**Metric**  
Language: English | 1 entry

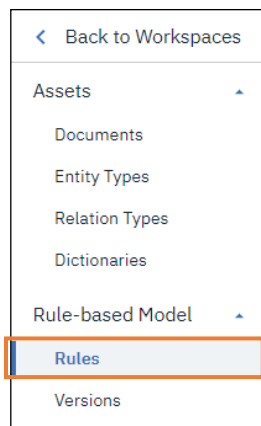
Entity type: None Rule class: None

Add Entry Delete Entries Upload Download

<input type="checkbox"/> Lemma	Surface Forms	Part of Speech	Action
<input checked="" type="checkbox"/> Revenue	Revenue, Revenues, revenue, revenues	Noun	<a href="#">Edit</a> <a href="#">Delete</a>

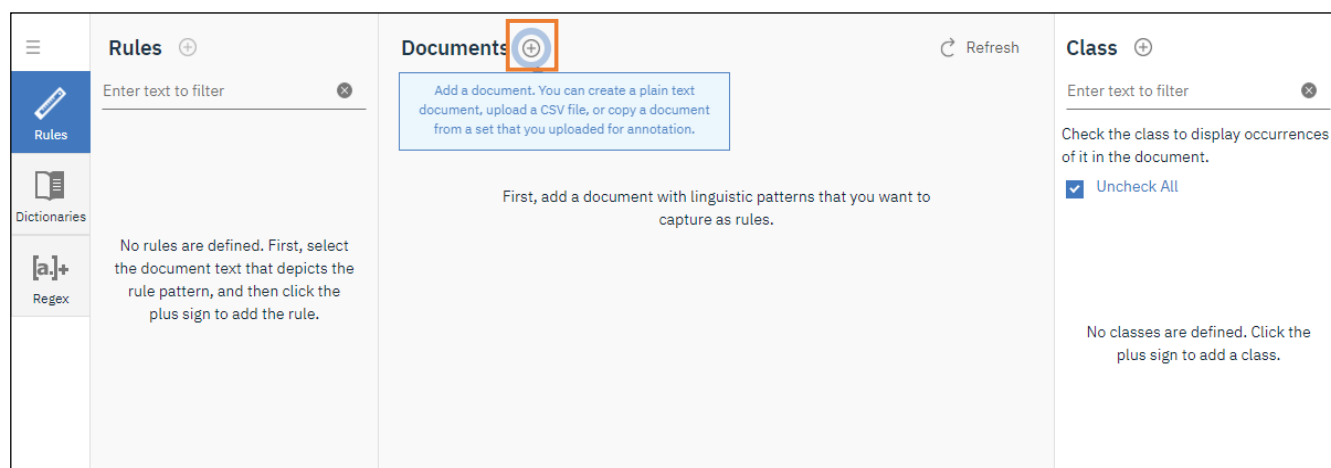
## Step 5: Upload documents in Rule editor

1. Select [Rule-based Model] – [Rules] in left navigation pane to open Rules page.

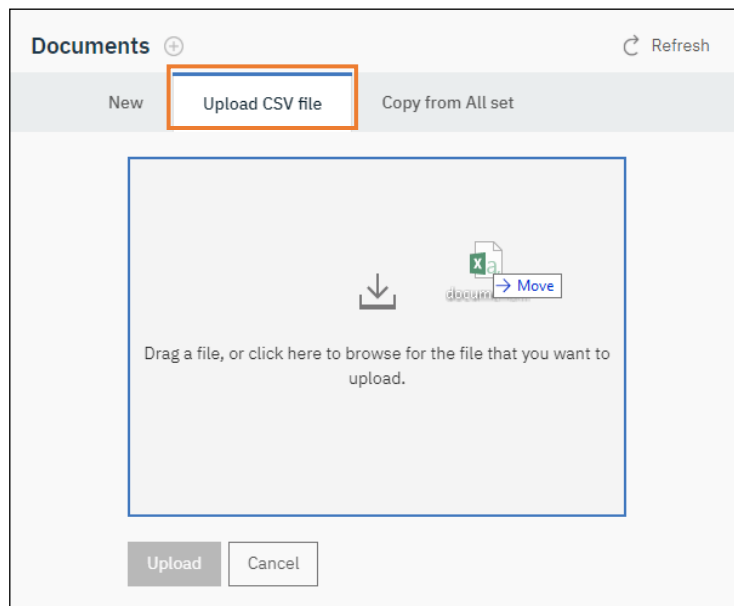


\* If Reminder notification covers the page, click **OK** button to close it.

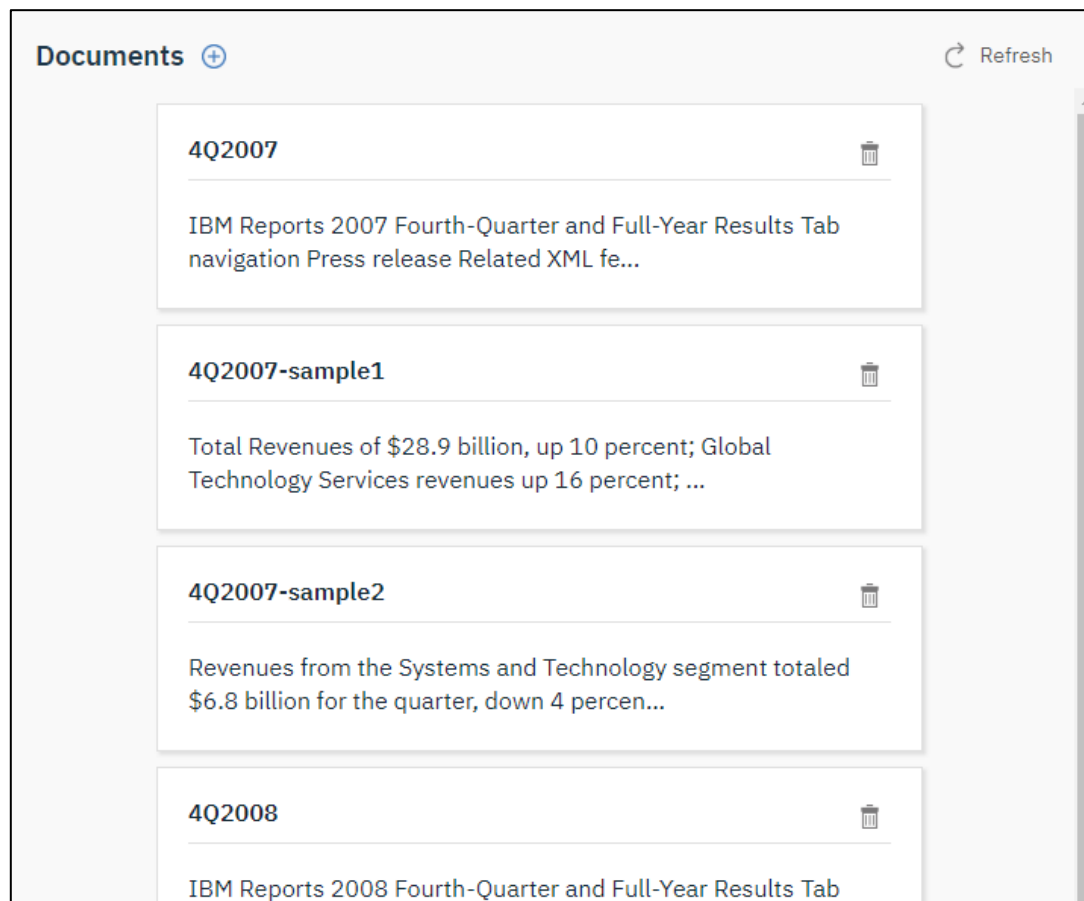
2. Click the plus sign icon  to upload documents in Document pane (center pane) of rule editor.



3. Select **Upload CSV file** tab and drag and drop “documents\_results.csv” file into drop file area.

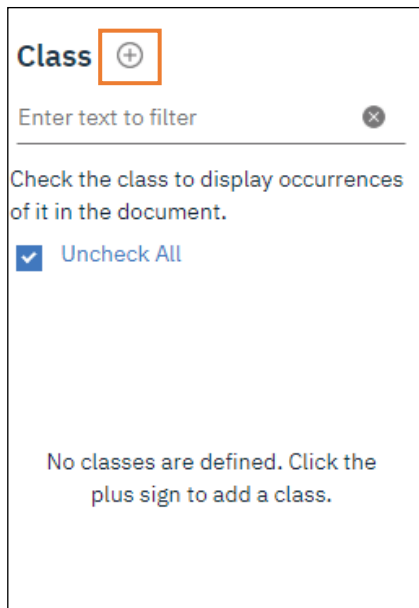


4. Click **Upload** button at the bottom of the drop file area. Make sure that 6 documents are uploaded successfully in Documents pane.

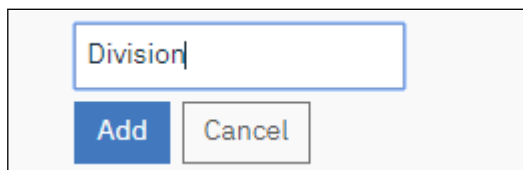



## Step 6: Create classes and mapping them with dictionaries

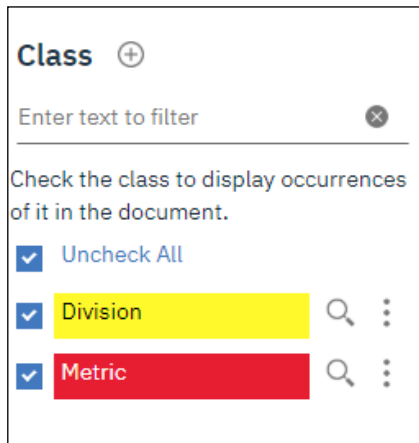
1. Click the plus sign icon  in Class pane (right side pane) of rule editor.



2. Enter “**Division**” in class name field, then click **Add** button to save it.



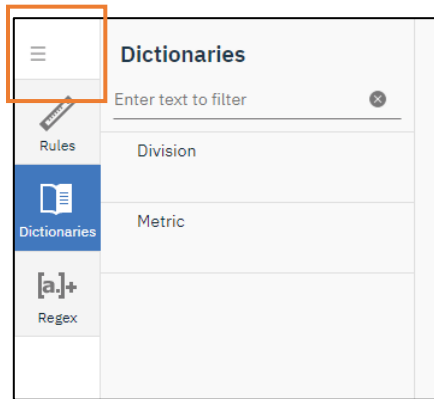
3. Click the plus sign icon  again and enter “**Metric**” in class name field, then click **Add** button to save it.
4. Make sure that “Division (Yellow)” and ”Metric (Red)” are listed in Class pane.



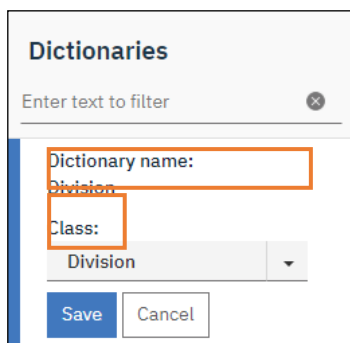


5. Select **Dictionaries** tab at the top left side of the rule editor to open Dictionaries editor.

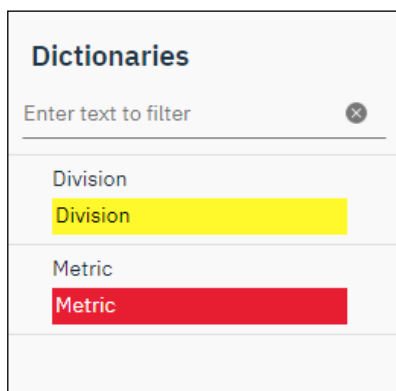
Confirm that Division and Metric dictionaries are listed in Dictionaries pane (right side pane). If none of them are listed here, revisit Step 2 and Step 3.



6. Select **Division** dictionary in Dictionaries pane, then select “**Division**” class from class dropdown menu. After that, click **Save** button.



7. Select **Metric** dictionary in Dictionaries pane, then select “**Metric**” class from class dropdown menu. After that, click **Save** button.
8. Make sure that **Division** dictionary is mapped with **Division** (Yellow) class and **Metric** dictionary is mapped with **Metric** (Red) class in Dictionaries pane.



9. In Document pane, select “**4Q2007-sample1**” document and confirm that all the terms (e.g. Global Technology Services) in Division dictionary are labeled as **Division** class and all the terms (e.g. Revenues) which were added in Metric dictionary are labeled as **Metric** class on the document.

EditClose

### 4Q2007-sample1

Total **Revenues** of \$28.9 billion, up 10 percent;

**Global Technology Services** **revenues** up 16 percent; pre-tax income up 26 percent;

**Global Business Services** **revenues** up 17 percent; pre-tax income up 9 percent;

Services signings of \$15.4 billion; short-term signings up 8 percent;

**Software** **revenues** up 12 percent; pre-tax income up 21 percent

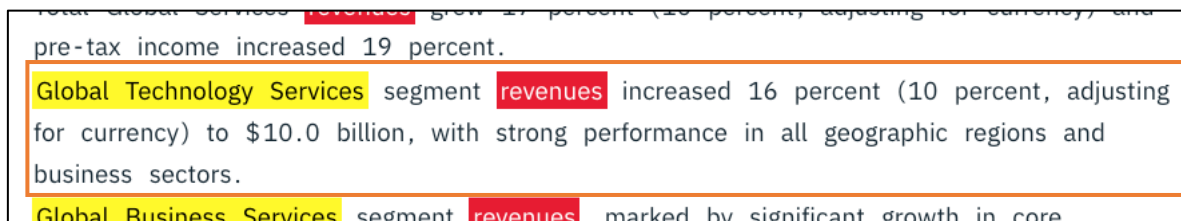
## Step 7: Defining a Rule for “RevenueOfDivision” - Part 1 -

We now create a rule to capture instances such as “Software segment revenue”

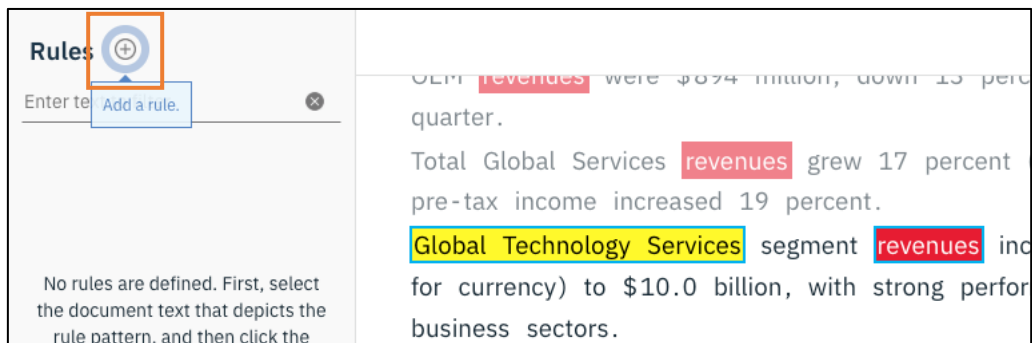
1. Add a class named “**RevenueOfDivision**” in the Class pane on the right side.



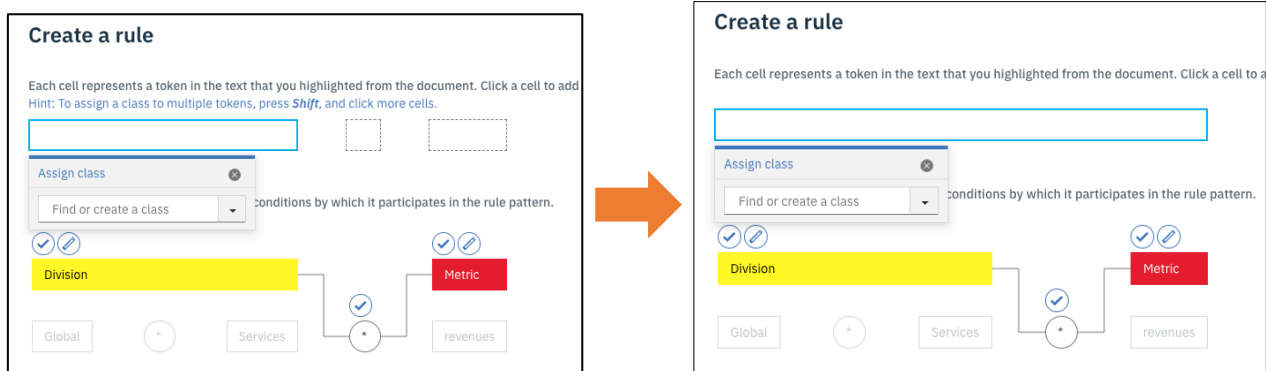
2. Select **Rules** tab at the top left side of the rule editor.
3. Find a sentence starting with “*Global Technology Services segment revenues...*” (It should be the 8<sup>th</sup> sentence from the top). Make sure that “Global Technology Services” is labeled as **Division** class and “revenues” as **Metric** class.



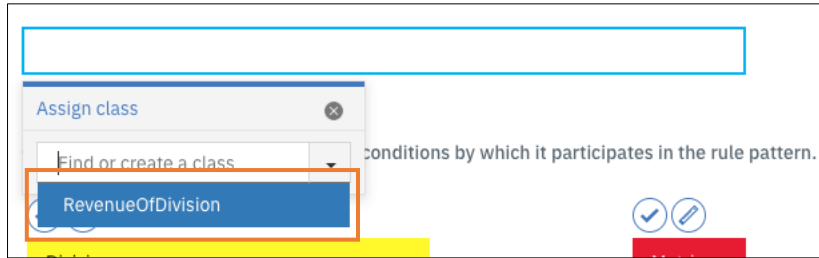
4. Click on “**Global Technology Services**” and “**revenues**” and click the plus sign icon  to add a rule.



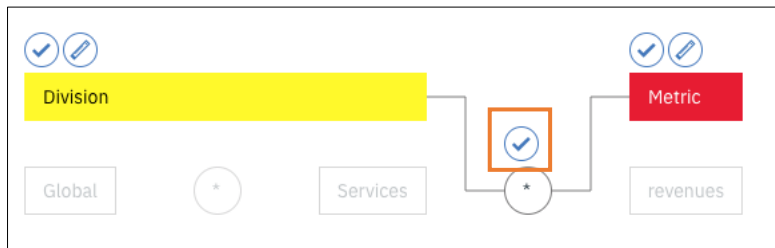
5. Select one of the dashed-line boxes on the top, hold down the Shift key then select the remaining boxes.



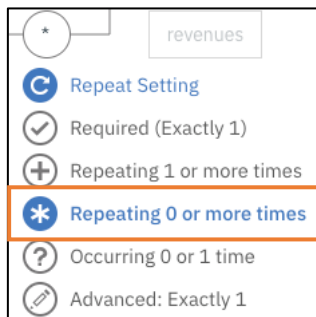
6. On “Assign a class” pop-up, select “**RevenueOfDivision**” class from the combo box.



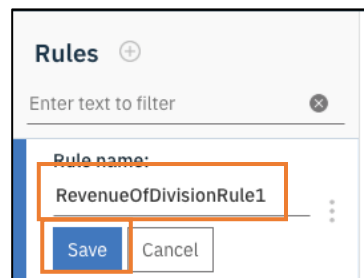
7. Click on the check mark icon  on “\*” to set repeat setting.



8. Select “**Repeating 0 or more times**”.



9. On the Rules pane, enter the rule name as “**RevenueOfDivisionRule1**”, then click **Save** button.



10. Confirm that all the patterns which match with the rule you created are labeled as **RevenueOfDivision** class.

Global Technology Services

revenues

 up 16 percent; pre-tax income up 26 percent;

Global Business Services

revenues

 up 17 percent; pre-tax income up 9 percent;

Global Technology Services

 segment 

revenues

 increased 16 percent (10 percent, adjusting for currency) to \$10.0 billion, with strong performance in all geographic regions and business sectors.

Global Business Services

 segment 

revenues

, marked by significant growth in core consulting and application management services, increased 17 percent (10 percent, adjusting for currency) to \$4.9 billion.

11. Click **Close** button to close the current document.

Edit

Close

4Q2007-sample1

Total 

Revenues

 of \$28.9 billion, up 10 percent;

## Step 8: Defining a Rule for “RevenueOfDivision” - Part 2 -

We now create another rule to capture instances such as “Revenues from the Systems and Technology Group”.

1. Select “**4Q2007-sample2**” document.
2. Find a sentence starting with “Revenues from the System and Technology segment...” (It should be the 1<sup>st</sup> sentence at the top). Make sure that “Revenues” is labeled as **Metric** class and “Systems and Technology” is labeled as **Division** class

**4Q2007-sample2**

Revenues

 from the 

Systems and Technology


 segment totaled \$6.8 billion for the quarter, down 4 percent (8 percent, adjusting for currency).

Revenues

 were flat excluding the year-to-year impact of the Printing Systems Division divestiture in June 2007.

Pre-tax income increased 18 percent.

3. Select “**Revenues**”, “**Systems and Technology**” and “**segment**”, then click the plus sign icon .

**Rules** 

Enter text to filter

RevenueOfDivisionRule1  
RevenueOfDivision

**4Q2007-sample2**

Revenues

 from the 

Systems and Technology

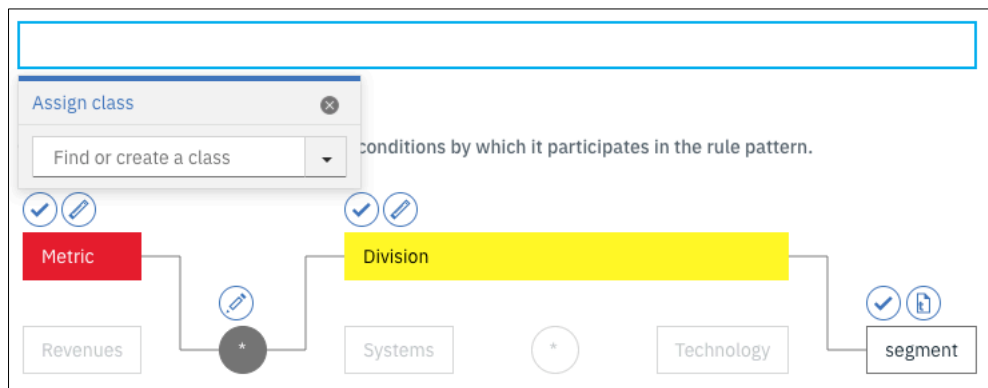
segment

 totaled \$6.8 billion for the quarter, (8 percent, adjusting for currency).

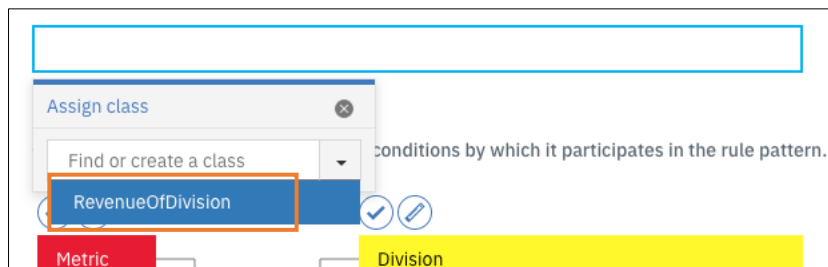
Revenues


 were flat excluding the year-to-year impact of the Printing Systems Division div 2007.

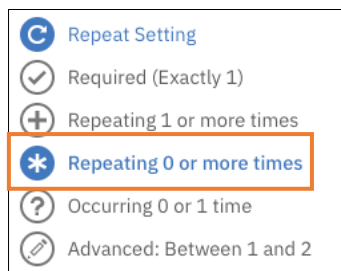
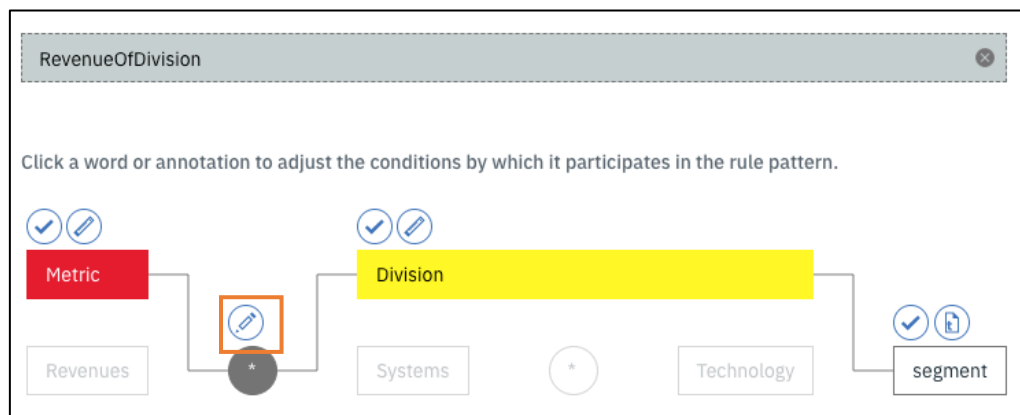
4. Select all the boxes using Shift key as described in the Step 6.5.



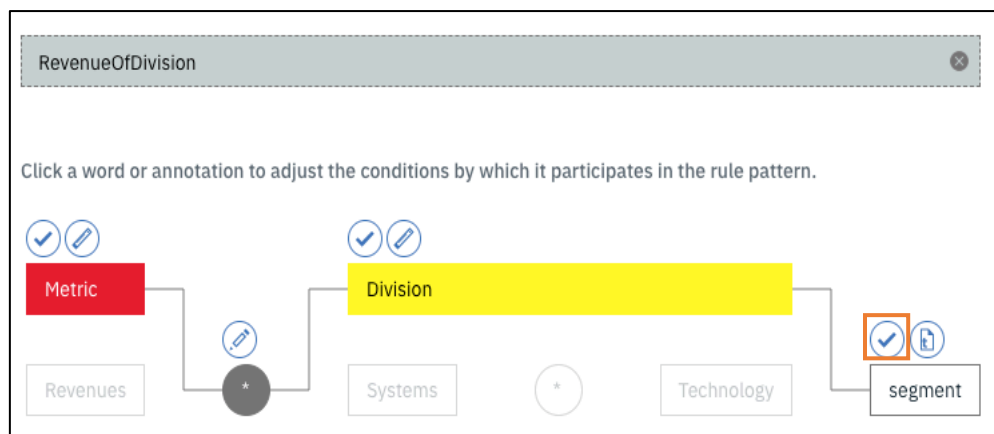
5. On “Assign a class” pop-up, select “**RevenueOfDivision**” class from the combo box.



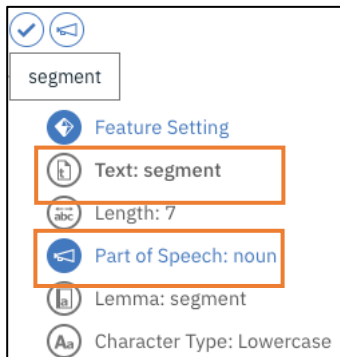
6. Click on the check mark icon  on “\*” and select “**Repeating 0 or more times**”.



7. Click on the check mark icon  on “segment” and select “**Repeating 0 or more times**”.



8. Click on the text icon  on “segment”. Select “**Part of Speech: noun**” and deselect “**Text: segment**”.



9. On the **Rules** pane, enter the rule name as “**RevenueOfDivisionRule2**”, then click **Save** button.



10. Make sure that all the patterns which match with the rule you created are labeled as **RevenueOfDivision** class.

**Revenues** from the **Systems and Technology** segment totaled \$6.8 billion for the quarter, down 4 percent (8 percent, adjusting for currency).

**Revenues** from the **Software** segment were \$6.3 billion, an increase of 12 percent (6 percent, adjusting for currency) compared with the fourth quarter of 2006; pre-tax income increased 21 percent.

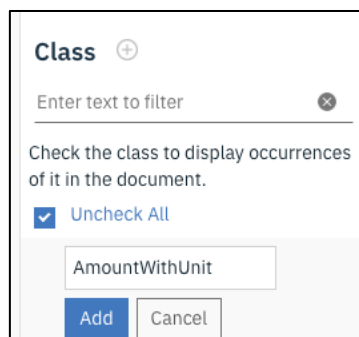
11. Click **Close** button to close the current document.




## Step 9: Defining a regular expression

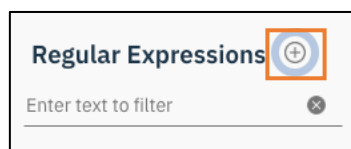
We now create a regular expression to capture instances such as “\$6.3 billion”.

1. Add a class named “**AmountWithUnit**” in the Class pane on the right side.



The screenshot shows a 'Class' dialog box. At the top, it says 'Class' with a plus icon. Below is a search bar 'Enter text to filter' with a close icon. A message says 'Check the class to display occurrences of it in the document.' There is a checkbox labeled 'Uncheck All' which is checked. Below that is a text input field containing 'AmountWithUnit'. At the bottom are 'Add' and 'Cancel' buttons.

2. Select **Regex** tab at the top left side of the rule editor and then click the plus sign icon  next to the Regular Expressions heading.



The screenshot shows the 'Regular Expressions' tab selected. It has a search bar 'Enter text to filter' with a close icon. A plus icon in a circle is highlighted with an orange box.


3. Click **Add Entry** button.



The screenshot shows the 'Regular Expressions' pane. The 'Add Entry' button is highlighted with an orange box. Below it is a table with columns: 'Regular Expression', 'Minimum Word Tokens', 'Maximum Word Tokens', and 'Action'.

Regular Expression	Minimum Word Tokens	Maximum Word Tokens	Action
--------------------	---------------------	---------------------	--------

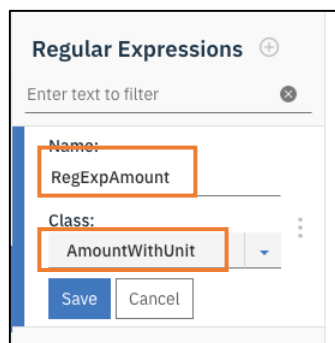
4. In the **Regular Expression** field, enter the following expression, which finds “\$” with amount of money.  
(\\$\d{1,2}(\.|\.)(billion))  
Note that there is a space inside the bracket before (billion)



The screenshot shows the 'Regular Expression' field with the expression `(\$\d{1,2}(\.|\.)(billion))` entered. The field is highlighted with an orange box. To the right are input fields for 'Minimum Word Tokens' (value 1) and 'Maximum Word Tokens' (value 5). At the bottom right are 'Cancel' and 'Add' buttons.

5. Click **Add** button to save the regular expression entry.

6. Enter “**RegExpAmount**” as the regular expression name, and then, from the **Class** menu, select the “**AmountWithUnit**” class that you defined earlier.



Regular Expressions

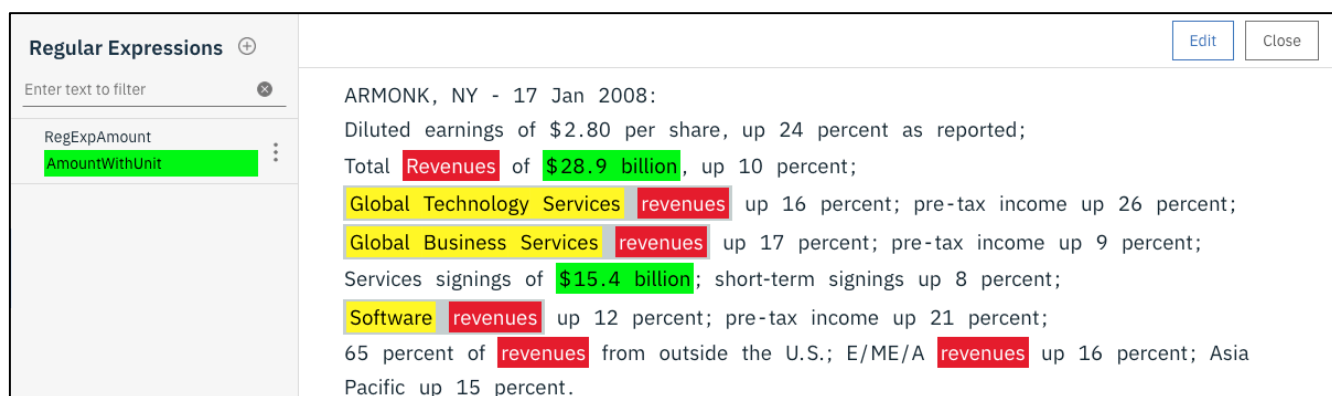
Enter text to filter

Name: RegExpAmount

Class: AmountWithUnit

Save Cancel

7. Click **Save** button.
8. Select “**4Q2007**” document. Make sure that all the patterns which match with the rule you created are labeled as **AmountWithUnit** class.



Regular Expressions

Enter text to filter

RegExpAmount

AmountWithUnit

ARMONK, NY - 17 Jan 2008:

Diluted earnings of \$2.80 per share, up 24 percent as reported;

Total Revenues of \$28.9 billion, up 10 percent;

Global Technology Services revenues up 16 percent; pre-tax income up 26 percent;

Global Business Services revenues up 17 percent; pre-tax income up 9 percent;

Services signings of \$15.4 billion; short-term signings up 8 percent;

Software revenues up 12 percent; pre-tax income up 21 percent;

65 percent of revenues from outside the U.S.; E/ME/A revenues up 16 percent; Asia Pacific up 15 percent.

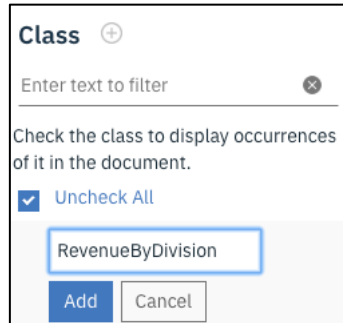
Edit Close


9. Click **Close** button to close the current document.


## Step 10: Defining a rule for “RevenueByDivision”

You are now ready to create a pattern to capture cases such as “Revenues from the Software segment were \$6.3 billion”.

1. Add a class named “**RevenueByDivision**” in the Class pane on the right side.



Class 

Enter text to filter 

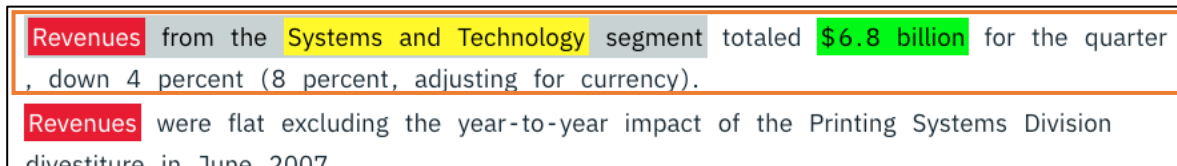
Check the class to display occurrences of it in the document.

☒ Uncheck All

RevenueByDivision

Add Cancel

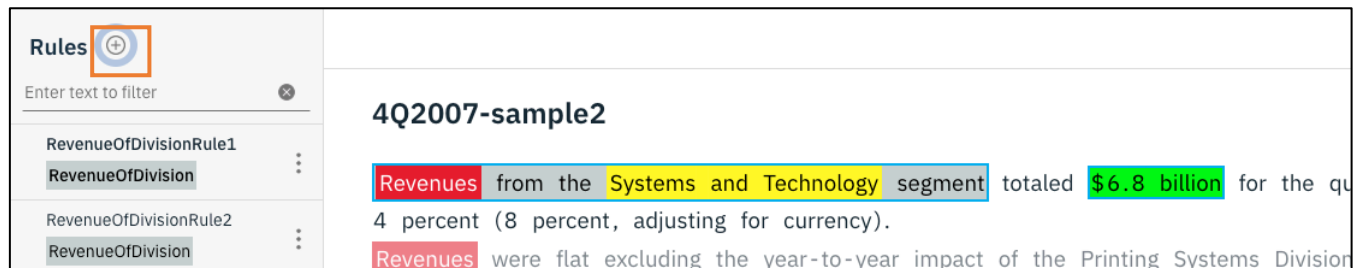
2. Select **Rules** tab at the top left side of the rule editor.
3. Select “**4Q2007-sample2**”
4. Find a sentence starting with “Revenues from the System and Technology segment...” (It should be the 1<sup>st</sup> sentence at the top). Make sure that “Revenues from the System and Technology segment” is labeled as **RevenueOfDivision** class and “\$6.8 billion” as **AmountWithUnit** class.





Revenues from the Systems and Technology segment totaled \$6.8 billion for the quarter, down 4 percent (8 percent, adjusting for currency).

Revenues were flat excluding the year-to-year impact of the Printing Systems Division divestiture in June 2007.

5. Select “**Revenues from the System and Technology segment**” (click on the label of “**RevenueOfDivision**” ) and “**\$6.8 billion**” and click on the plus sign icon .



Rules 

Enter text to filter 

RevenueOfDivisionRule1  
RevenueOfDivision

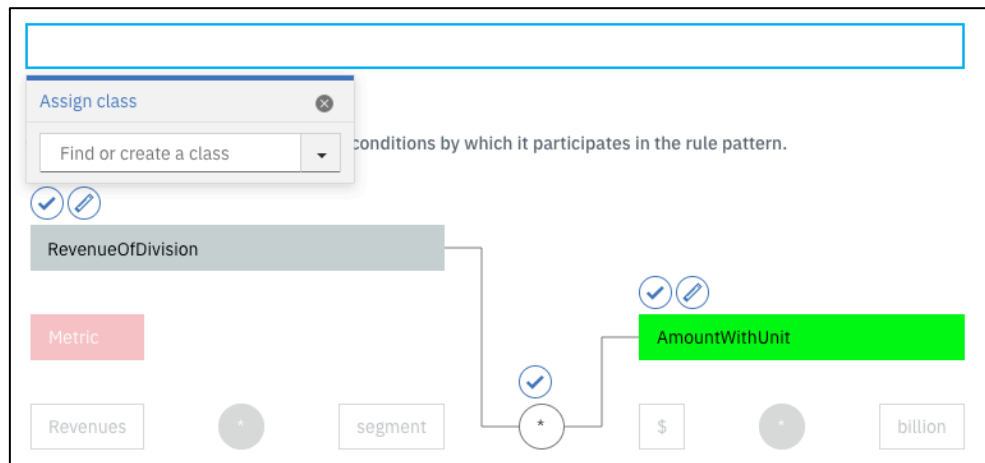
RevenueOfDivisionRule2  
RevenueOfDivision

4Q2007-sample2

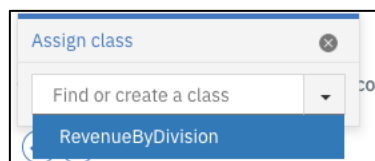
Revenues from the Systems and Technology segment totaled \$6.8 billion for the quarter, down 4 percent (8 percent, adjusting for currency).


Revenues were flat excluding the year-to-year impact of the Printing Systems Division divestiture in June 2007.

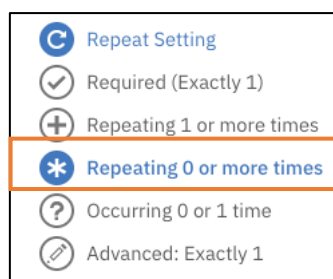
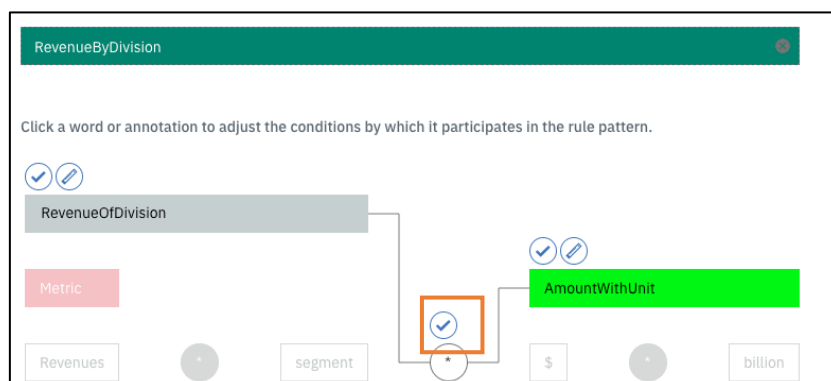
6. Select all the boxes using Shift key as described in Step 6.5.



7. On "Assign a class" pop-up, select "RevenueByDivision" class from the combo box.



8. Click on the check mark icon  on "\*" and select "Repeating 0 or more times".

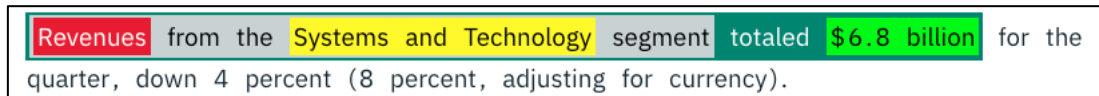


9. On the **Rules** pane, enter the rule name as “**RevenueByDivisionRule**”, then click **Save** button.

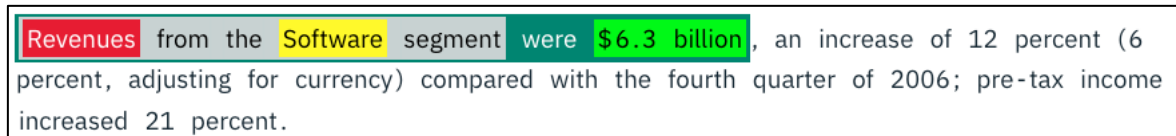


A screenshot of a dialog box for naming a rule. It has a label 'Rule name:' followed by a text input field containing 'RevenueByDivisionRule'. Below the input field are two buttons: 'Save' and 'Cancel'.

10. Make sure that all the patterns which match with the rule you created are labeled as **RevenueByDivision** class.



A text snippet with colored highlights: 'Revenues' (red), 'from the' (light blue), 'Systems and Technology' (yellow), 'segment' (light blue), 'totaled' (dark green), '\$6.8 billion' (green), 'for the' (light blue), 'quarter, down 4 percent (8 percent, adjusting for currency).' (light blue).



A text snippet with colored highlights: 'Revenues' (red), 'from the' (light blue), 'Software' (yellow), 'segment' (light blue), 'were' (dark green), '\$6.3 billion' (green), ', an increase of 12 percent (6 percent, adjusting for currency) compared with the fourth quarter of 2006; pre-tax income increased 21 percent.' (light blue).

11. Click **Close** button to close the current document, and check other documents (4Q2007, 4Q2008, 4Q2009, and 4Q2010) to see all the rules and a regular expression are working fine.

**Congratulations!**  
**You have completed the Watson Knowledge Studio Lab!**