

# **IBM Training**

## **Student Exercises**

**Lab-1: Develop Socioeconomic  
Annotators for COVID-19**

**Hands-On Lab**

Legal Copyright: © Copyright IBM Corp. 2020  
*Course materials may not be reproduced in whole or in part without the  
prior written permission of IBM*

## **Table of Contents**

<b>Introduction .....</b>	<b>3</b>
<b>Objectives.....</b>	<b>3</b>
<b>Exercise 1: Create a Watson Knowledge Studio Instance .....</b>	<b>3</b>
<b>Exercise 2: Create a Watson Discovery Instance.....</b>	<b>6</b>
<b>Exercise 3: Create a Type System .....</b>	<b>7</b>
<b>Exercise 4: Create a Dictionary .....</b>	<b>11</b>
<b>Exercise 5: Upload a corpus of documents.....</b>	<b>13</b>
<b>Exercise 6: Perform Manual Annotation.....</b>	<b>16</b>
<b>Exercise 7: Train and create a machine learning (ML) annotator .....</b>	<b>22</b>
<b>Exercise 8: Save and Deploy the ML Annotator to Discovery .....</b>	<b>26</b>

## Introduction

This lab will cover the development of socioeconomic annotators for COVID-19 in order to create a COVID-19 vulnerability index. IBM Watson Knowledge Studio will be used to develop the socioeconomic annotators.

## Objectives

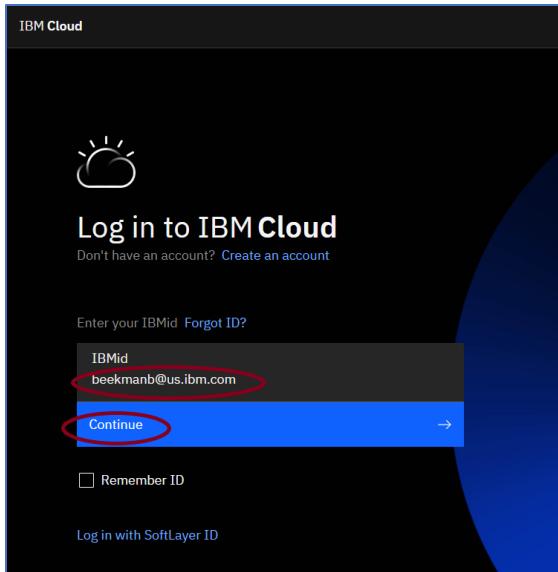
The goal of this lab is to familiarize the user with the Watson Knowledge Studio service. Watson Knowledge Studio lets you build a machine learning annotator by applying a type system, dictionary pre-annotator and human annotation on a training corpus of unstructured documents. Upon training and evaluation, the machine learning annotator can be saved and deployed to Watson Discovery for automated entity extraction. Watson Discovery is an enterprise AI search technology that leverages machine learning, including natural language processing, to retrieve specific answers to your questions and analyze trends and relationships buried in enterprise data; by integrating a machine learning annotator from Watson Knowledge Studio, Watson Discovery can be trained on the language of your domain. Both Watson Knowledge Studio and Watson Discovery can be deployed on any cloud or on-premises environment.

After completing this lab, you will be able to perform the following exercises:

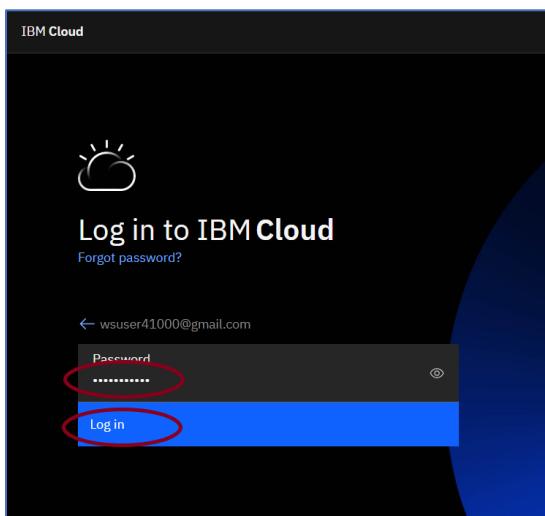
1. Provision an instance of Watson Knowledge Studio
2. Provision an instance of Watson Discovery
3. Create a type system
4. Create a dictionary
5. Upload a corpus of documents
6. Perform manual annotation
7. Train and create a machine learning (ML) annotator
8. Save and deploy the ML annotator to Watson Discovery

### Exercise 1: Create a Watson Knowledge Studio Instance

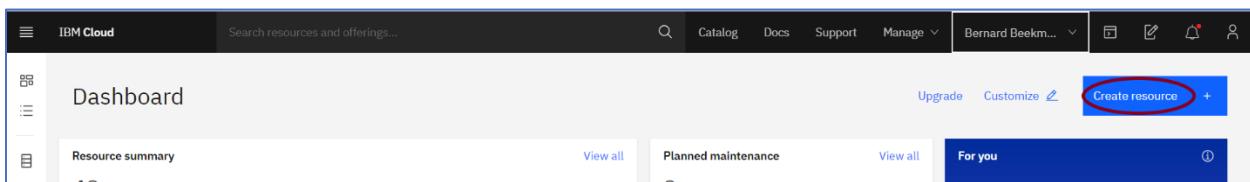
1. Log into your IBM Cloud account by typing **cloud.ibm.com** into the URL address bar of your Firefox or Chrome browser.
2. Enter your **IBMid** and click **Continue**.



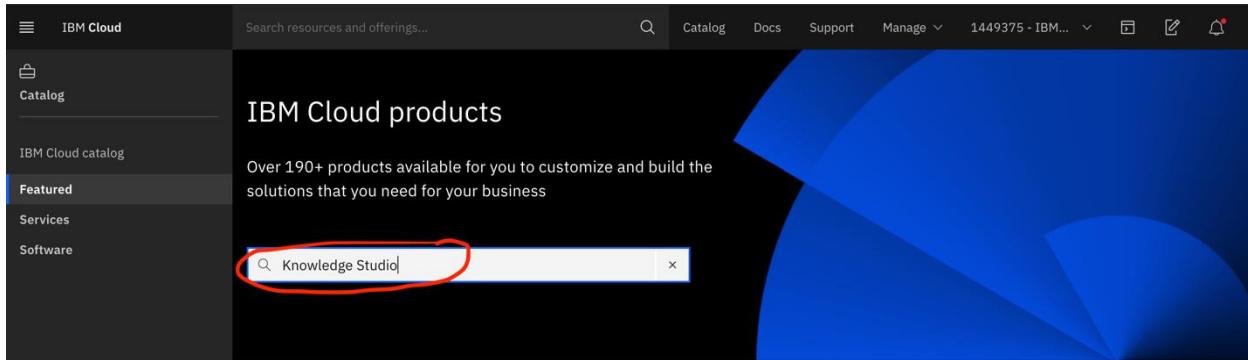
3. Enter your **Password** and click **Log in**.



4. Click **Create Resource**.



5. Enter **Knowledge Studio** and click the <Enter> key.



## 6. Click on Knowledge Studio.

A screenshot showing the search results for 'Knowledge Studio'. The results section is titled 'Search results for 'Knowledge Studio'' with '1 result'. The single result is 'Knowledge Studio' by IBM, categorized under 'Services • AI'. It has a small icon, the title 'Knowledge Studio' (circled in red), and a brief description: 'Teach Watson the language of your domain.' Below the description are the words 'Lite • Free • IAM-enabled'.

## 7. Click on the Lite plan and click Create.

A screenshot of the 'Knowledge Studio' creation page. At the top, there are tabs for 'Create' (which is active) and 'About'. Below that, a dropdown menu 'Select a region' is set to 'Dallas'. Under 'Select a pricing plan', it says 'Displayed prices do not include tax. Monthly prices shown are for country or region: United States'. A table compares three plans: 'Lite', 'Standard', and 'Premium'. The 'Lite' plan is circled in red. The 'Create' button at the bottom right of the page is also circled in red.

## Exercise 2: Create a Watson Discovery Instance

1. Enter **Discovery** into the *Search resources and offerings* bar and click on **Discovery** under *Catalog Results*.

The screenshot shows the IBM Cloud catalog interface. In the top navigation bar, the search bar contains the text "Discovery". Below the search bar, the "Catalog Results" section is visible, showing a list of services. The "Discovery" service is highlighted with a red circle, indicating it is the selected item.

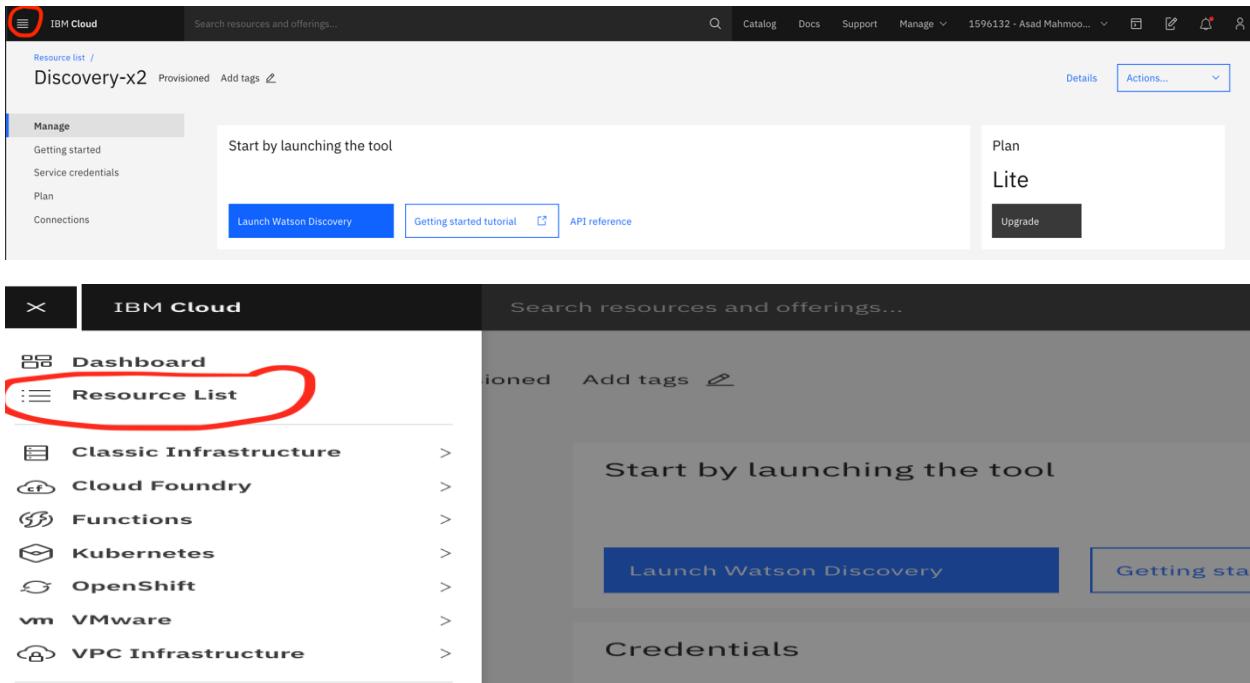
2. Select the **Lite** plan and click **Create**.

The screenshot shows the "Create Watson Discovery" page. On the left, there is a table comparing different plan tiers: **Lite**, **Advanced**, and **Premium**. The **Lite** plan is circled in red. The **Summary** section on the right provides details about the selected plan: Region: Dallas, Plan: Lite, Service name: Discovery-gr, Resource group: default. At the bottom right, the "Create" button is circled in red.

Although we will be using this Watson Discovery instance in Lab 2, we need to provision this instance in order to link it to the deployed machine learning annotator, which we will create by the end of this lab. The machine learning annotator will be used by Watson Discovery to perform entity extraction in Lab 2.

## Exercise 3: Create a Type System

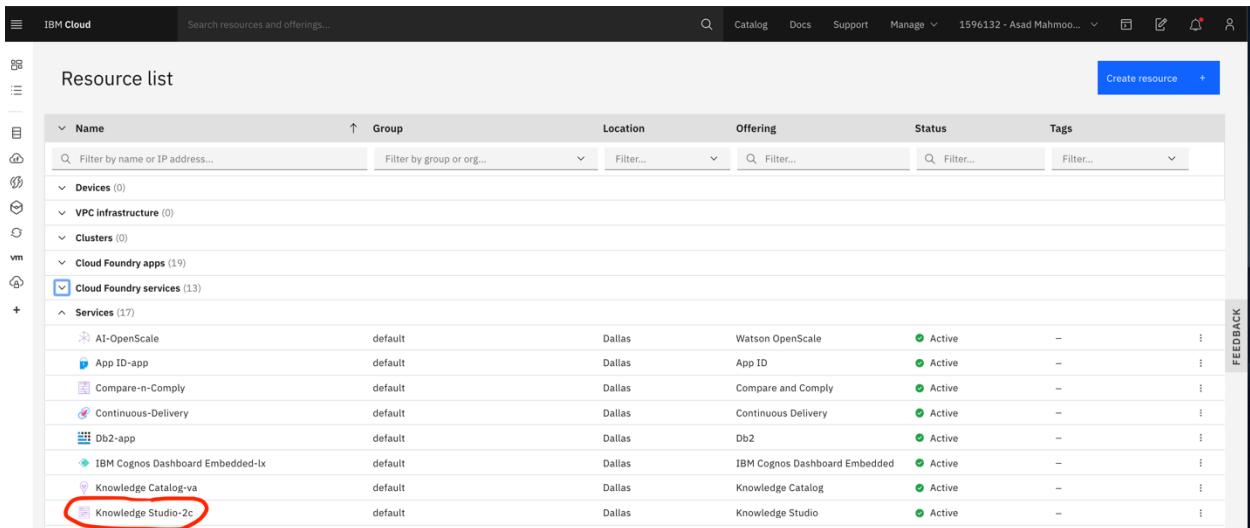
1. Select the Navigation Menu icon on the top left corner of the screen (the hamburger icon) and click **Resource List** on the drop down menu.



The top screenshot shows the 'Discovery-x2' service details page. The 'Manage' tab is selected, showing options like 'Getting started', 'Service credentials', 'Plan', and 'Connections'. Below this is a central panel with a 'Start by launching the tool' button, a 'Launch Watson Discovery' button, a 'Getting started tutorial' link, and an 'API reference' link. To the right is a 'Plan' section showing 'Lite' and an 'Upgrade' button. The top navigation bar includes 'IBM Cloud', a search bar, and account information.

The bottom screenshot shows the 'Resource List' page. The 'Resource List' item in the sidebar is circled in red. The main area displays a table of resources with columns for Name, Group, Location, Offering, Status, and Tags. A 'Create resource' button is at the top right. The sidebar also lists other service categories like Classic Infrastructure, Cloud Foundry, Functions, Kubernetes, OpenShift, VMware, and VPC Infrastructure.

2. Under Services, click on your Knowledge Studio instance (for a new IBM Cloud account, you should only see Knowledge Studio and Discovery listed here).



This screenshot shows the 'Resource list' page with the 'Services' section expanded. A specific entry, 'Knowledge Studio-2c', is circled in red. The table columns are identical to the previous screenshot: Name, Group, Location, Offering, Status, and Tags. The 'Knowledge Studio-2c' row shows it is located in Dallas, offered by Watson OpenScale, and is active. The sidebar on the left shows the navigation tree, and the top navigation bar is visible.

3. Click **Launch Knowledge Studio** to start your instance of Watson Knowledge Studio.

The screenshot shows the IBM Cloud interface with the title "Knowledge Studio-2c". The "Manage" section includes links for "Getting started" and "Plan". A central panel says "Start by launching the tool" with a blue button "Launch Watson Knowledge Studio" circled in red. To the right, a "Plan" section shows "Lite" and a "Upgrade" button.

4. Select **Create entities and relations workspace**.

The screenshot shows the "IBM Watson Knowledge Studio" interface. It features a "Create a workspace" section with two options: "Identify custom entities and relations in your data" (with a blue button "Create entities and relations workspace" circled in red) and "Extract entities with advanced rules" (with a blue button "Create advanced rules workspace"). Below these options is a link "Not sure which to pick? Learn more".

5. Type **COVID19-Vulnerability** for the Workspace name and click **Create**.

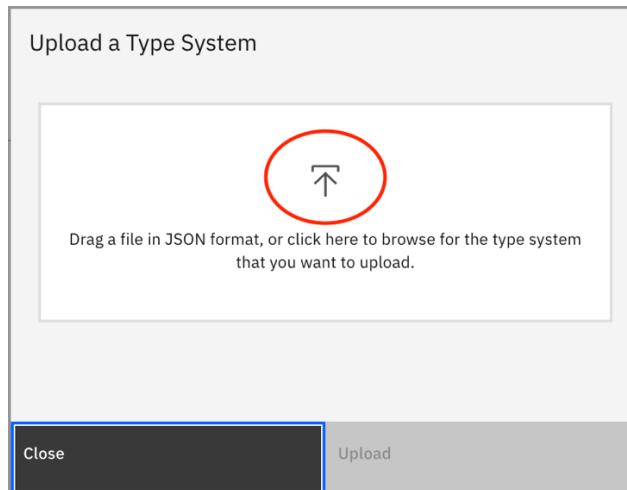
The screenshot shows the "Create Workspace" dialog. The "Workspace name" field contains "COVID19-Vulnerability" (circled in red). The "Language of documents" dropdown is set to "English". At the bottom, there are "Cancel" and "Create" buttons, with "Create" also circled in red.

Inside of this workspace, we will create a type system consisting of the custom entities of the COVID-19 vulnerability index, create a dictionary, perform manual annotation and upload a training corpus for the development of the entity recognition machine learning model.

- Although we can manually enter the entity types for our type system, we will instead upload the type system CSV file downloaded from the Box folder or the GitHub repository.

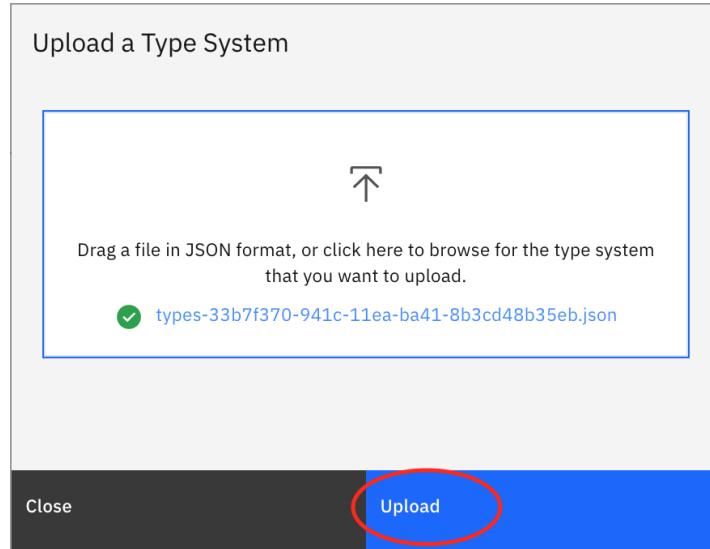
On the Entity Types screen, click **Upload**.

- Click on the upload icon and select the **types-33b7f370-941c-11ea-ba41-8b3cd48b35eb.json**.



Name	Date Modified	Size	Kind
Lab 1 - Watson Knowledge Studio	Today at 6:08 PM	1.8 MB	Micros...(docx)
Home_Owner_1589693231974.csv	Today at 1:27 AM	170 bytes	CSV Document
Car_Owner_1589693226813.csv	Today at 1:27 AM	224 bytes	CSV Document
No_Vehicle_1589693223234.csv	Today at 1:27 AM	146 bytes	CSV Document
Mobile_Home_Owner_1589693216310.csv	Today at 1:26 AM	116 bytes	CSV Document
Crowded_Living_1589693209295.csv	Today at 1:26 AM	191 bytes	CSV Document
Apartment_Renter_1589693204808.csv	Today at 1:26 AM	211 bytes	CSV Document
ESL_Speaker_1589693200465.csv	Today at 1:26 AM	212 bytes	CSV Document
Minority_1589693196515.csv	Today at 1:26 AM	324 bytes	CSV Document
Single_Parent_1589693192264.csv	Today at 1:26 AM	196 bytes	CSV Document
University_Student_1589693188097.csv	Today at 1:26 AM	304 bytes	CSV Document
No_High_School_Diploma_1589693182889.csv	Today at 1:26 AM	225 bytes	CSV Document
High_School_Student_1589693176460.csv	Today at 1:26 AM	217 bytes	CSV Document
Disabled_1589693170302.csv	Today at 1:26 AM	134 bytes	CSV Document
Minor_1589693163782.csv	Today at 1:26 AM	169 bytes	CSV Document
Senior_Citizen_1589693158519.csv	Today at 1:25 AM	264 bytes	CSV Document
Full_Time_Employment_1589693149530.csv	Today at 1:25 AM	161 bytes	CSV Document
Medically_Insured_1589693143233.csv	Today at 1:25 AM	225 bytes	CSV Document
Hourly_Wage_Employment_1589693134280.csv	Today at 1:25 AM	337 bytes	CSV Document
No_Health_Insurance_1589693125605.csv	Today at 1:25 AM	286 bytes	CSV Document
Unemployed_1589693113791.csv	Today at 1:25 AM	486 bytes	CSV Document
types-33b7f370-941c-11ea-ba41-8b3cd48b35eb.json	Today at 1:23 AM	12 KB	JSON Document

8. Click on **Upload**.



You should now see 20 entity types on your screen. These entity types directly pertain to social vulnerability to COVID-19 and will be used to annotate a corpus of social media posts from citizens living in New York City, Washington DC, Los Angeles, Seattle and Chicago – 5 cities that are among the most populous in the U.S. and most affected by the COVID-19 pandemic.

The screenshot shows the "Entity Types" page in IBM Watson Knowledge Studio. The left sidebar includes sections for Assets, Documents, Entity Types (which is selected), Relation Types, Dictionaries, Rule-based Model, Machine Learning Model, Settings, and Help. The main content area has tabs for "Entity Types" (selected), "Mention Classes", and "Mention Types". Below these are buttons for "Add Entity Type" and "Upload". A search bar says "Enter text to filter". A table lists 20 entity types:

Entity Type Name	Roles	Subtypes	Action
No_Health_Insurance	No_Health_Insurance		Edit Delete
Medically_Insured	Medically_Insured		Edit Delete
No_Vehicle	No_Vehicle		Edit Delete
Car_Owner	Car_Owner		Edit Delete
Mobile_Home_Owner	Mobile_Home_Owner		Edit Delete
Crowded_Living	Crowded_Living		Edit Delete
Apartment_Renter	Apartment_Renter		Edit Delete
Home_Owner	Home_Owner		Edit Delete
ESL_Speaker	ESL_Speaker		Edit Delete
Minority	Minority		Edit Delete

Pagination at the bottom shows "First" and "Last" with page number "1" highlighted.

## Exercise 4: Create a Dictionary

To help with manual annotation (which we will tackle in the next exercise), we will create a dictionary for each of the entity types in our type system. Each dictionary will contain a list of terms and key phrases pertaining to each entity type.

Although we can manually create a dictionary for each entity type, we will instead upload dictionary files for all 20 entity types in this exercise.

### 1. Under Assets, click **Dictionaries**.

The screenshot shows the Entity Types page in Watson Knowledge Studio. The left sidebar has links for Back to Workspaces, Assets, Documents, Entity Types (which is highlighted with a blue bar), Relation Types, and Dictionaries. The main area shows a table with columns for Entity Types, Mention Classes, and Mention Types. There are buttons for Add Entity Type, Upload, and Download Types. A search bar at the bottom right says 'Enter text to filter'.

### 2. On the Dictionaries page, we can upload the zip file containing dictionaries for all of our entity types. Click on the **vertical dots icon** and select **Upload Dictionary**.

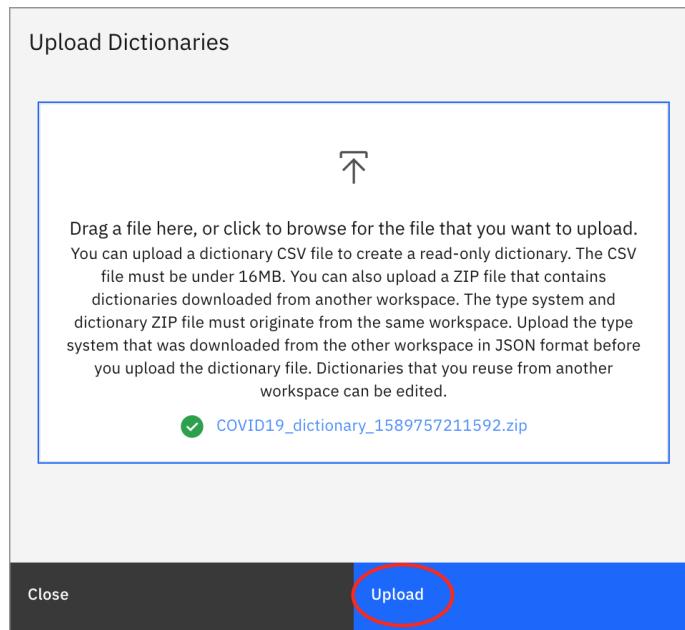
The screenshot shows the Dictionaries page in Watson Knowledge Studio. The left sidebar has links for Back to Workspaces, Assets, Documents, Entity Types, Relation Types, and Dictionaries. The main area shows a table with a 'Create Dictionary' button. A vertical ellipsis icon in the toolbar has a red circle around it. A context menu is open with options: Create Dictionary, Upload Dictionary (highlighted with a red oval), and Download Dictionaries.

### 3. Click on the **Upload icon** and select the **COVID19\_dictionary\_1589757211592.zip** file.

The screenshot shows the 'Upload Dictionaries' dialog box. It has a large input field with a red circle around the upload icon. Below the input field is text: 'Drag a file here, or click to browse for the file that you want to upload. You can upload a dictionary CSV file to create a read-only dictionary. The CSV file must be under 16MB. You can also upload a ZIP file that contains dictionaries downloaded from another workspace. The type system and dictionary ZIP file must originate from the same workspace. Upload the type system that was downloaded from the other workspace in JSON format before you upload the dictionary file. Dictionaries that you reuse from another workspace can be edited.' At the bottom are 'Close' and 'Upload' buttons.

Favorites	Name	Date Modified	Size	Kind
Recents	COVID19 dictionary 1589757211592	Today at 7:13 PM	--	Folder
Desktop	COVID19_dictionary_1589757211592.zip	Today at 7:13 PM	7 KB	ZIP archive

#### 4. Click on Upload.



You should now be able to see dictionaries for each entity type. We will use these dictionaries to pre-annotate a sample set of the social media posts prior to manual annotation.

In order to save these dictionaries as a pre-annotator, we need to remember to match each dictionary with its corresponding entity type. For example, the Unemployed dictionary, which currently has an entity type of None needs to be matched to the Unemployed entity type.

To fix this, we will have to update the Entity type for each dictionary. For the Unemployed dictionary:

5. Click the drop-down menu under Entity type and select **Unemployed**.

Repeat the same process for each dictionary until all 20 dictionaries are matched to their corresponding entity type (none of the dictionaries should have an Entity type of None).

## Exercise 5: Upload a corpus of documents

In this exercise, we will upload a corpus of social media posts to which we will apply a dictionary pre-annotator and perform manual annotation. This is a small set of social media posts containing first-hand narratives from citizens living in New York City, Washington D.C., Los Angeles, Seattle and Chicago.

1. Under Assets, select **Documents**.

The screenshot shows the 'Dictionaries' page in IBM Watson Knowledge Studio. The left sidebar under 'Assets' has 'Documents' selected, indicated by a red circle. The main area displays a list of dictionaries: 'Unemployed' (9 entries), 'No\_Health\_Insura...' (5 entries), 'Hourly\_Wage\_Emp...' (9 entries), and 'Medically\_Insured' (6 entries). A blue button 'Create Dictionary' is at the top left. On the right, there's a section for 'Unemployed' with a language note 'Language: English | 9 entries' and a dropdown for 'Entity type: None'. Buttons for 'Add Entry' and 'Upload' are at the bottom right.

2. Click **Upload Document Sets**.

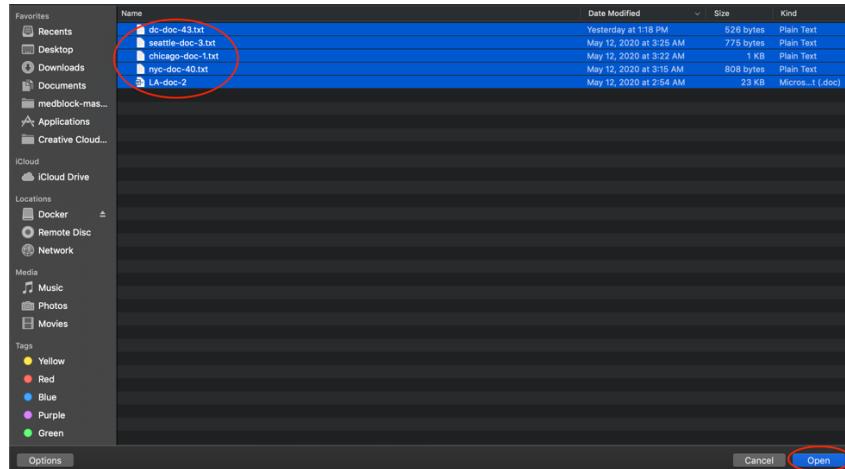
The screenshot shows the 'Documents' page in IBM Watson Knowledge Studio. The left sidebar under 'Assets' has 'Documents' selected, indicated by a blue highlight. The main area shows 'Document Sets (1)' and 'Documents (All, 0)'. A blue button 'Upload Document Sets' is highlighted with a red circle. A tooltip 'Add documents to the corpus' is shown above it. Below the button, there are options 'Name' and 'All'. At the bottom, there are 'Close' and 'Upload' buttons.

3. Click on the **Upload icon** and double-click on the **SampleDocs** folder.

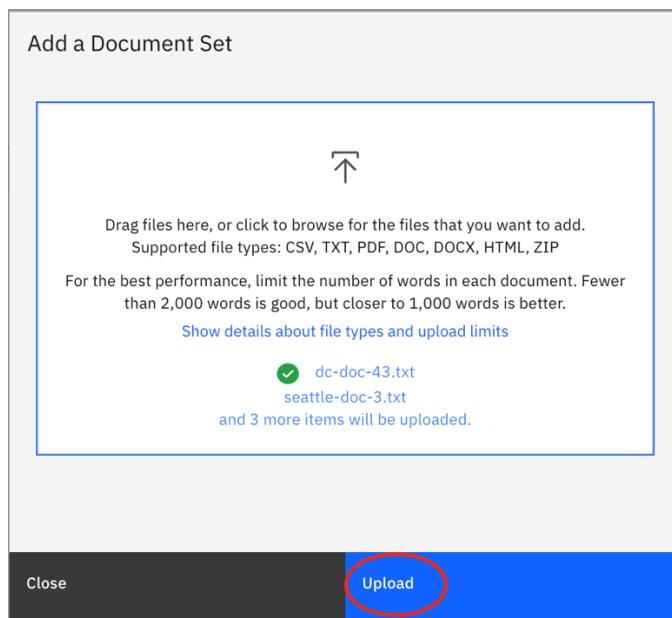
The screenshot shows the 'Add a Document Set' dialog box. It has a large central area for dragging files or clicking to browse, with an upward arrow icon circled in red. Below it, text says 'Drag files here, or click to browse for the files that you want to add. Supported file types: CSV, TXT, PDF, DOC, DOCX, HTML, ZIP'. It also notes '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.' and a link 'Show details about file types and upload limits'. At the bottom are 'Close' and 'Upload' buttons. Below the dialog, a file explorer window shows a folder structure. The 'SampleDocs' folder in the 'Desktop' section is circled in red. The file list shows:

File/Folder	Last Modified	Size	Type
Desktop			
Downloads			
Documents			
medblock-mas...			
Applications			
Creative Cloud			
SampleDocs.zip	Today at 9:06 PM	10 KB	ZIP archive
SampleDocs	Today at 7:59 PM	--	Folder
nyc-docs	Today at 7:55 PM	--	Folder
LA-docs	Today at 7:53 PM	--	Folder
COVID19_dictionary_1589757211592	Today at 7:13 PM	--	Folder
COVID19_dictionary_1589757211592.zip	Today at 7:13 PM	7 KB	ZIP archive
Dictionaries.zip	Today at 6:58 PM	19 KB	ZIP archive
Dictionaries	Today at 6:57 PM	--	Folder

4. Shift select all 5 documents in the folder and click **Open**.



## 5. Click Upload.



You should now be able to see a set of five documents named dc-doc-43.txt\_set to which we will apply a dictionary pre-annotator as well as manually annotate in the next exercise.

The screenshot shows the 'Documents' page in a workspace interface. The sidebar on the left includes sections for 'Assets' (with 'Entity Types', 'Relation Types', 'Dictionaries', 'Rule-based Model', and 'Machine Learning Model'), and 'Documents' (selected). The main content area has a heading 'Document Sets (2)' and a sub-heading 'Documents (All, 5)'. It contains two buttons: 'Upload Document Sets' (highlighted by a red box) and 'Download Document Sets'. Below these buttons is a message: 'To begin annotating documents, go to [Annotations](#) page.' A table lists the document sets:

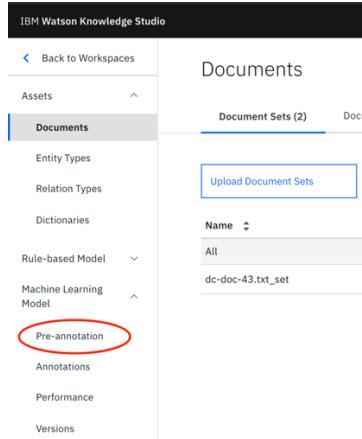
Name	Documents	Last Modified	Action
All	5	-	<a href="#">Rename</a> <a href="#">Delete</a>
dc-doc-43.txt_set	5	05/17/2020	<a href="#">Rename</a> <a href="#">Delete</a>

## Exercise 6: Perform Manual Annotation

In order to create an entity recognition model, we will need to teach Watson about our custom entity types by manually annotating a sample corpus of documents.

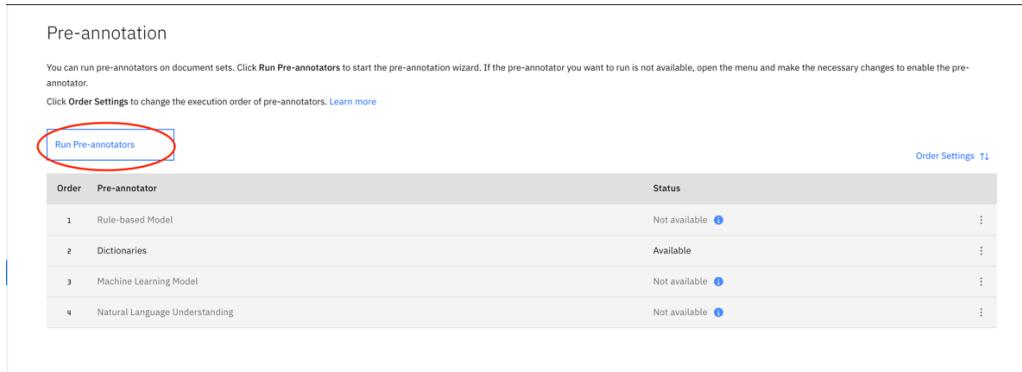
We will start by pre-annotating the document set with our dictionaries. This will allow Watson to quickly annotate our documents using the terms defined in each entity type dictionary.

1. Under Machine Learning Model, click **Pre-annotation**.



The screenshot shows the 'Documents' section of the IBM Watson Knowledge Studio. On the left, there's a sidebar with various options like Entity Types, Relation Types, Dictionaries, Rule-based Model, Machine Learning Model, and Pre-annotation. The 'Pre-annotation' link is highlighted with a red circle. The main area shows 'Document Sets (2)' and an 'Upload Document Sets' button. Below it, there's a search bar with 'Name' and 'All' selected, and a dropdown menu showing 'dc-doc-43.txt\_set'.

2. Click **Run Pre-annotators**.



The screenshot shows the 'Pre-annotation' wizard. At the top, there's a note about running pre-annotators on document sets. Below it, there's a link to 'Order Settings'. A red circle highlights the 'Run Pre-annotators' button. The main part is a table titled 'Pre-annotator' with columns 'Order', 'Pre-annotator', and 'Status'. The table lists four items: 'Rule-based Model' (Not available), 'Dictionaries' (Available), 'Machine Learning Model' (Not available), and 'Natural Language Understanding' (Not available). There's also a 'Order Settings' link at the top right of the table.

Order	Pre-annotator	Status
1	Rule-based Model	Not available
2	Dictionaries	Available
3	Machine Learning Model	Not available
4	Natural Language Understanding	Not available

You should be able to see that Dictionaries is available as a pre-annotator. If you do not see any available pre-annotators in the table, please revisit Exercise 4, step 5 to match each dictionary with its corresponding entity type.

3. Under Select pre-annotators, click the **checkbox** next to Dictionaries and click **Next**.

Run Pre-annotators

Select pre-annotators

Select the pre-annotators that you want to use.

Pre-annotator
<input checked="" type="checkbox"/> Dictionaries

**Close** **Next**



4. Under Select document sets, click the **checkbox** next to dc-doc-43.txt\_set and click **Run**.

Run Pre-annotators

Select document sets

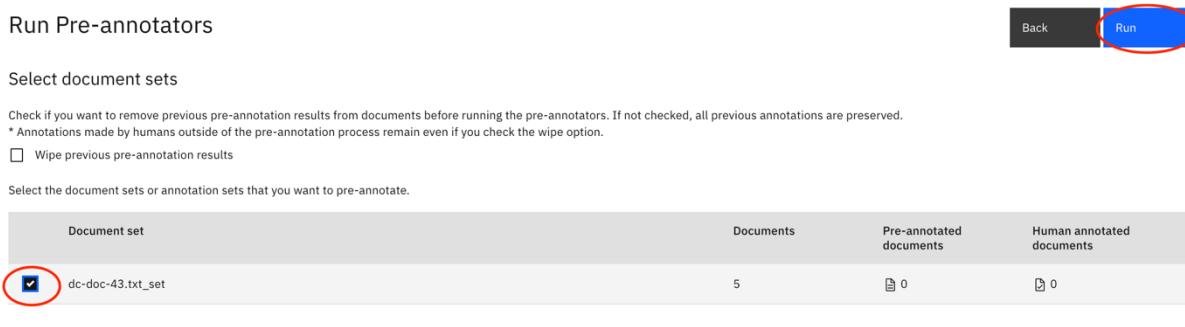
Check if you want to remove previous pre-annotation results from documents before running the pre-annotators. If not checked, all previous annotations are preserved.  
\* Annotations made by humans outside of the pre-annotation process remain even if you check the wipe option.

Wipe previous pre-annotation results

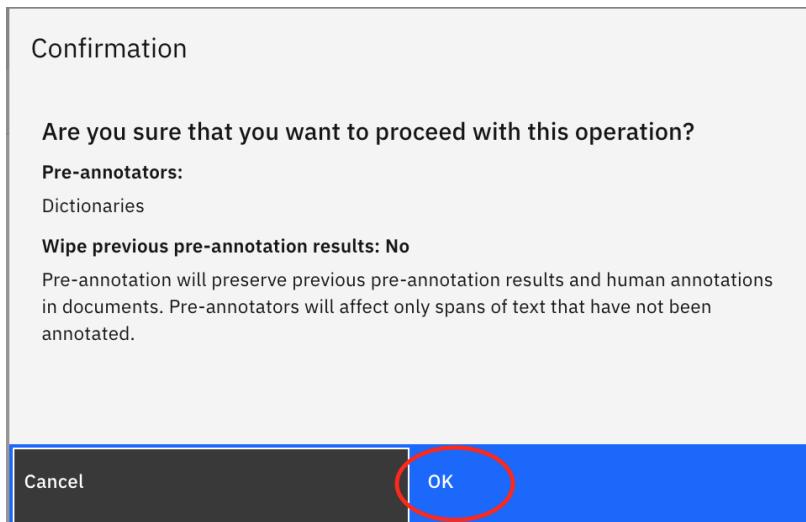
Select the document sets or annotation sets that you want to pre-annotate.

Document set	Documents	Pre-annotated documents	Human annotated documents
<input checked="" type="checkbox"/> dc-doc-43.txt_set	5	0	0

**Back** **Run**



5. Click **OK**.



After a few seconds, pre-annotation will be complete and you will see the following success message:

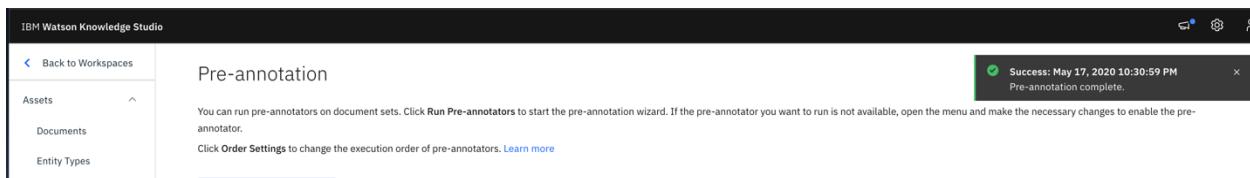
IBM Watson Knowledge Studio

Pre-annotation

You can run pre-annotators on document sets. Click Run Pre-annotators to start the pre-annotation wizard. If the pre-annotator you want to run is not available, open the menu and make the necessary changes to enable the pre-annotator.

Click Order Settings to change the execution order of pre-annotators. [Learn more](#)

**Success: May 17, 2020 10:30:59 PM**  
Pre-annotation complete.



6. Under Machine Learning Model, click **Annotations**.

The screenshot shows the 'Pre-annotation' section of the Watson Knowledge Studio interface. On the left, there's a sidebar with 'Machine Learning Model', 'Pre-annotation' (which is highlighted and has its 'Annotations' link circled in red), 'Performance', and 'Versions'. The main area displays a table with four rows:

	Document	Status	Actions
1	Rule-based Model	Not available	⋮
2	Dictionaries	Available	⋮
3	Machine Learning Model	Not available	⋮
4	Natural Language Understanding	Not available	⋮

On the Annotations screen, you will see that Watson used the dictionary pre-annotator to annotate 3 of the 5 documents. We will now manually annotate all 5 documents. When annotating each document, we will highlight any mention of the custom entity types in each social media post.

#### 7. Click the second **Annotate** link on the dc-doc-43.txt\_set row.

The screenshot shows the 'Annotations' page in Watson Knowledge Studio. The left sidebar includes 'Back to Workspaces', 'Assets', 'Documents', 'Entity Types', 'Relation Types', 'Dictionaries', 'Rule-based Model', 'Machine Learning Model', and 'Pre-annotation' (which is highlighted). The main area has tabs for 'Ground Truth' and 'Annotation Tasks'. Below is a note about creating ground truth by annotating documents. A table lists document sets:

Document Set	Last Modified	Documents (Annotated/Total)	Action
All	-	3 / 5	<a href="#">Annotate</a>
dc-doc-43.txt_set	05/17/2020	3 / 5	<a href="#">Annotate</a>

#### 8. On the Select Document screen, click the first **Open** link (on the dc-doc-43.txt row).

The screenshot shows a 'Select Document' dialog box. It says 'Document Set: dc-doc-43.txt\_set'. At the bottom right is a 'Close' button. The main area shows a table with one row:

Document Name	Status	Last Modified	Actions
dc-doc-43.txt		May 17, 2020 10:30:48 PM	<a href="#">Open</a>

At the bottom right of the dialog box, it says 'Showing 1-5 of 5'.

#### 9. To annotate the dc-doc-43.txt, we will skim through the post and find any mention of the custom entity types. We see that this post mentions one entity type in particular – **University\_Student**. The following sentences can be highlighted with this entity type: “UDC incoming student,” “I will be in the speech program at UDC” and “off campus student housing.”

To highlight a sentence with an entity type, click on the first word of the sentence and then the last word before clicking on the corresponding entity type on the right, which in this post is only University\_Student.

Manually annotate the above sentences with the **University\_Student** entity type so that you get the following annotated post:

The screenshot shows the annotation interface for the document 'dc-doc-43.txt'. The right side of the screen features a sidebar titled 'Entity' with a color-coded legend for various entity types. Some entities have subtypes or roles listed below them.

Type	Subtype	Role
Apartment_Renter		
Car_Owner		
Crowded_Living		
Disabled		
ESL_Speaker		
Full_Time_Employment		
High_School_Student		
Home_Owner		
Hourly_Wage_Employment		
Medically_Insured		
Minor		
Minority		
Mobile_Home_Owner		
No_Health_Insurance		
No_High_School_Diploma		
No_Vehicle		
Senior_Citizen		
Single_Parent		
Unemployed		
University_Student		

10. Click **Save** and click **Open document list** to return to the list of documents.

The screenshot shows the annotation interface for the document 'dc-doc-43.txt'. The top navigation bar includes links for 'Back to Annotations', 'Open document list', 'View Details', 'Replace', 'Concordance', 'Attribute View', and 'Save'. The 'Save' button is highlighted with a red circle.

11. Click the **Open** link on the seattle-doc-3.txt row.

The screenshot shows a 'Select Document' dialog box. It lists several documents with columns for 'Document Name', 'Status', 'Last Modified', and 'Actions'. The 'Actions' column contains 'Open' links, which are highlighted with red circles.

Document Name	Status	Last Modified	Actions
seattle-doc-3.txt		May 17, 2020 10:30:48 PM	<a href="#">Open</a>
chicago-doc-1.txt		May 17, 2020 10:30:48 PM	<a href="#">Open</a>
nyc-doc-40.txt		May 17, 2020 10:30:48 PM	<a href="#">Open</a>
LA-doc-2.doc		May 17, 2020 10:30:48 PM	<a href="#">Open</a>
dc-doc-43.txt		May 18, 2020 12:02:21 AM	<a href="#">Open</a>

12. Annotate the seattle-doc-3.txt and compare your result with the annotated post below. Don't forget to **Save** your annotation and click **Open document list** when you're done annotating this post.

The screenshot shows the annotation interface for the document 'seattle-doc-3.txt'. The top navigation bar includes links for 'Back to Annotations', 'Open document list', 'View Details', 'Replace', 'Concordance', 'Attribute View', and 'Save'. The 'Save' button is highlighted with a red circle.

13. Click the **Open** link on the chicago-doc-1.txt row.

Select Document

Document Set: dc-doc-43.txt\_set

Showing 1-5 of 5

Document Name	Status	Last Modified	Actions
chicago-doc-1.txt		May 17, 2020 10:30:48 PM	<a href="#">Open</a>

14. Annotate the chicago-doc-1.txt post and compare your result with the annotated post below. Don't forget to **Save** your annotation and click **Open document list** when you're done annotating this post.

Back to Annotations [Open document list](#)

View Details Replace Concordance Attribute View

Save Entity Mention

Type Subtype Role

Apartment\_Renter  
Car\_Owner  
Crowded\_Living  
Disabled  
ESL\_Speaker  
Full\_Time\_Employment  
High\_School\_Student  
Home\_Owner  
Hourly\_Wage\_Employment  
Medically\_Insured  
Minor  
Minority  
Mobile\_Home\_Owner  
No\_Health\_Insurance  
No\_High\_School\_Diploma  
No\_Vehicle  
Senior\_Citizen  
Single\_Parent  
Unemployed  
University\_Student

chicago-doc-1.txt

1 { "title": "Are **landlords** really allowed to enter occupied **apartments** to film new virtual tours?"  
2 ", "text": "As the questions asks, are **landlords** actually allowed to do this?  
3 I had my alderman refer me to **tenant** rights but haven't heard anything back.  
4 Speaking with a lawyer they said they recording a virtual tour is a nice compromise and didn't seem aware that in person showing if occupied **units** is actually prohibited.  
5 \\n\\nTo add insult to injury, they are posting these videos publicly on YouTube with the **unit** numbers along with the name of the **apartment**, so on top of potentially getting sick we are being forced to publish our private property online for anyone to see.  
6 \\n\\nPrior to knowing the very public way they were distributing these videos we offered to record one only to be told the format would have to be perfect or we'd have to keep recording until we got it correct.  
7 \\n\\nI see a lot of posts about how Chicago is a very **tenant** friendly city, but I don't see it right now.  
8 ", " subreddit": "chicago", "created": "2020-04-24T07:42:46.000Z" }

15. Click the **Open** link on the nyc-doc-40.txt row.

Select Document

Document Set: dc-doc-43.txt\_set

Showing 1-5 of 5

Document Name	Status	Last Modified	Actions
nyc-doc-40.txt		May 17, 2020 10:30:48 PM	<a href="#">Open</a>

16. Annotate the nyc-doc-40.txt post and compare your result with the annotated post below. Don't forget to **Save** your annotation and click **Open document list** when you're done annotating this post.

The screenshot shows the annotation interface for the document 'nyc-doc-40.txt'. At the top, there are buttons for 'Back to Annotations', 'Open document list', 'View Details', 'Replace', 'Concordance', and 'Attribute View'. The 'Save' button is highlighted with a red circle. On the left, there's a sidebar with icons for Mention, Relation, Conference, and Coreference. The main area contains numbered annotations:

1. {"text": "I know most of the world is **laid off** right now and apparently most of NYC but i can't get through to **unemployment** at all!! I've been calling for hours i don't understand why they couldn't complete my claim online."}
2. Is there a center i can go to in person?
3. I'd rather wait in line then to call back to back to get some automated system that hangs up on me or actually get through to the menu, enter all my info and have it hang up on me AGAIN after I'm supposed to be transfers to a rep.
4. I have to pay my **rent** and my partner is also **laid off**.
5. He got approved but his benefits aren't going to come for 2-3 weeks it says .. and i can't even get through to get mine approved.
6. ANY advice seriously I'm spinning out here", "author\_fullname": "t2\_4qqx83ci", "title": "**Unemployment**"}

A sidebar on the right lists entity types with 'Unemployed' selected. Other types listed include Apartment\_Renter, Car\_Owner, Crowded\_Living, Disabled, ESL\_Speaker, Full\_Time\_Employment, High\_School\_Student, Home\_Owner, Hourly\_Wage\_Employment, Medically\_Insured, Minor, Minority, Mobile\_Home\_Owner, No\_Health\_Insurance, No\_High\_School\_Diploma, No\_Vehicle, Senior\_Citizen, Single\_Parent, and University\_Student.

17. Click the **Open** link on the LA-doc-2.doc row.

The screenshot shows a 'Select Document' dialog. At the top, there's a 'Close' button. Below it, it says 'Showing 1-5 of 5'. A table lists documents:

Document Name	Status	Last Modified	Actions
LA-doc-2.doc		May 17, 2020 10:30:48 PM	<b>Open</b>

18. Annotate the nyc-doc-40.txt post and compare your result with the annotated post below.  
Don't forget to **Save** your annotation and click **Open document list** when you're done  
annotating this post.

The screenshot shows the annotation interface for the document 'LA-doc-2.doc'. At the top, there are buttons for 'Back to Annotations', 'Open document list', 'View Details', 'Replace', 'Concordance', and 'Attribute View'. The 'Save' button is highlighted with a red circle. On the left, there's a sidebar with icons for Mention, Relation, Conference, and Coreference. The main area contains numbered annotations:

- 1 {
- 2 "title": "Isolated outdoors spot?"
- 3 ",
- 4 "text": "It's my wife's birthday this weekend and was wondering if anyone knew of any places within and around the city that are isolated where you could **bring** and park your **car** to enjoy outdoors for a picnic or something?"
- 5 I don't want to endanger anyone or break any county rules but was just hoping to get us outside of the **house** for a couple hours to make it at least a little memorable.
- 6 ",
- 7 " subreddit": "LosAngeles",
- 8 "created": "2020-04-22T02:49:34.000Z"
- 9 }

A sidebar on the right lists entity types with 'House' selected. Other types listed include Apartment\_Renter, Car\_Owner, Crowded\_Living, Disabled, ESL\_Speaker, Full\_Time\_Employment, High\_School\_Student, Home\_Owner, Hourly\_Wage\_Employment, Medically\_Insured, Minor, Minority, Mobile\_Home\_Owner, No\_Health\_Insurance, No\_High\_School\_Diploma, No\_Vehicle, Senior\_Citizen, Single\_Parent, and University\_Student.

All 5 documents have now been manually annotated. However, we will need a much larger set of documents in order to train and create a machine learning model. In the next exercise, we will upload the complete corpus of documents and create an entity recognition model.

19. Click **Annotations** to return to the Annotations screen.

The screenshot shows the 'Annotations' tab selected in the sidebar. The main area displays a table titled 'Select Document' with a document set named 'dc-doc-43.txt\_set'. The table has columns for 'Document Name', 'Status', and 'Last Modified'. The data includes:

Document Name	Status	Last Modified
dc-doc-43.txt		May 18, 2020 12:02:21 AM
seattle-doc-3.txt		May 18, 2020 12:33:04 AM
chicago-doc-1.txt		May 18, 2020 12:40:47 AM
nyc-doc-40.txt		May 18, 2020 12:47:26 AM
LA-doc-2.doc		May 18, 2020 1:01:13 AM

## Exercise 7: Train and create a machine learning (ML) annotator

As stated above, we will require a much larger set of documents in order to create a machine learning annotator. Although we can provide a folder with all of the social media posts extracted for each of the 5 cities and instruct you annotate each post one by one, we have already done all of the hard work for you and have prepared a zip file containing the entire corpus of documents called Lab1-WKS.zip. Let's upload this zip file to our workspace.

- Under Assets, click **Documents**.

The screenshot shows the 'Documents' tab selected in the sidebar. The main area displays a table titled 'Annotations' with tabs for 'Ground Truth' and 'Annotation Tasks'. The table has columns for 'Document Set', 'Last Modified', 'Documents (Annotated/Total)', and 'Action'. The data includes:

Document Set	Last Modified	Documents (Annotated/Total)	Action
All	-	5 / 5	Annotate
dc-doc-43.txt_set	05/17/2020	5 / 5	Annotate

- On the Documents screen, click **Upload Document Sets**.

The screenshot shows the 'Documents' tab selected in the sidebar. The main area displays a table titled 'Documents' with tabs for 'Document Sets (2)' and 'Documents (All, 5)'. A red circle highlights the 'Upload Document Sets' button. The table has columns for 'Name', 'Documents', 'Last Modified', and 'Action'. The data includes:

Name	Documents	Last Modified	Action
All	5	-	Rename Delete
dc-doc-43.txt_set	5	05/17/2020	Rename Delete

- Click on the **Upload icon** and select the **Lab1-WKS.zip** file and click **Open**.

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

**Close** **Upload**

**Lab1-WKS.zip**

Lab1-WKS.zip	Today at 1:26 AM	5.4 MB	ZIP archive
SampleDocs.zip	Yesterday at 9:06 PM	10 KB	ZIP archive
SampleDocs	Yesterday at 7:59 PM	--	Folder
nyc-docs	Yesterday at 7:55 PM	--	Folder
LA-docs	Yesterday at 7:53 PM	--	Folder
COVID19_dictionary_1589757211592	Yesterday at 7:13 PM	--	Folder
COVID19_dictionary_1589757211592.zip	Yesterday at 7:13 PM	7 KB	ZIP archive
Dictionaries.zip	Yesterday at 6:58 PM	19 KB	ZIP archive
Dictionaries	Yesterday at 6:57 PM	--	Folder

- Click the box next to **Upload corpus documents and include ground truth (upload the original workspace's type system first)** 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

**Lab1-WKS.zip**

Upload corpus documents and include ground truth (upload the original workspace's type system first)

Upload documents in UIMA CAS XMI format

**Close** **Upload**

You should now see several new document sets on the Documents screen including an Import document set consisting of 368 posts that were just now added to the workspace. We will be using these newly uploaded documents to train and create a ML annotator.

IBM Watson Knowledge Studio

Back to Workspaces

Assets

**Documents**

Entity Types

Relation Types

Dictionaries

Rule-based Model

Machine Learning Model

Pre-annotation

Annotations

**Performance**

Versions

Import

Settings

Help

Documents

Document Sets (8) Documents (All, 373)

Upload Document Sets To begin annotating documents, go to [Annotations](#) page.

Name	Documents	Last Modified	Action
All	373	-	Rename Delete
LA-doc-6.doc_set	32	05/12/2020	Rename Delete
nyc-doc-1.txt_set	96	05/12/2020	Rename Delete
chicago-doc-1.txt_set	26	05/12/2020	Rename Delete
seattle-doc-1.txt_set	113	05/12/2020	Rename Delete
dc-doc-1.txt_set	91	05/13/2020	Rename Delete
dc-doc-43.txt_set	5	05/18/2020	Rename Delete
Import	368	05/18/2020	Rename Delete

Download Document Sets

## 5. Under Machine Learning Model, click on Performance.

IBM Watson Knowledge Studio

Back to Workspaces

Assets

**Documents**

Entity Types

Relation Types

Dictionaries

Rule-based Model

Machine Learning Model

Pre-annotation

Annotations

**Performance**

Versions

Import

Settings

Help

Documents

Document Sets (8) Documents (All, 373)

Upload Document Sets To begin annotating documents, go to [Annotations](#) page.

Name
All
LA-doc-6.doc_set
nyc-doc-1.txt_set
chicago-doc-1.txt_set
seattle-doc-1.txt_set
dc-doc-1.txt_set
dc-doc-43.txt_set
Import

## 6. On the Performance screen, click on Train and evaluate.

IBM Watson Knowledge Studio

Back to Workspaces

Assets

Documents

Entity Types

Relation Types

Dictionaries

Rule-based Model

Machine Learning Model

Pre-annotation

Annotations

**Performance**

Performance

COVID19-Vulnerability

Language of documents English

Number of documents per set

0 Training Set  
0 Test Set  
0 Blind Set

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

Last trained on: Last evaluated on:

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

Document set evaluation

Model over time

View Log Mention Precision: -- Recall: --

7. On the Select Training/Test/Blind Sets screen, choose **Import**, change the **Training Set** percentage to 85%, **Test Set** to 10% and **Blind Set** to 5%. Click **Train & Evaluate**.

← | Training / Test / Blind Sets

Select Training/ Test/ Blind Sets

**Train**   **Train & Evaluate** (highlighted with a red circle)

Document Set	Task Status
<input type="checkbox"/> All	
<input type="checkbox"/> LA-doc-6.doc_set	
<input type="checkbox"/> nyc-doc-1.txt_set	
<input type="checkbox"/> chicago-doc-1.txt_set	
<input type="checkbox"/> seattle-doc-1.txt_set	
<input type="checkbox"/> dc-doc-1.txt_set	
<input type="checkbox"/> dc-doc-43.txt_set	
<input checked="" type="checkbox"/> Import	

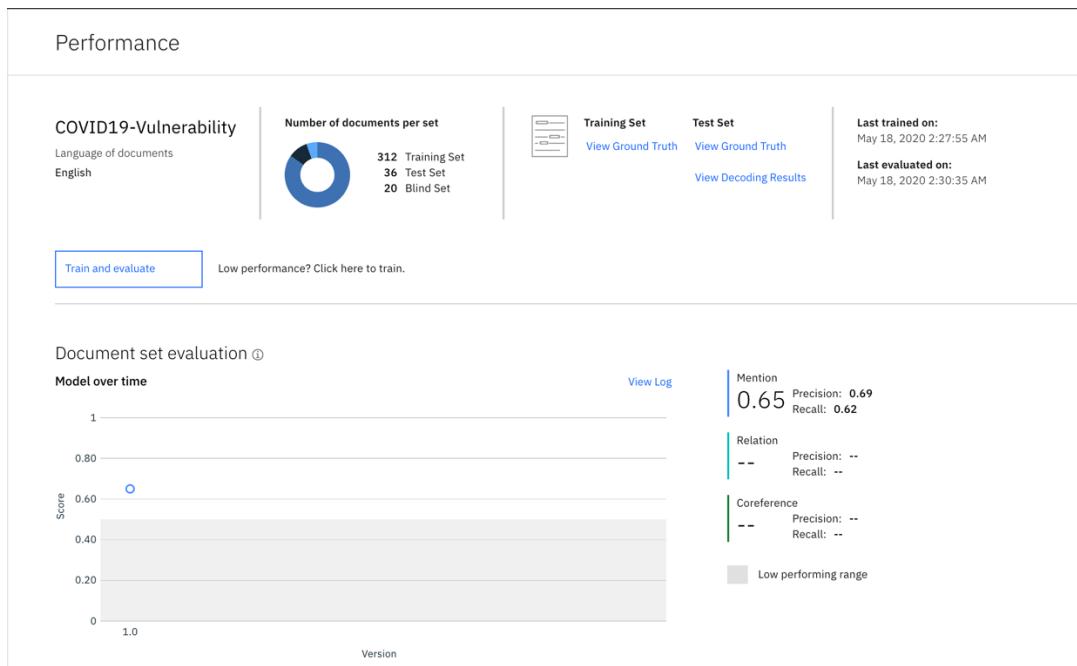
Create new sets by splitting the selected document sets

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

85	Training Set (70% Recommended)
10	Test Set (23% Recommended)
5	Blind Set (7% Recommended)

This will start the process of training and evaluating a machine learning annotator, which should take approximately 14 minutes to complete. You will see a progress message on the top right corner of the screen detailing the current phase – training or evaluation – and the amount of time elapsed.

Once the model is created, you should see the following on your Performance screen:

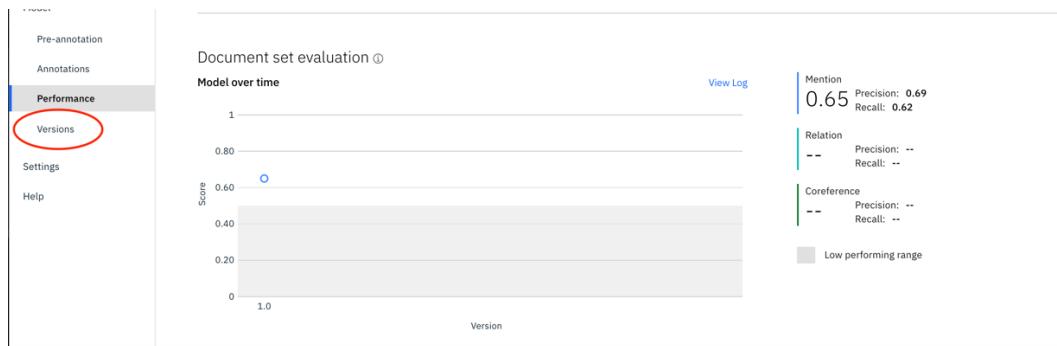


## Exercise 8: Save and Deploy the ML Annotator to Discovery

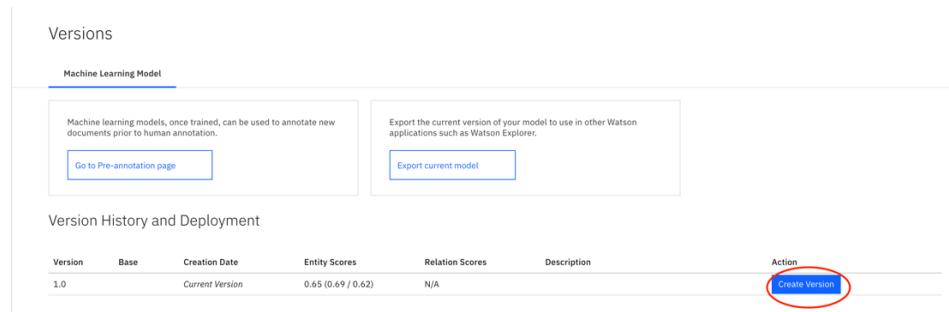
Now that we have a machine learning annotator, we can use it to automatically perform entity extraction inside of Watson Discovery. The automated entity extraction of social media posts for all 5 cities will get us closer to determining the social vulnerability index of each city.

Let's save this machine learning model and deploy it to the Discovery instance that we created at the beginning of this lab.

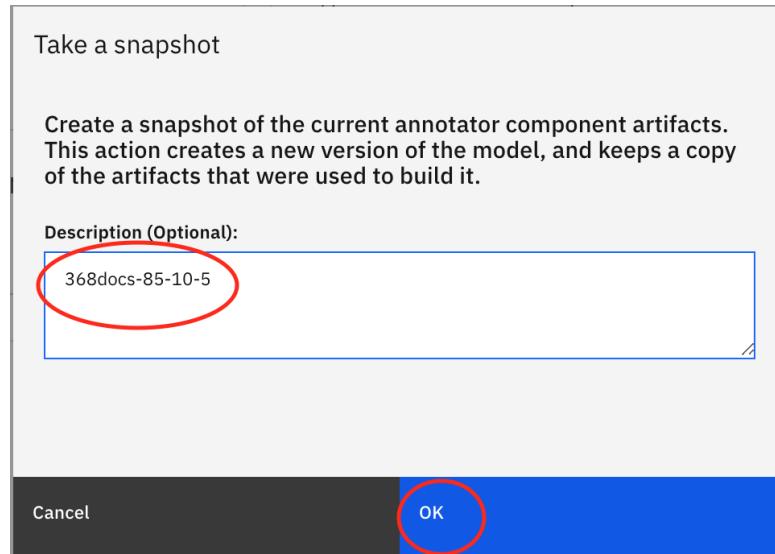
1. Under Machine Learning Model, click on **Versions**.



2. On the Versions page, click **Create Version**.



3. Type **368docs-85-10-5** (to distinguish this as an entity model using 368 docs with an 85-10-5 split) under Description and click **OK**.

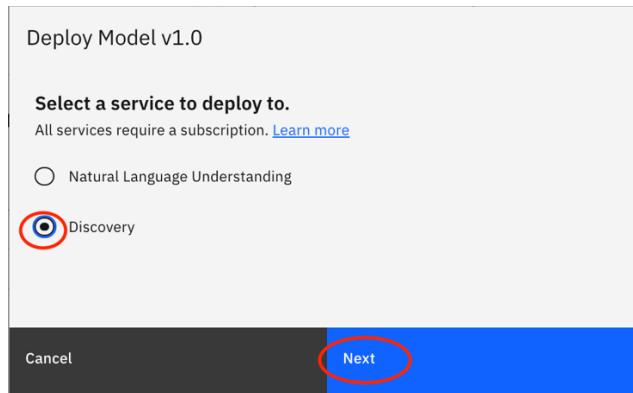


#### 4. In the Version 1.0 row, click **Deploy**.

Version History and Deployment

Version	Base	Creation Date	Entity Scores	Relation Scores	Description	Action
1.1	Current Version		0.65 (0.69 / 0.62)	N/A		<a href="#">Create Version</a>
1.0	05/18/2020		0.65 (0.69 / 0.62)	N/A	368docs-85-10-5	<a href="#">Promote</a> <a href="#">Delete</a> <a href="#">Deploy</a>

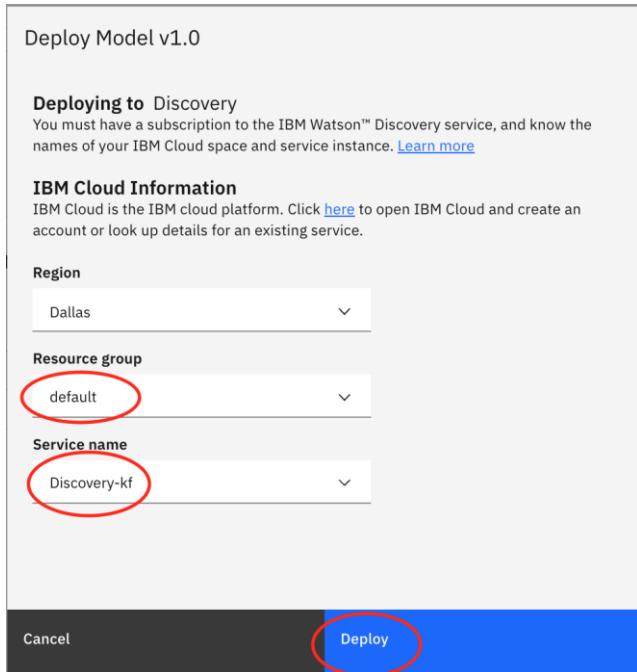
#### 5. Select **Discovery** and click **Next**.



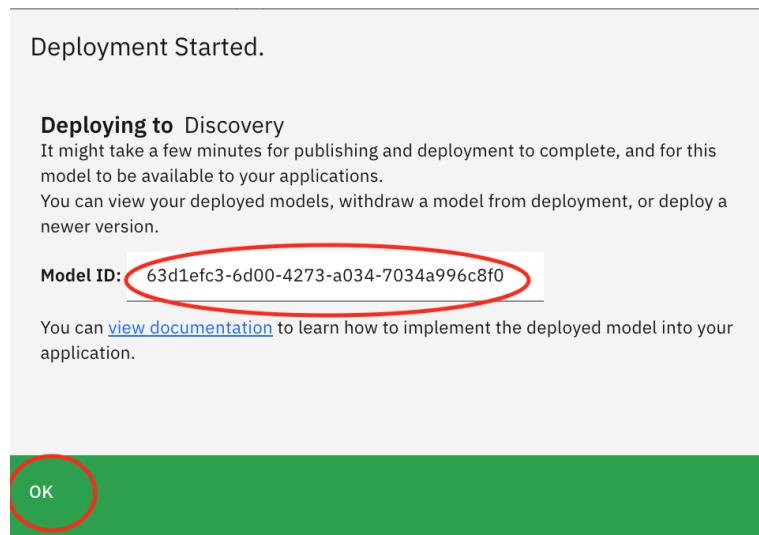
#### 6. In order to deploy this model to your Discovery instance, you will need to select the resource group containing your instance as well as the Service name of the instance that you created.

If this is your first time working with the Watson APIs on the IBM Cloud, you should only have one instance of Discovery currently provisioned.

Select **default** from the drop-down menu under **Resource group** and the **name of the Discovery instance** under **Service name**.



7. Copy the **Model ID** displayed on the screen to use in the next lab and click **OK**.



8. Underneath Deployed Models, you should now see the Model ID number for your newly deployed model. This deployed model will be used to perform entity extraction within Watson Discovery in Lab 2.

## Versions

### Machine Learning Model

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

[Go to Pre-annotation page](#)

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

[Export current model](#)

## Version History and Deployment

Version	Base	Creation Date	Entity Scores	Relation Scores	Description	Action
1.1	<i>Current Version</i>	0.65 (0.69 / 0.62)	N/A			<a href="#">Create Version</a>
1.0	05/18/2020	0.65 (0.69 / 0.62)	N/A	368docs-85-10-5		<a href="#">Promote</a> <a href="#">Delete</a> <a href="#">Deploy</a>

**You have completed Lab 1!**