Exercise 7. Creating a machine learning model with Watson Knowledge Studio

Estimated time

01:15

Overview

This exercise helps you understand the process for building a machine learning model that you can later deploy and use with other Watson services.

Objectives

After completing this exercise you should be able to:

- Create a workspace for Watson Knowledge Studio.
- Configure the workspace resources.
- Create document sets
- Pre-annotate documents
- Create tasks for human annotators
- Analyze inter-annotator agreement and adjudicate conflicts in annotated documents
- Create machine learning models.

Introduction

Use IBM Watson™ Knowledge Studio to create a machine learning model that understands the linguistic nuances, meaning, and relationships specific to a certain industry or domain. Knowledge Studio provides easy-to-use tools for annotating unstructured domain literature and uses those annotations to create a custom machine learning model that understands the language of the domain.

Requirements

IBM Cloud account.

Exercise instructions

In this exercise you will complete the following tasks:

__ 1. Create a Knowledge Studio service.

__ 2. Create a workspace.

__ 3. Create a type system.

__ 4. Add a dictionary.

__ 5. Add documents for annotation

__ 6. Create annotation sets

__ 7. Pre-annotate with a dictionary-based annotator

__ 8. Create an annotation task

__ 9. Annotate documents

__ 10. Analyze inter-annotator agreement

__ 11. Adjudicate conflicts in annotated documents

__ 12. Create a machine learning model

Part 1. Creating a Knowledge Studio service instance

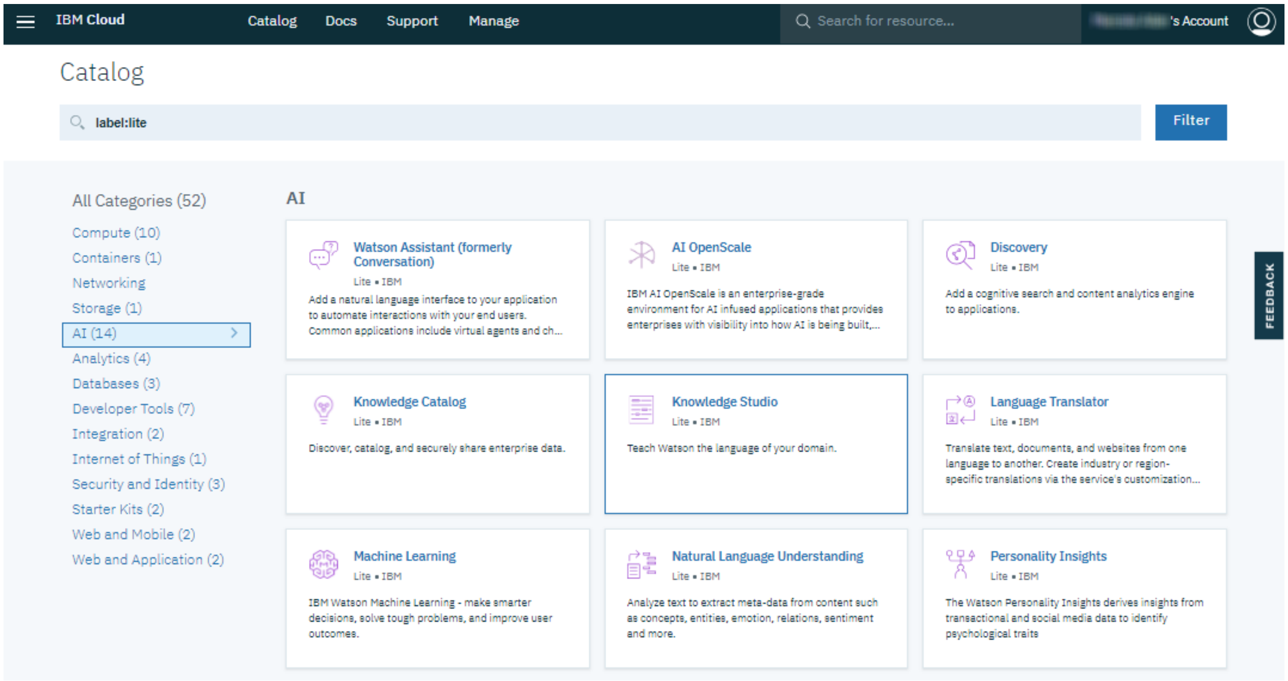
In this part, you will create a Knowledge Studio service instance on IBM Cloud.

Perform the following steps:

__ 1. Log in to IBM Cloud using your IBMid.

__ 2. Click **Catalog** > **AI**. All the AI services are listed.

__ 3. Click the **Knowledge Studio** tile.



Watson Assistant (formerly Conversation)
Lite • IBM
Add a natural language interface to your application to automate interactions with your end users. Common applications include virtual agents and ch...

AI OpenScale
Lite • IBM
IBM AI OpenScale is an enterprise-grade environment for AI infused applications that provides enterprises with visibility into how AI is being built,...

Discovery
Lite • IBM
Add a cognitive search and content analytics engine to applications.

Knowledge Catalog
Lite • IBM
Discover, catalog, and securely share enterprise data.

Knowledge Studio
Lite • IBM
Teach Watson the language of your domain.

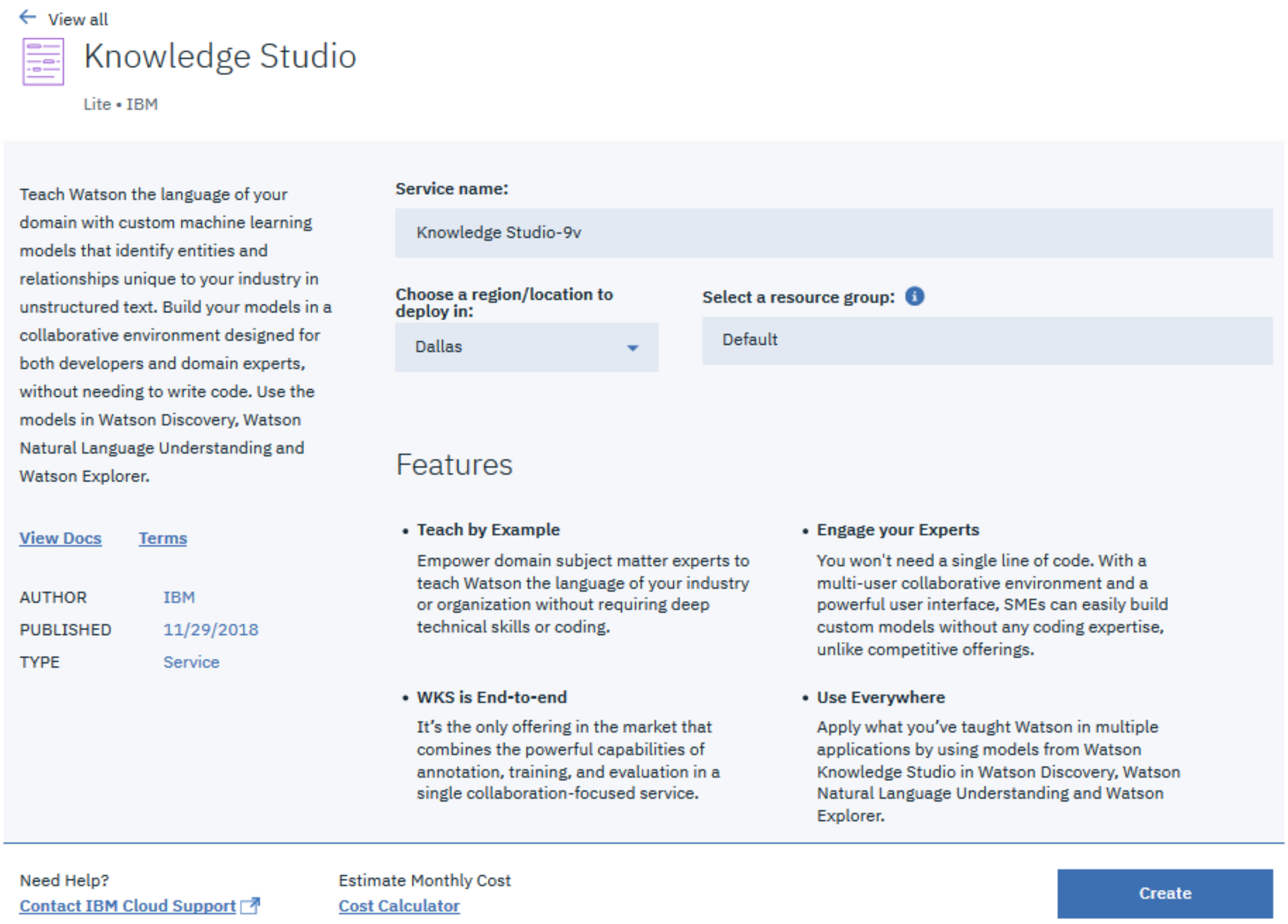
Language Translator
Lite • IBM
Translate text, documents, and websites from one language to another. Create industry or region-specific translations via the service's customization...

Machine Learning
Lite • IBM
IBM Watson Machine Learning - make smarter decisions, solve tough problems, and improve user outcomes.

Natural Language Understanding
Lite • IBM
Analyze text to extract meta-data from content such as concepts, entities, emotion, relations, sentiment and more.

Personality Insights
Lite • IBM
The Watson Personality Insights derives insights from transactional and social media data to identify psychological traits

__ 4. Accept the default values and click **Create** to create a Lite plan instance of the Knowledge Studio service.



← View all

Knowledge Studio

Lite • IBM

Teach Watson the language of your domain with custom machine learning models that identify entities and relationships unique to your industry in unstructured text. Build your models in a collaborative environment designed for both developers and domain experts, without needing to write code. Use the models in Watson Discovery, Watson Natural Language Understanding and Watson Explorer.

View Docs Terms

AUTHOR	IBM
PUBLISHED	11/29/2018
TYPE	Service

Service name: Knowledge Studio-9v

Choose a region/location to deploy in: Dallas

Select a resource group: Default

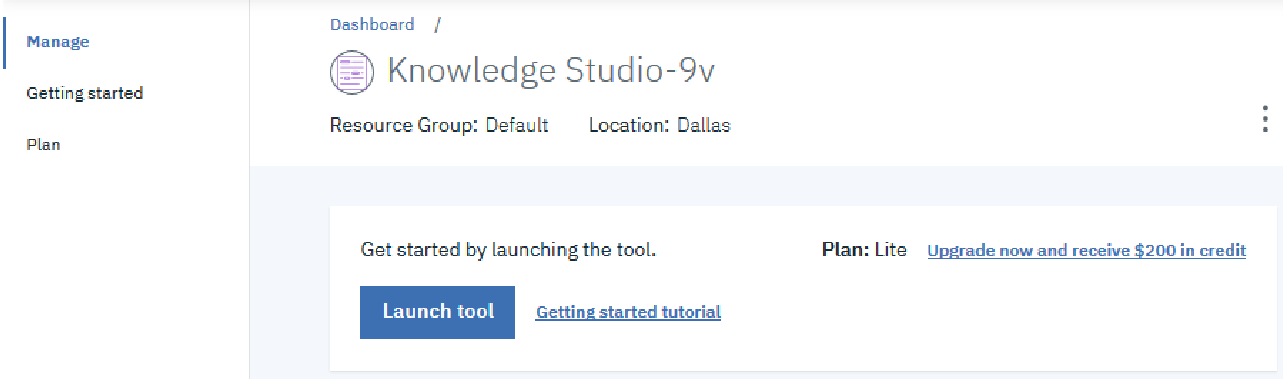
Features

- **Teach by Example**
  Empower domain subject matter experts to teach Watson the language of your industry or organization without requiring deep technical skills or coding.
- **Engage your Experts**
  You won't need a single line of code. With a multi-user collaborative environment and a powerful user interface, SMEs can easily build custom models without any coding expertise, unlike competitive offerings.
- **WKS is End-to-end**
  It's the only offering in the market that combines the powerful capabilities of annotation, training, and evaluation in a single collaboration-focused service.
- **Use Everywhere**
  Apply what you've taught Watson in multiple applications by using models from Watson Knowledge Studio in Watson Discovery, Watson Natural Language Understanding and Watson Explorer.

Need Help?
Contact IBM Cloud Support

Estimate Monthly Cost
Cost Calculator

__ 5. After the instance is created, click **Manage** (left menu), and then click **Launch Tool** to open the Watson Knowledge Studio tooling.



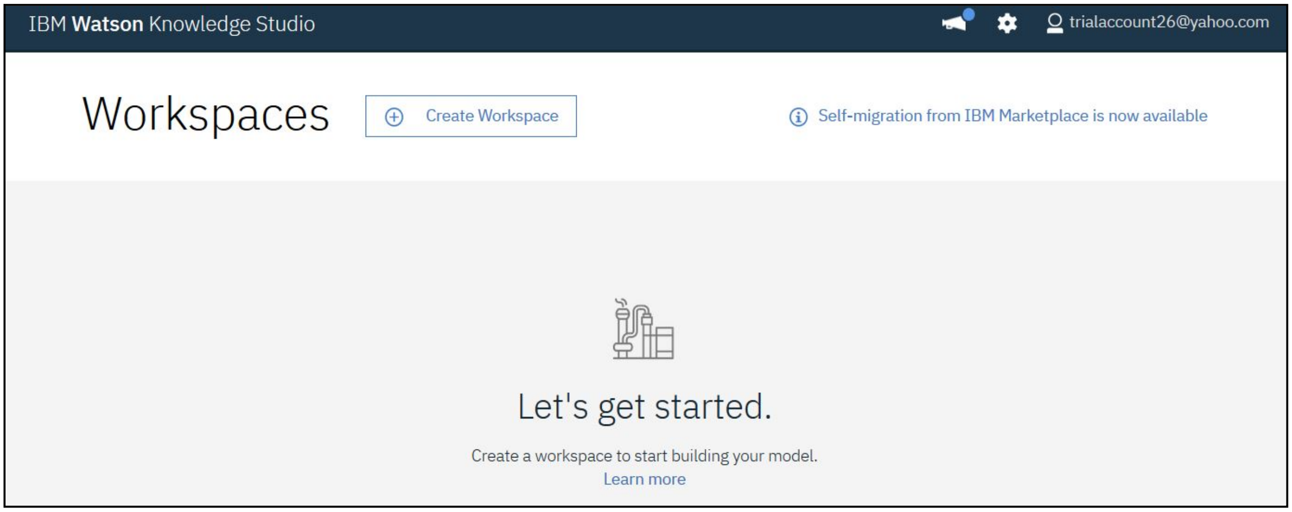
Knowledge Studio-9v

Resource Group: Default Location: Dallas

Get started by launching the tool.

Plan: Lite Upgrade now and receive $200 in credit

__ 6. The Workspaces page is displayed. (Dismiss or close any pop-ups).



IBM **Watson** Knowledge Studio

trialaccount26@yahoo.com

Workspaces

Create Workspace

Self-migration from IBM Marketplace is now available

Let's get started.

Create a workspace to start building your model.

Learn more

Part 2. Creating a workspace

In this part, you will learn how to create a workspace in Watson Knowledge Studio. A workspace defines all the resources that are required to create a machine learning model, including training documents, the type system, dictionaries, and annotations that are added by human annotators.

Perform the following steps:

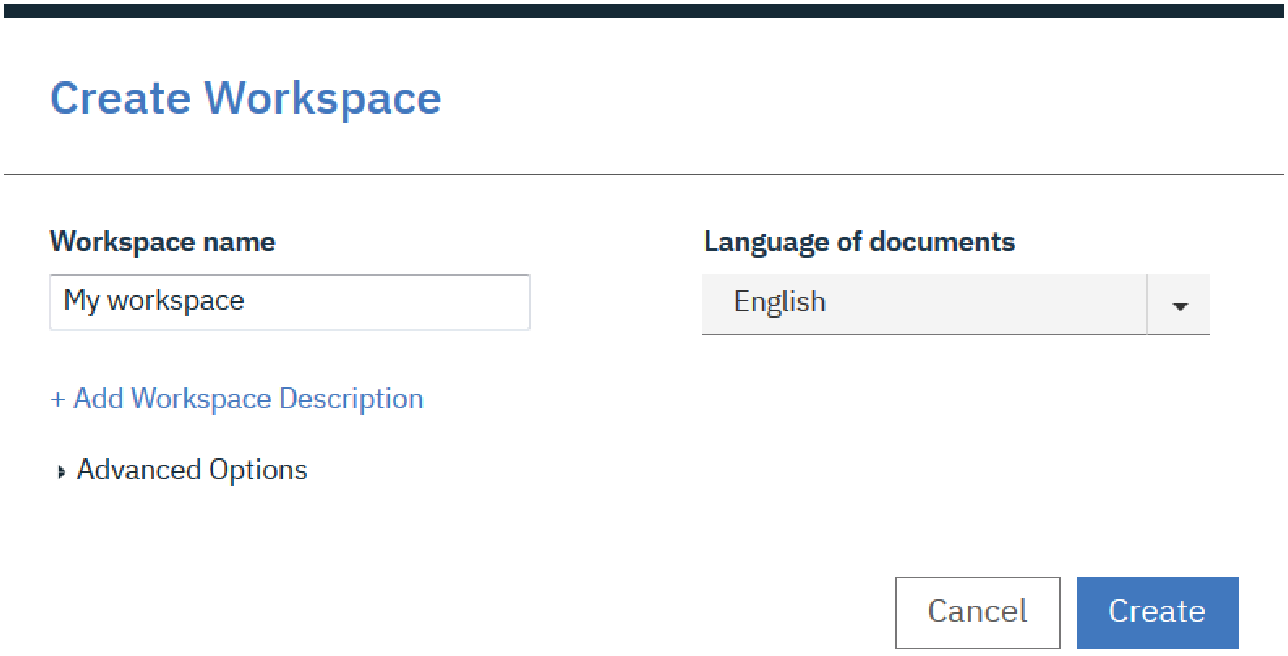
__ **1.** Click **+ Create Workspace**

__ 2. Specify the details for the new workspace:

a. In the **Workspace name** field, type *My workspace*.

b. In the **Language of documents** field, use the default value, **English**. The sample files you will be using for this tutorial are in English.

__ 3. Click **Create**.

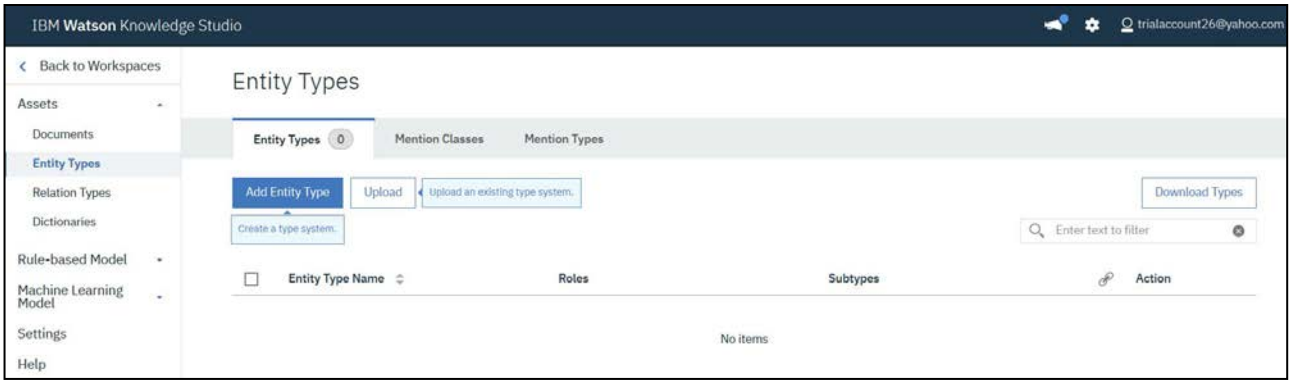


Create Workspace

Workspace name: My workspace

Language of documents: English

__ 4. After the workspace is successfully created it opens automatically and the Entity Types page is displayed as shown in the figure.



Part 3. Creating a type system

A type system defines things that are interesting in your domain content that you want to label with an annotation. The type system controls how content can be annotated by defining the types of entities that can be labeled and how relationships among different entities can be labeled. Typically subject matter experts for a domain help to define the type system.

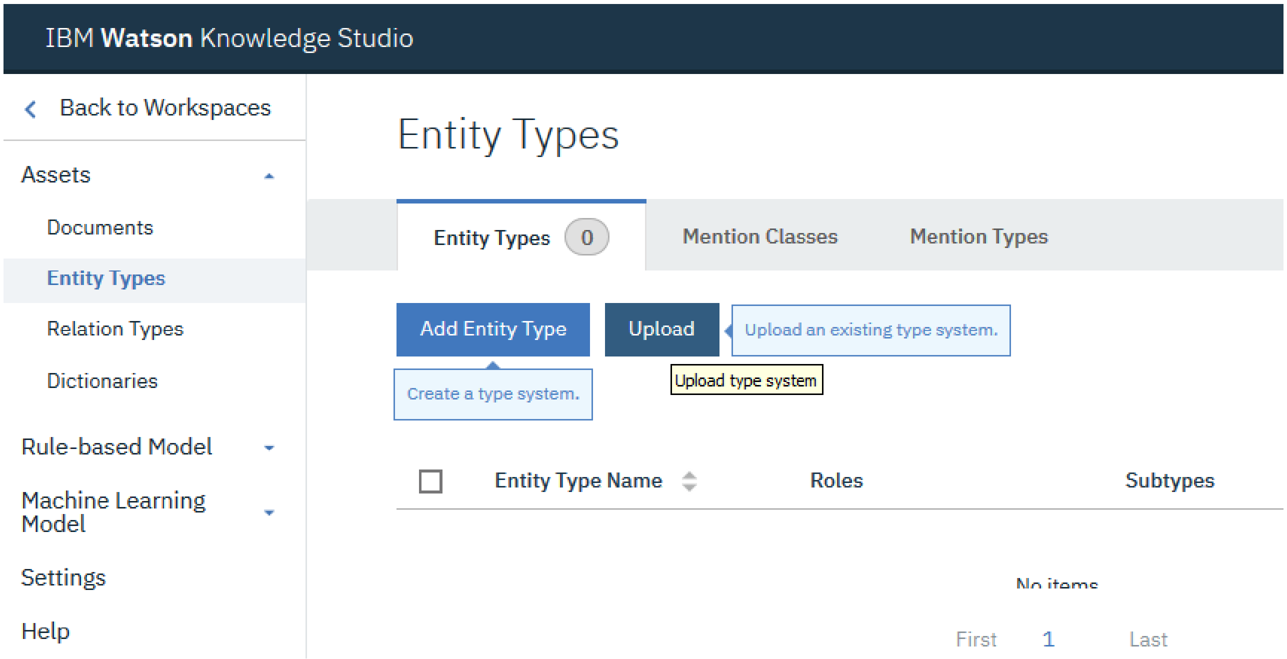
In this part, you will learn how to upload and modify a type system within Knowledge Studio. You must create or upload a type system before you begin any annotation tasks.

Perform the following steps:

__ 1. Download the en-klue2-types.json file to your computer. This file contains an example type system.

__ 2. Click **Assets** > **Entity Types**.

__ 3. On the Entity Types page, click **Upload**.



IBM **Watson** Knowledge Studio

< Back to Workspaces

Assets

Documents

Entity Types

Relation Types

Dictionaries

Rule-based Model

Machine Learning Model

Settings

Help

Entity Types

Entity Types 0 | Mention Classes | Mention Types

Add Entity Type | Upload

Upload an existing type system.

Create a type system.

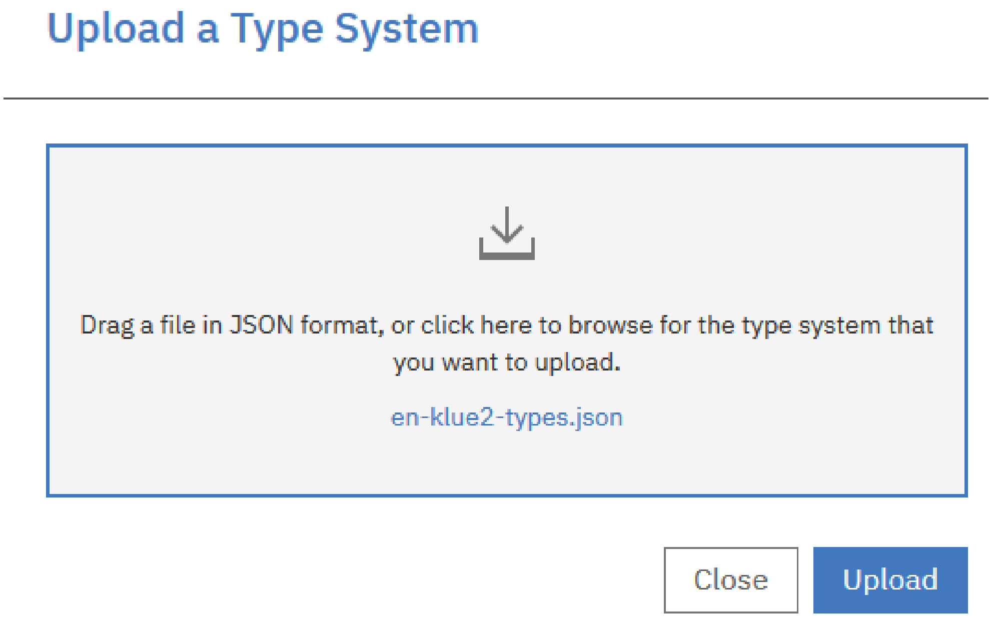
Upload type system

□	Entity Type Name	Roles	Subtypes

No items

First 1 Last

__ 4. Upload the *en-klue2-types.json* file from your computer. Click **Upload** after specifying the json file to upload.



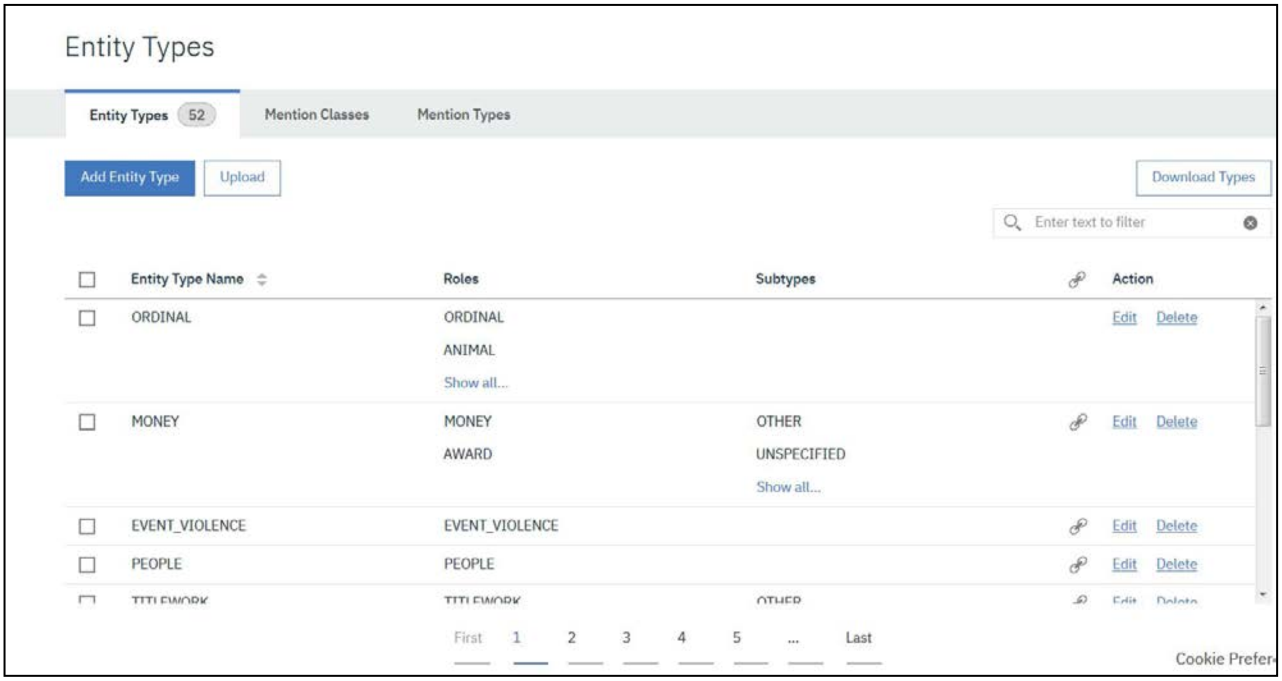
Upload a Type System

Drag a file in JSON format, or click here to browse for the type system that you want to upload.

en-klue2-types.json

Close | Upload

__ 5. The uploaded type system is displayed in the table as shown in the figure.



Entity Types

Entity Types 52 | Mention Classes | Mention Types

Add Entity Type | Upload | Download Types

Enter text to filter

□	Entity Type Name	Roles	Subtypes		Action
□	ORDINAL	ORDINAL ANIMAL Show all...			Edit Delete
□	MONEY	MONEY AWARD	OTHER UNSPECIFIED Show all...		Edit Delete
□	EVENT_VIOLENCE	EVENT_VIOLENCE			Edit Delete
□	PEOPLE	PEOPLE			Edit Delete
□	TITLEWORK	TITLEWORK	OTHER		Edit Delete

First 1 2 3 4 5 ... Last

Cookie Prefer

__ 6. Browse the type system so you can see the data that was uploaded.

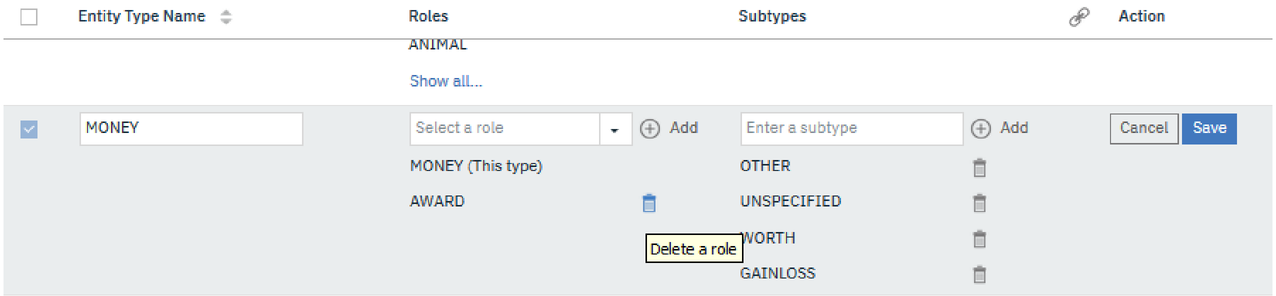
__ 7. Edit an entity type:

__ a. Locate the **MONEY** entity type.

__ b. Double-click anywhere in the table row to edit the entity type.

__ c. In the **Roles** column, click the delete icon next to the **AWARD** role.

__ d. Click **Save**.



□	Entity Type Name	Roles	Subtypes		Action
		ANIMAL Show all...			
☑	MONEY	Select a role ⊕ Add MONEY (This type) AWARD	Enter a subtype ⊕ Add OTHER UNSPECIFIED WORTH GAINLOSS		Cancel Save

After you finish making changes to the type system, you can begin adding documents to your workspace.

Part 4. Adding a dictionary

To help human annotators get started with their annotation tasks, you can create a dictionary and use it to pre-annotate documents that you add to the corpus.

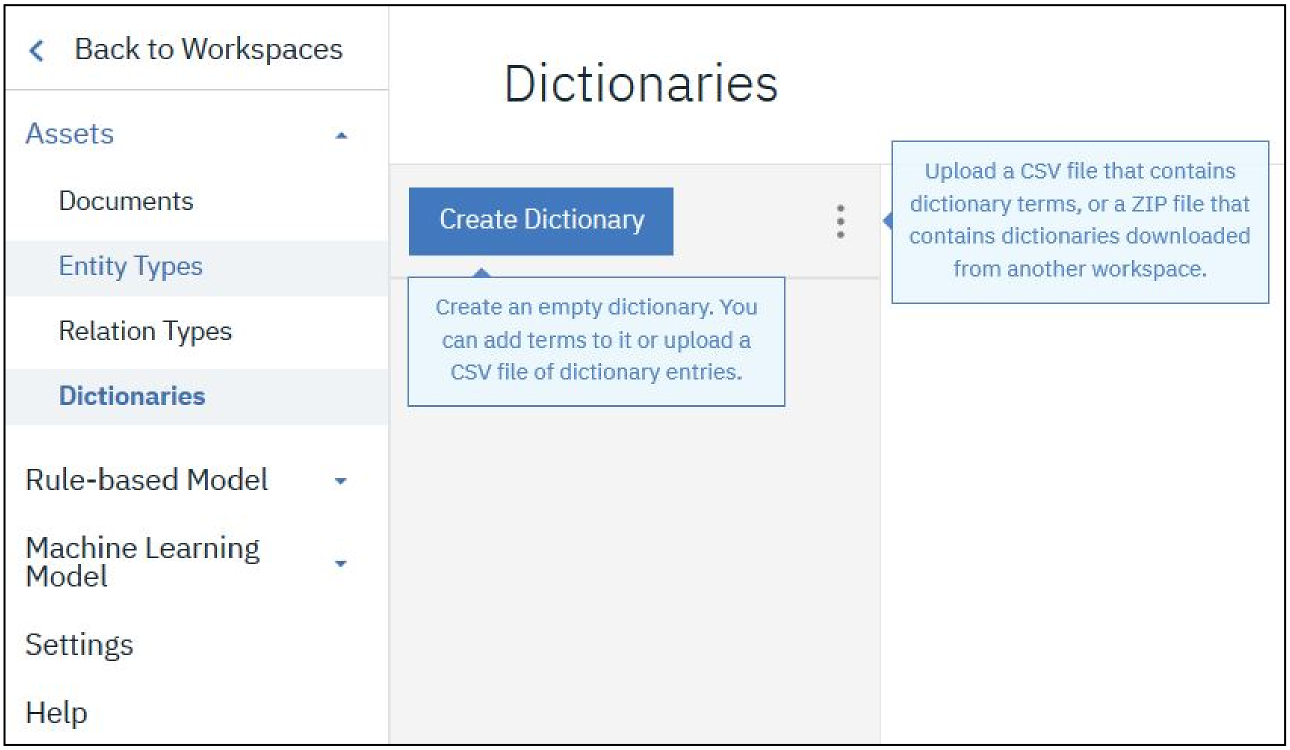
In this part, you will learn how to add a dictionary to a workspace in Knowledge Studio. Dictionaries are used for pre-annotating text when creating a machine learning model.

Perform the following steps:

__ 1. Download the file dictionary-items-organization.csv to your computer. This file contains dictionary terms in CSV format, suitable for uploading into a Knowledge Studio dictionary.

__ 2. Click **Assets** > **Dictionaries**.

__ 3. Click **Create Dictionary** to add a dictionary as shown in the figure.

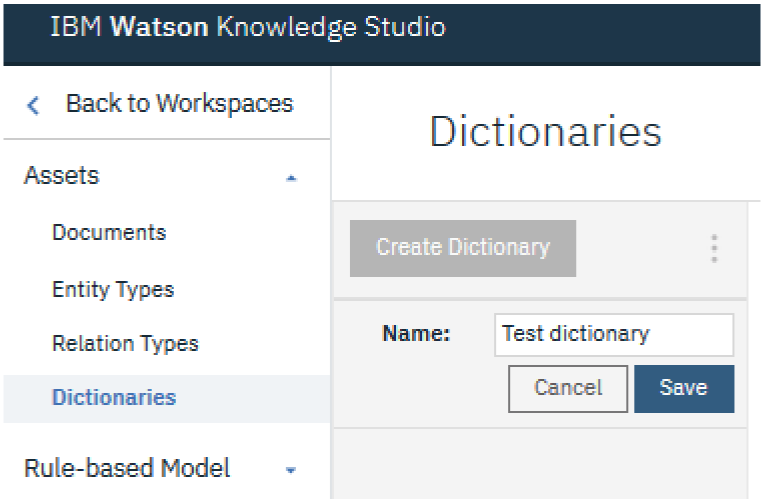


Create an empty dictionary. You can add terms to it or upload a CSV file of dictionary entries.

Upload a CSV file that contains dictionary terms, or a ZIP file that contains dictionaries downloaded from another workspace.

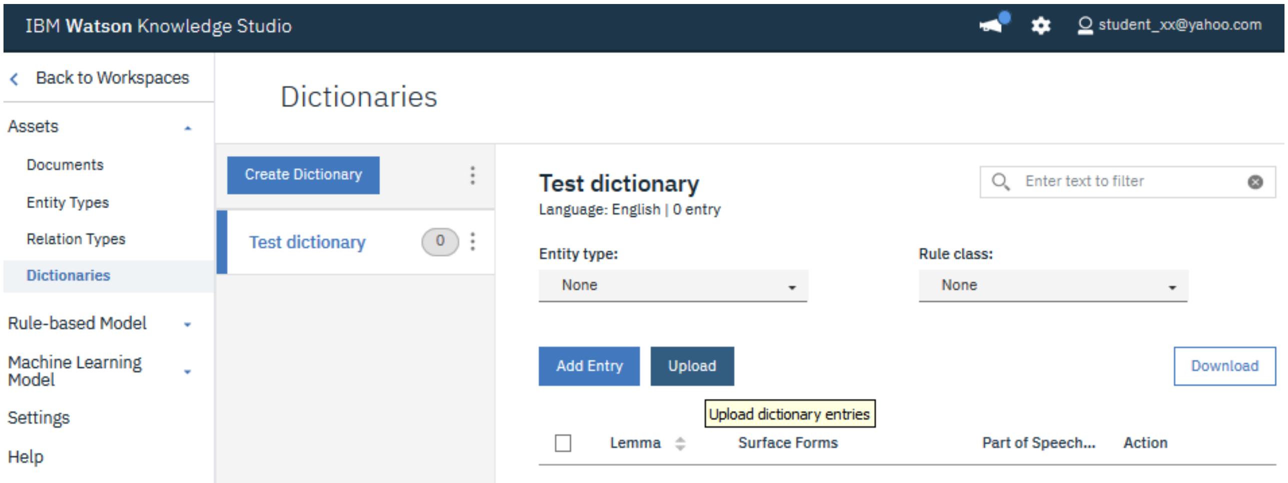
Note: Do not click **Upload Dictionary,** which is used to upload a dictionary that you want to use as-is. For this exercise, you will create a new editable dictionary and then upload terms into it.

__ 4. In the **Name** field, type *Test dictionary* and click **Save** to create the (empty) dictionary.



Name: Test dictionary

__ 5. The new dictionary is created and automatically opened for editing.



Test dictionary

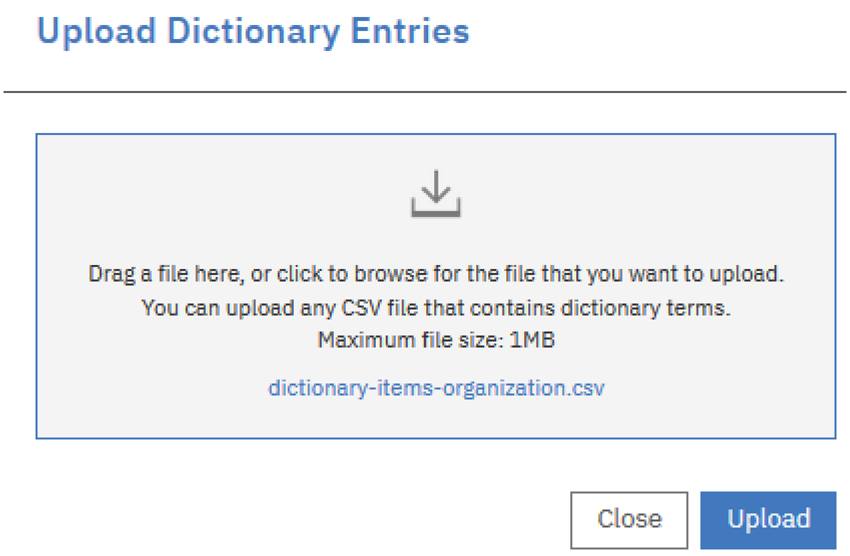
Language: English | 0 entry

Entity type: None

Rule class: None

__ 6. In the dictionary pane, click **Upload**.

__ 7. Upload the file *dictionary-items-organization.csv* from your computer.



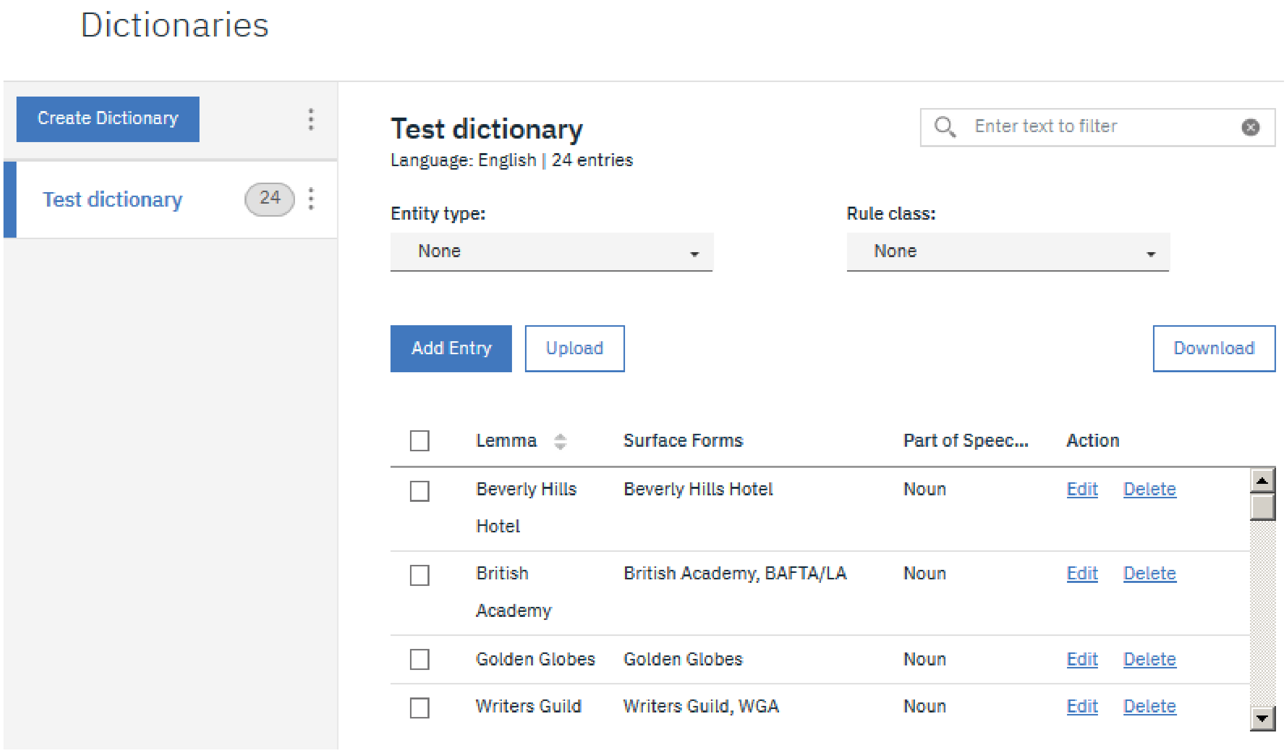
Upload Dictionary Entries

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

dictionary-items-organization.csv

The terms in the file are uploaded into the dictionary.

__ 8. Click **Add Entry** to create a new term. An editable row is added at the top of the table.



Dictionaries

Test dictionary

Language: English | 24 entries

Entity type: None

Rule class: None

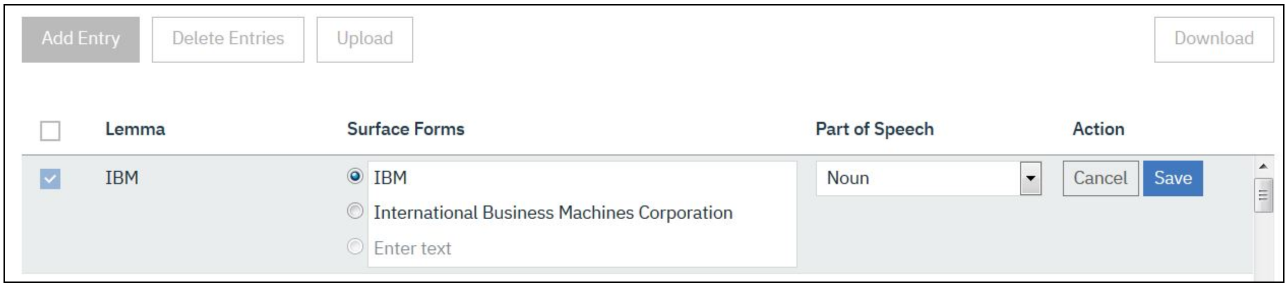
□	Lemma	Surface Forms	Part of Speec...	Action
□	Beverly Hills Hotel	Beverly Hills Hotel	Noun	Edit Delete
□	British Academy	British Academy, BAFTA/LA	Noun	Edit Delete
□	Golden Globes	Golden Globes	Noun	Edit Delete
□	Writers Guild	Writers Guild, WGA	Noun	Edit Delete

__ 9. In the **Surface Forms** column, type IBM and International Business Machines Corporation on separate lines. When you begin to type a new surface form, a space is added below for an additional surface form. Leave the radio button next to IBM selected, which indicates that IBM is the lemma.

Information: *Lemma* specifies the most representative word form for the entry and *surface forms* specify equivalent terms.

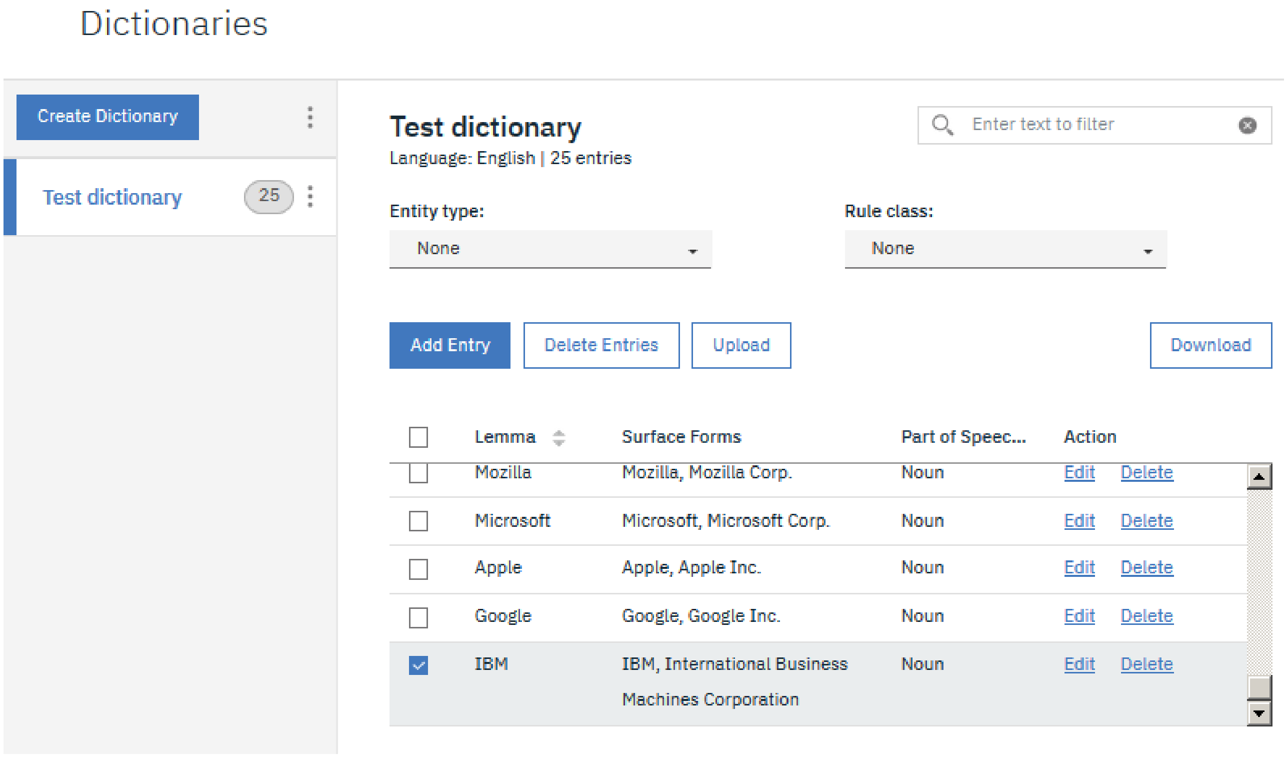
__ 10. In the **Part of Speech** column, select **Noun**.

__ 11. Click **Save**.



☐	Lemma	Surface Forms	Part of Speech	Action
☑	IBM	● IBM ○ International Business Machines Corporation ○ Enter text	Noun	Cancel Save

__ 13. Scroll through the entries to confirm that the new term has been added to the dictionary.



Test dictionary

Language: English | 25 entries

Entity type: None

Rule class: None

☐	Lemma	Surface Forms	Part of Speec...	Action
☐	Mozilla	Mozilla, Mozilla Corp.	Noun	Edit Delete
☐	Microsoft	Microsoft, Microsoft Corp.	Noun	Edit Delete
☐	Apple	Apple, Apple Inc.	Noun	Edit Delete
☐	Google	Google, Google Inc.	Noun	Edit Delete
☑	IBM	IBM, International Business Machines Corporation	Noun	Edit Delete

After you create a dictionary, you can use it to speed up human annotation tasks by pre-annotating the documents.

Part 5. Adding documents for annotation

To train a model, you must add documents that are representative of your domain content to your workspace. As a best practice, start with a relatively small collection of documents. Use these documents to train human annotators and to refine the annotation guidelines. As annotation accuracy improves, you can add more documents to the corpus to provide greater depth to the training effort.

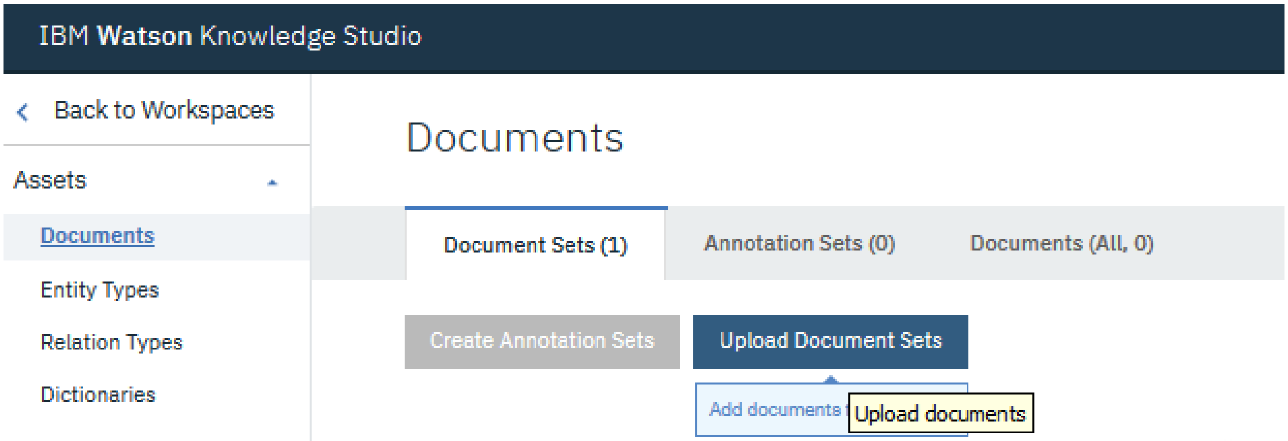
In this part, you will learn how to add documents to a workspace in Knowledge Studio that can be annotated by human annotators.

Perform the following steps:

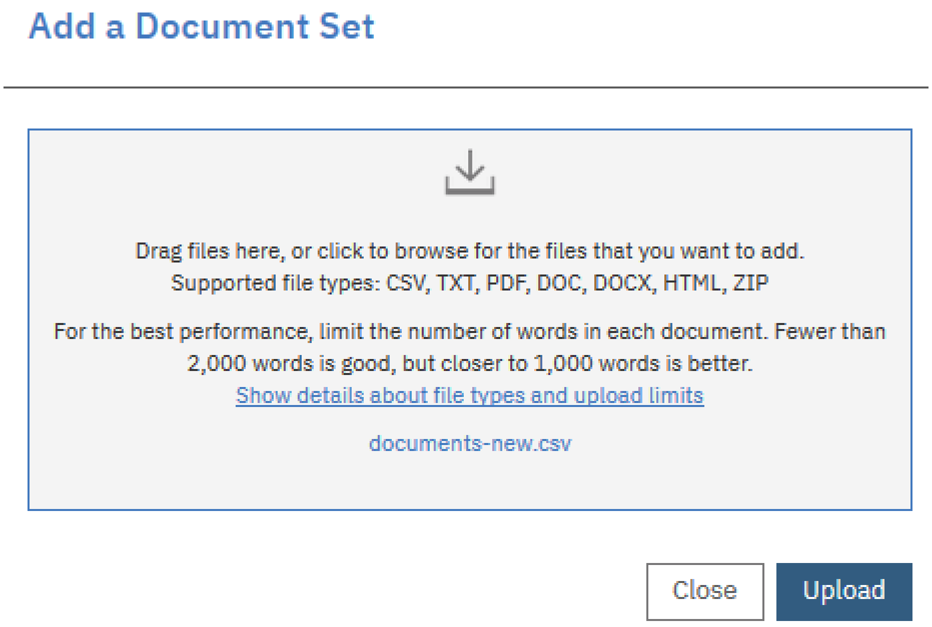
__ 1. Download the file documents-new.csv to your computer. This file contains example documents suitable for uploading.

__ 2. Within your workspace, click **Assets** > **Documents**.

__ 3. On the Documents page, click **Upload Document Sets.**



__ 4. Upload the file *documents-new.csv* from your computer. Specify the file to upload and click **Upload**.



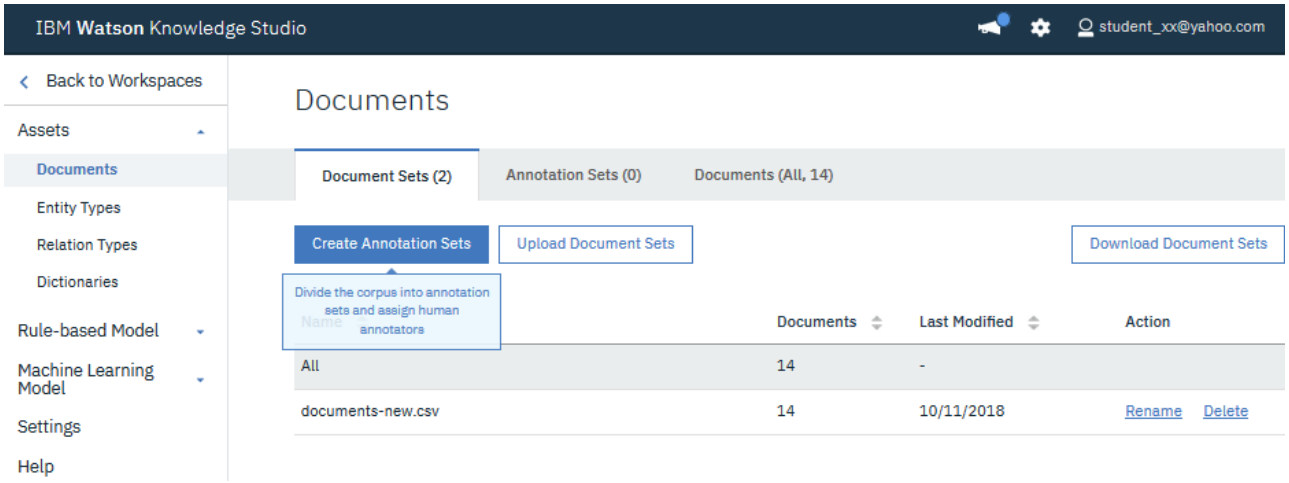
Add a Document Set

Drag files here, or click to browse for the files that you want to add.
Supported file types: CSV, TXT, PDF, DOC, DOCX, HTML, ZIP

For the best performance, limit the number of words in each document. Fewer than 2,000 words is good, but closer to 1,000 words is better.
Show details about file types and upload limits

documents-new.csv

__ 5. The uploaded file is displayed in the table.

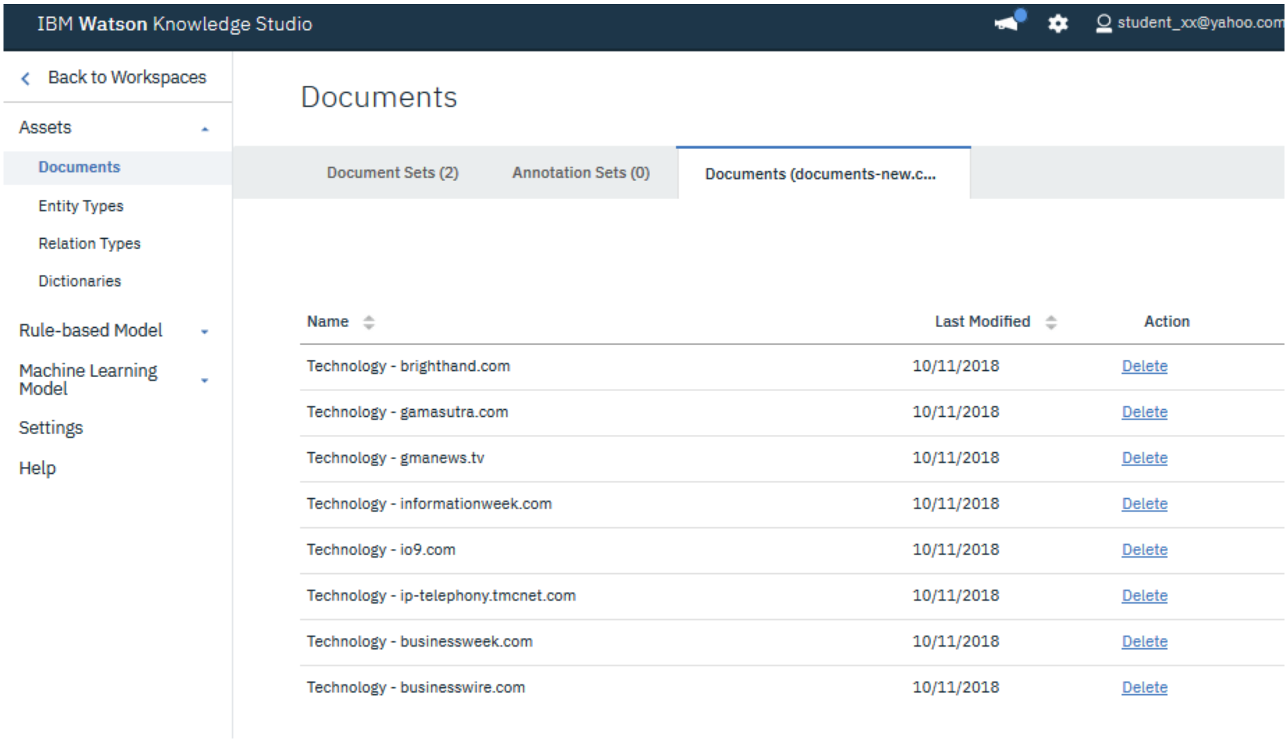


Documents

Divide the corpus into annotation sets and assign human annotators

Name	Documents	Last Modified	Action
All	14	-	
documents-new.csv	14	10/11/2018	Rename Delete

__ 6. Click **documents-new.csv** to browse the documents that were uploaded.



Name	Last Modified	Action
Technology - brighthand.com	10/11/2018	Delete
Technology - gamasutra.com	10/11/2018	Delete
Technology - gmanews.tv	10/11/2018	Delete
Technology - informationweek.com	10/11/2018	Delete
Technology - io9.com	10/11/2018	Delete
Technology - ip-telephony.tmcnet.com	10/11/2018	Delete
Technology - businessweek.com	10/11/2018	Delete
Technology - businesswire.com	10/11/2018	Delete

__ 7. Click on a document to display the text.

You can now divide the corpus into multiple document sets and assign the document sets to human annotators.

Part 6. Creating annotation sets

An annotation set is a subset of documents from an uploaded document set that you assign to a human annotator. The human annotator annotates the documents in the annotation set. To later use inter-annotator scores to compare the annotations that are added by each human annotator, you must assign at least two human annotators to different annotation sets. You must also specify that some percentage of documents overlap between the sets.

In this part, you will learn how to create annotation sets in Knowledge Studio

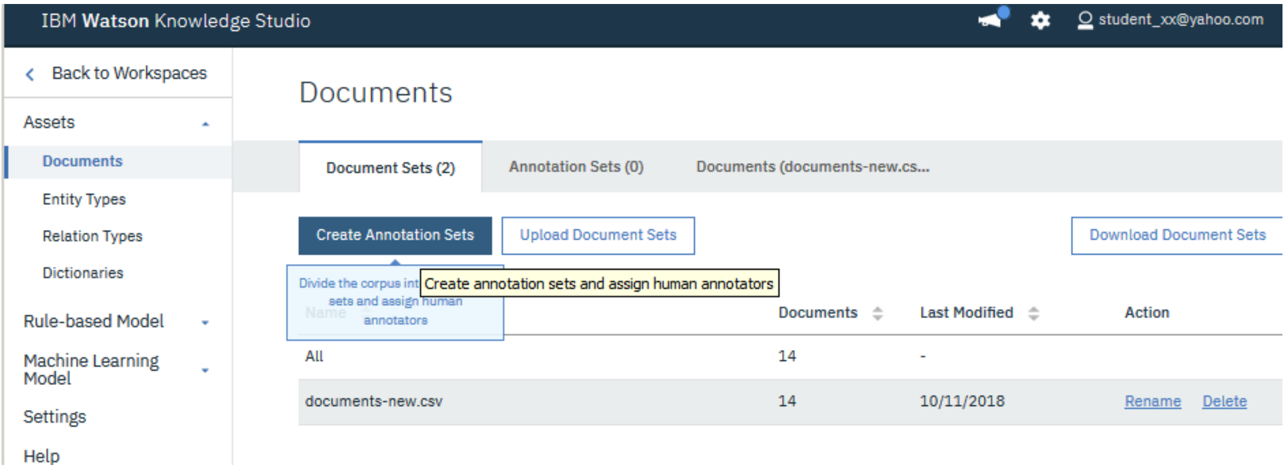
Note: In a realistic scenario, many users with different roles can have access to one workspace to collaborate, the different roles are Admin, Project Manager, and Human Annotator. In this exercise you are using a Lite plan for Watson Knowledge Studio which allows only one user in the workspace (you) with the Administrator role. Normally, you would create as many annotation sets as needed, based on the number of human annotators who are working in the workspace. In this exercise, you will create two annotation sets and you will assign both annotation sets to the same user (you).

Perform the following steps:

__ 1. Within your workspace, click **Assets** > **Documents**.

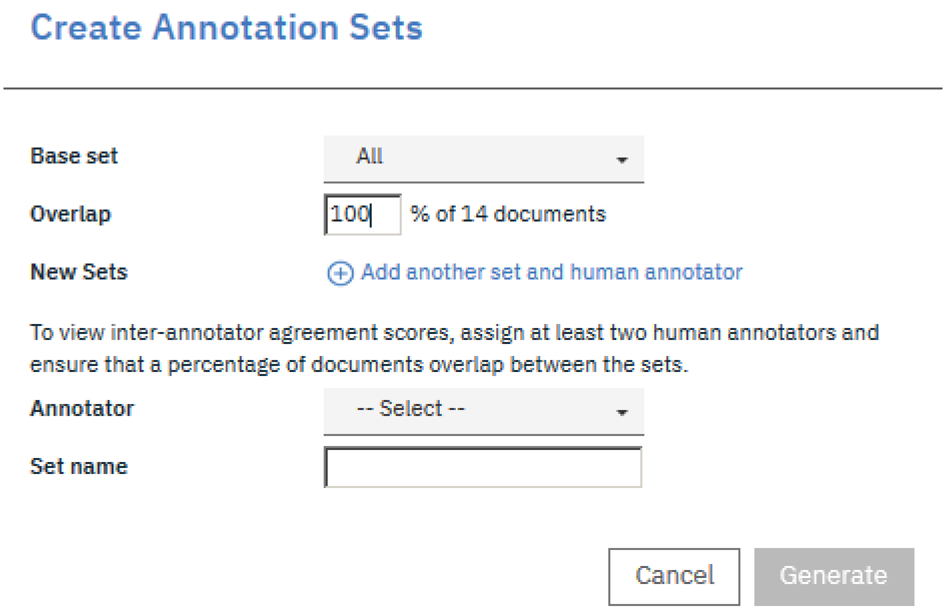
__ 2. Click the **Document Sets** tab.

__ 3. Click **Create Annotation Sets**.



Name	Documents	Last Modified	Action
All	14	-	
documents-new.csv	14	10/11/2018	Rename Delete

The Create Annotation Sets window opens. By default, this window shows the base set, which contains all documents, and fields where you can specify the information for a new annotation set.


Create Annotation Sets

Base set: All

Overlap: 100 % of 14 documents

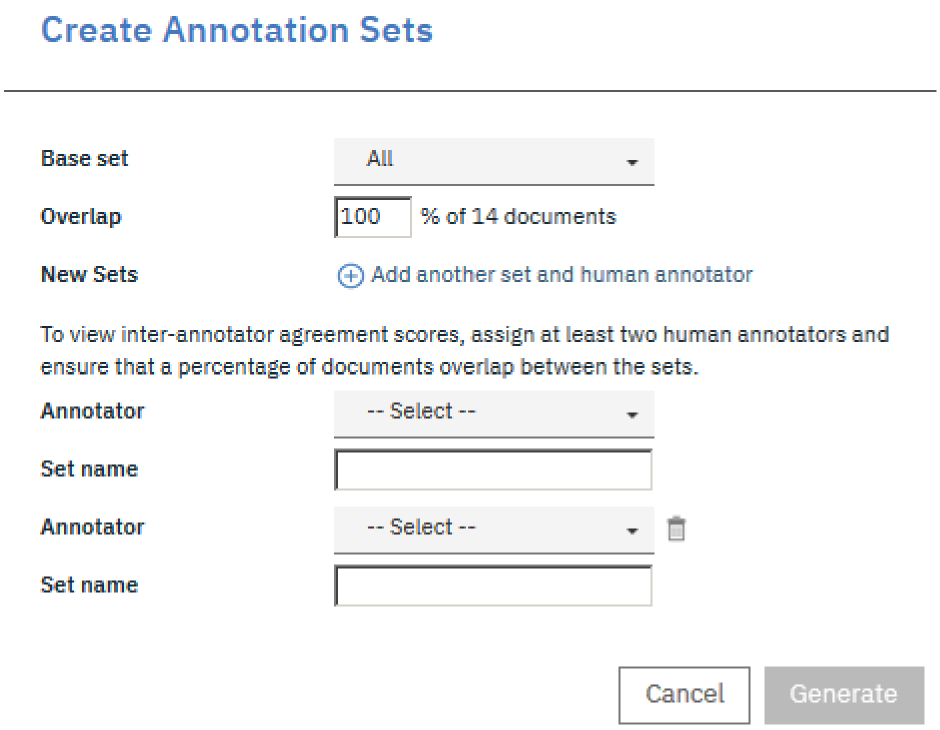
New Sets: Add another set and human annotator

To view inter-annotator agreement scores, assign at least two human annotators and ensure that a percentage of documents overlap between the sets.

Annotator: -- Select --

Set name:

__ 4. Click **Add another set and human annotator** to add fields for an additional annotation set. You can click to add as many annotation sets as you want to create. For this exercise, you need only two annotation sets.



Create Annotation Sets

Base set: All

Overlap: 100 % of 14 documents

New Sets: ⊕ Add another set and human annotator

To view inter-annotator agreement scores, assign at least two human annotators and ensure that a percentage of documents overlap between the sets.

Annotator: -- Select --

Set name:

Annotator: -- Select --

Set name:

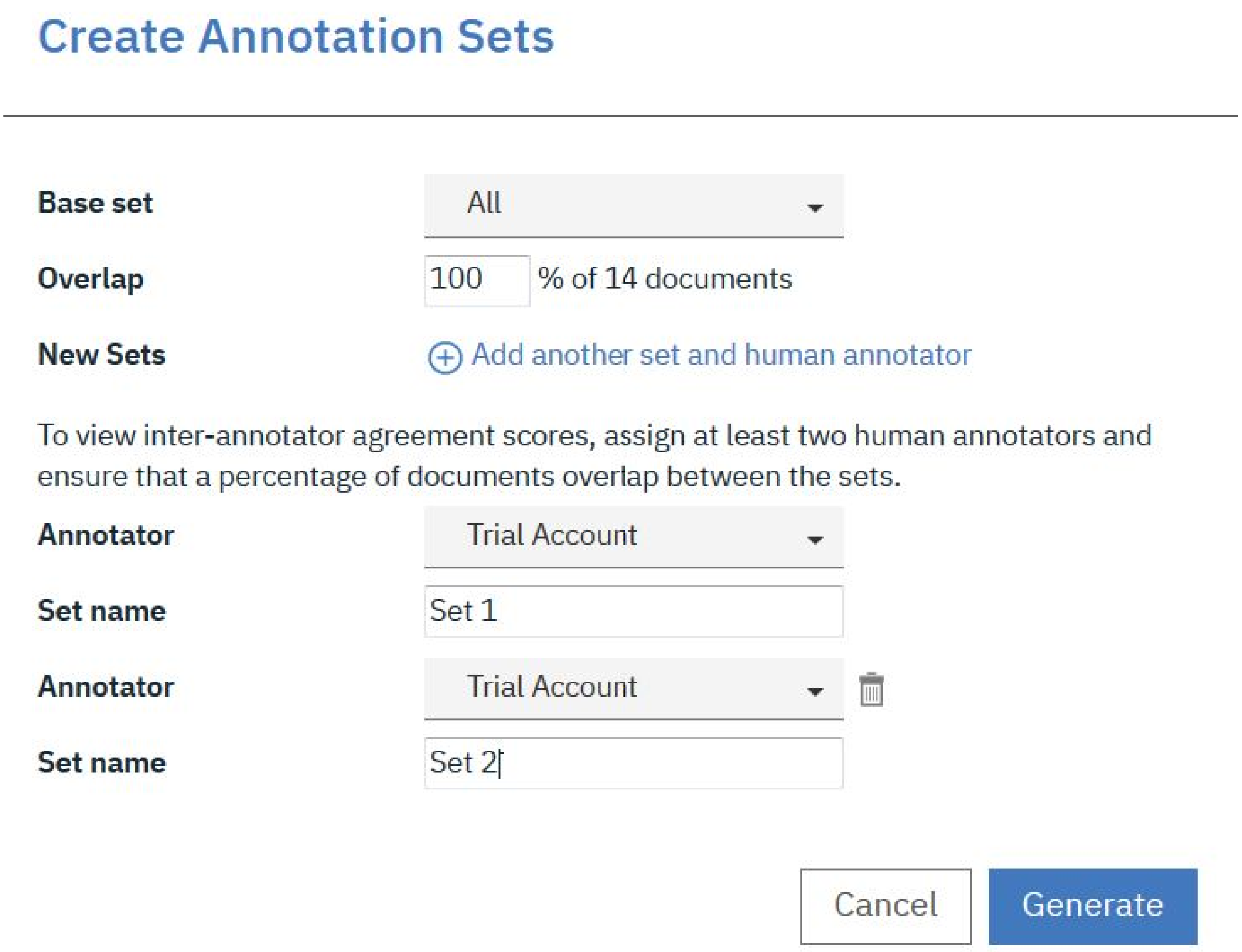
__ 5. In the **Overlap** field, specify 100. This value specifies that you want 100 percent of the documents in the base set to be included in all the new annotation sets so they can be annotated by all human annotators.

__ 6. For each new annotation set, specify the required information.

- In the **Annotator** field, select a human annotator user ID to assign to the new annotation set. For this exercise, the administrator will act as human annotator.

Note: In a realistic scenario, each annotation set is assigned to a different human annotator, as you would have multiple human annotators.

- In the **Set name** field, specify a descriptive name for the annotation set. For this exercise, you can use the names*, Set 1* and *Set 2.*



Create Annotation Sets

Base set: All

Overlap: 100 % of 14 documents

New Sets: ⊕ Add another set and human annotator

To view inter-annotator agreement scores, assign at least two human annotators and ensure that a percentage of documents overlap between the sets.

Annotator: Trial Account

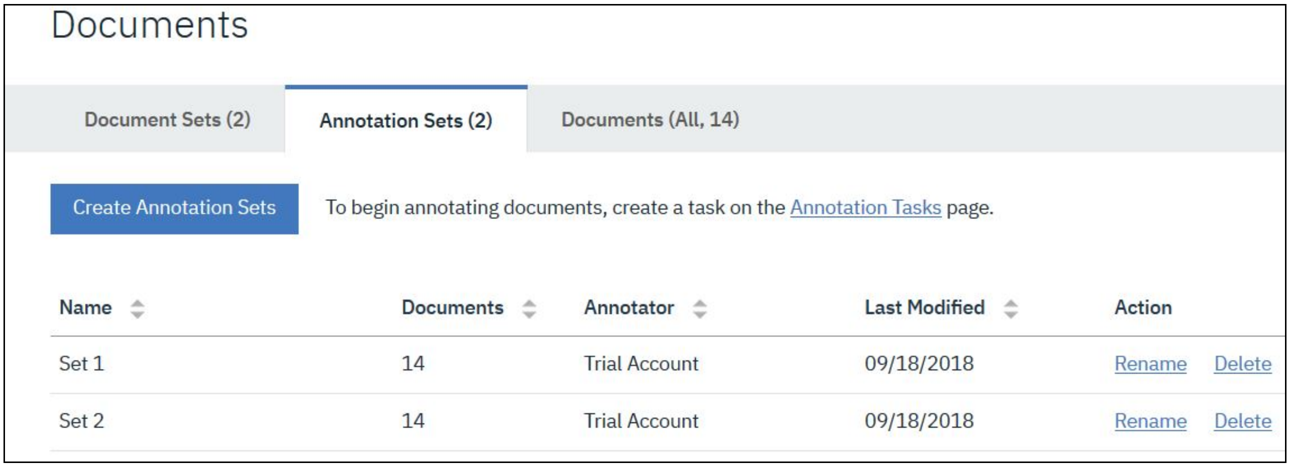
Set name: Set 1

Annotator: Trial Account

Set name: Set 2

__ **7.** Click **Generate**

The new annotation sets are created as shown in the figure.



Documents

To begin annotating documents, create a task on the Annotation Tasks page.

Name	Documents	Annotator	Last Modified	Action
Set 1	14	Trial Account	09/18/2018	Rename Delete
Set 2	14	Trial Account	09/18/2018	Rename Delete

Part 7. Pre-annotating with a dictionary-based annotator

Pre-annotating documents bootstraps the annotation effort of the human annotators.

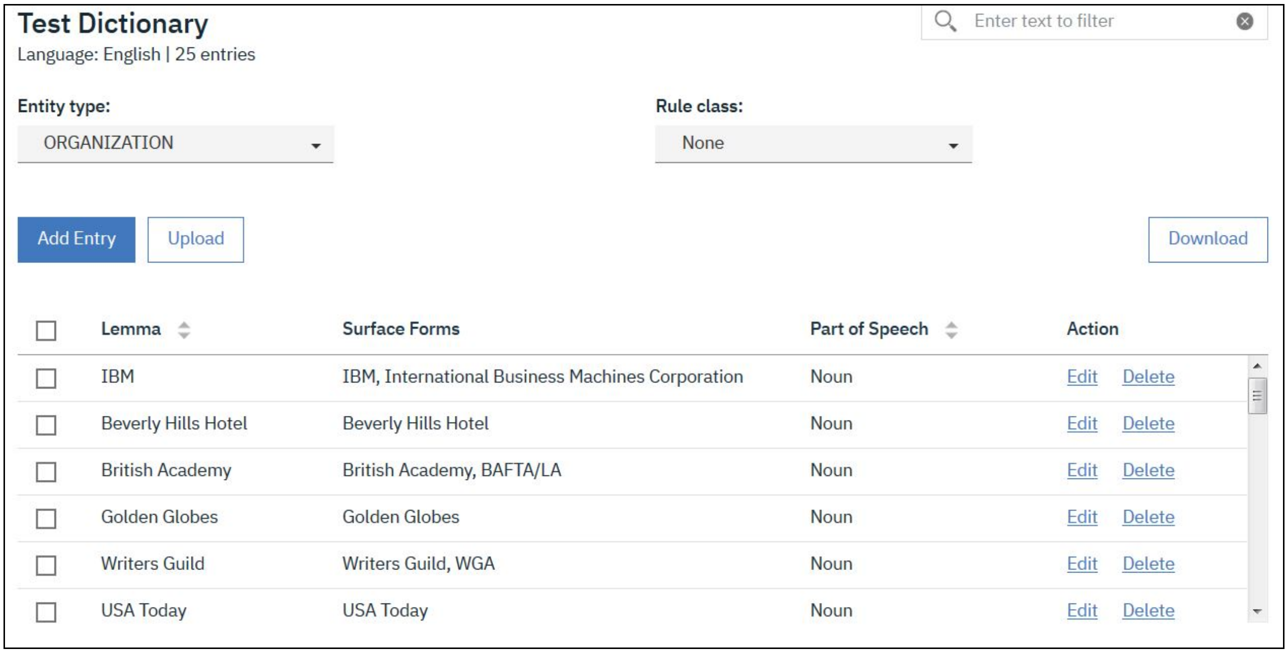
In this part, you will learn how to use a dictionary-based annotator to pre-annotate documents in Knowledge Studio. Pre-annotating documents is an optional step. However, it is a worthwhile step because it makes the job of human annotators easier later.

Perform the following steps:

__ 1. Within your workspace, click **Assets** > **Dictionaries**.

The dictionary *Test Dictionary* that was created in Part 4 Adding a dictionary opens.

__ 2. From the **Entity type** list, select the *ORGANIZATION* entity type to map it to the dictionary *Test dictionary*. Inspect the entries to ensure that they all represent organizations. The *ORGANIZATION* entity type is part of the type system that was created in Part 3 Creating a type system.



Test Dictionary

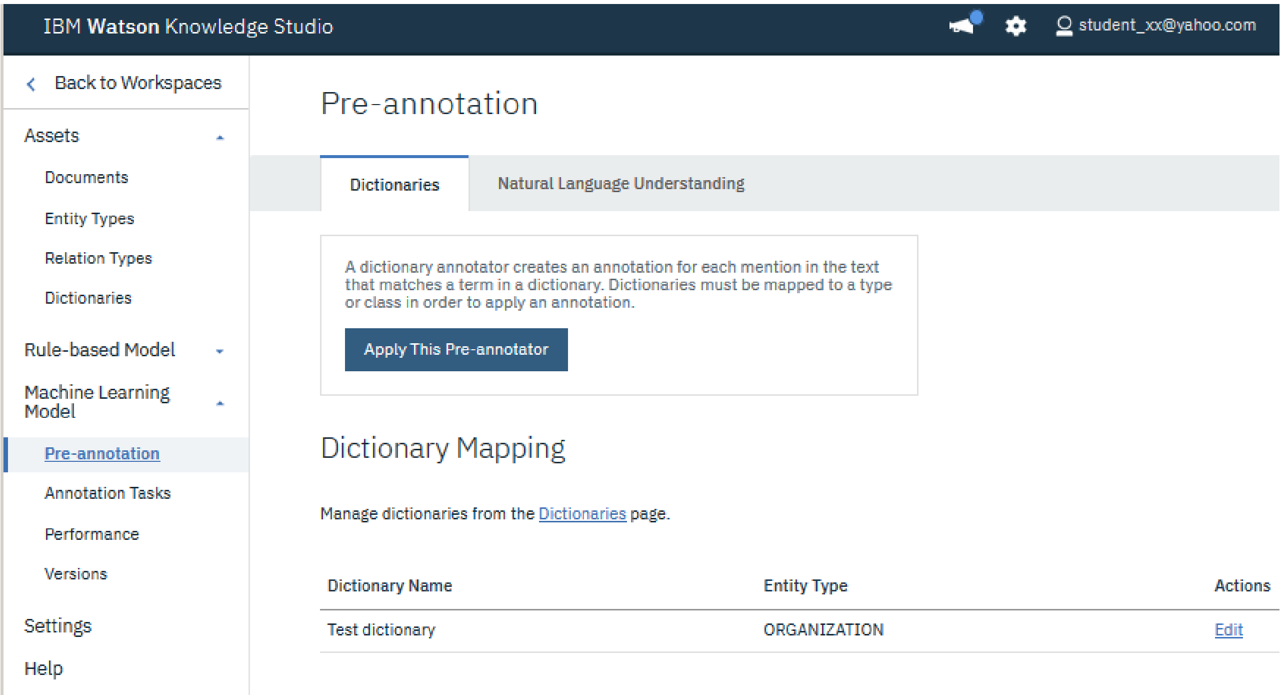
Language: English | 25 entries

Entity type: ORGANIZATION

Rule class: None

□	Lemma	Surface Forms	Part of Speech	Action
□	IBM	IBM, International Business Machines Corporation	Noun	Edit Delete
□	Beverly Hills Hotel	Beverly Hills Hotel	Noun	Edit Delete
□	British Academy	British Academy, BAFTA/LA	Noun	Edit Delete
□	Golden Globes	Golden Globes	Noun	Edit Delete
□	Writers Guild	Writers Guild, WGA	Noun	Edit Delete
□	USA Today	USA Today	Noun	Edit Delete

__ 3. On the **Machine Learning Model** > **Pre-annotation** > **Dictionaries** tab, click **Apply This Pre-annotator**.



Pre-annotation

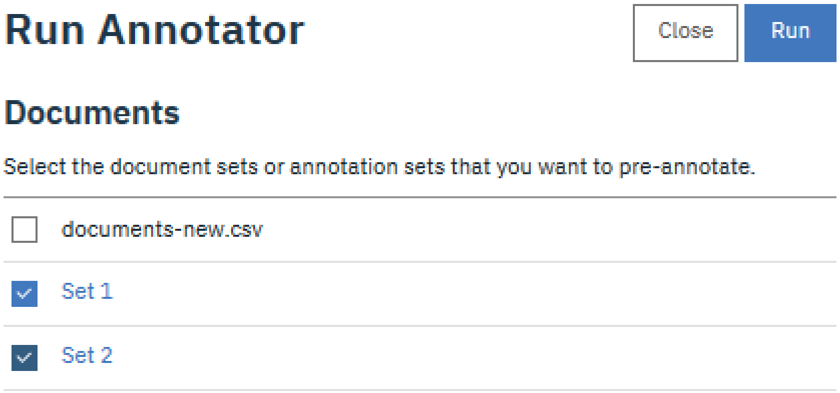
A dictionary annotator creates an annotation for each mention in the text that matches a term in a dictionary. Dictionaries must be mapped to a type or class in order to apply an annotation.

Dictionary Mapping

Manage dictionaries from the Dictionaries page.

Dictionary Name	Entity Type	Actions
Test dictionary	ORGANIZATION	Edit

__ **4.** Select the annotation sets that you created named *Set 1* and *Set 2* and then click **Run**



Run Annotator

Documents

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

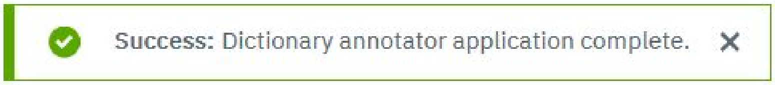
☐ documents-new.csv

☑ Set 1

☑ Set 2

.

__ 5. A notification reporting the annotator completion results is displayed. The documents in the selected sets are pre-annotated by using the dictionary that you created.



Success: Dictionary annotator application complete.

Part 8. Creating an annotation task

Before human annotators can begin adding annotations to documents, the annotation process manager must create an annotation task.

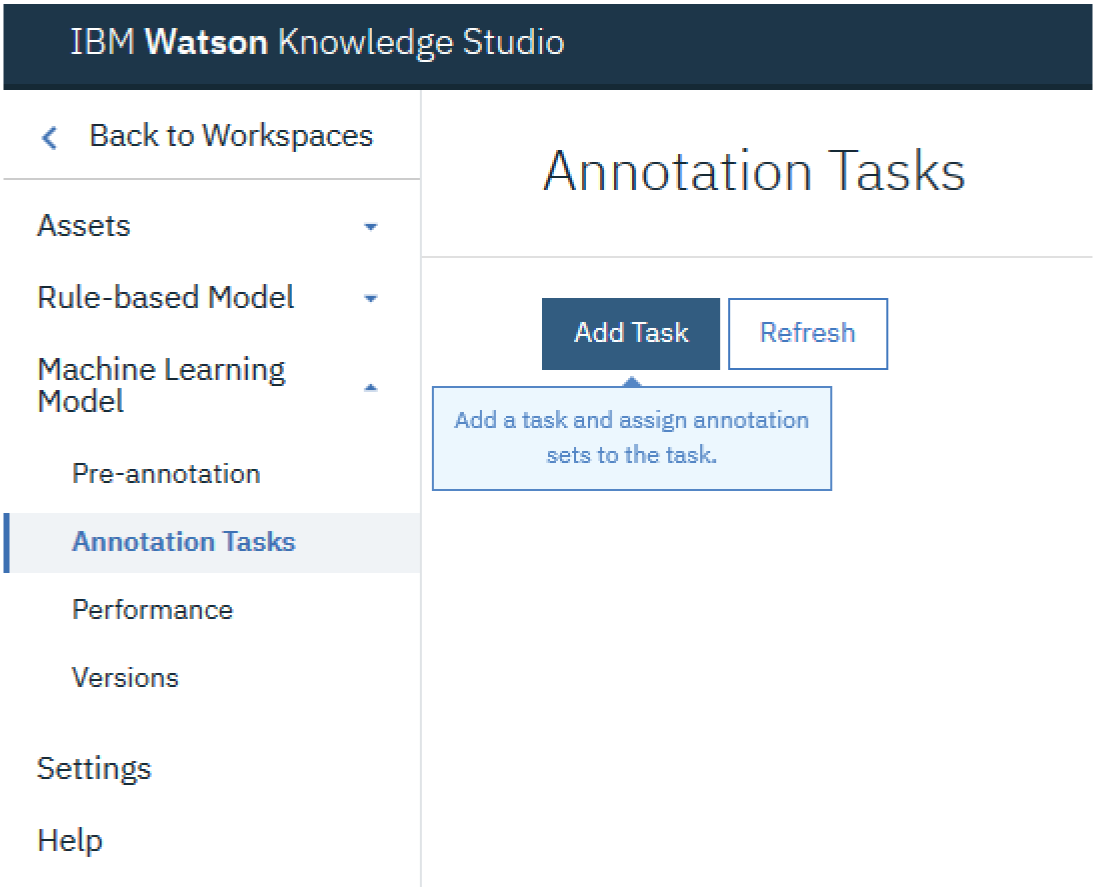
The annotation task specifies which documents are to be annotated. To compare how well the human annotators perform, and to see how consistently they apply the annotation guidelines, you must include at least two human annotators in the task. In addition, some percentage of documents must occur in all of the annotation sets that you add to the task (you specify the overlap percentage when you create the annotation sets).

In this part, you will learn how to use annotation tasks to track the work of human annotators in Knowledge Studio.

Perform the following steps:

__ **1.** Within your workspace, click **Machine Learning Model** > **Annotation Tasks**.

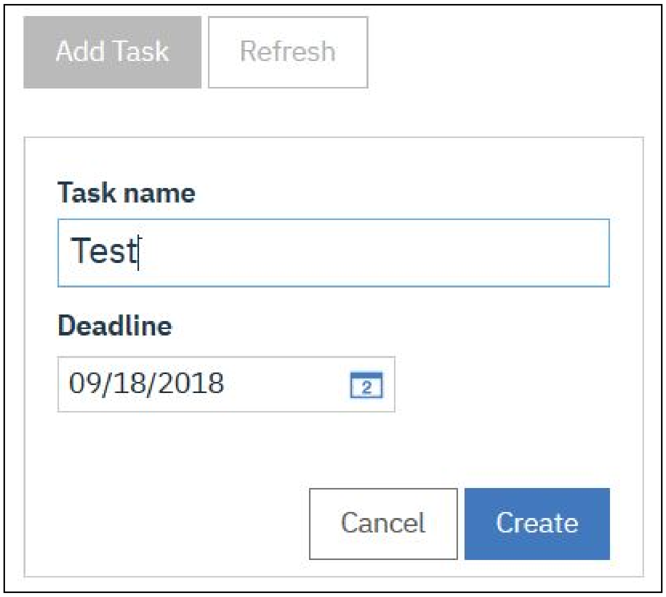
__ 2. On the Tasks page, click **Add Task**.



__ 3. Specify the details for the task:

- In the **Task name** field, enter *Test*.
- In the **Deadline** field, select a date in the future.

__ 4. Click **Create**.



Task name

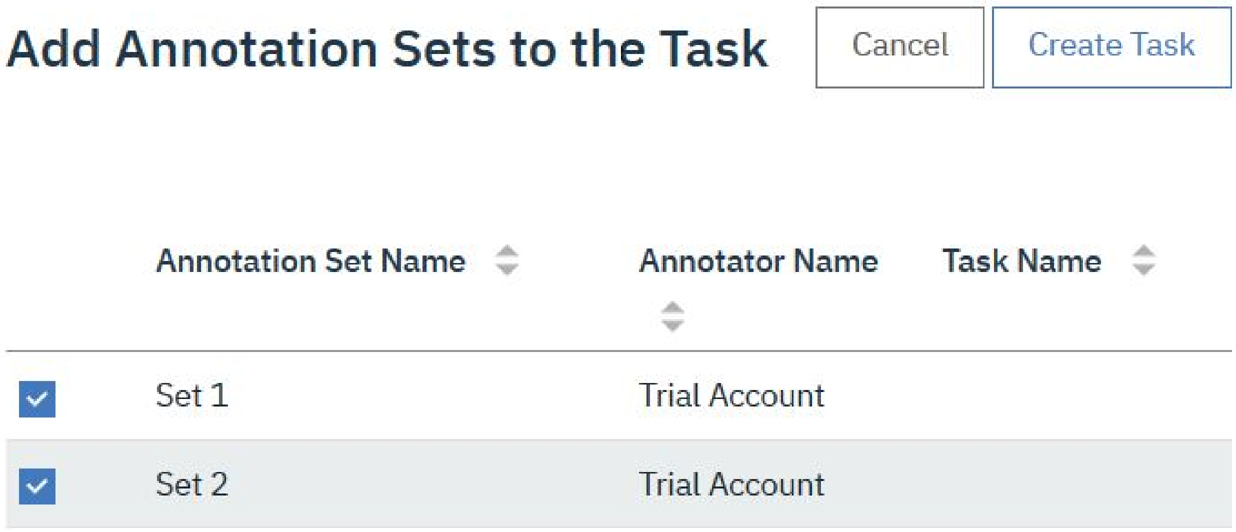
Test

Deadline

09/18/2018

__ 5. Select the annotation sets that you created previously *Set 1* and *Set 2*.

Selecting both annotation sets specifies that both sets must be annotated by their assigned human annotators to complete this task.

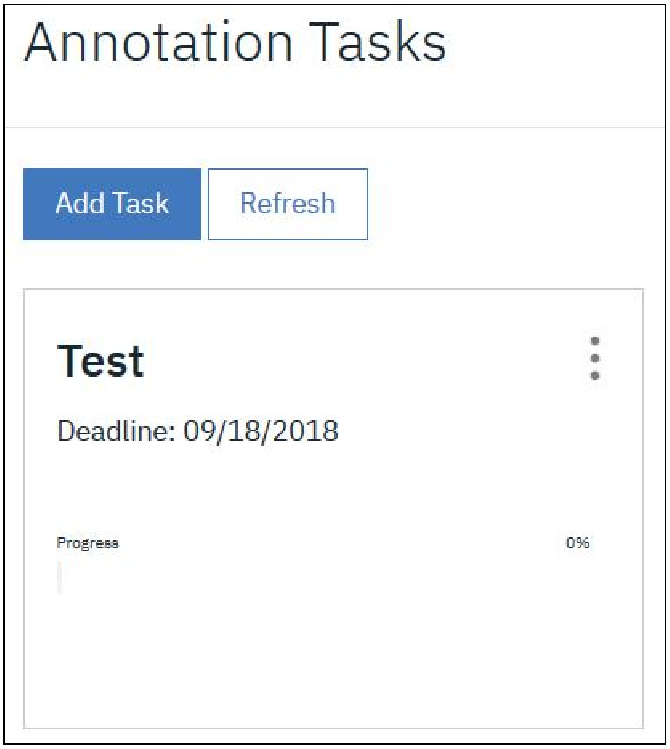


Add Annotation Sets to the Task

	Annotation Set Name	Annotator Name	Task Name
☑	Set 1	Trial Account	
☑	Set 2	Trial Account	

__ 6. Click **Create Task**.

The Test annotation task is added.



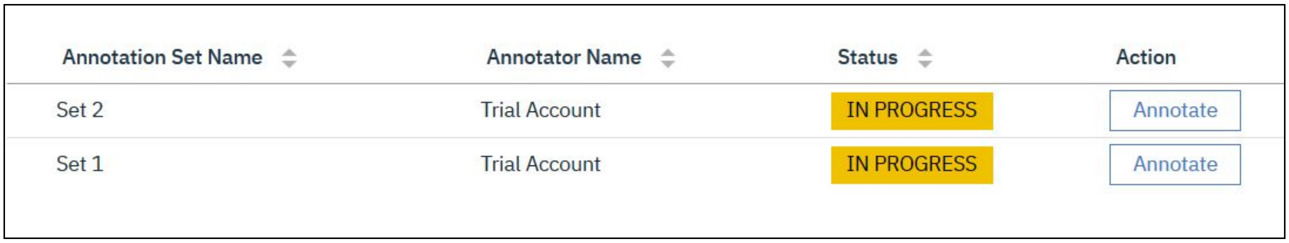
Annotation Tasks

Test

Deadline: 09/18/2018

Progress 0%

__ 8. Click **Test**. As human annotators begin annotating documents, you can open tasks to see their progress.



Annotation Set Name	Annotator Name	Status	Action
Set 2	Trial Account	IN PROGRESS	Annotate
Set 1	Trial Account	IN PROGRESS	Annotate

Part 9. Annotating documents

When a human annotator annotates a document, the document is opened in the *ground truth editor*. The ground truth editor is a visual tool that human annotators use to apply labels to text.

The goal of human annotation is to label mentions, relations, and coreferenced mentions so that the machine learning model can be trained to detect these patterns in unseen text.

Information: Ground truth is the collection of vetted data that is used to adapt Watson to a particular domain. In Knowledge Studio, human annotators, who are typically experts in the subject matter of the target domain, play a major role in determining ground truth.

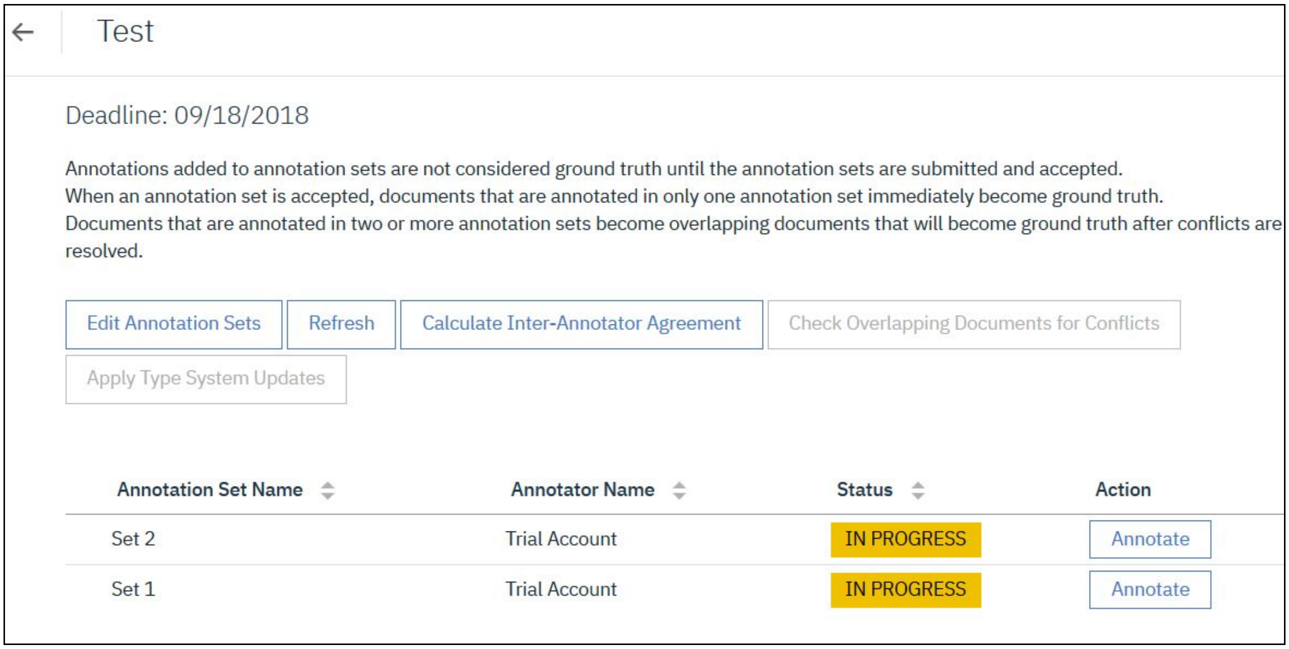
In this part, you will learn how to use the *ground truth editor* to annotate documents in Knowledge Studio.

Perform the following steps:

__ 1. Within your workspace, click **Machine Learning Model** > **Annotation Tasks.**

__ 2. Open the *Test* annotation task you just created in Part 8 Creating an annotation task.

__ 3. Click **Annotate** for one of the assigned annotation sets.



← Test

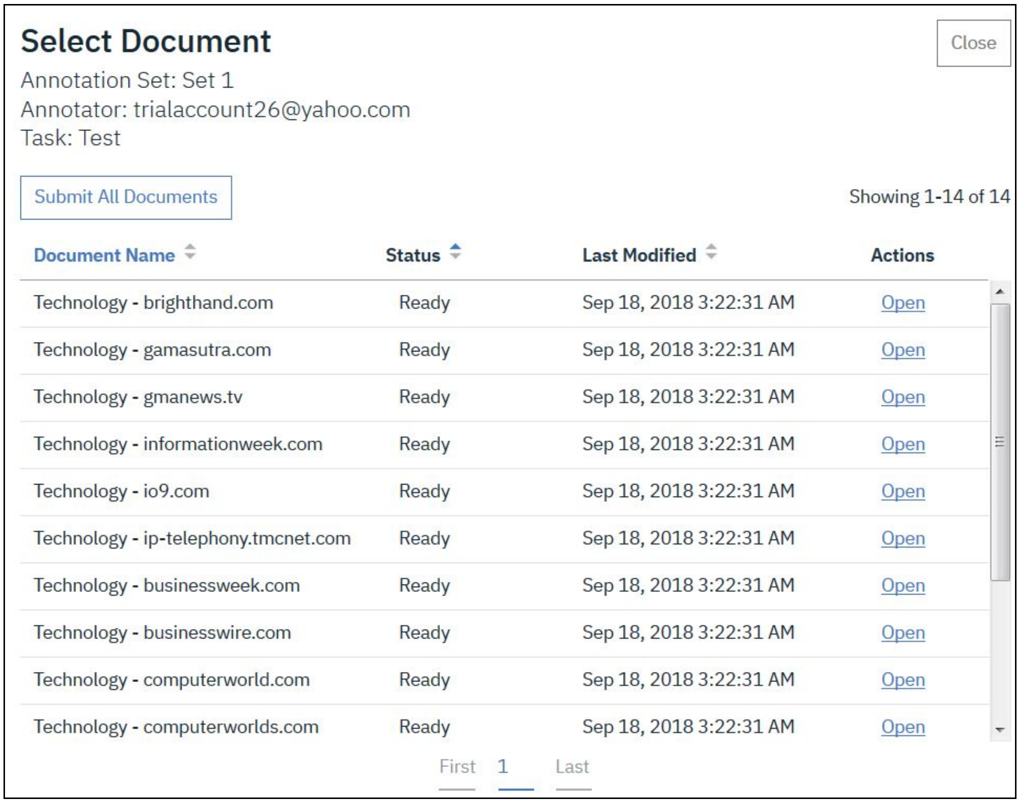
Deadline: 09/18/2018

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
Set 2	Trial Account	IN PROGRESS	Annotate
Set 1	Trial Account	IN PROGRESS	Annotate

__ 4. From the list of documents, find the **Technology - gmanews.tv** document and open it.



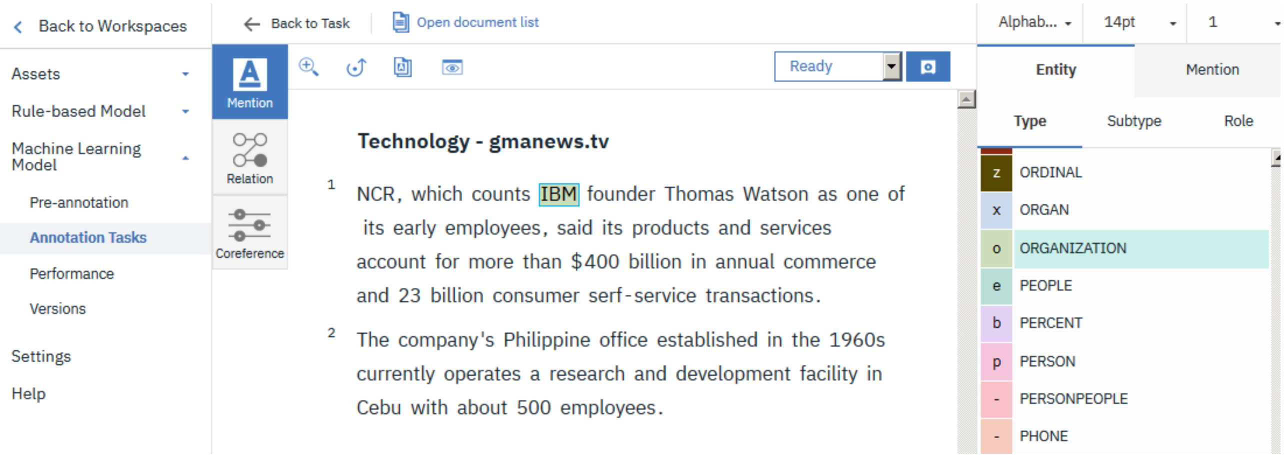
Select Document

Annotation Set: Set 1
Annotator: trialaccount26@yahoo.com
Task: Test

Showing 1-14 of 14

Document Name	Status	Last Modified	Actions
Technology - brighthand.com	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - gamasutra.com	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - gmanews.tv	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - informationweek.com	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - io9.com	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - ip-telephony.tmcnet.com	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - businessweek.com	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - businesswire.com	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - computerworld.com	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - computerworlds.com	Ready	Sep 18, 2018 3:22:31 AM	Open

Notice in the following figure that the term *IBM* was already annotated with the *ORGANIZATION* entity type. This annotation was added by the dictionary pre-annotator that was applied in Part 7 Pre-annotating with a dictionary-based annotator. This pre-annotation is correct, so it does not need to be modified (colors may be different for you).



Technology - gmanews.tv

1 NCR, which counts IBM founder Thomas Watson as one of its early employees, said its products and services account for more than $400 billion in annual commerce and 23 billion consumer serf-service transactions.

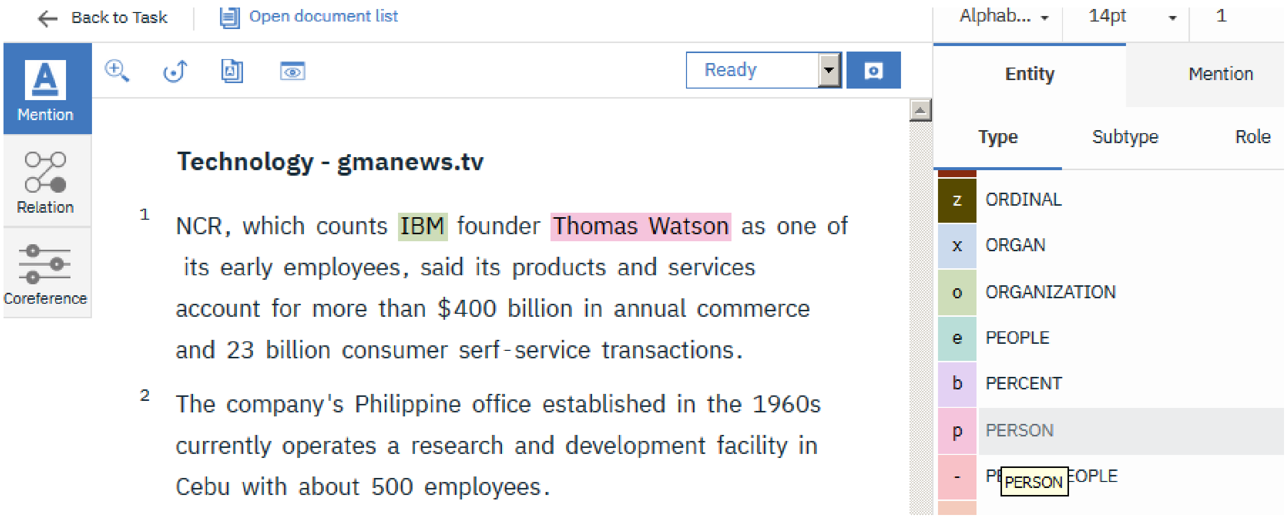
2 The company's Philippine office established in the 1960s currently operates a research and development facility in Cebu with about 500 employees.

__ 5. Annotate a mention.

__ a. Click the **Entity** tab.

__ b. In the document body, select the text *Thomas Watson.*

__ c. In the list of entity types, click **PERSON**. The entity type PERSON is applied to the selected mention.



Technology - gmanews.tv

1 NCR, which counts IBM founder Thomas Watson as one of its early employees, said its products and services account for more than $400 billion in annual commerce and 23 billion consumer serf-service transactions.

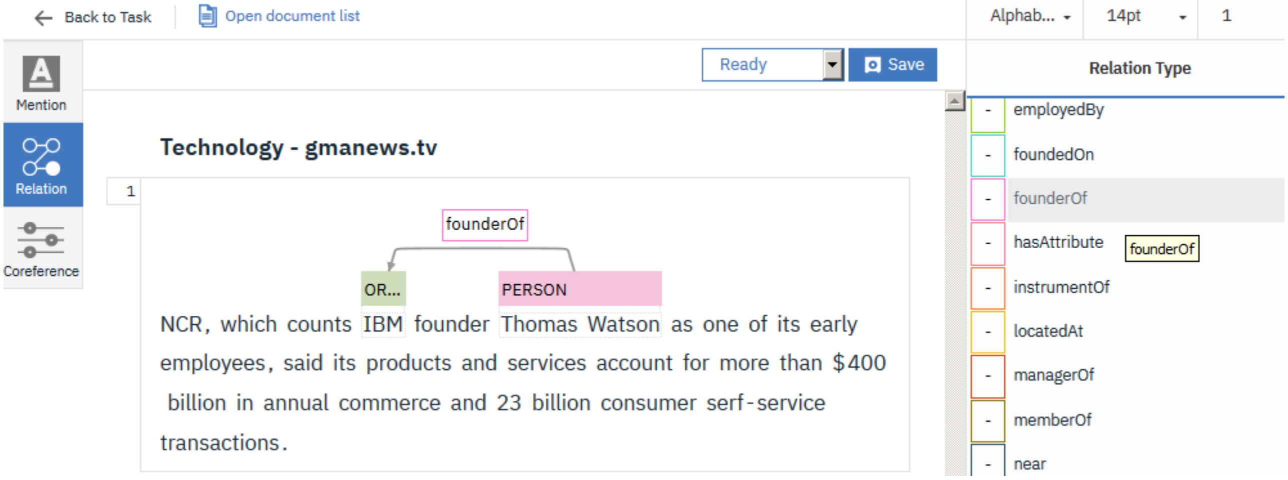
2 The company's Philippine office established in the 1960s currently operates a research and development facility in Cebu with about 500 employees.

__ 6. Annotate a relation.

__ a. Click the **Relation** tab.

__ b. Select the *Thomas Watson* and *IBM* mentions (in that order). To select a mention, click the entity type label above the text.

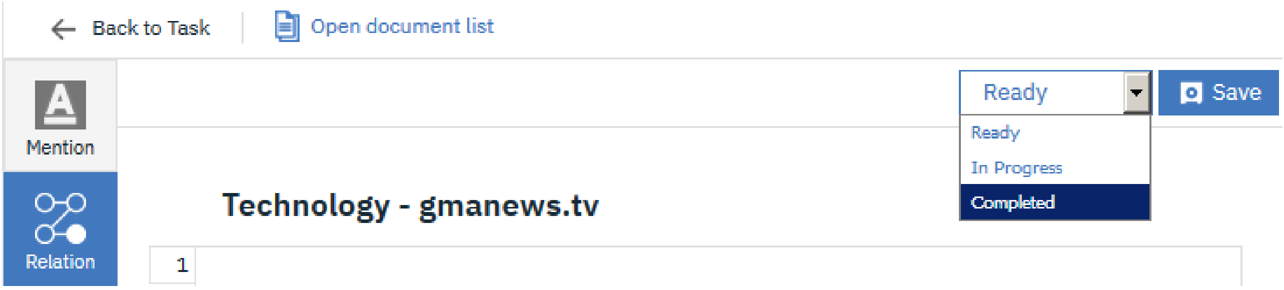
__ c. In the list of relation types, click **founderOf**. The two mentions are connected with a *founderOf* relationship.



Technology - gmanews.tv

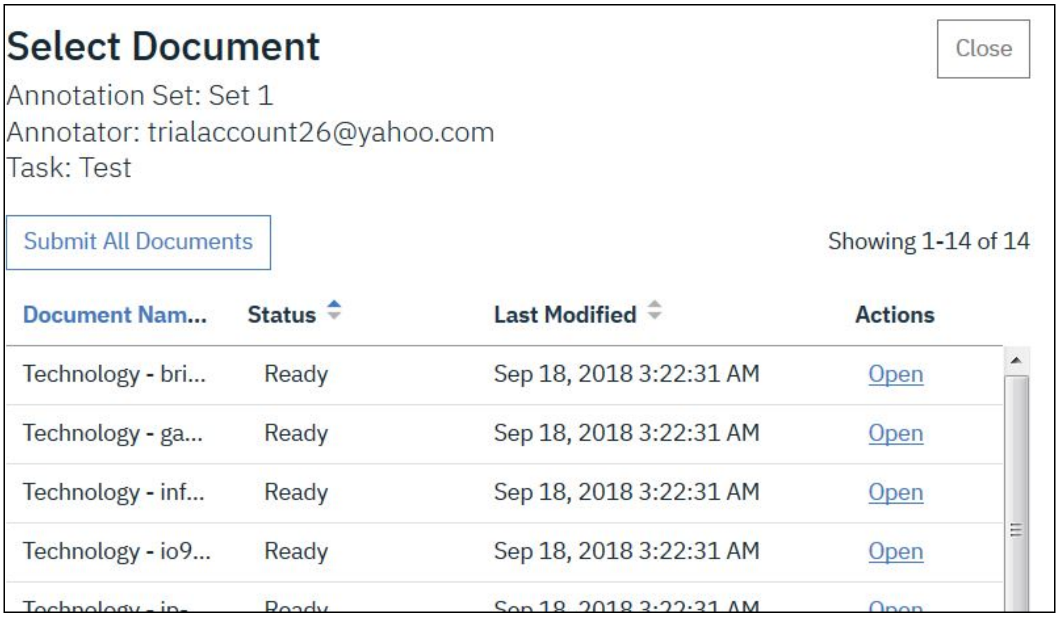
1 NCR, which counts IBM founder Thomas Watson as one of its early employees, said its products and services account for more than $400 billion in annual commerce and 23 billion consumer serf-service transactions.

__ 7. From the status menu, select **Completed**, and then click **Save**.



__ 8. Repeat the previous steps to create more annotations from the documents in the set to practice the annotation process.

__ 9. Click **Open document list** to return to the list of documents for this task and click **Submit All Documents** to submit the documents for approval.



Select Document

Close

Annotation Set: Set 1
Annotator: trialaccount26@yahoo.com
Task: Test

Submit All Documents

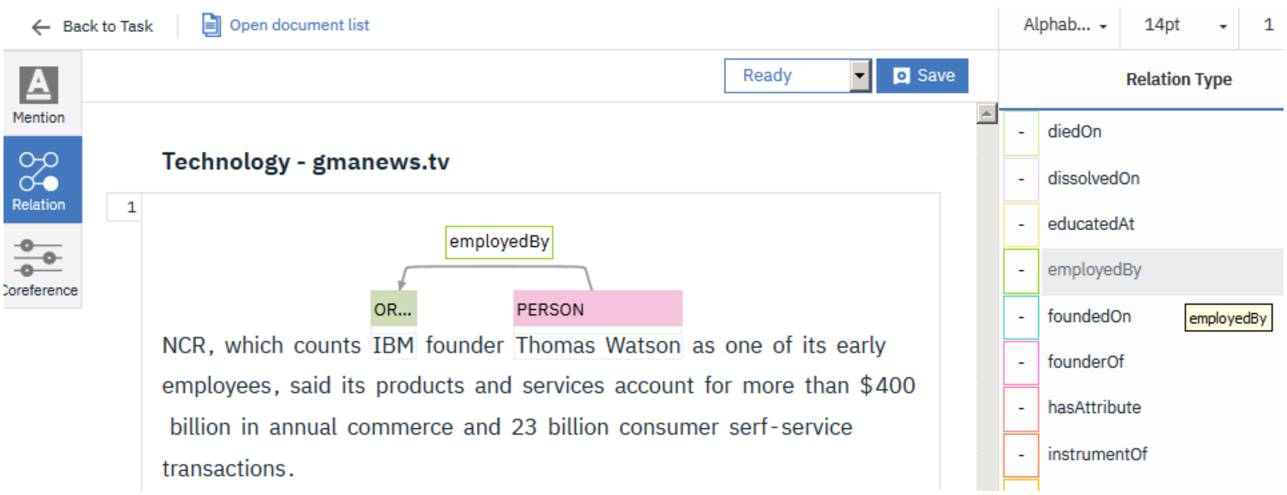
Showing 1-14 of 14

Document Nam...	Status	Last Modified	Actions
Technology - bri...	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - ga...	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - inf...	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - io9...	Ready	Sep 18, 2018 3:22:31 AM	Open
Technology - in...	Ready	Sep 18, 2018 3:22:31 AM	Open

__ 10. At the confirmation prompt click **OK**.

__ 11. Close this annotation set, and then open the other annotation set in the *Test* task.

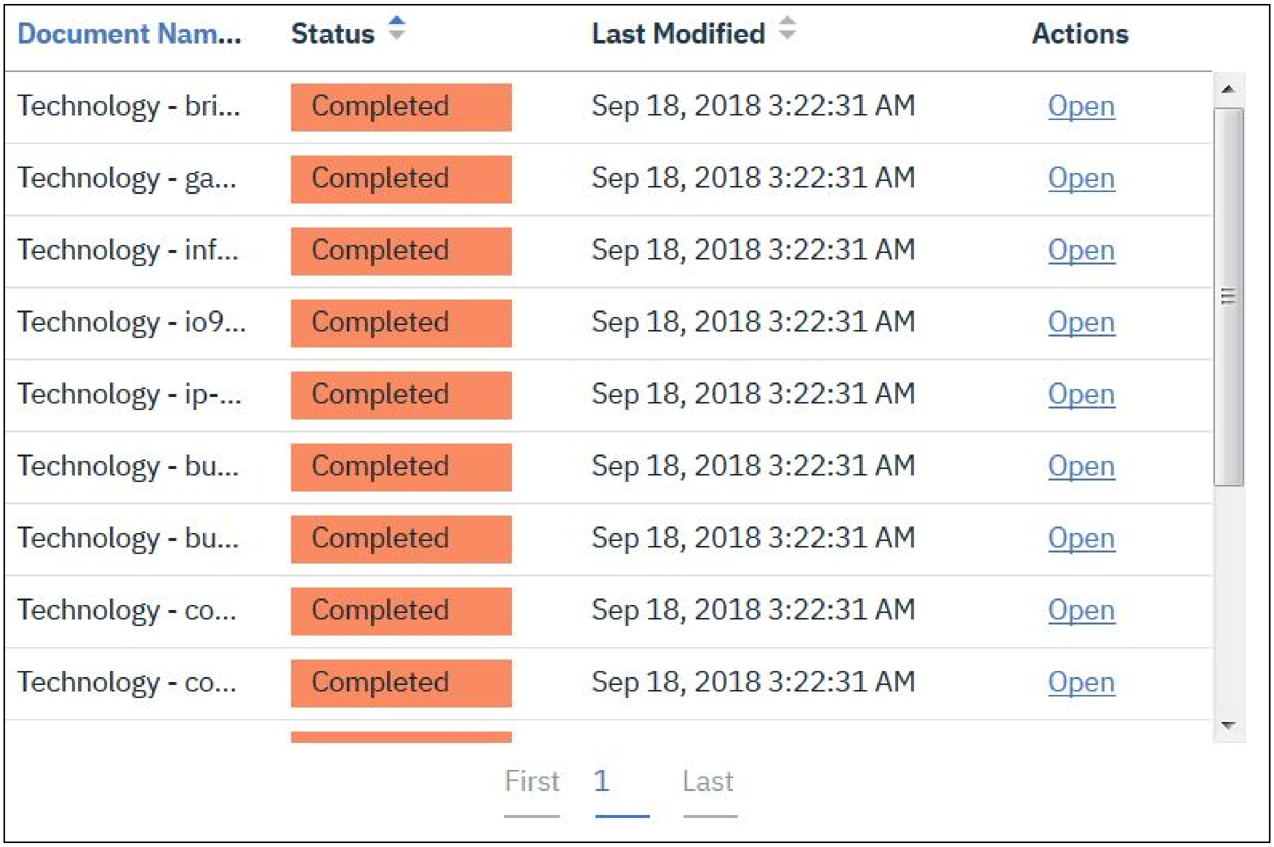
__ 12. Repeat the same annotations done in the previous example in the *Technology - gmanews.tv* document, except this time, use the ***employedBy*** relation instead of the ***founderOf*** relation, when creating the relation annotation.



Technology - gmanews.tv

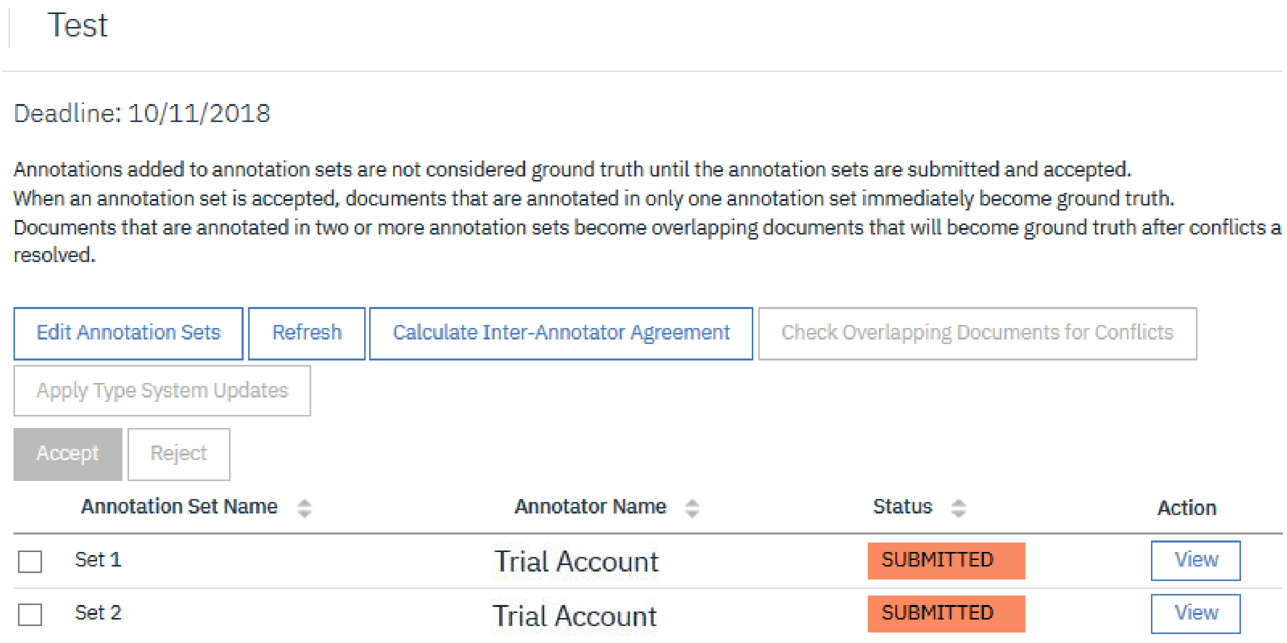
1 NCR, which counts IBM founder Thomas Watson as one of its early employees, said its products and services account for more than $400 billion in annual commerce and 23 billion consumer serf-service transactions.

__ 13. After you complete the annotations for the second annotation set, click **Submit All Documents**.



Document Nam...	Status	Last Modified	Actions
Technology - bri...	Completed	Sep 18, 2018 3:22:31 AM	Open
Technology - ga...	Completed	Sep 18, 2018 3:22:31 AM	Open
Technology - inf...	Completed	Sep 18, 2018 3:22:31 AM	Open
Technology - io9...	Completed	Sep 18, 2018 3:22:31 AM	Open
Technology - ip-...	Completed	Sep 18, 2018 3:22:31 AM	Open
Technology - bu...	Completed	Sep 18, 2018 3:22:31 AM	Open
Technology - bu...	Completed	Sep 18, 2018 3:22:31 AM	Open
Technology - co...	Completed	Sep 18, 2018 3:22:31 AM	Open
Technology - co...	Completed	Sep 18, 2018 3:22:31 AM	Open

Both annotation sets should now be in status SUBMITTED.



Test

Deadline: 10/11/2018

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 a
resolved.

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

	Annotation Set Name	Annotator Name	Status	Action
□	Set 1	Trial Account	SUBMITTED	View
□	Set 2	Trial Account	SUBMITTED	View

Part 10. Analyzing inter-annotator agreement

To determine whether different human annotators are annotating overlapping documents consistently, review the inter-annotator agreement (IAA) scores.

Knowledge Studio calculates IAA scores by examining all overlapping documents in all document sets in the task, regardless of the status of the document sets. The IAA scores show how different human annotators annotated mentions, relations, and coreference chains. It is a good idea to check IAA scores periodically and verify that human annotators are consistent with each other.

At the end of Part 9 Annotating documents, the human annotators submitted all the document sets for approval. If the inter-annotator agreement scores are acceptable, you can approve the document sets. If you reject a document set, it is returned to the human annotator for improvement.

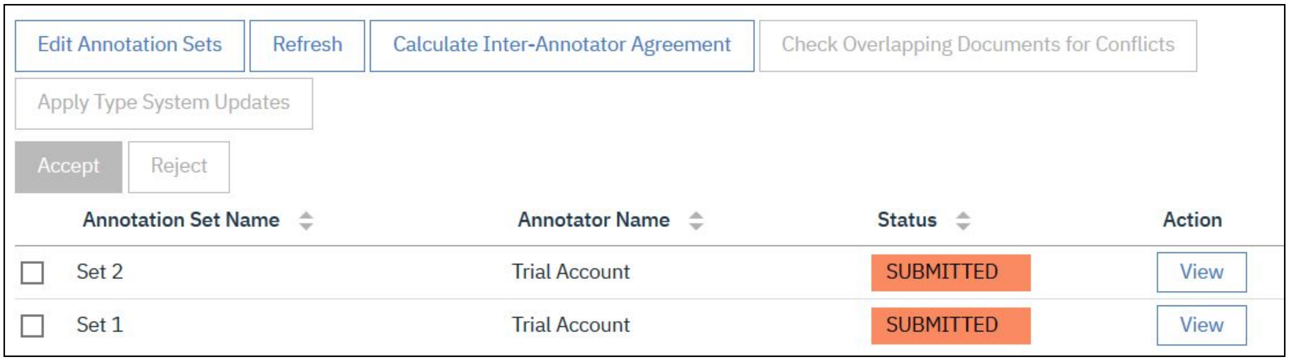
In this part, you will learn how to compare the work of multiple human annotators in Knowledge Studio.

Perform the following steps:

__ 1. Within your workspace, click **Machine Learning Model** > **Annotation Tasks,** and click the *Test* task.

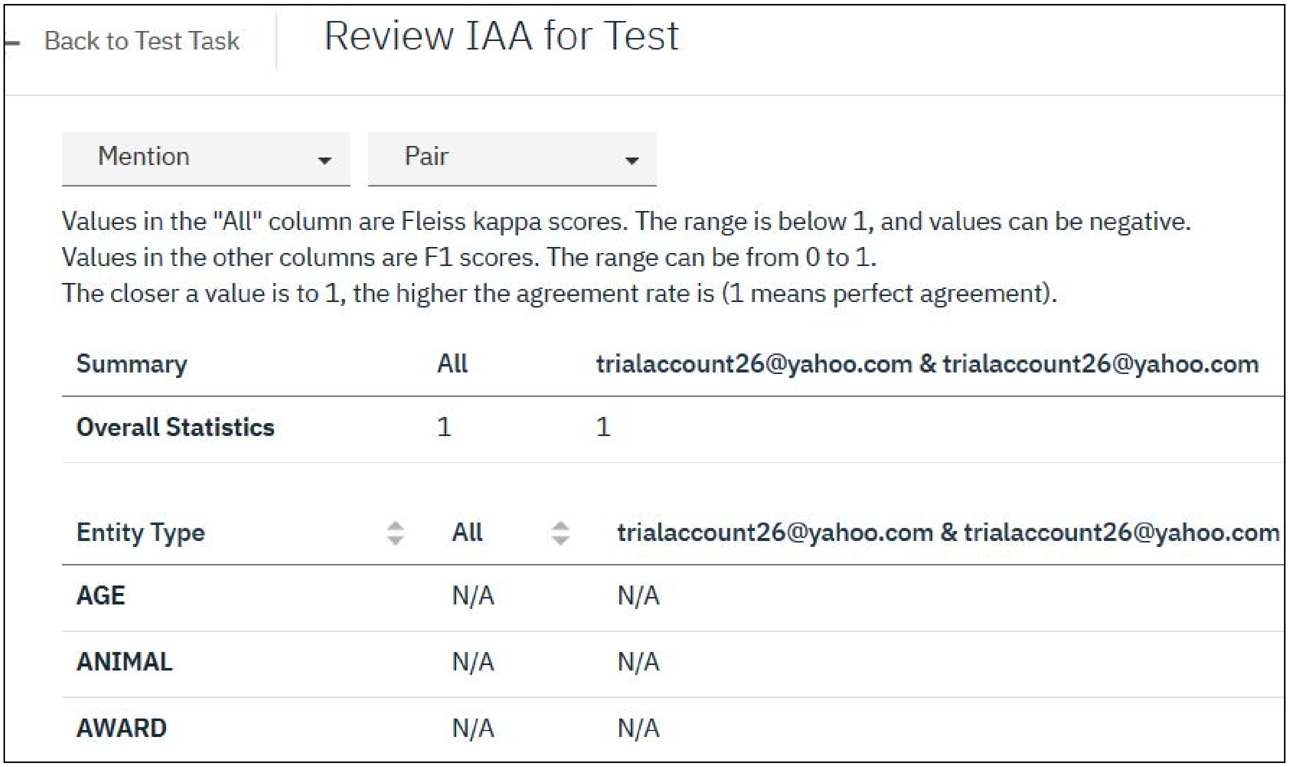
In the **Status** column, you can see that the document sets are submitted.

__ 2. Click **Calculate Inter-Annotator Agreement**.


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

	Annotation Set Name	Annotator Name	Status	Action
□	Set 2	Trial Account	SUBMITTED	View
□	Set 1	Trial Account	SUBMITTED	View

__ 3. View IAA scores for mention, relations, and coreference chains by clicking the first menu. You can also view agreement by pairs of human annotators. You can also view agreement by specific documents. In general, aim for a score of 0.8 out of 1, where 1 means perfect agreement. Because you annotated only two entity types in this exercise, most of the entity type scores are **N/A** (not applicable), which means no information is available to give a score.



Review IAA for Test

Values in the "All" column are Fleiss kappa scores. The range is below 1, and values can be negative.
Values in the other columns are F1 scores. The range can be from 0 to 1.
The closer a value is to 1, the higher the agreement rate is (1 means perfect agreement).

Summary	All	trialaccount26@yahoo.com & trialaccount26@yahoo.com
Overall Statistics	1	1

Entity Type	All	trialaccount26@yahoo.com & trialaccount26@yahoo.com
AGE	N/A	N/A
ANIMAL	N/A	N/A
AWARD	N/A	N/A

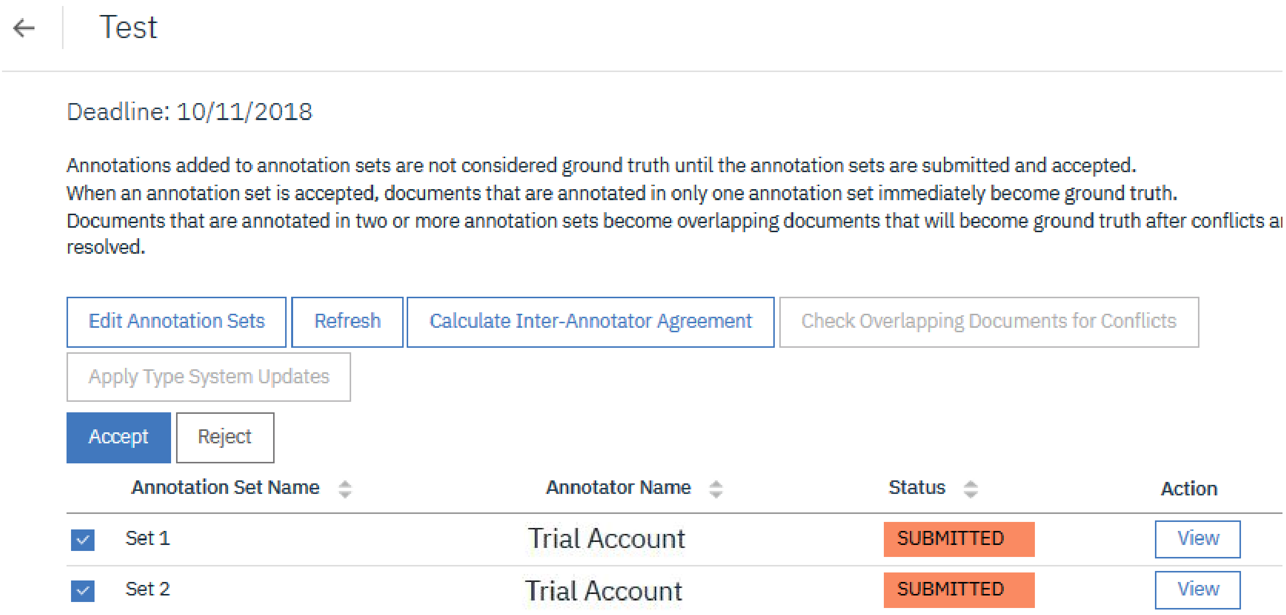
__ 4. Ensure that **Mention** is selected and scroll through the list and check the score for the ORGANIZATION and PERSON entities. You will find that they have a score of 1 to show complete agreement.

__ 5. Select Relation and locate the *employedBy* and *founderOf* relations. You will find a score of 0 to show complete disagreement.

__ 6. After you review the scores, you can decide whether you want to approve or reject document sets that are in the SUBMITTED status. Take one of these actions:

- If the scores are acceptable for an annotation set, select the check box and click **Accept**. Documents that do not overlap with other document sets are promoted to ground truth. Documents that do overlap must first be reviewed through adjudication (discussed in Part 11) so that conflicts can be resolved.
- If the scores are not acceptable for an annotation set, select the check box and click **Reject**. The document set needs to be revisited by the human annotator to improve the annotations.

For this exercise, click **Back to Test Task**, make sure both annotation sets are selected, and then click **Accept** to accept both document sets.


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

Deadline: 10/11/2018

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 a resolved.

	Annotation Set Name	Annotator Name	Status	Action
☑	Set 1	Trial Account	SUBMITTED	View
☑	Set 2	Trial Account	SUBMITTED	View

At the Confirmation prompt, click **OK.**

When you evaluate the inter-annotator agreement scores, you can see how different pairs of human annotators annotated the same document. If the inter-annotator agreement score is acceptable, you accept the document set, otherwise, you reject it to be revisited by the human annotator for improvements.

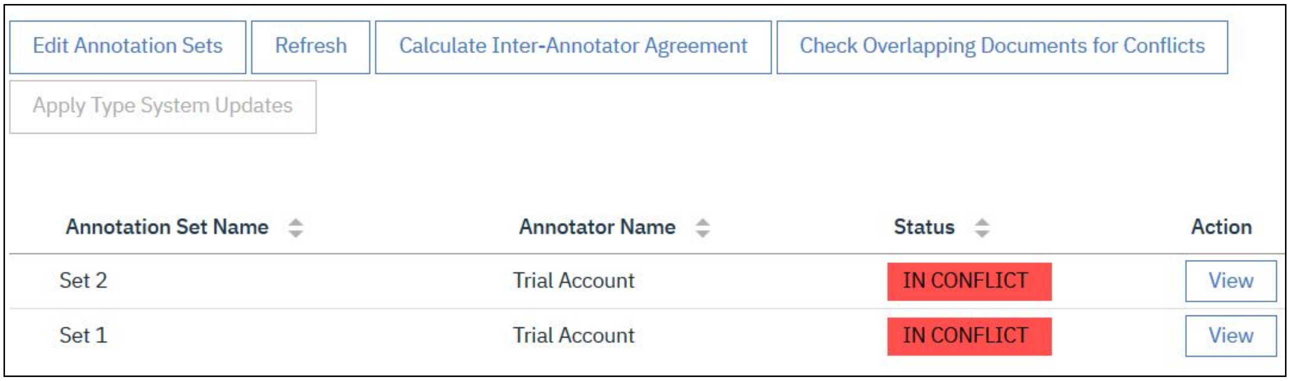
Part 11. Adjudicating conflicts in annotated documents

When you approve a document set, only the documents that do not overlap with other document sets are promoted to ground truth. If a document is part of the overlap between multiple document sets, you must adjudicate any annotation conflicts before the document can be promoted to ground truth.

In this part, you will learn how to adjudicate conflicts in documents that overlap between document sets in Knowledge Studio.

Perform the following steps:

__ 1. Within your workspace, click **Machine Learning Model** > **Annotation Tasks,** and click the *Test* task.

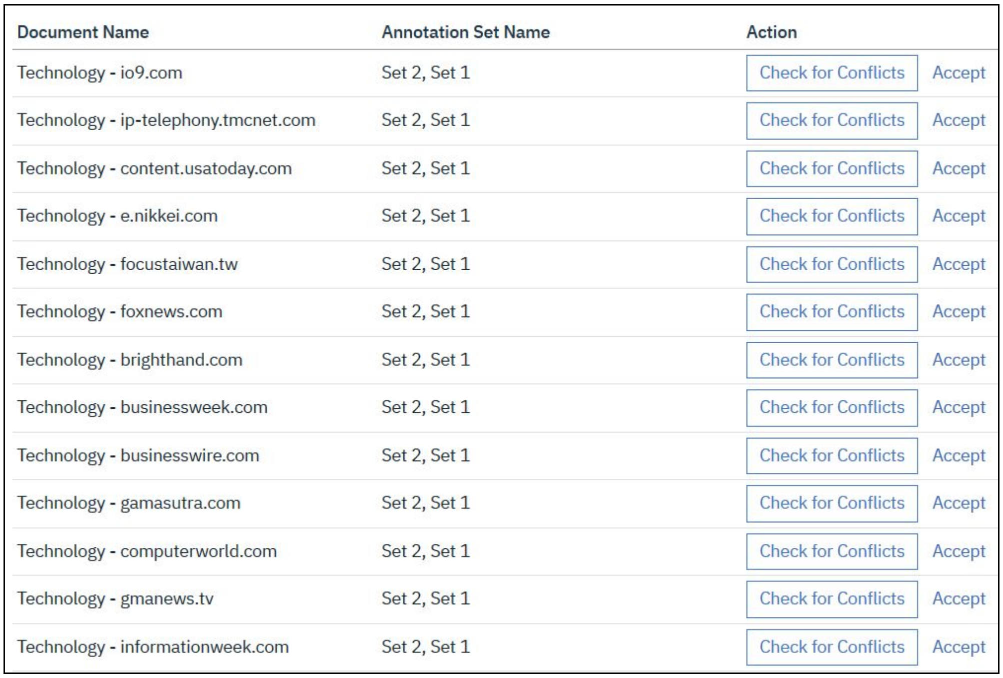

Edit Annotation Sets
Refresh
Calculate Inter-Annotator Agreement
Check Overlapping Documents for Conflicts
Apply Type System Updates

Annotation Set Name	Annotator Name	Status	Action
Set 2	Trial Account	IN CONFLICT	View
Set 1	Trial Account	IN CONFLICT	View

__ 2. Click **Check Overlapping Documents for Conflicts**.

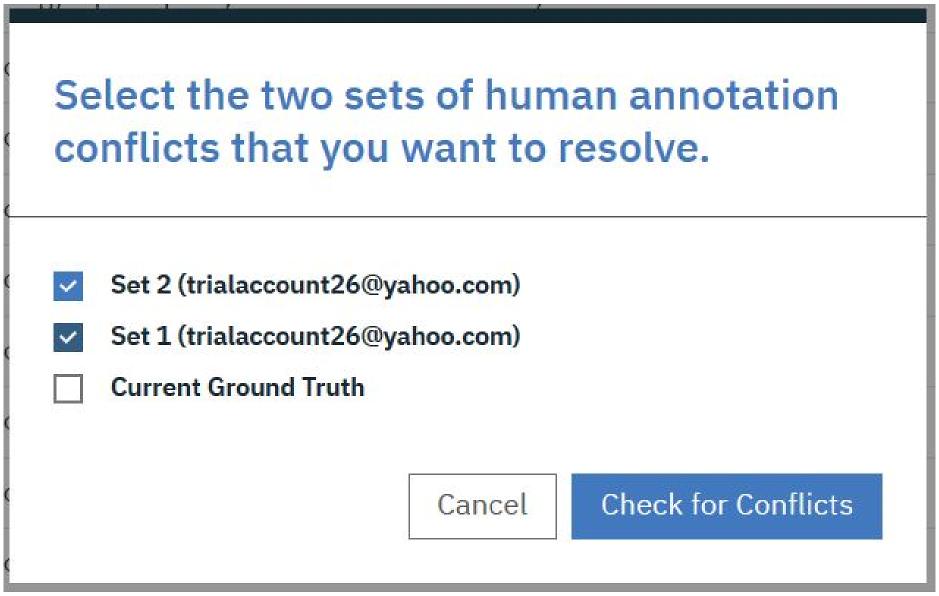
You can see the overlapping documents that were annotated by more than one human annotator. In this exercise, all documents overlap therefore the overlap percentage is 100%.

__ 3. Because the exercise instructed you to create a conflicting relation for the *Technology - gmanews.tv* document, find that document in the list and click **Check for Conflicts**.



Document Name	Annotation Set Name	Action	
Technology - io9.com	Set 2, Set 1	Check for Conflicts	Accept
Technology - ip-telephony.tmcnet.com	Set 2, Set 1	Check for Conflicts	Accept
Technology - content.usatoday.com	Set 2, Set 1	Check for Conflicts	Accept
Technology - e.nikkei.com	Set 2, Set 1	Check for Conflicts	Accept
Technology - focustaiwan.tw	Set 2, Set 1	Check for Conflicts	Accept
Technology - foxnews.com	Set 2, Set 1	Check for Conflicts	Accept
Technology - brighthand.com	Set 2, Set 1	Check for Conflicts	Accept
Technology - businessweek.com	Set 2, Set 1	Check for Conflicts	Accept
Technology - businesswire.com	Set 2, Set 1	Check for Conflicts	Accept
Technology - gamasutra.com	Set 2, Set 1	Check for Conflicts	Accept
Technology - computerworld.com	Set 2, Set 1	Check for Conflicts	Accept
Technology - gmanews.tv	Set 2, Set 1	Check for Conflicts	Accept
Technology - informationweek.com	Set 2, Set 1	Check for Conflicts	Accept

__ 4. Select the two conflicting annotation sets and click **Check for Conflicts**.



Select the two sets of human annotation conflicts that you want to resolve.

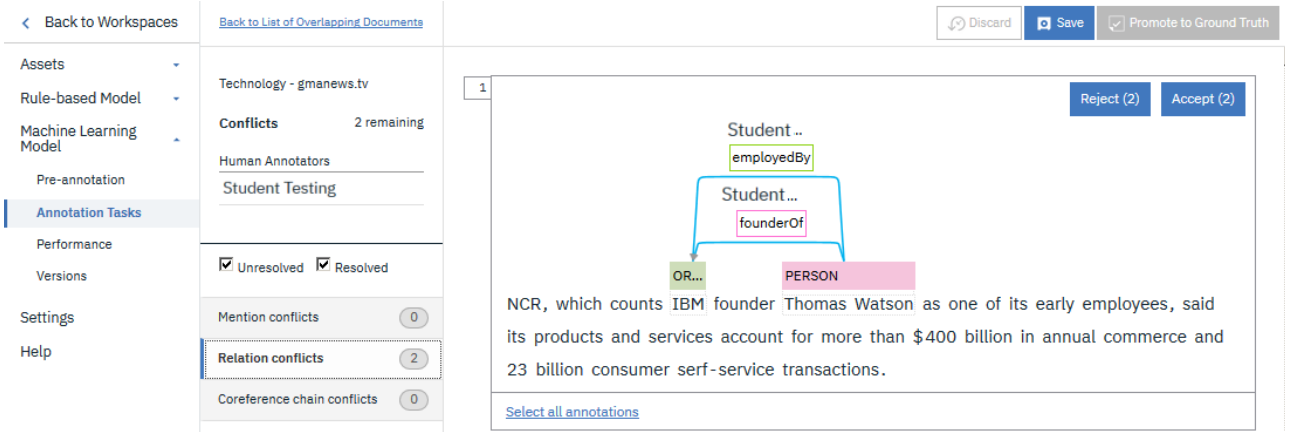
☑ Set 2 (trialaccount26@yahoo.com)

☑ Set 1 (trialaccount26@yahoo.com)

□ Current Ground Truth

Adjudication mode opens. In adjudication mode, you can view overlapping documents, check for conflicts, and remove or replace annotations before you promote the documents to ground truth.

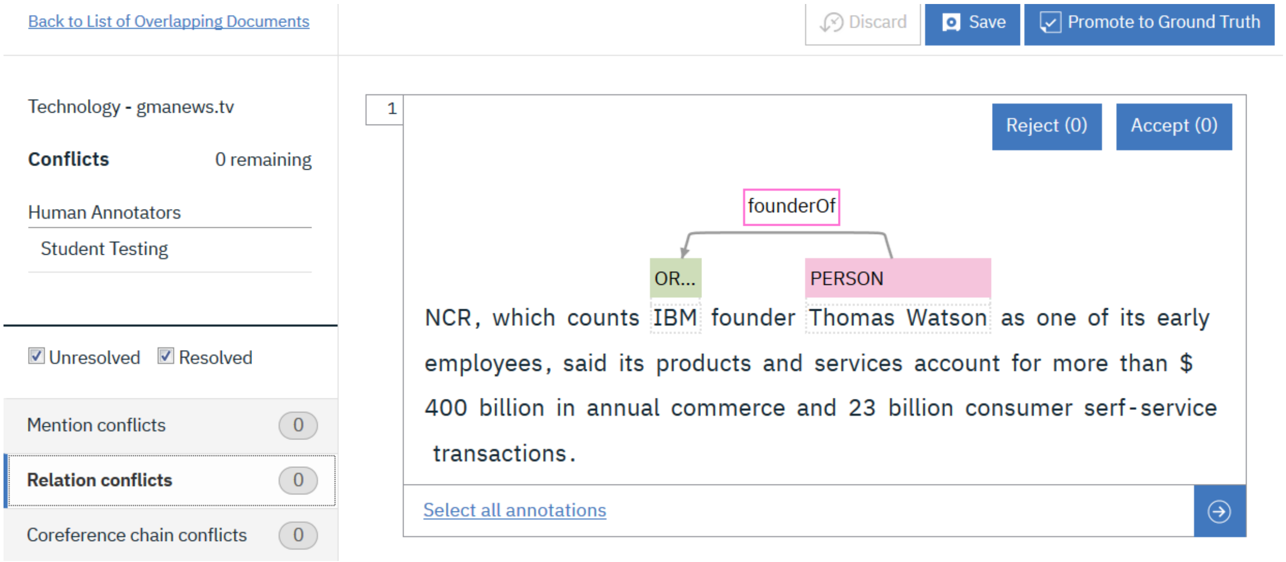
__ 5. Select **Relation conflicts**, to accept the *founderOf* relation, and reject the *employedBy* relation.



NCR, which counts IBM founder Thomas Watson as one of its early employees, said its products and services account for more than $400 billion in annual commerce and 23 billion consumer serf-service transactions.

__ a. Select the **founderOf** relation and click **Accept**.

__ b. Select the **employedBy** relation and click **Reject**.



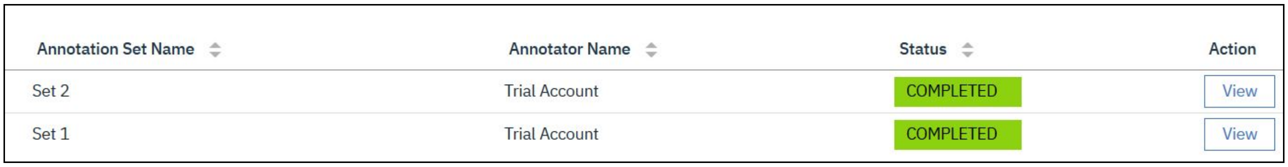
NCR, which counts IBM founder Thomas Watson as one of its early employees, said its products and services account for more than $ 400 billion in annual commerce and 23 billion consumer serf-service transactions.

__ 6. Click **Promote to Ground Truth**.

__ 7. Click **Check for Conflicts** and **Promote to Ground Truth** for all remaining documents in the list.

__ 8. Click **Back to Test Task**.

After you resolve all annotation conflicts and promote the documents to ground truth, the status of the annotation sets is COMPLETED. And now you can use the documents to train the machine learning model.



Annotation Set Name	Annotator Name	Status	Action
Set 2	Trial Account	COMPLETED	View
Set 1	Trial Account	COMPLETED	View

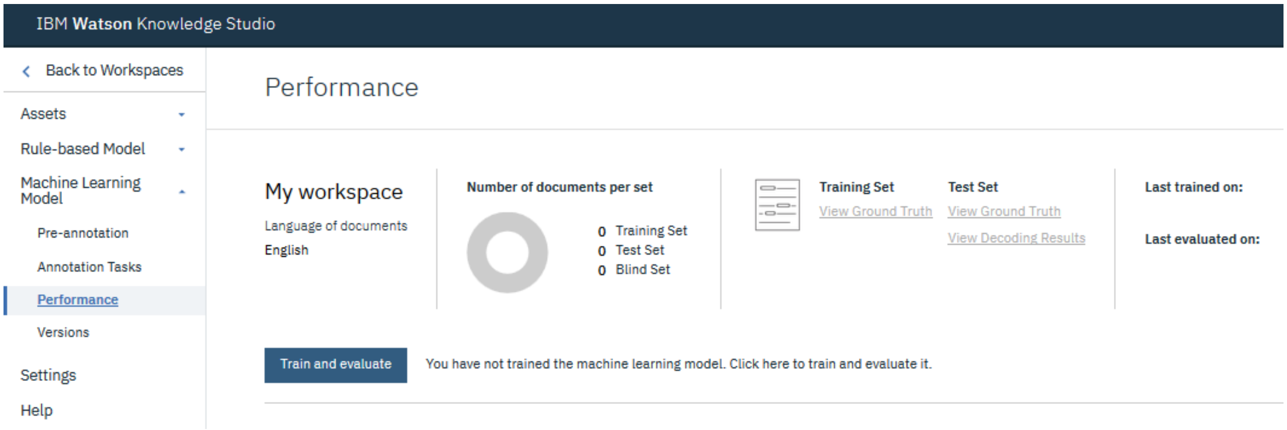
Part 12. Creating a machine learning model

When you create a machine learning model, you select the document sets that you want to use to train it. You also specify the percentage of documents that are to be used as training data, test data, and blind data. Only documents that became ground truth through approval or adjudication can be used to train the machine learning model.

In this part, you will learn how to create a machine learning model in Knowledge Studio.

Perform the following steps:

__ 1. Click **Machine Learning Model** > **Performance** > **Train and evaluate**.



Performance

My workspace

Language of documents
English

Number of documents per set

0 Training Set
0 Test Set
0 Blind Set

Training Set

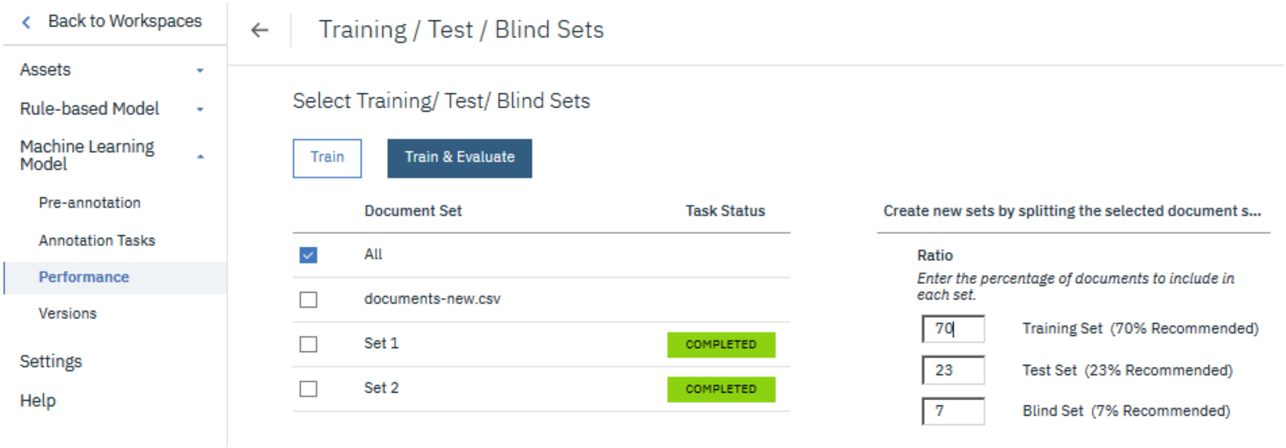
Test Set

Last trained on:

Last evaluated on:

You have not trained the machine learning model. Click here to train and evaluate it.

__ 2. Select **All**, and then click **Train & Evaluate**.



Training / Test / Blind Sets

Select Training/ Test/ Blind Sets

	Document Set	Task Status
☑	All	
□	documents-new.csv	
□	Set 1	COMPLETED
□	Set 2	COMPLETED

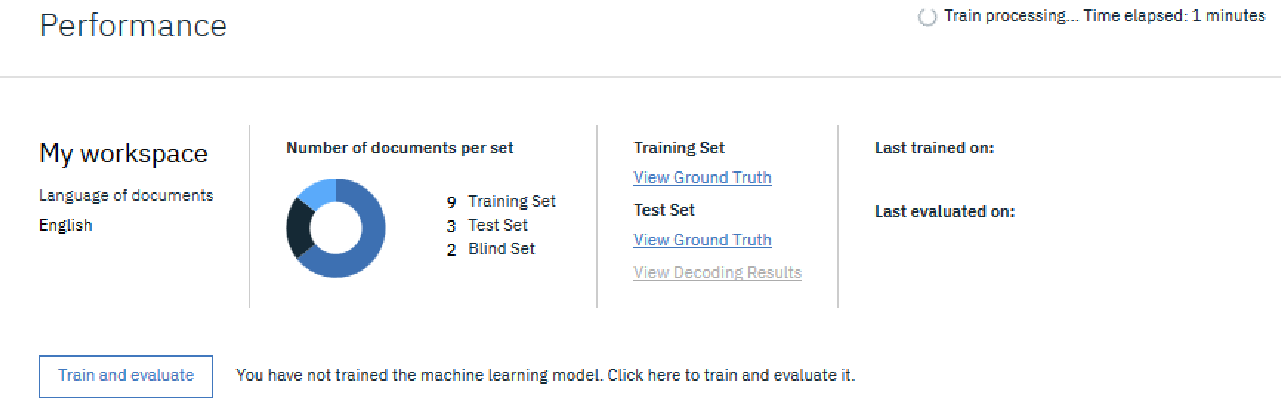
Create new sets by splitting the selected document s...

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)

__ 3. Training might take more than ten minutes, or even hours, depending on the number of human annotations and the number of words in all the documents. The Train processing status is displayed.



Performance

Train processing... Time elapsed: 1 minutes

My workspace

Language of documents
English

Number of documents per set

9 Training Set
3 Test Set
2 Blind Set

Training Set

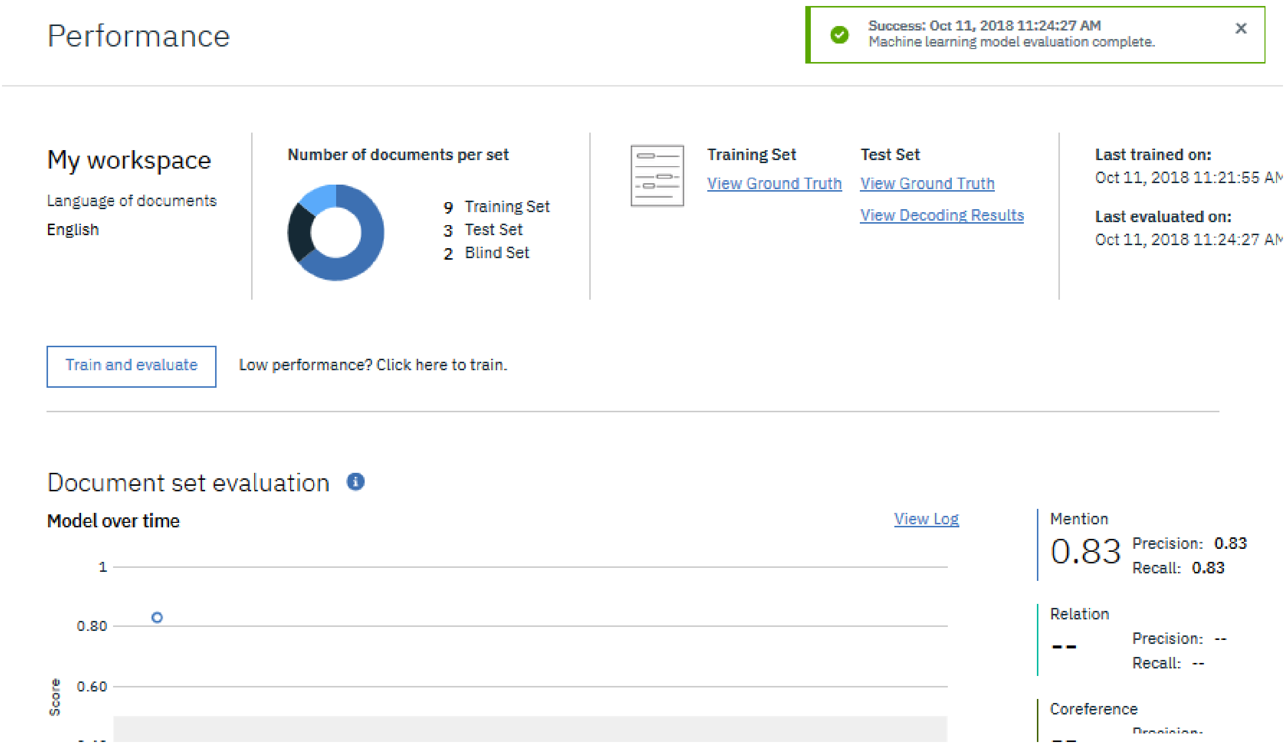
Test Set

Last trained on:

Last evaluated on:

You have not trained the machine learning model. Click here to train and evaluate it.

__ 4. At the end of this step, you receive a notification indicating that the machine learning model evaluation is completed. Scroll-down to review the Performance page.



Performance

Success: Oct 11, 2018 11:24:27 AM
Machine learning model evaluation complete.

My workspace

Language of documents
English

Number of documents per set

9 Training Set
3 Test Set
2 Blind Set

Training Set
View Ground Truth

Test Set
View Ground Truth
View Decoding Results

Last trained on:
Oct 11, 2018 11:21:55 AM

Last evaluated on:
Oct 11, 2018 11:24:27 AM

Low performance? Click here to train.

Document set evaluation

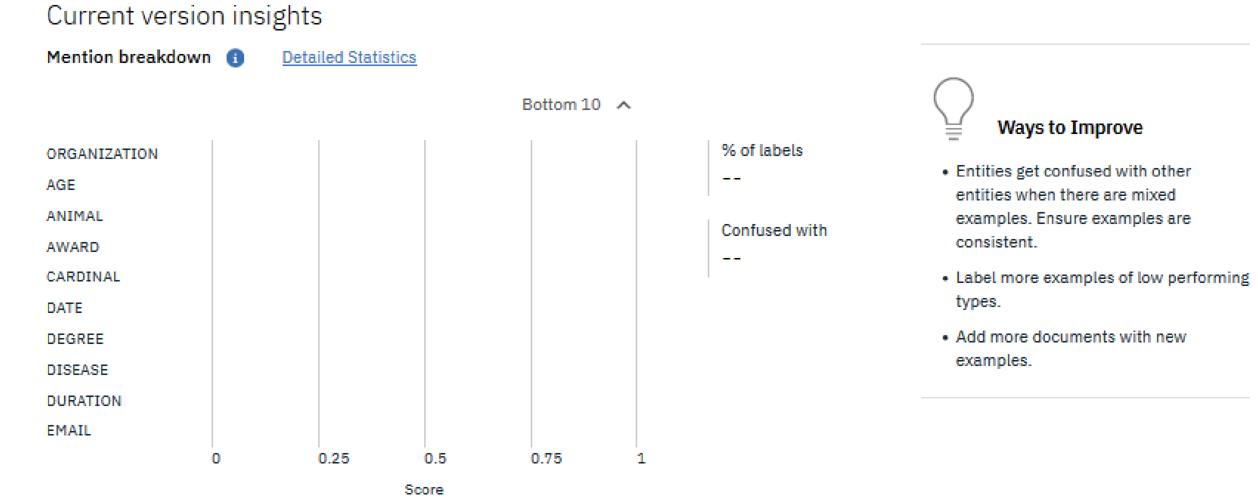
Model over time

Mention
0.83 Precision: 0.83
Recall: 0.83

Relation
-- Precision: --
Recall: --

Coreference

__ 5. Click the Detailed Statistics links above each of the graphs to see detailed information about the machine learning model performance. On these Statistics pages, you can view the scores for mentions, relations, and coreference chains by using the radio buttons.



Current version insights

Mention breakdown

% of labels
--

Confused with
--

Ways to Improve

- Entities get confused with other entities when there are mixed examples. Ensure examples are consistent.
- Label more examples of low performing types.
- Add more documents with new examples.

- To view the Training / Test / Blind Sets page, click the **Train and evaluate** button.
- To see the documents that human annotators worked on, click **View Ground Truth**.

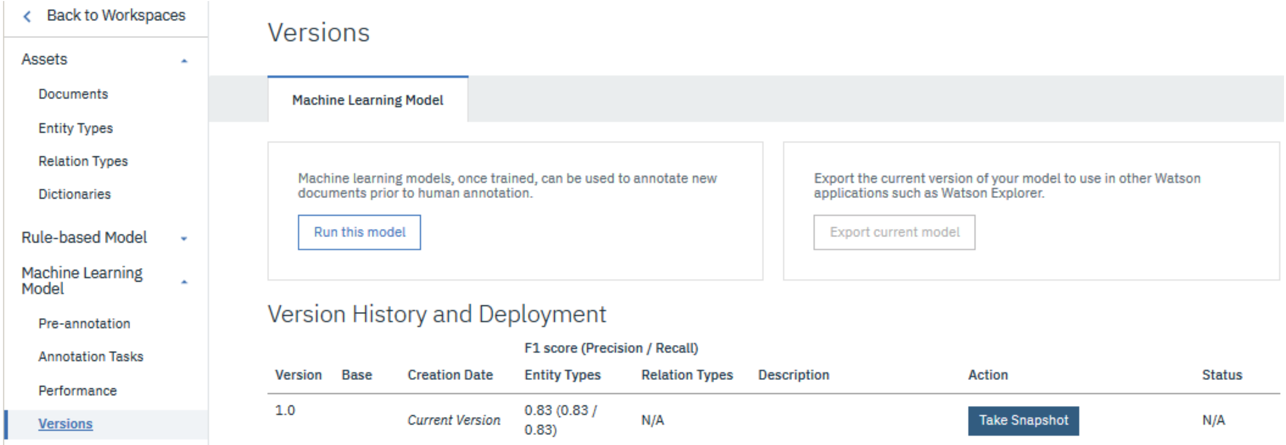
- To see the annotations that the trained machine learning model created on that same set of documents, click **View Decoding Results**.
- To view details about the precision, recall, and F1 scores for the machine learning model, click the **Performance** page.
- You can analyze performance by viewing a summary of statistics for entity types, relation types, and coreference chains.

__ 9. When you are satisfied with the performance of the model, you can export the current version of the machine learning model to use it in other Watson services such as Watson Discovery, and Watson Natural Language Understanding, or Watson software such as Watson Explorer.

This feature enables your applications to use the deployed machine learning model to enrich the insights that you get from your data to include the recognition of concepts and relations that are relevant to your domain and analyze semantic features of text input, including entities and relations.

To export a version of the machine learning model:

__ a. Click **Machine Learning** > **Versions** > **Take Snapshot**.



Versions

Machine learning models, once trained, can be used to annotate new documents prior to human annotation.

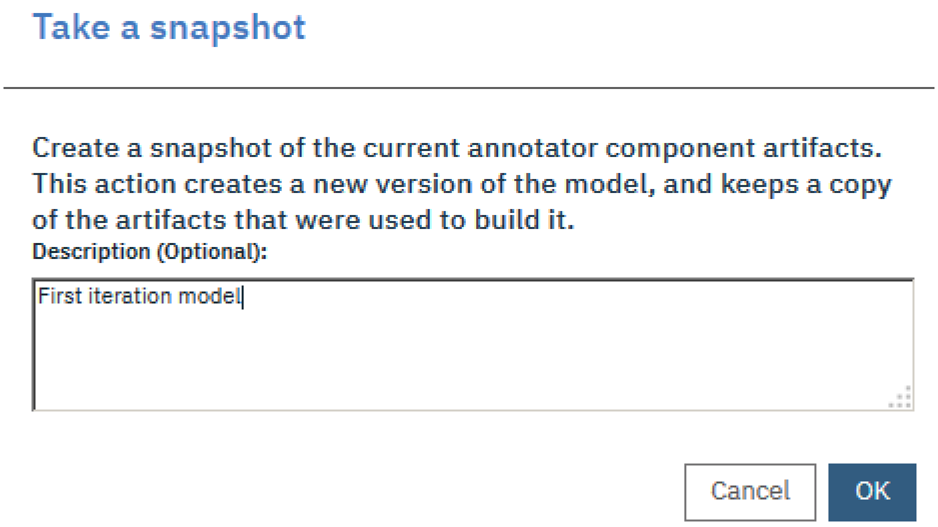
Export the current version of your model to use in other Watson applications such as Watson Explorer.

Version History and Deployment

			F1 score (Precision / Recall)				
Version	Base	Creation Date	Entity Types	Relation Types	Description	Action	Status
1.0		*Current Version*	0.83 (0.83 / 0.83)	N/A		Take Snapshot	N/A

Note: The Export current model option is not available for Lite plans.

__ b. Enter a description (optional) and click **OK**.



Take a snapshot

Create a snapshot of the current annotator component artifacts. This action creates a new version of the model, and keeps a copy of the artifacts that were used to build it.

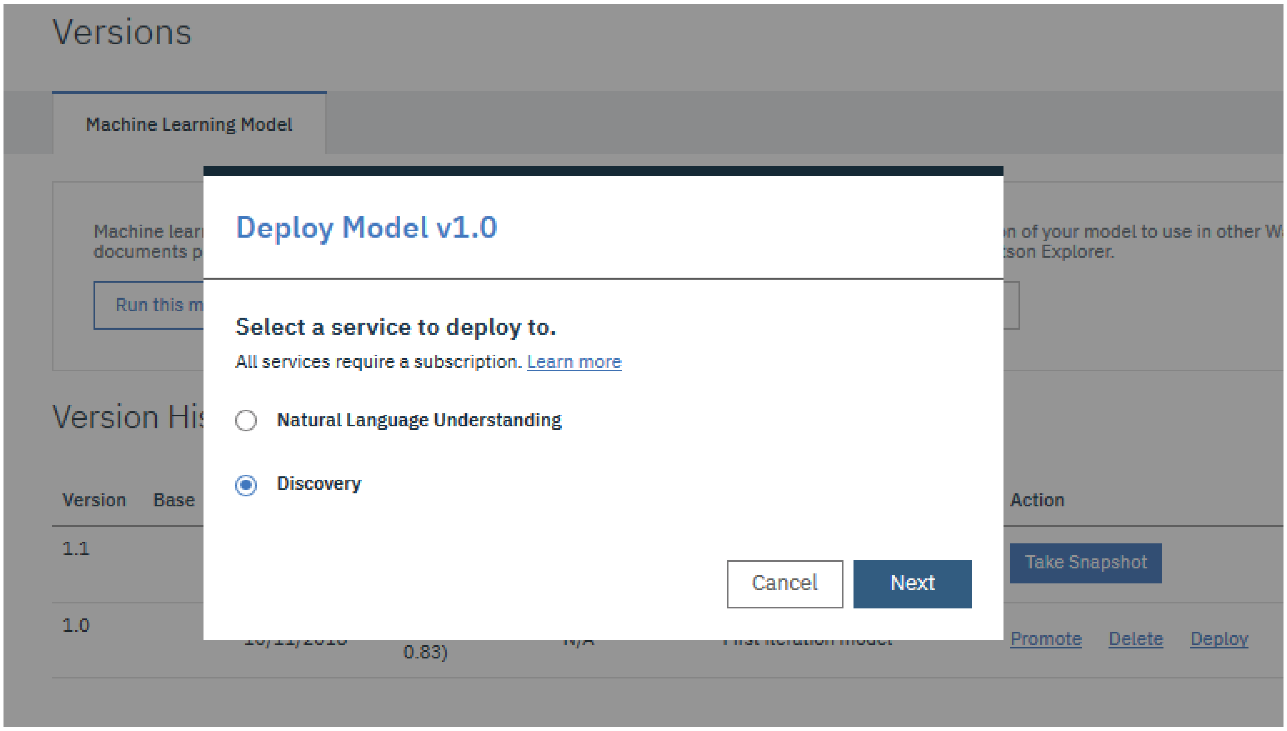
Description (Optional):

First iteration model

__ c. Choose the version of the model that you want to deploy.

__ d. Click Deploy.

__ e. Select the Watson service instance to deploy to.



Machine lear… documents p… …n of your model to use in other W… …tson Explorer.

Deploy Model v1.0

Select a service to deploy to.

All services require a subscription. Learn more

○ Natural Language Understanding

● Discovery

__ f. For this exercise, just click **Cancel**. You will deploy a Watson Studio machine learning model to Discovery in a future exercise.

In this part, you created a machine learning model, trained it, and evaluated how well it performed when annotating test data and blind data. By exploring the performance metrics, you can identify ways to improve the accuracy of the machine learning model.

Optional exercise: Creating a rule-based model

Estimated time:

00:30

Overview

This exercise helps you understand how to create a rule-based model that you can use to find text patterns that you define in documents.

Objectives

After completing this exercise, you should be able to:

- Create classes
- Add documents for defining rules
- Associate dictionaries with classes
- Define regular expressions to capture sequences of characters
- Define rules

Introduction

You will build a model that can find text in documents that matches the pattern month day, year. For example, the model would find the date reference *May 1, 2010*. Before you define the rule pattern itself, you will create artifacts that will help you build the pattern, including a dictionary class that recognizes month mentions and a regular expression class that recognizes year mentions in text.

Exercise instructions

In this exercise you will complete the following tasks:

__ 1. Add a dictionary of months

__ 2. Add sample documents

__ 3. Create classes.

__ 4. Associate a dictionary with a class.

__ 5. Find class annotations in documents.

__ 6. Define a regular expression.

__ 7. Define a rule.

__ 8. Create a rule-based model.

Part 1. Adding a dictionary of months

In this part you will add a dictionary to your workspace in Knowledge Studio. The dictionary contains terms related to the months of the year. This task is similar to the task that you performed in Part 4 Adding a dictionary but this is a different dictionary related to the months of the year. Continue using the same workspace in Knowledge Studio.

Perform the following steps:

__ 1. Download the file dictionary-items-month.csv to your computer. This file contains dictionary terms in CSV format, suitable for uploading into a Knowledge Studio dictionary.

__ 2. Click **Assets** > **Dictionaries**.

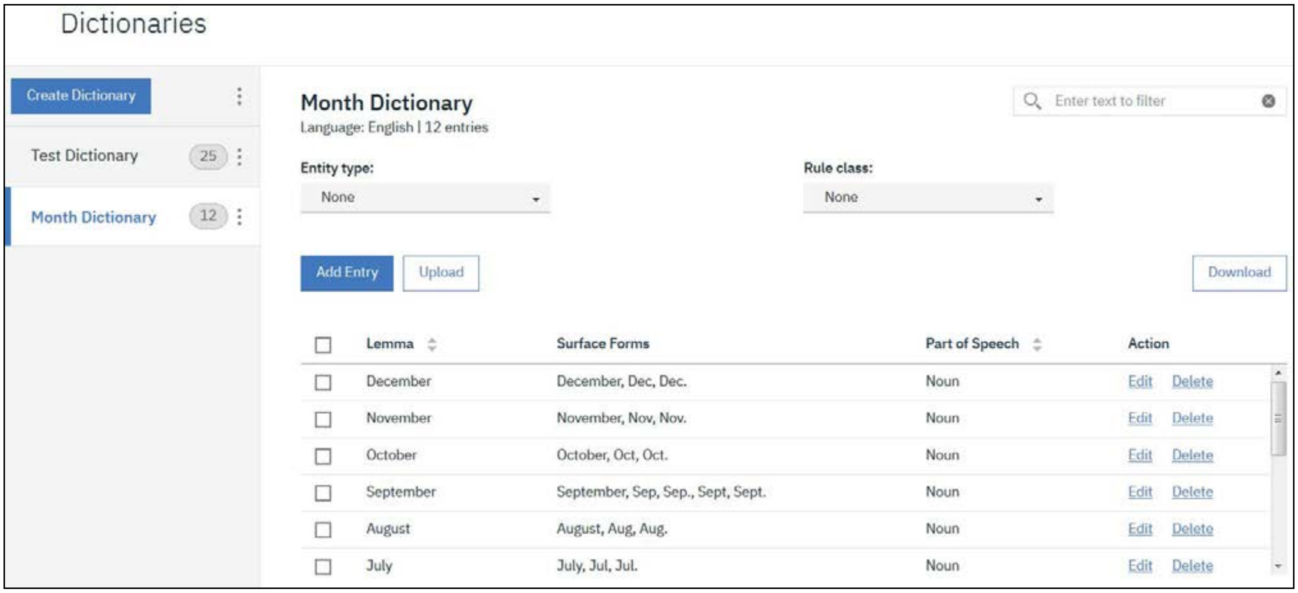
__ 3. Click **Create Dictionary** to add a dictionary.

__ 4. In the **Name** field, type **Month dictionary** and click **Save** to create the dictionary. The new dictionary is created and automatically opened for editing.

__ 5. In the dictionary pane, click **Upload**.

__ 6. Select the file **dictionary-items-month.csv** from your computer and click **Upload**.

The terms from the file are imported into the dictionary as shown in the figure.



Month Dictionary

Language: English | 12 entries

Entity type: None

Rule class: None

☐	Lemma	Surface Forms	Part of Speech	Action
☐	December	December, Dec, Dec.	Noun	Edit Delete
☐	November	November, Nov, Nov.	Noun	Edit Delete
☐	October	October, Oct, Oct.	Noun	Edit Delete
☐	September	September, Sep, Sep., Sept, Sept.	Noun	Edit Delete
☐	August	August, Aug, Aug.	Noun	Edit Delete
☐	July	July, Jul, Jul.	Noun	Edit Delete

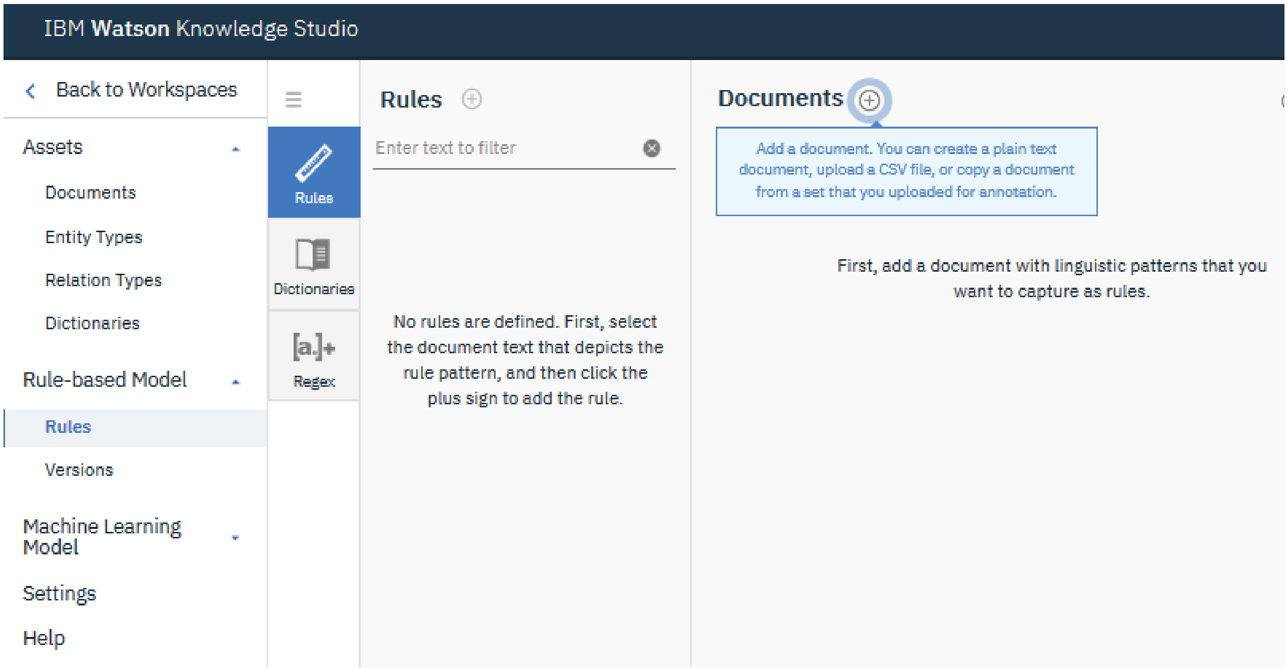
Part 2. Adding sample documents

In this part, you will learn how to add documents with linguistic patterns that illustrate the types of rules you want to define. In this exercise, the documents should include dates in the format that you want to capture.

Perform the following steps:

__ 1. Download the documents-new.csv file to your computer. This file contains example documents suitable for uploading. These are the same document that were used before.

__ 2. Click **Rule-based Model** > **Rules**.



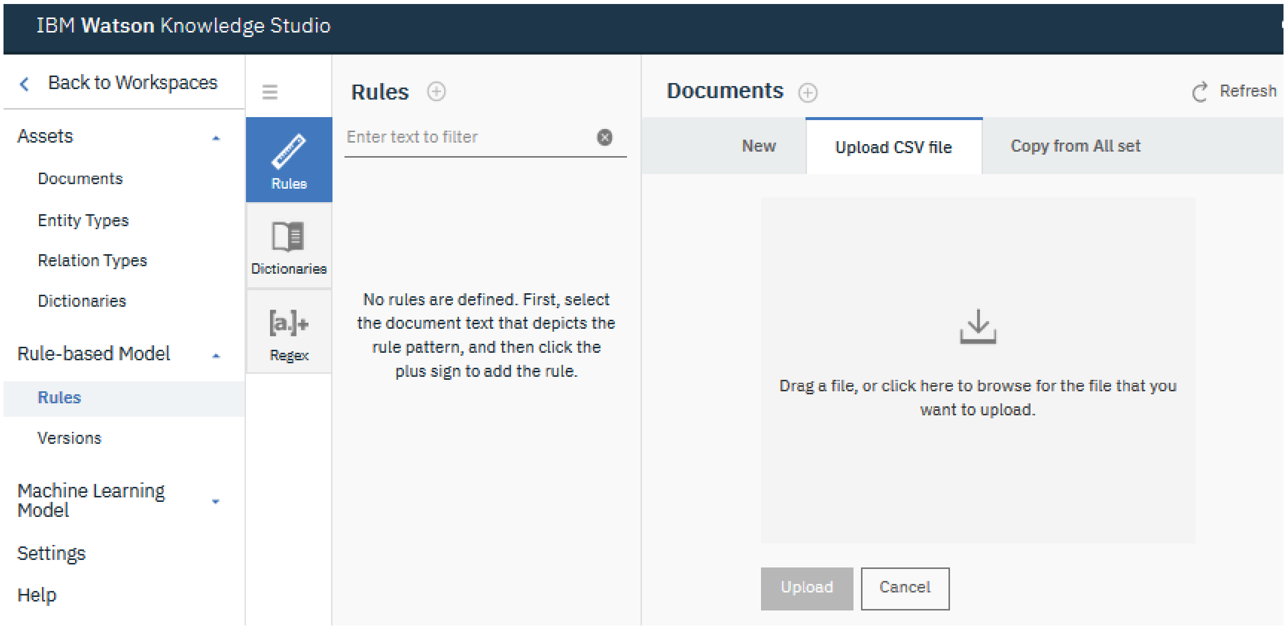
Add a document. You can create a plain text document, upload a CSV file, or copy a document from a set that you uploaded for annotation.

First, add a document with linguistic patterns that you want to capture as rules.

No rules are defined. First, select the document text that depicts the rule pattern, and then click the plus sign to add the rule.

__ 3. Click the **Add a document** icon, which is next to the **Documents** page heading.

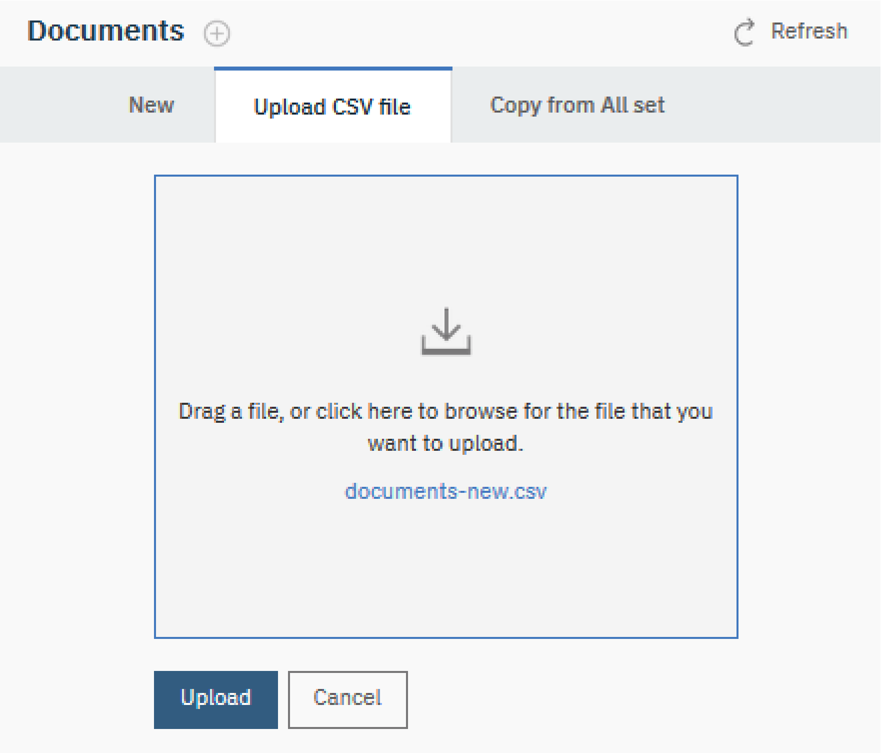
__ 4. Click the **Upload CSV file** tab.



No rules are defined. First, select the document text that depicts the rule pattern, and then click the plus sign to add the rule.

Drag a file, or click here to browse for the file that you want to upload.

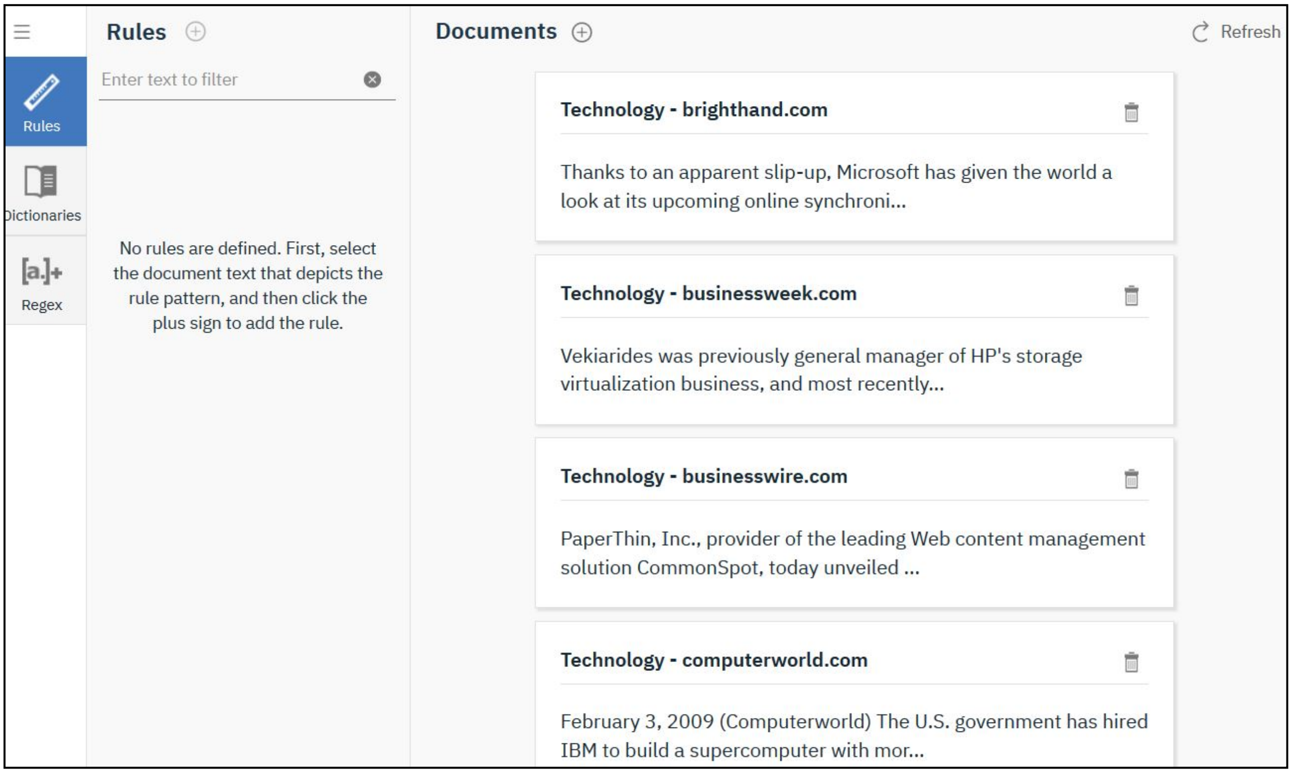
__ 5. Click to browse for the *documents-new.csv* file that you downloaded to your computer earlier, and then click **Upload**.



Drag a file, or click here to browse for the file that you want to upload.

documents-new.csv

__ 6. A set of documents will be displayed in the main Documents page as shown in the figure.



No rules are defined. First, select the document text that depicts the rule pattern, and then click the plus sign to add the rule.

Technology - brighthand.com

Thanks to an apparent slip-up, Microsoft has given the world a look at its upcoming online synchroni...

Technology - businessweek.com

Vekiarides was previously general manager of HP's storage virtualization business, and most recently...

Technology - businesswire.com

PaperThin, Inc., provider of the leading Web content management solution CommonSpot, today unveiled ...

Technology - computerworld.com

February 3, 2009 (Computerworld) The U.S. government has hired IBM to build a supercomputer with mor...

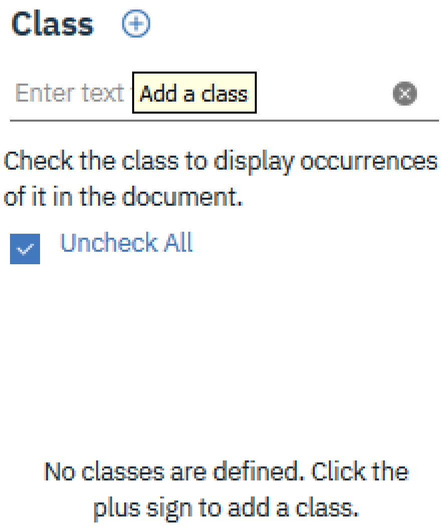
Part 3. Creating classes

When you construct a rule, you use classes to represent types of information. As you build rules, you can define intermediate classes that are used only to build other more complex classes.

In this part, you will learn how to define classes that you will use later when you define a rule.

Perform the following steps:

__ 1. From the **Rules** page of your workspace, click the **Add a class** icon next to the **Class** heading in the right panel as shown in the figure.



Class ⊕

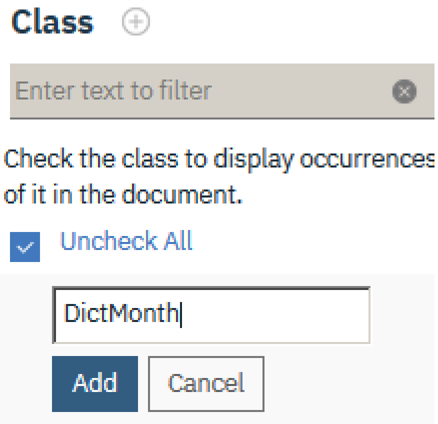
Enter text Add a class

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

☑ Uncheck All

No classes are defined. Click the plus sign to add a class.

__ 2. Enter **DictMonth** as the class name, and then click **Add**.



Class ⊕

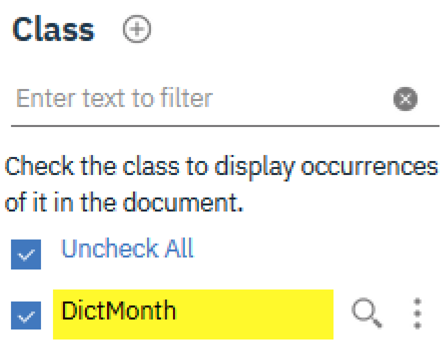
Enter text to filter

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

☑ Uncheck All

DictMonth

__ 3. The new class is displayed in the Class panel.



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

☑ Uncheck All

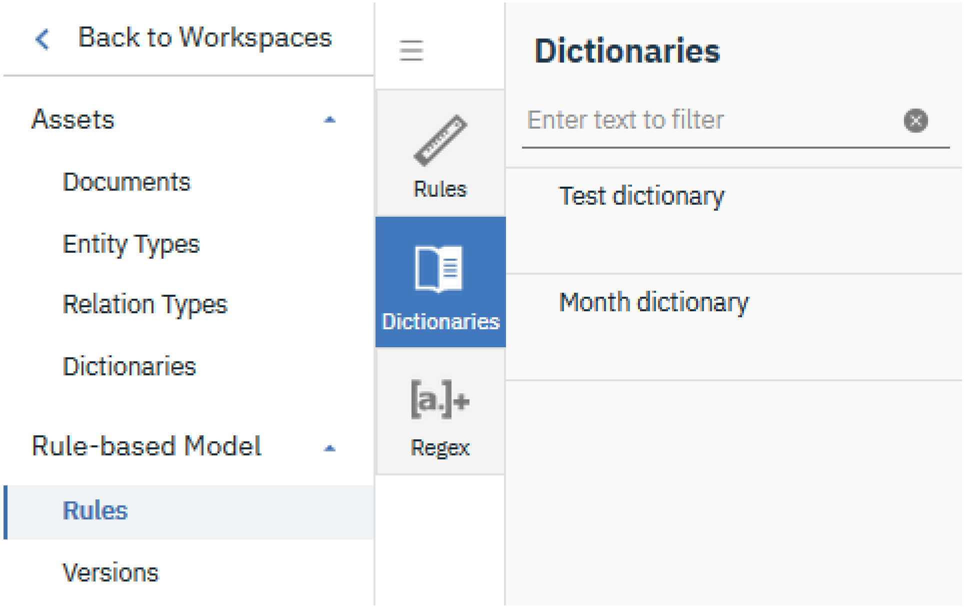
☑ DictMonth

Part 4. Associating a dictionary with a class

In this part, you will learn how to use a dictionary in the rule editor.

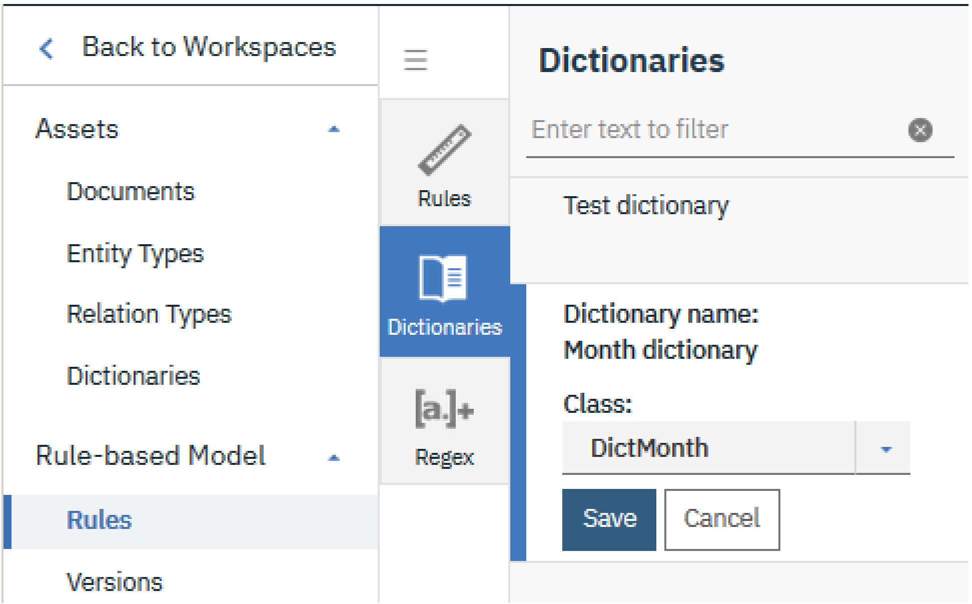
Perform the following steps:

__ 1. Click **Rule-based Model** > **Rules**, and then click the **Dictionaries** tab.



__ 2. Select **Month dictionary** that you created previously.

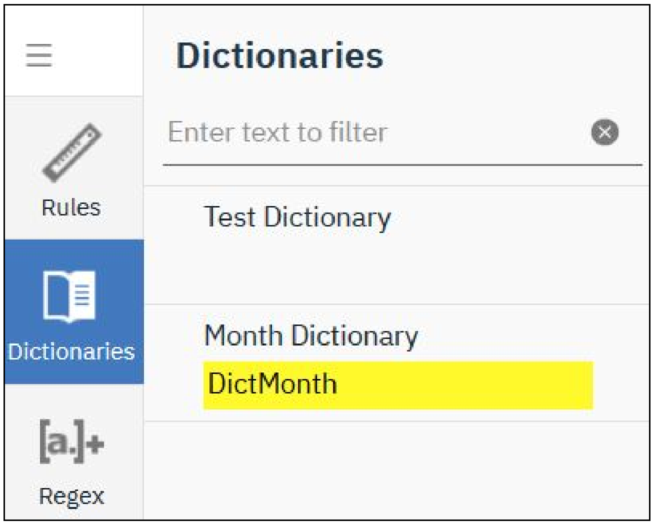
__ 3. From the **Class** list, select the **DictMonth** class and then click **Save**.



Dictionary name:
Month dictionary

Class:
DictMonth

The class is now associated with the dictionary.



Test Dictionary

Month Dictionary
DictMonth

For documents that are associated with the rule editor, any references to terms in the dictionary are annotated as DictMonth class mentions.

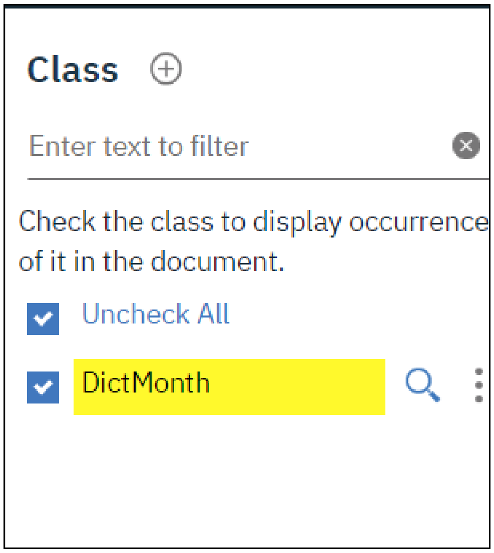
Part 5. Finding class annotations in documents

In this part, you will learn how to find class annotations in rule editor documents.

Perform the following steps:

__ 1. Select **Rule-based Model** > **Rules**.

From the **Class** panel, find the **DictMonth** class that you defined earlier, and click the **Search annotations in documents** Icon beside the class name.



Class ⊕

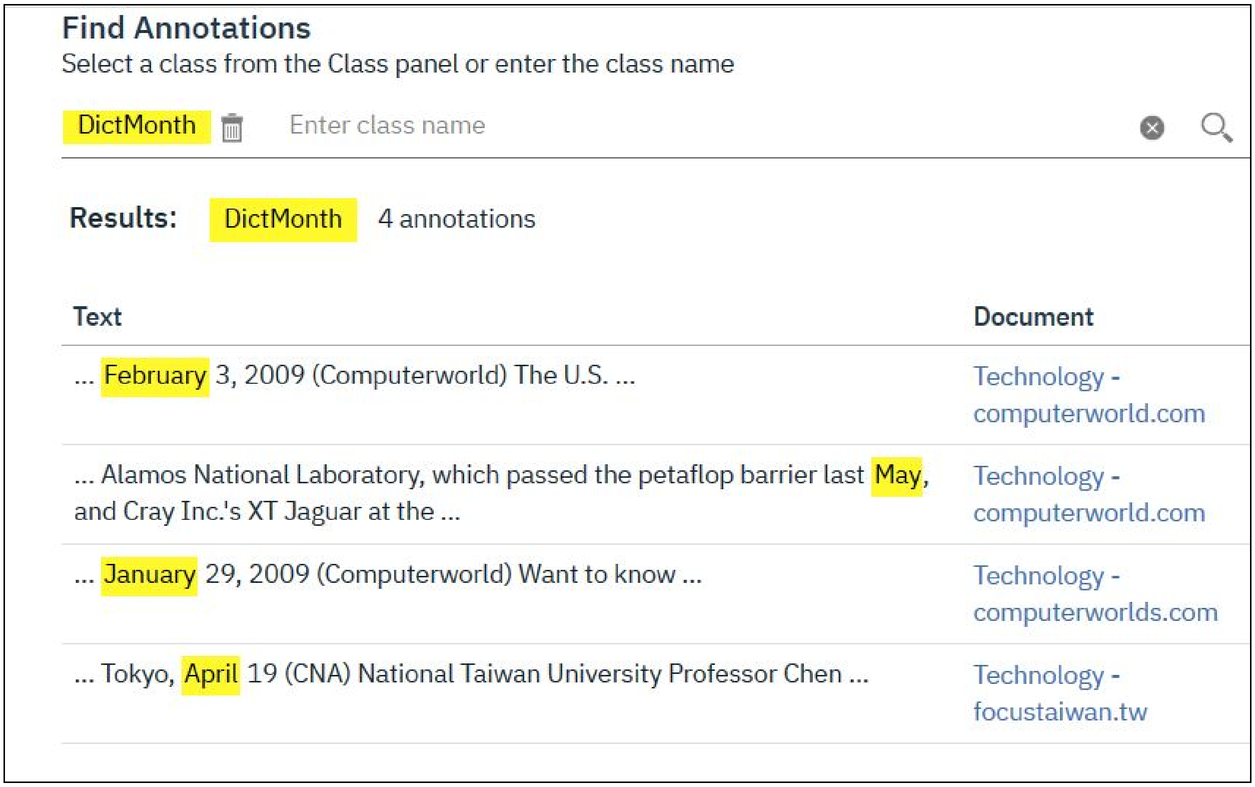
Enter text to filter

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

☑ Uncheck All

☑ DictMonth

__ 2. The **Find Annotations** page is displayed and shows all the documents that contain text references to months.



Find Annotations

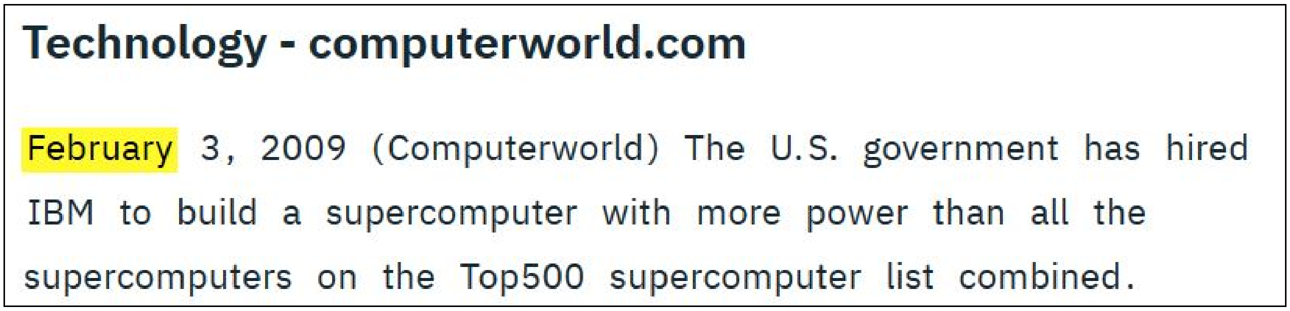
Select a class from the Class panel or enter the class name

DictMonth | Enter class name

Results: DictMonth 4 annotations

Text	Document
... February 3, 2009 (Computerworld) The U.S. ...	Technology - computerworld.com
... Alamos National Laboratory, which passed the petaflop barrier last May, and Cray Inc.'s XT Jaguar at the ...	Technology - computerworld.com
... January 29, 2009 (Computerworld) Want to know ...	Technology - computerworlds.com
... Tokyo, April 19 (CNA) National Taiwan University Professor Chen ...	Technology - focustaiwan.tw

__ 3. Click the **Technology - computerworld.com** document to view the full document. Notice that the text *February* is highlighted, which means it was annotated as a mention of the *DictMonth* class. If you scroll you will notice that *May* is highlighted also.



Technology - computerworld.com

February 3, 2009 (Computerworld) The U.S. government has hired IBM to build a supercomputer with more power than all the supercomputers on the Top500 supercomputer list combined.

Part 6. Define a regular expression

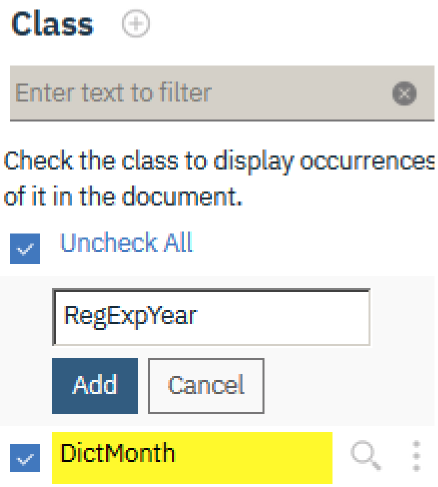
At this point, you can capture month mentions by using the dictionary that you uploaded earlier and associated a class with it.

In this part you will define a regular expression to capture year patterns like *2009*.

Perform the following steps:

__ 1. From the **Rules** page, click the **Add a class** icon ⊕ next to **Class** from the right panel.

__ 2. Enter **RegExpYear** as the class name and click **Add**.



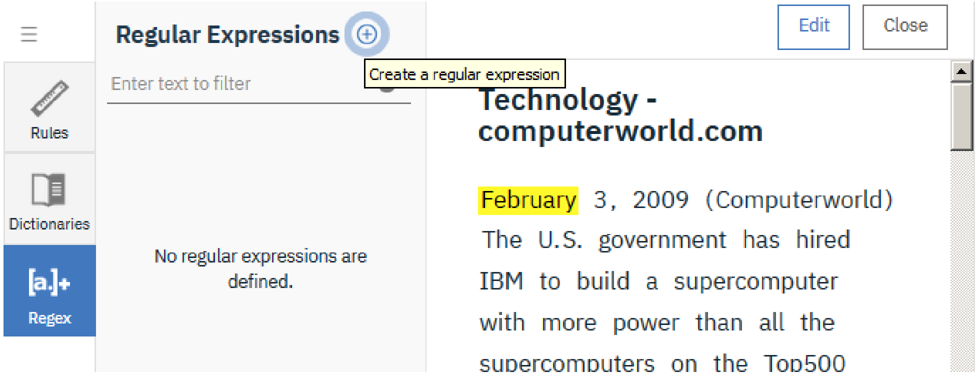
Check the class to display occurrences of it in the document.

☑ Uncheck All

RegExpYear

☑ DictMonth

__ 3. Click the **Regex** tab, and then click the **Create a regular expression** icon next to the Regular Expressions heading.



Technology - computerworld.com

February 3, 2009 (Computerworld) The U.S. government has hired IBM to build a supercomputer with more power than all the supercomputers on the Top500

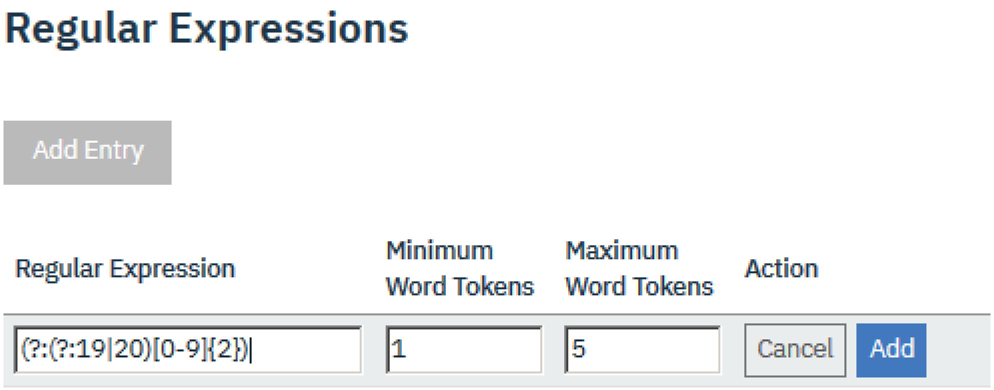
__ 4. Click **Add Entry**.

__ 5. In the **Regular Expression** field, enter the following expression, which finds years between *1900* and *2099:*

```
(?:(?:19|20)[0-9]{2})
```

__ 6. Set **Minimum Word Tokens** to 1 and **Maximum Word Tokens** to 1.

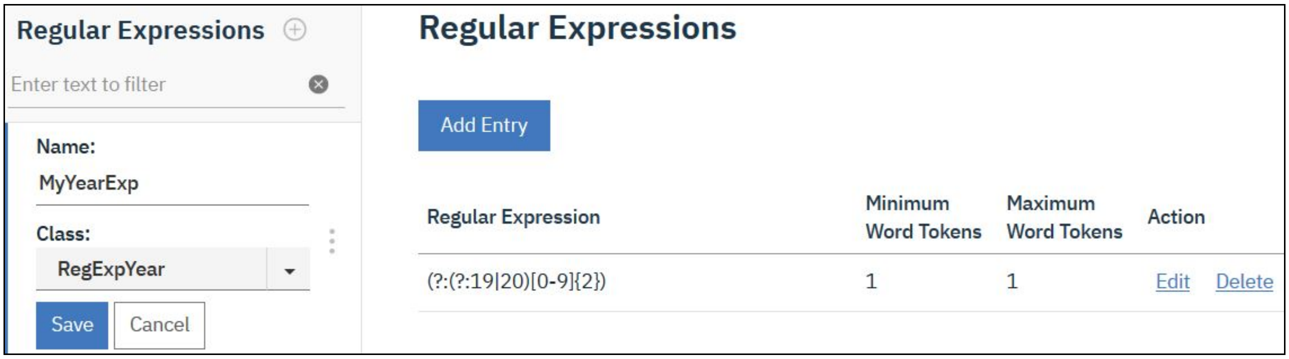
__ 7. Click **Add** to save the regular expression.



Regular Expression	Minimum Word Tokens	Maximum Word Tokens	Action
(?:(?:19|20)[0-9]{2})	1	5	Cancel Add

__ 8. Enter **MyYearExp** as the regular expression name, and then, from the **Class** menu, select the **RegExpYear** class that you defined earlier.

__ 9. Click **Save**.



Name: MyYearExp

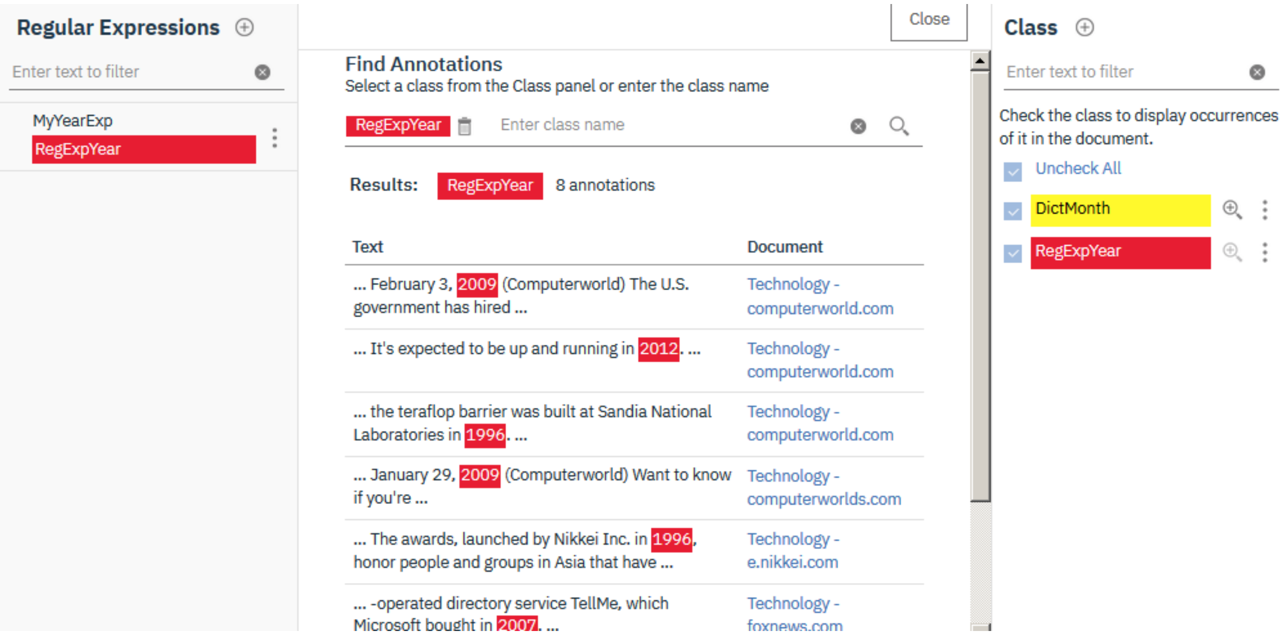
Class: RegExpYear

Regular Expression	Minimum Word Tokens	Maximum Word Tokens	Action
(?:(?:19|20)[0-9]{2})	1	1	Edit Delete

__ 10. After you save the regular expression, it is automatically applied to the sample documents. Any text strings that follow the pattern that you defined in the regular expression are annotated as mentions of the *RegExpYear* class.

__ 11. To check whether the expression you defined is capturing time occurrences correctly, you can search for mentions. Click the **Search annotations in documents** icon 🔍 beside the class name *RegExpYear* in the Class panel.

__ 12. The **Find Annotations** page is displayed. Occurrences of year mentions are highlighted in the sample documents in which they occur.



Results: RegExpYear 8 annotations

Text	Document
... February 3, 2009 (Computerworld) The U.S. government has hired ...	Technology - computerworld.com
... It's expected to be up and running in 2012. ...	Technology - computerworld.com
... the teraflop barrier was built at Sandia National Laboratories in 1996. ...	Technology - computerworld.com
... January 29, 2009 (Computerworld) Want to know if you're ...	Technology - computerworlds.com
... The awards, launched by Nikkei Inc. in 1996, honor people and groups in Asia that have ...	Technology - e.nikkei.com
... -operated directory service TellMe, which Microsoft bought in 2007. ...	Technology - foxnews.com

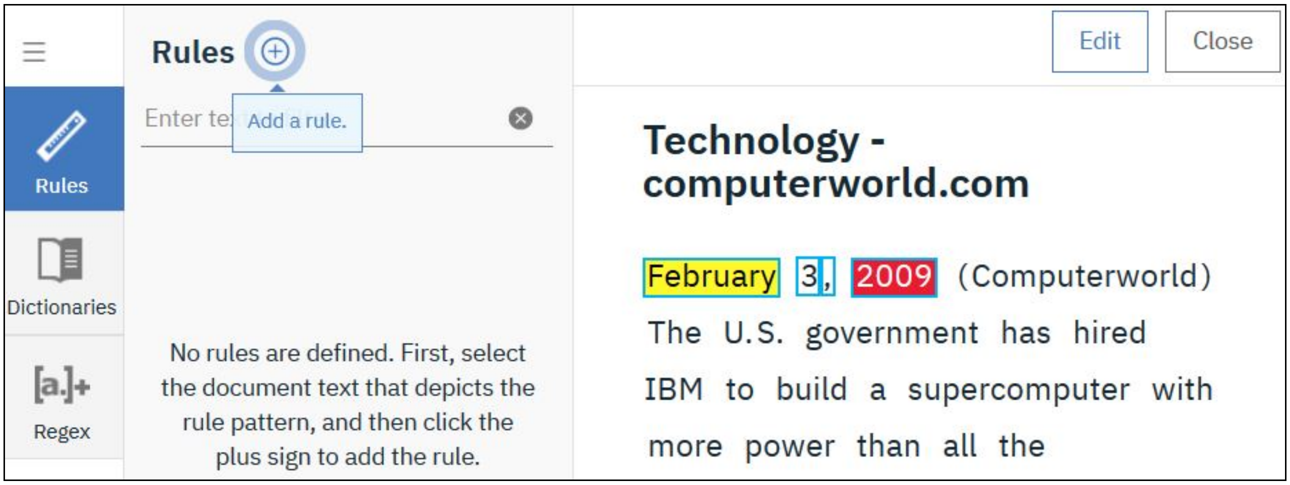
Part 7. Defining a rule

You already defined a dictionary-based class for annotating month mentions. You also defined a regular expression that finds numeric values which represent a year. In this part, you will define a rule that captures the sequence of a month followed by a number, a comma, and then a year. You will define a rule for date expressions like *September 21, 2016*.

Perform the following steps:

__ 1. Select **Rule-based Model** > **Rules** and open the *Technology - computerworld.com* document.

__ 2. Select the text *February 3, 2009* in the document. Make sure you select the comma, too as shown in the figure.



No rules are defined. First, select the document text that depicts the rule pattern, and then click the plus sign to add the rule.

Technology - computerworld.com

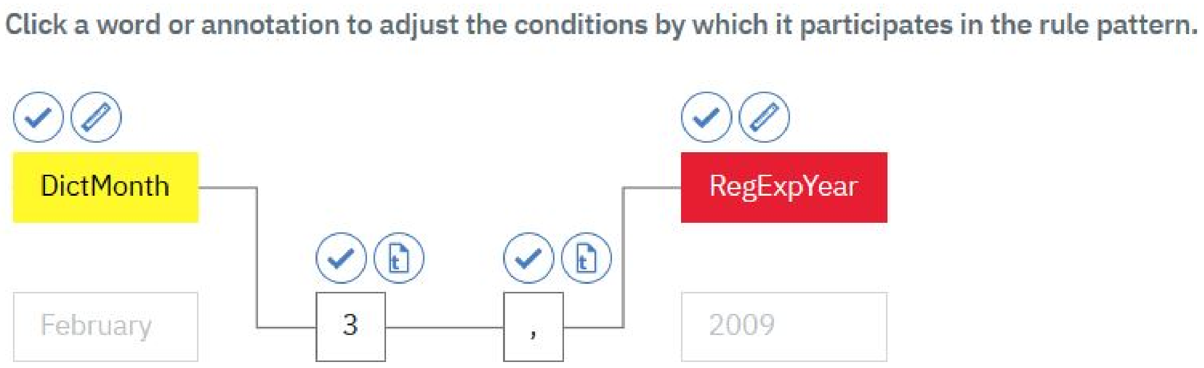
February 3, 2009 (Computerworld) The U.S. government has hired IBM to build a supercomputer with more power than all the

__ 3. Click the **Add a rule** icon.

The rule editor shows a depiction of the rule pattern that you identified.

The text *February 3, 2009* is visible. A solid line that connects the cells in the depiction identifies which cells are currently part of the pattern as in the figure.

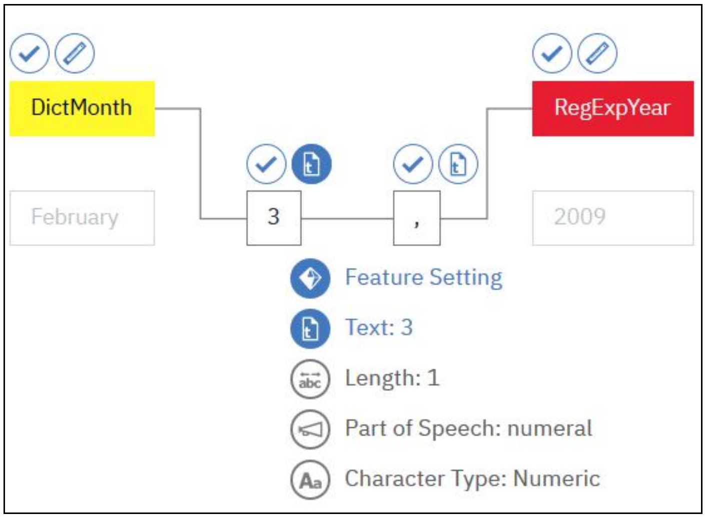
- The `DictMonth` class is part of the rule pattern instead of the text `February`. This selection is preferred because you want the model to find any month that is annotated by the `DictMonth` class as the first token in the date pattern instead of the text `February` only.
- At the end of the rule, the year *2009* is already annotated as being a mention of the `RegExpYear` class. The `RegExpYear` class is part of the rule pattern instead of the number `2009`. This selection is also preferred because you want the model to find any year that is annotated by the `RegExpYear` class as the last token in the date pattern instead of the specific text `2009` only.
- The number 3 and the comma (,) after it are shown as the second and third tokens in the pattern. As the pattern is currently specified, the model will find only occurrences of dates that specify the 3rd day of a month. We want the model to find dates that specify any day of the month, so next you will change the feature settings for the day token.



Click a word or annotation to adjust the conditions by which it participates in the rule pattern.

__ 4. Above the day ***3*** cell, click the **Text** icon to open the feature settings for the token.

Currently, the rule is set to match the exact text, ***3***. Instead, we want it to match any number.



Feature Setting

Text: 3

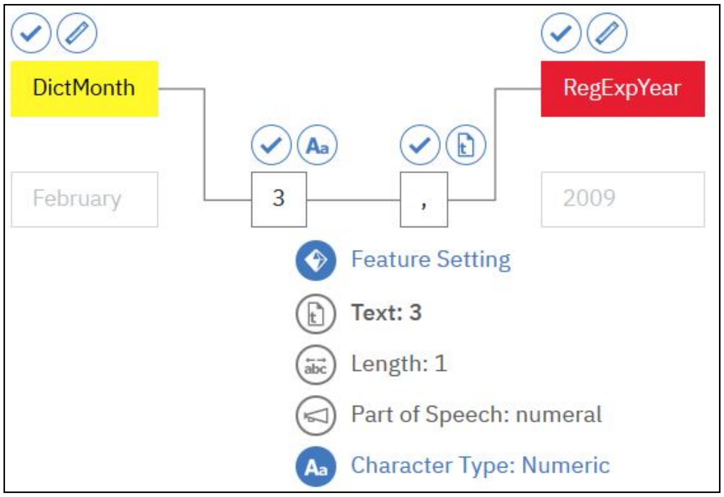
Length: 1

Part of Speech: numeral

Character Type: Numeric

__ 5. Change the feature setting to be numeric by selecting **Character Type : Numeric**, and then clearing the selection, **Text : 3**.

You changed the definition for the number ***3*** cell. The **Character Type** icon indicates that instead of requiring the number to be equal to 3 exactly, it can be any number.



Feature Setting

Text: 3

Length: 1

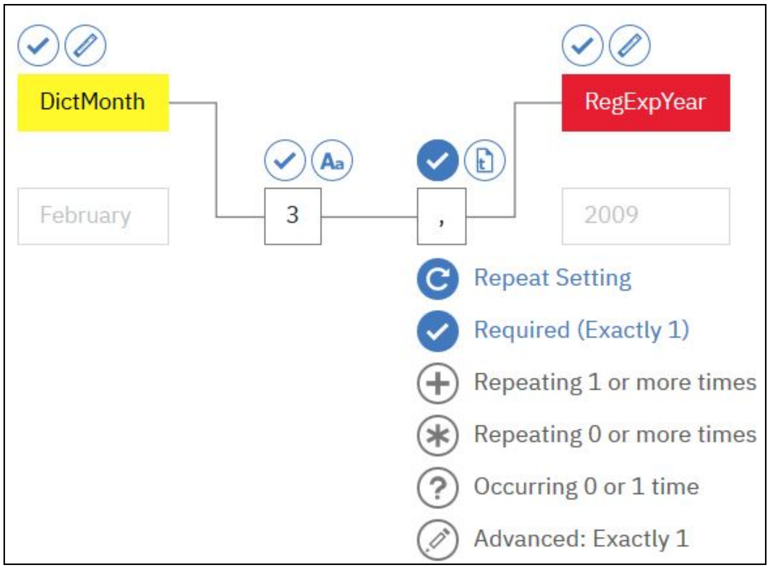
Part of Speech: numeral

Character Type: Numeric

__ 6. Do not change any settings for the comma token.

You want the third token in the pattern to be a comma, so the current feature setting of **text : ,** is appropriate.

In addition to a feature setting, each token has a repeat setting. The repeat setting specifies how many times the token can be repeated in the text for it to match the pattern. The current repeat setting of **Required (Exactly 1)** is appropriate.



Repeat Setting

Required (Exactly 1)

Repeating 1 or more times

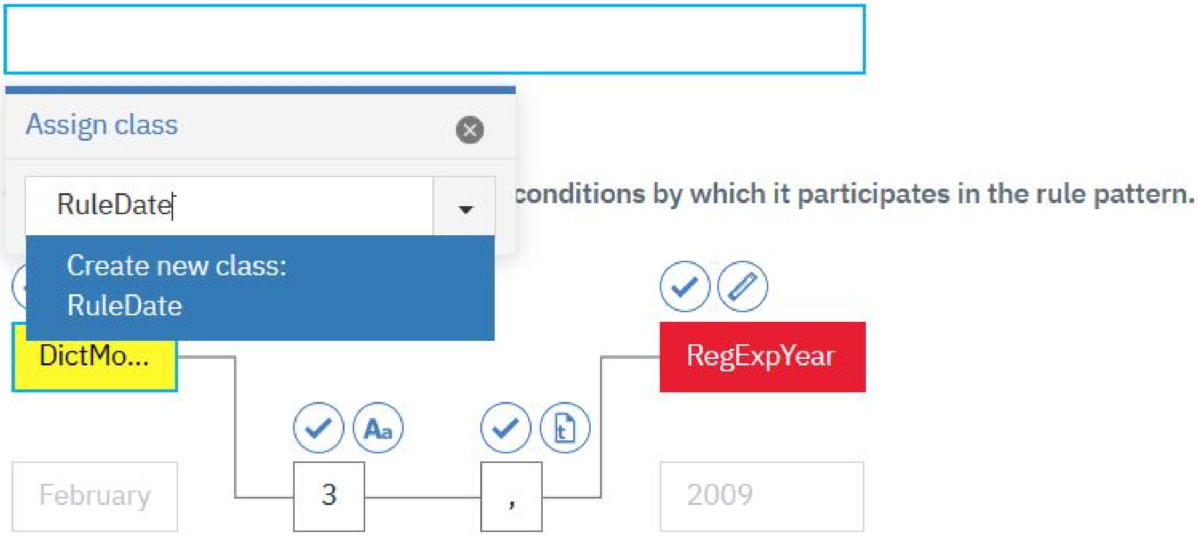
Repeating 0 or more times

Occurring 0 or 1 time

Advanced: Exactly 1

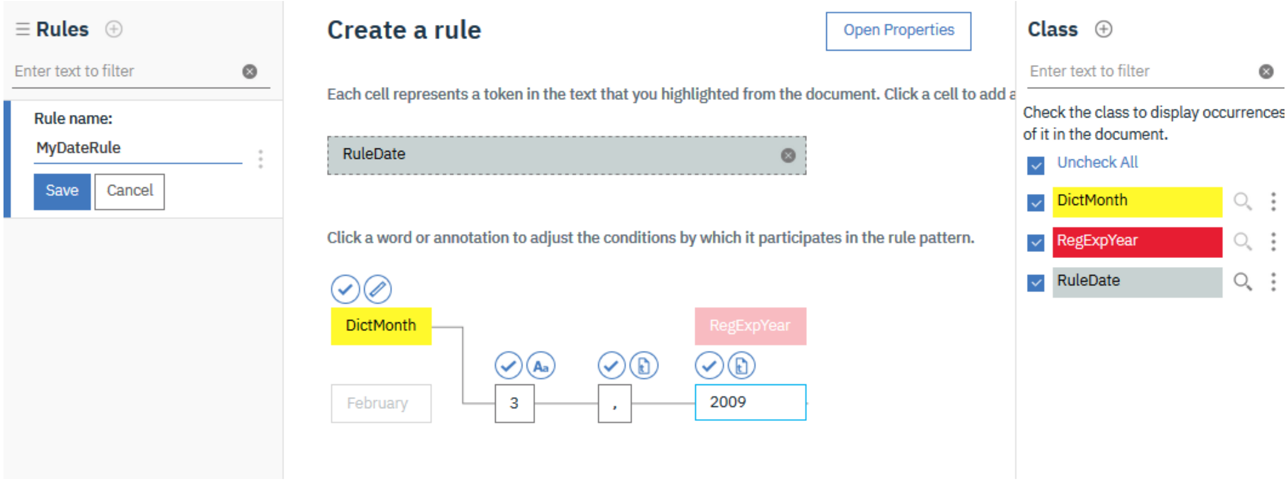
__ 7. Assign a class to represent the pattern `DictMonth + numeric token + comma + RegExpYear`.

Notice the four empty cells that represent the four tokens that you selected from the document. To select all the cells, select the first cell, and then press **Shift** + click each additional cell. Enter **RuleDate** as the class name, and then click it to create the new class.



Assign class
RuleDate
Create new class: RuleDate
conditions by which it participates in the rule pattern.
DictMo...
RegExpYear
February
3
,
2009

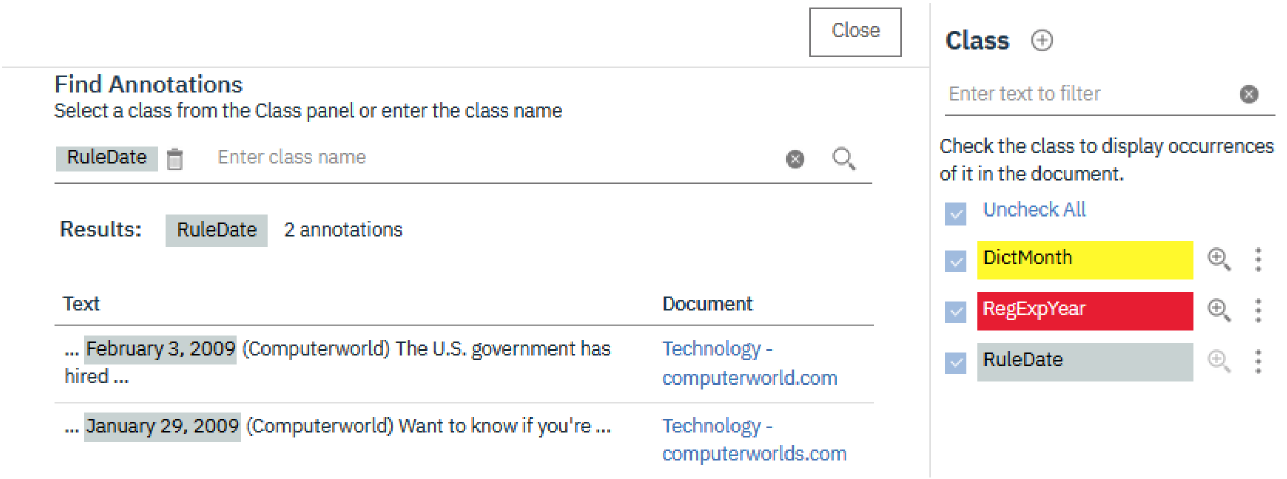
__ 8. In the **Rule name** field, enter **MyDateRule** and click **Save**.



Rules
Enter text to filter
Rule name:
MyDateRule
Save
Cancel
Create a rule
Open Properties
Each cell represents a token in the text that you highlighted from the document. Click a cell to add a
RuleDate
Click a word or annotation to adjust the conditions by which it participates in the rule pattern.
DictMonth
RegExpYear
February
3
,
2009
Class
Enter text to filter
Check the class to display occurrences of it in the document.
Uncheck All
DictMonth
RegExpYear
RuleDate

After you save the rule, it is automatically applied to the sample documents.

__ 9. You can search for all occurrences of `RuleDate` class mentions in the sample documents by clicking the **Search annotation in documents** icon 🔍 next to the `RuleDate` class from the Class panel. It is a good practice to check that all dates are captured properly to confirm that you defined the pattern correctly as in the figure.



Find Annotations

Select a class from the Class panel or enter the class name

RuleDate Enter class name

Results: RuleDate 2 annotations

Text	Document
... February 3, 2009 (Computerworld) The U.S. government has hired ...	Technology - computerworld.com
... January 29, 2009 (Computerworld) Want to know if you're ...	Technology - computerworlds.com

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

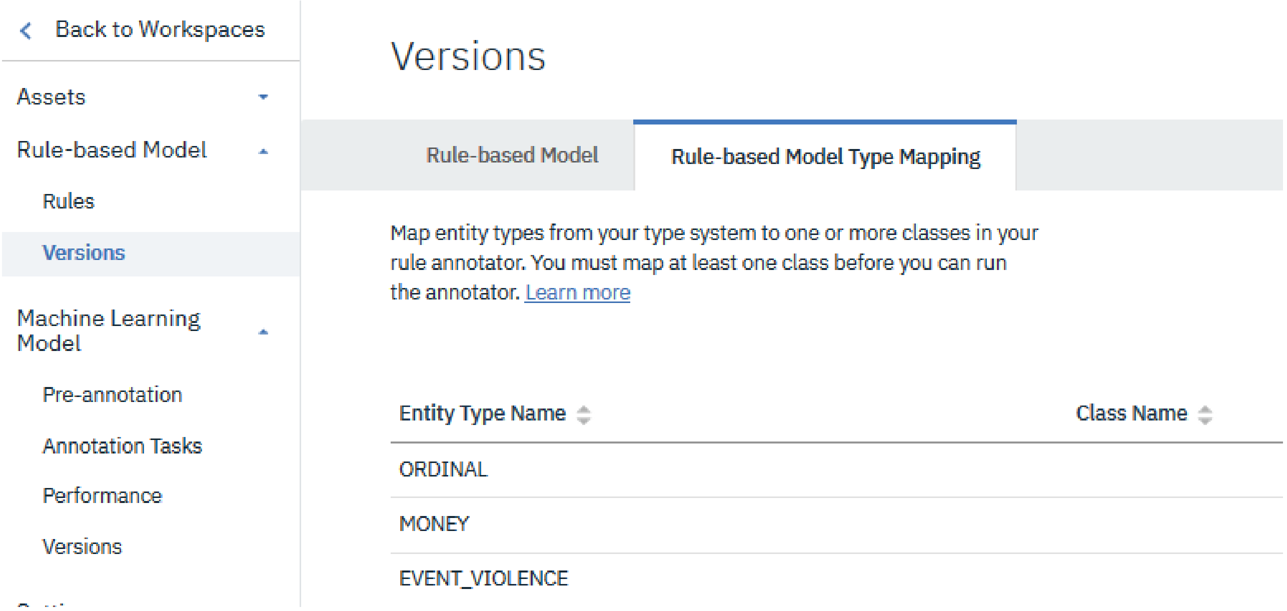
Part 8. Creating a rule-based model

After defining rules, you can create a rule-based model. The rule-based model can be run as a pre-annotator only on documents that were not already annotated by humans.

In this part, you will learn how to create a rule-based model.

Perform the following steps:

__ 1. Select **Rule-based Model** > **Versions** and click the **Rule-based model type mapping** tab.

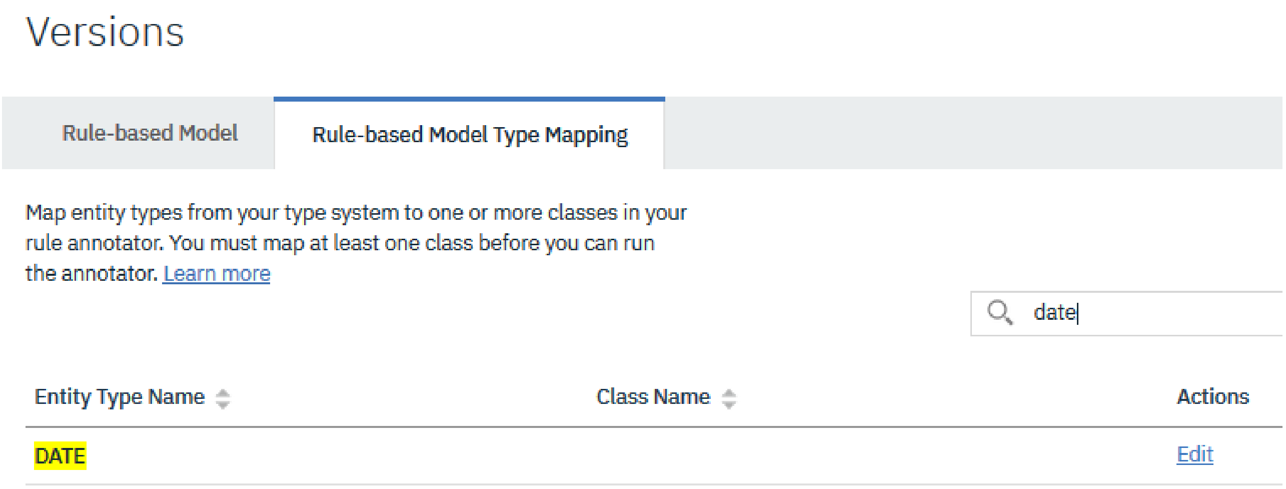


Map entity types from your type system to one or more classes in your rule annotator. You must map at least one class before you can run the annotator. Learn more

Entity Type Name	Class Name
ORDINAL	
MONEY	
EVENT_VIOLENCE	

__ 2. Map the `RuleDate` class to the DATE entity from the type system:

__ a. Find the **DATE** entity (use the filter to find it quickly) and click **Edit**.

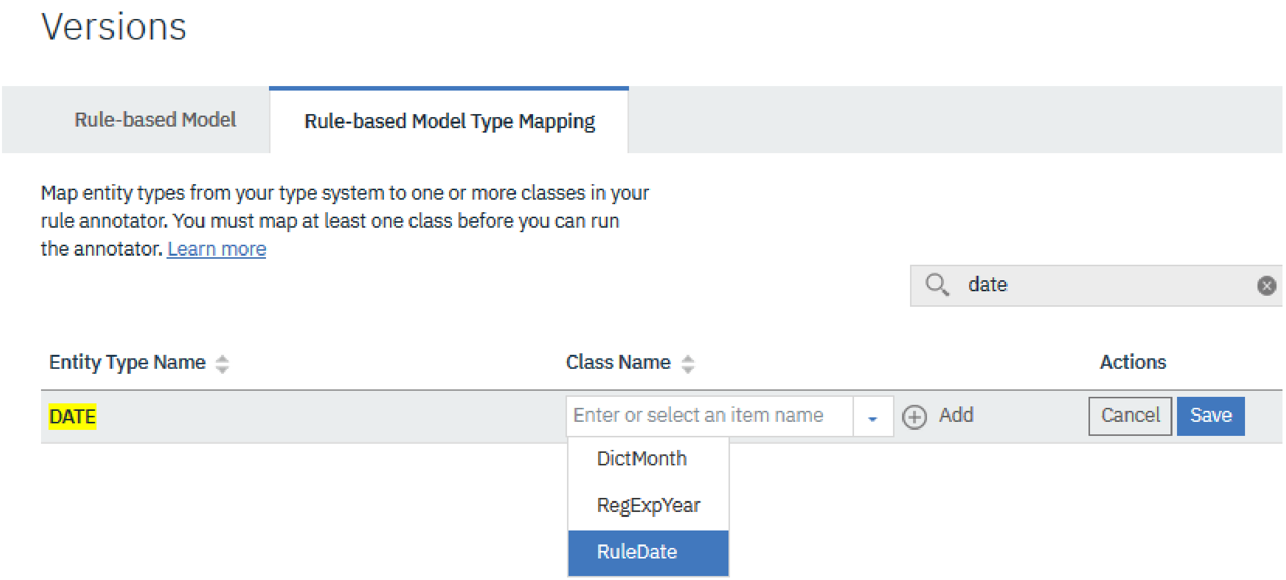


Versions

Map entity types from your type system to one or more classes in your rule annotator. You must map at least one class before you can run the annotator. Learn more

Entity Type Name	Class Name	Actions
DATE		Edit

__ b. Choose the `RuleDate` class from the list and click **Save**.



Versions

Map entity types from your type system to one or more classes in your rule annotator. You must map at least one class before you can run the annotator. Learn more

Entity Type Name	Class Name	Actions
DATE	Enter or select an item name ⊕ Add	Cancel Save

__ 3. To pre-annotate document sets or annotation sets with the rule-based model, select the **Rule-based Model** tab and click **Run this model.**

Attention: To be able to run the rule-based model successfully, run the rule-based model as a pre-annotator only on documents that were not already annotated by humans, so in order to try this out you will need to add new documents and try running the newly created rule-based model on it.

Exercise review and wrap-up

In this exercise, while learning about Knowledge Studio, you created a workspace and added artifacts to it. You then created a machine learning model, trained it, and evaluated how well it performed when annotating test data and blind data. By exploring the performance metrics, you can identify ways to improve the accuracy of the machine learning model. You created a custom machine learning model that you can use with other Watson services.

By completing this exercise, you learned about the following concepts:

- Workspaces
- Type systems
- Dictionaries
- Document sets
- Machine learning models
- Human annotation tasks
- Inter-annotator agreement and adjudication

If you performed the optional exercise, you created a rule-based model.

By completing the optional exercise, you learned about the following concepts:

- Classes
- Regular expressions
- Rules
- Ruled-based models