

IBM Training

Student Exercises

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

Hands-On Lab

Legal Copyright: © *Copyright IBM Corp. 2021*

Course materials may not be reproduced in whole or in part without the prior written permission of IBM

Table of Contents

| | |
|---|-----------|
| Prerequisites | 3 |
| Introduction..... | 3 |
| Objectives..... | 3 |
| Exercise 1: Create a Watson Knowledge Studio Instance | 4 |
| Exercise 2: Create a Watson Discovery Instance..... | 6 |
| Exercise 3: Create a Type System | 8 |
| Exercise 4: Create a Dictionary | 13 |
| Exercise 5: Upload a corpus of documents | 16 |
| Exercise 6: Perform Manual Annotation..... | 19 |
| Exercise 7: Train and create a machine learning (ML) annotator | 28 |
| Exercise 8: Save and Deploy the ML Annotator to Discovery | 32 |

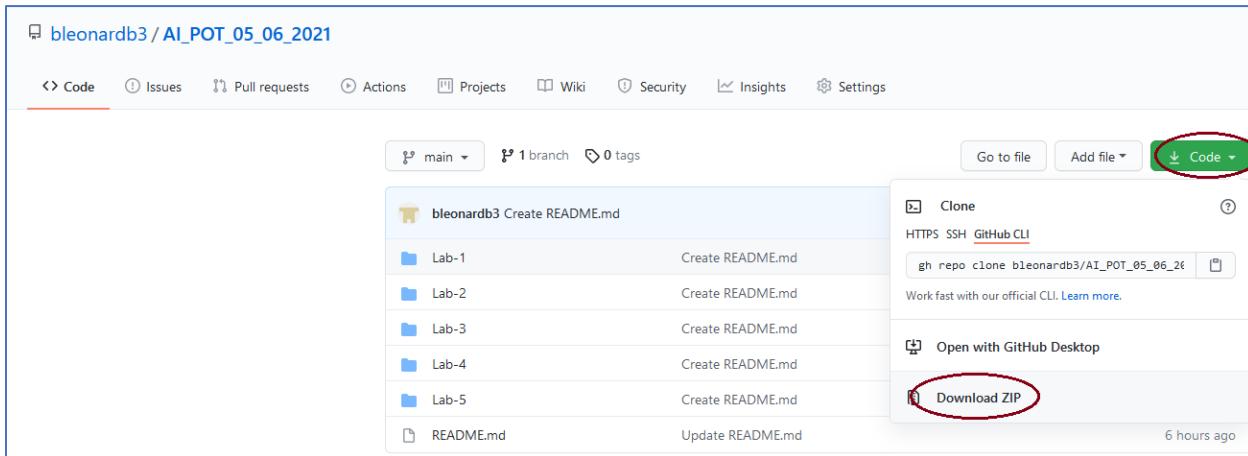
Prerequisites

1. Make sure you have an IBM Cloud account before starting these labs.

Please make sure that you have already registered and currently have a working IBM Cloud account before starting these labs. If you do not have an IBM Cloud account, please inform the instructors right away.

2. Download all lab files to Desktop

- Go to the GitHub repository for this workshop:
https://github.com/bleonardb3/AI_POT_05-06-2021
- Click the down arrow on the **Code button** and select **Download ZIP**. This will download all of the folders for this workshop as a zip file to your computer. You will use these folders with the instruction guides to complete each lab.



- Extract the downloaded zip file and save it on your Desktop.

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 also can

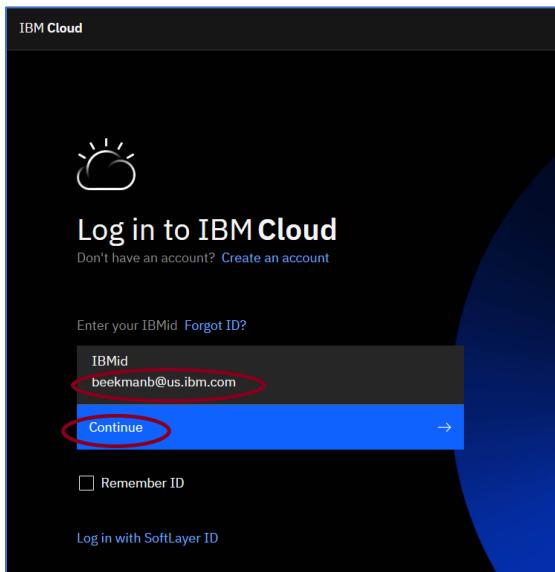
analyze trends and relationships buried in enterprise data. By integrating a machine learning annotator from Watson Knowledge Studio, Watson Discovery can be trained on the language of your domain. Both Watson Knowledge Studio and Watson Discovery can be deployed on any cloud or on-premises environment.

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

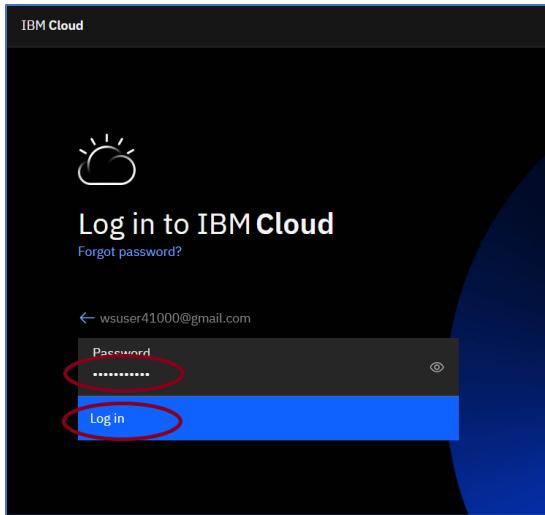
1. Provision an instance of Watson Knowledge Studio
2. Provision an instance of Watson Discovery
3. Create a type system
4. Create a dictionary
5. Upload a corpus of documents
6. Perform manual annotation
7. Train and create a machine learning (ML) annotator
8. Save and deploy the ML annotator to Watson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

Exercise 1: Create a Watson Knowledge Studio Instance

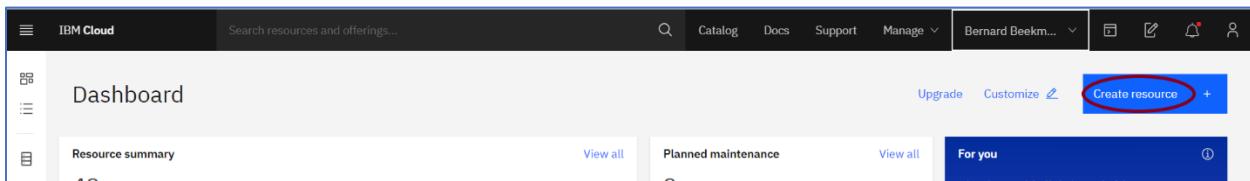
1. Log into your IBM Cloud account by typing **cloud.ibm.com** into the URL address bar of your Firefox or Chrome browser.
2. Enter your **IBMid** and click **Continue**.



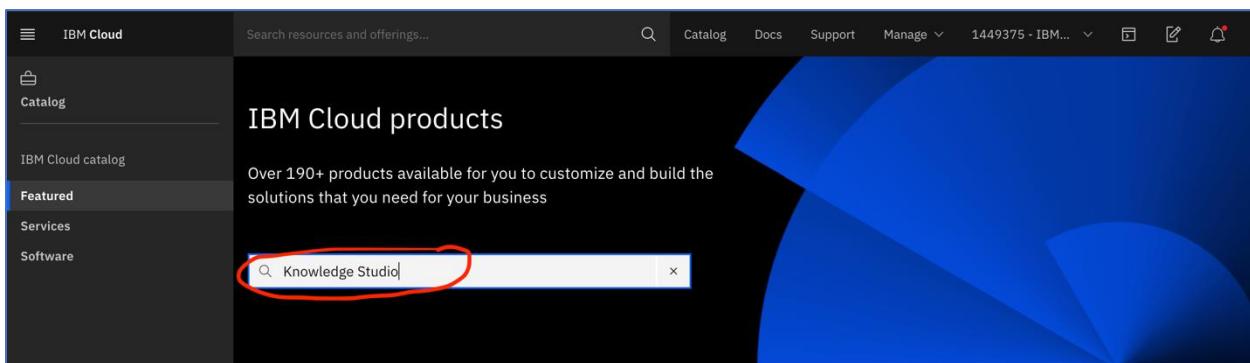
3. Enter your **Password** and click **Log in**.



4. Click **Create Resource**.



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



6. Click on **Knowledge Studio**.

Search results for 'Knowledge Studio' 2 results

Knowledge Studio
 IBM • Services • AI / Machine Learning

Teach Watson the language of your domain.

Lite • Free • IAM-enabled

Domain Name Registration
 IBM • Services • Developer Tools

IBM Cloud offers domain registration services complete with dedicated support staff, knowledgeable customer service, and...

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

| Plan | Features | Pricing |
|-------------|---|--|
| Lite | 1 user 5 GB of storage included Up to 5 workspaces 30 machine learning training sessions per month Deploy models directly to Watson Natural Language Understanding and Watson Discovery services | Free |
| Standard | 5 to 10 users 10 GB to 100 GB of storage Up to 50 workspaces 100 machine learning training sessions per month Deploy models directly to Watson Natural Language Understanding and Watson Discovery services High availability and service level uptime guarantee Export models for use with Watson Explorer | \$100.00 USD/per user per month \$50.00 USD/per 10 GB storage per month |
| Premium | 10 to 100 users | |

Summary

Knowledge Studio **Free**
 Region: Dallas
 Plan: Lite
 Service name: Knowledge Studio-2c
 Resource group: default

Exercise 2: Create a Watson Discovery Instance

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

The screenshot shows the IBM Cloud interface with the search bar containing 'Discovery'. The results are categorized into 'Resource Results' and 'Catalog Results'. In the 'Catalog Results' section, the 'Watson Discovery' service is listed under 'Knowledge Studio' and is highlighted with a red oval. Below the catalog results, there are links to search support cases and documentation.

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

The screenshot shows the creation page for the Watson Discovery service. It displays three plan options: Lite, Advanced, and Premium. The 'Lite' plan is selected and highlighted with a red oval. On the right side, a summary panel shows the service details: Region: Dallas, Plan: Lite, Service name: Discovery-gr, and Resource group: default. At the bottom right, the 'Create' button is highlighted with a red oval.

An instance of the Watson Discovery service will be created. We will link this instance to the machine learning annotator that we create and deploy in this lab. Watson Discovery will use this annotator to perform entity extraction in Lab-2.

Exercise 3: Create a Type System

A type system defines entities 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.

In Knowledge Studio, you can create a type system from scratch or upload an existing type system. To jump-start a workspace, you might want to upload a type system that was created for a similar domain. You can then edit the type system to add or remove entity types or redefine the relationship types.

You must create or upload a type system before you begin any annotation tasks.

Mentions

A mention is any span of text that you consider relevant in your domain data. For example, in a type system about automotive vehicles, occurrences of terms like **airbag**, **Ford Explorer**, and **child restraint system** might be relevant mentions.

Entity Types

An entity type is how you categorize a real-world thing. An entity mention is an example of a thing of that type. For example, the mention President Obama can be annotated as a PERSON entity type. The mention IBM can be annotated as an ORGANIZATION entity type. Entities are often nouns, but can also be verbs, as long as the verb is important to capture for the purposes of the application that will use the type system. For example, EVENT_CRASH might be a valid entity type for a type system about automotive vehicles, so that the word hit in the sentence, The car hit the barrier. can be annotated.

The goal of your annotation workspace is to annotate each mention in a document with the type of thing that it is. After a mention is classified by entity type, the labeled span of text is referred to as an entity.

A best practice is to keep the entity type names sort and representative, so human annotators can remember them easily. In addition, try to define enough entity types to capture the key concepts that you want to annotate, but not so many entity types that it becomes cumbersome for human annotators to apply the labels accurately.

Relationship Types

A relation type defines a binary, ordered relationship between two entities. For a relation mention to exist, text must explicitly define the relation and bind mentions of the two entities together and must do so within a single sentence. For example, the sentence **Mary** works for **IBM** is textual evidence of the **employedBy** relation type.

For some relation types, the order of entity mentions matters. For example, the **employedBy** relation type allows the entity type PERSON or PEOPLE as the first mention in the relationship, and ORGANIZATION or GPE as the second mention, but not the other way around. Mary **employedBy** IBM is a valid relationship. IBM **employedBy** Mary is not. For some relation types, such as **spouseOf**, **colleague**, or **sibling**, order does not matter. When you define a relation type where order is not important, a best practice is to add information to the annotation guidelines to regularize how the relation type is used. A convention for noting such symmetrical relations is to say that the entity mention that occurs first in the text should be the first one in the relation.

Steps to create the 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 image contains two screenshots of the IBM Cloud interface. The top screenshot shows the 'Discovery-x2' resource details page. It features a 'Manage' sidebar with links for 'Getting started', 'Service credentials', 'Plan', and 'Connections'. Below the sidebar is a central area with a 'Start by launching the tool' button, a 'Launch Watson Discovery' button, and links for 'Getting started tutorial' and 'API reference'. To the right, there's a 'Plan' section showing 'Lite' and a 'Upgrade' button. The top navigation bar includes 'IBM Cloud', a search bar, and various account-related links. The bottom screenshot shows the main navigation menu with 'Dashboard' and 'Resource List' highlighted. Other menu items include 'Classic Infrastructure', 'Cloud Foundry', 'Functions', 'Kubernetes', 'OpenShift', 'VMware', and 'VPC Infrastructure'. The 'Resource List' item is circled in red.

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).

The screenshot shows the IBM Cloud Resource list interface. On the left, there's a sidebar with icons for VPC infrastructure, Clusters, Cloud Foundry apps, and Cloud Foundry services. Under 'Cloud Foundry services', 'Knowledge Studio-2c' is listed. A red circle highlights this entry. The main table lists various services with columns for Name, Group, Location, Offering, Status, and Tags.

| Name | Group | Location | Offering | Status | Tags |
|----------------------------------|---------|----------|-------------------------------|--------|------|
| AI-OpenScale | default | Dallas | Watson OpenScale | Active | |
| App ID-app | default | Dallas | App ID | Active | |
| Compare-n-Comply | default | Dallas | Compare and Comply | Active | |
| Continuous-Delivery | default | Dallas | Continuous Delivery | Active | |
| Db2-app | default | Dallas | Db2 | Active | |
| IBM Cognos Dashboard Embedded-ix | default | Dallas | IBM Cognos Dashboard Embedded | Active | |
| Knowledge Catalog-va | default | Dallas | Knowledge Catalog | Active | |
| Knowledge Studio-2c | default | Dallas | Knowledge Studio | Active | |

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

The screenshot shows the resource details page for 'Knowledge Studio-2c'. It includes sections for 'Manage' (Getting started, Plan), 'Start by launching the tool' (Launch Watson Knowledge Studio, Getting started tutorial), and 'Plan' (Lite, Upgrade). The 'Launch Watson Knowledge Studio' button is highlighted with a red circle.

4. Select Create entities and relations workspace.

The screenshot shows the 'Create a workspace' interface for IBM Watson Knowledge Studio. It features two main options: 'Identify custom entities and relations in your data' (Create entities and relations workspace) and 'Extract entities with advanced rules' (Create advanced rules workspace). The 'Create entities and relations workspace' button is highlighted with a red oval.

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

Create Workspace

Workspace name

Language of documents

English

+ Add Workspace Description

► Advanced Options

Cancel **Create**

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 file downloaded from the GitHub repository.
- On the Entity Types screen, click **Upload**.

IBM Watson Knowledge Studio

Entity Types

Entity Types 0

Add Entity Type Upload

Upload an existing type system.

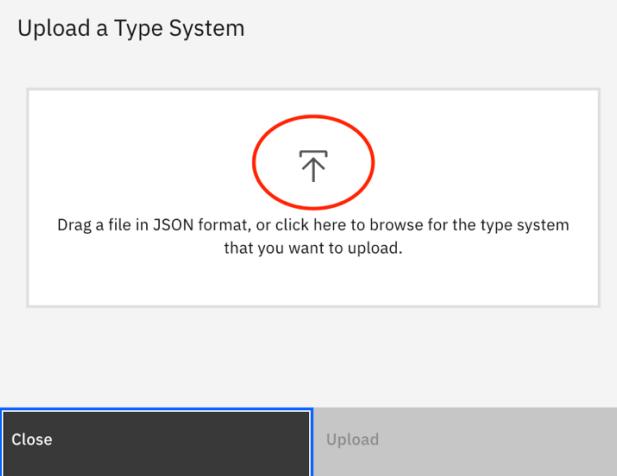
Download Types

Enter text to filter

Entity Type Name Roles Subtypes Action

No items

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



Favorites

- Recents
- Desktop
- Downloads
- Documents
- medblock-mas...
- Applications
- Creative Cloud...

iCloud

- iCloud Drive

Locations

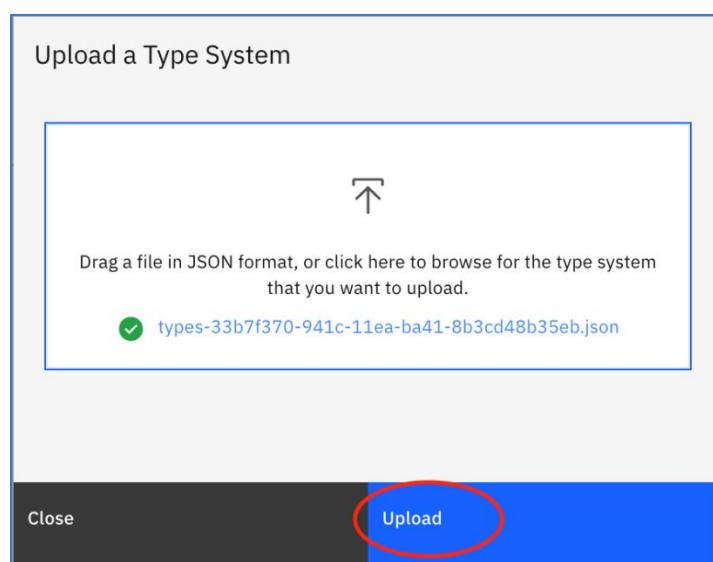
- Docker
- Remote Disc
- Network

Media

- Music
- Photos

| 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_15896932040808.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_1589693188087.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_1589693163792.csv | Today at 1:26 AM | 169 bytes | CSV Document |
| Senior_Citizen_1589693168519.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_1589693134260.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 were most affected by the COVID-19 pandemic.

| 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 |

Exercise 4: Create a Dictionary

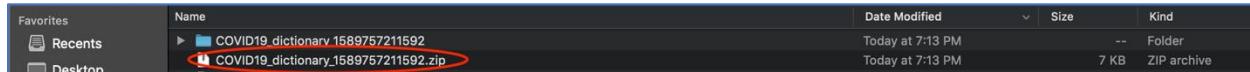
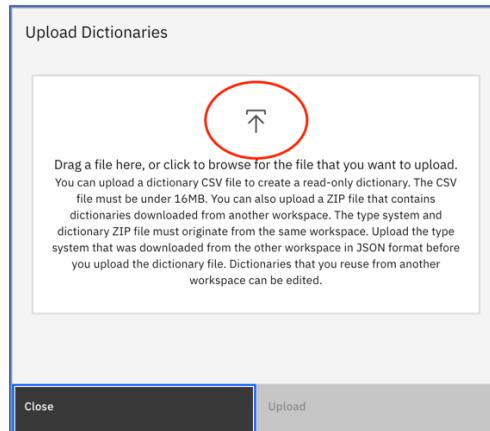
To help with manual annotation (which we will tackle in the next exercise), we will create a dictionary for each of the entity types in our type system. A dictionary is a list of words or phrases that are equivalent for information-extraction purposes, meaning that they are interchangeable for the purposes of identifying entity and relation mentions. Each dictionary will contain a list of terms and key phrases pertaining to each entity type. Dictionaries help the Knowledge Studio machine learning models to understand the language of the domain. You can create dictionaries in Knowledge Studio by manually adding individual entries. Knowledge Studio also supports the ability to upload several types of dictionary files. We will use this capability to upload dictionary files for all 20 entity types in this exercise.

- Under Assets, click **Dictionaries**.

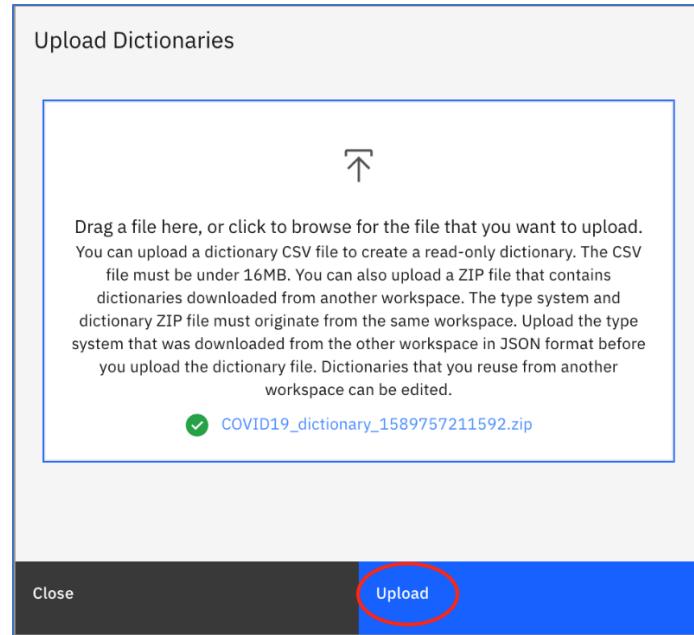
2. On the Dictionaries page, we can upload the zip file containing dictionaries for all our entity types. Click on the **vertical dots icon** and select **Upload Dictionary**.

The screenshot shows the 'Dictionaries' page in Watson Studio. The left sidebar has a tree view with 'Assets' expanded, showing 'Documents', 'Entity Types', 'Relation Types', 'Dictionaries' (which is selected and highlighted in grey), 'Rule-based Model', and 'Machine Learning'. The main area is titled 'Dictionaries' and contains a 'Create Dictionary' button and a text box with instructions: 'Create an empty dictionary. You can add terms to it or upload a CSV file of dictionary entries.' To the right of this is a context menu with three options: 'Upload a CSV file that contains dictionary terms, or a ZIP file that contains dictionaries downloaded from another workspace.', 'Upload Dictionary' (which is circled in red), and 'Download Dictionaries'.

3. Click on the **Upload icon** and select the **COVID19_dictionary_1589757211592.zip** file.



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.

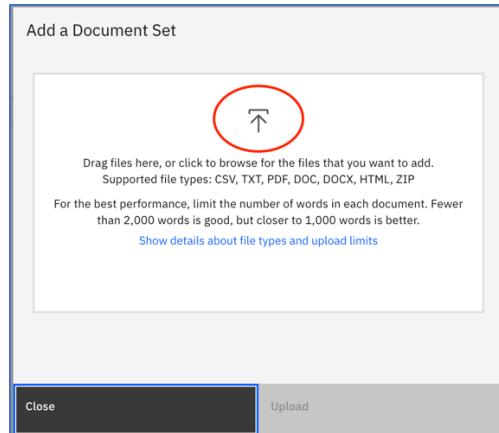
1. Under Assets, select **Documents**.

The screenshot shows the 'Dictionaries' page in IBM Watson Knowledge Studio. On the left, there's a sidebar with 'Assets' expanded, showing 'Documents' (circled in red), 'Entity Types', 'Relation Types', and 'Dictionaries' (selected). Below that are 'Rule-based Model' and 'Machine Learning Model'. The main area is titled 'Dictionaries' and shows a list of dictionaries: 'Unemployed' (9 entries), 'No_Health_Insura...' (5 entries), 'Hourly_Wage_Emp...' (9 entries), and 'Medically_Insured' (6 entries). To the right, there's a section for 'Unemployed' with 'Language: English | 9 entries', an 'Entity type:' dropdown set to 'None', and buttons for 'Add Entry' and 'Upload'.

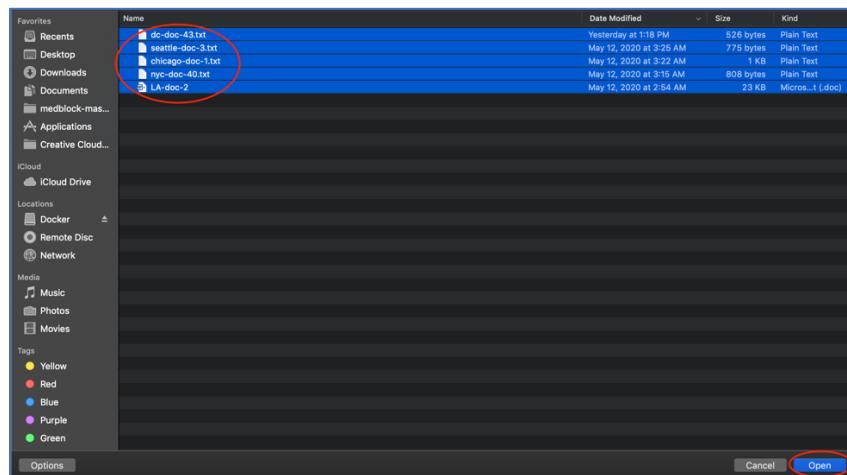
2. Click **Upload Document Sets**.

The screenshot shows the 'Documents' page in IBM Watson Knowledge Studio. The sidebar has 'Assets' expanded, with 'Documents' selected (circled in red) and 'Entity Types', 'Relation Types', 'Dictionaries', and 'Rule-based Model' listed below. The main area shows 'Document Sets (1)' and 'Documents (All, 0)'. A large button labeled 'Upload Document Sets' is highlighted with a red oval. A tooltip below it says 'Add documents to the corpus'.

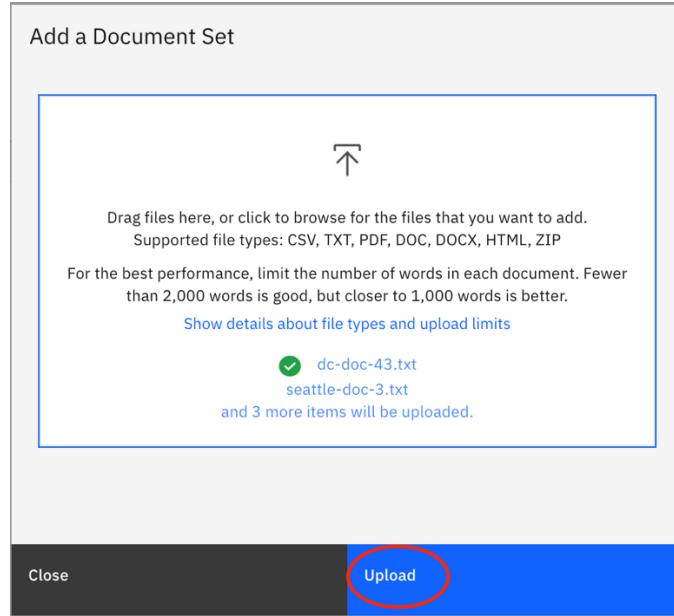
3. Click on the **Upload icon** and double-click on the **SampleDocs** folder.



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 Chicago-doc-1.txt_set to which we will apply a dictionary pre-annotator as well as manually annotate in the next exercise.

| Name | Documents |
|---------------------------------------|-----------|
| All | 5 |
| chicago-doc-1.txt_set | 5 |

Exercise 6: Perform Manual Annotation

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**.

2. Click Run Pre-annotators.

| 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**.

4. Under Select document sets, click the **checkbox** next to chicago-doc-1.txt_set and click **Run**.

Run Pre-annotators

Select document sets

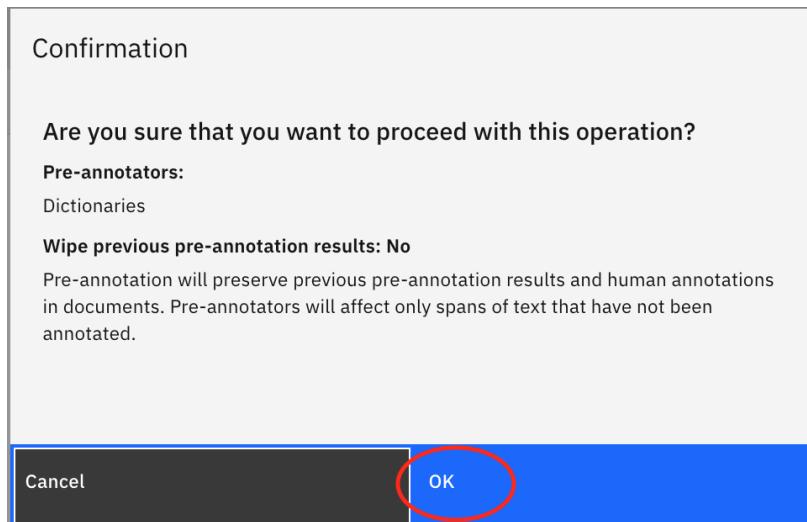
Check if you want to remove previous pre-annotation results from documents before running the pre-annotators. If not checked, all previous annotations are preserved.
 * Annotations made by humans outside of the pre-annotation process remain even if you check the wipe option.

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"/> chicago-doc-1.txt_set | 5 | 0 | 0 |

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**.

Machine Learning Model

- Pre-annotation
- Annotations** (highlighted with a red circle)
- Performance
- Versions

| 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 **Annotate** link on the chicago-doc-1.txt_set row.

The screenshot shows the Annotations interface. The 'Ground Truth' tab is selected. A note says: 'Create ground truth by annotating your documents. Click **Annotate** under the Action column to start annotating a document set.' Another note says: 'To create and manage annotation tasks for a team, go to [Annotation Tasks](#). Annotations added to annotation sets are not considered ground truth until the annotation sets are submitted and accepted by a designated reviewer.' A third note says: 'Note: Annotations added to documents from the Ground Truth tab will override annotations added in annotation tasks completed on the Annotation Tasks page. Annotation tasks that you complete in the future will override annotations added to documents from the Ground Truth tab.' Below is a table:

| Document Set | Last Modified | Documents (Annotated/Total) | Actions |
|-----------------------|---------------|-----------------------------|--------------------------|
| All | - | 3 / 5 | Annotate |
| chicago-doc-1.txt_set | 04/26/2021 | 3 / 5 | Annotate |

8. On the Select Document screen, click on chicago-doc-1.txt.

The screenshot shows the 'Select Document' screen. It says 'Document Set: chicago-doc-1.txt_set' and has a 'Close' button. Below is a table of documents:

| Document Name | Status | Last Modified |
|-------------------|--------|-------------------------|
| chicago-doc-1.txt | | Apr 26, 2021 9:29:17 PM |
| dc-doc-43.txt | | Apr 26, 2021 9:29:17 PM |
| LA-doc-2.doc | | Apr 26, 2021 9:29:17 PM |
| nyc-doc-40.txt | | Apr 26, 2021 9:29:17 PM |
| seattle-doc-3.txt | | Apr 26, 2021 9:29:17 PM |

Showing 1-5 of 5

Items per page: 50 ▾ 1-5 of 5 items 1 ▾ of 1 pages ◀ ▶

9. To annotate the chicago-doc-1.txt, we will skim through the post and find any mention of the custom entity types. Note that several mentions have already been pre-annotated. The annotation color matches the entity type on the right. When we find a word or phrase that we want to annotate, we click on the first word of the phrase and then the last word of the phrase, and then click on the corresponding entity type on the right. Annotate the chicago-doc-1.txt post and compare your result with the annotated post below.

[Back to Annotations](#) | [Open document list](#)

Mention

[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 | |

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" }

```

10. The **property** pre-annotation is not accurate in this case. Select the property annotation and click the delete button on the keyboard to remove this annotation. Click **Save** and then **Open document list** to annotate the next document.

[Back to Annotations](#) | [Open document list](#)

Mention

[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 | |

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" }

```

11. On the **Select Document** panel, click on **dc-doc-43.txt**.

Select Document

Document Set: chicago-doc-1.txt_set

Showing 1-5 of 5

| Document Name | Status | Last Modified |
|-------------------|--------|-------------------------|
| dc-doc-43.txt | | Apr 26, 2021 9:29:17 PM |
| LA-doc-2.doc | | Apr 26, 2021 9:29:17 PM |
| nyc-doc-40.txt | | Apr 26, 2021 9:29:17 PM |
| seattle-doc-3.txt | | Apr 26, 2021 9:29:17 PM |
| chicago-doc-1.txt | | Apr 27, 2021 1:53:42 PM |

Items per page: 50 ▾ 1-5 of 5 items

1 ▾ of 1 pages ◀ ▶

12. 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**.”

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

Back to Annotations | Open document list

View Details ⌂ Replace ⌂ Concordance ⌂ Attribute View ⌂ Save ⌂

Alphab... 14pt 1

Mention

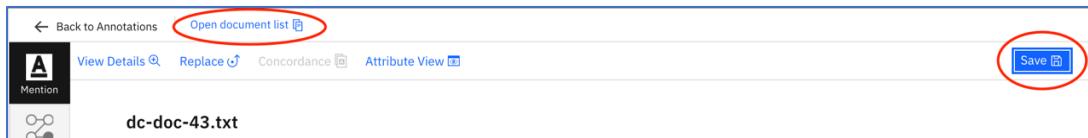
dc-doc-43.txt

1 { "title": "UDC incoming student", "text": "Hi, I am wondering if there are any students out there who will be attending any graduate programs at UDC this fall."}
2 I will be in the speech program at UDC beginning in the fall and think it would be nice to meet people that will also be there.
3 I am from Maryland and attended school in Baltimore.
4 \nI am also wondering if there are any off campus student housing options that anyone knows about?
5 ", "subreddit": "washingtongdc", "created": "2020-04-23T22:24:27.000Z" }

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 | |

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



14. On the **Select Document** panel, click on **LA-doc-2.doc**

A screenshot of a 'Select Document' panel. It shows a table of documents with columns for 'Document Name', 'Status', and 'Last Modified'. The table contains five rows: 'LA-doc-2.doc' (circled in red), 'nyc-doc-40.txt', 'seattle-doc-3.txt', 'chicago-doc-1.txt', and 'dc-doc-43.txt'. At the bottom, there are pagination controls: 'Items per page: 50', '1-5 of 5 items', '1 of 1 pages', and navigation arrows.

15. Annotate as shown below, and then click **Save** and **Open document list**.

A screenshot of the annotation interface for the document 'LA-doc-2.doc'. The left side shows the document content with some words highlighted in green. The right side features a sidebar with a tree view of entity types under 'Entity'. The 'Car_Owner' type is currently selected and highlighted in grey. Other types listed include Apartment_Renter, 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, and Senior_Citizen. The 'Save' button and the 'Open document list' link at the top are both circled in red.

16. On the **Select Document** panel, click on **nyc-doc-40.txt**.

Select Document

Document Set: chicago-doc-1.txt_set

Close

Showing 1-5 of 5

| Document Name | Status | Last Modified |
|-------------------|--------|-------------------------|
| nyc-doc-40.txt | | Apr 26, 2021 9:29:17 PM |
| seattle-doc-3.txt | | Apr 26, 2021 9:29:17 PM |
| chicago-doc-1.txt | | Apr 27, 2021 1:53:42 PM |
| dc-doc-43.txt | | Apr 27, 2021 2:30:58 PM |
| LA-doc-2.doc | | Apr 27, 2021 2:47:27 PM |

Items per page: 50 ▾ 1-5 of 5 items

1 ▾ of 1 pages ◀ ▶

17. Annotate the nyc-doc-40.txt post and compare your result with the annotated post below.
 Click **Save** to save your annotation and click **Open document list** when you're done
 annotating this post to move to the next document.

[Back to Annotations](#) | [Open document list](#)

[View Details](#) [Replace](#) [Concordance](#) [Attribute View](#)

[Save](#)

nyc-doc-40.txt

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 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**"}

Entity

| Type | Subtype | Role |
|------------------------|---------|------|
| 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 | | |

18. On the **Select Document** panel, click on **seattle-doc-3.txt**

Select Document

Document Set: chicago-doc-1.txt_set

Close

Showing 1-5 of 5

| Document Name | Status | Last Modified |
|-------------------|--------|-------------------------|
| seattle-doc-3.txt | | Apr 26, 2021 9:29:17 PM |
| chicago-doc-1.txt | | Apr 27, 2021 1:53:42 PM |
| dc-doc-43.txt | | Apr 27, 2021 2:30:58 PM |
| LA-doc-2.doc | | Apr 27, 2021 2:47:27 PM |
| nyc-doc-40.txt | | Apr 27, 2021 2:56:59 PM |

Items per page: 50 ▾ 1-5 of 5 items 1 ▾ of 1 pages ◀ ▶

19. 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.

[← Back to Annotations](#) [Open document list](#) [Save](#)

Mention

Relation

Coreference

seattle-doc-3.txt

```

1 {   "title": "Etiquette Question - Sheltering in Place, Apartments & Music",   "text": "Just getting a
read of general feelings on this.

2 Given that many of us are staying at home/working from home during the pandemic, there's a lot
more opportunity for grating on each other unintentionally.

3 I live in an apartment complex with fairly thin walls.

4 At what point, in your personal opinion, is it reasonable to start playing music, watch action movies,
or other entertainment activities that involve a degree of noise?

5 Personally I don't turn on music my neighbors may hear before 9:00am (I start work early each
morning), but is this a good rule of thumb?

6 Too early?

7 Curious to hear folks' thoughts.

8 ",   " subreddit": "Seattle",   "created": "2020-04-25T01:45:48.000Z" }

```

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.

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

| Document Name | Status | Last Modified |
|-------------------|--------|-------------------------|
| chicago-doc-1.txt | | Apr 27, 2021 1:53:42 PM |
| dc-doc-43.txt | | Apr 27, 2021 2:30:58 PM |
| LA-doc-2.doc | | Apr 27, 2021 2:47:27 PM |
| nyc-doc-40.txt | | Apr 27, 2021 2:56:59 PM |
| seattle-doc-3.txt | | Apr 28, 2021 2:03:47 PM |

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

As stated above, we will require a much larger set of documents 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 to 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.

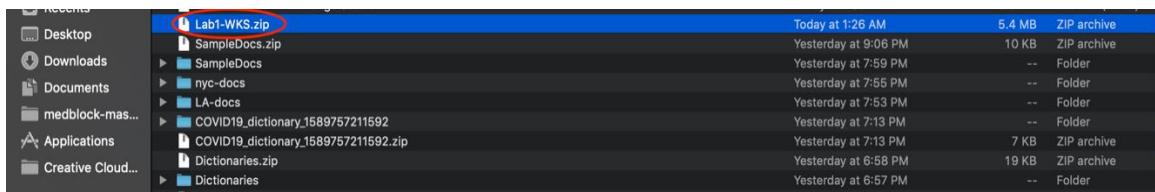
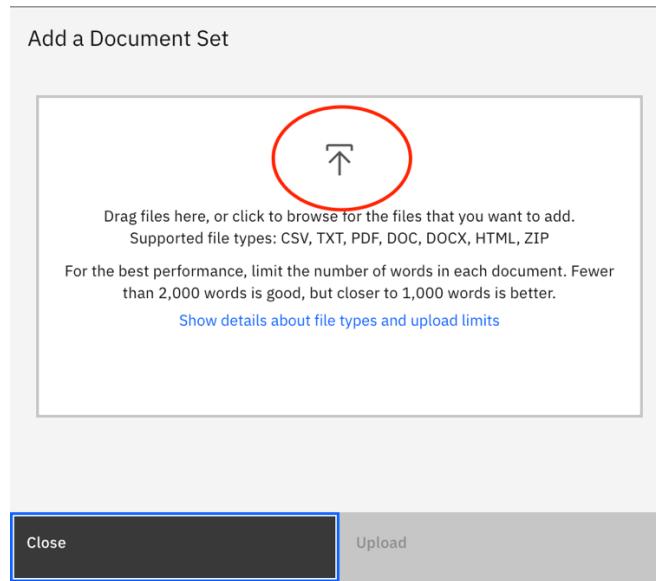
1. Under Assets, click **Documents**.

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

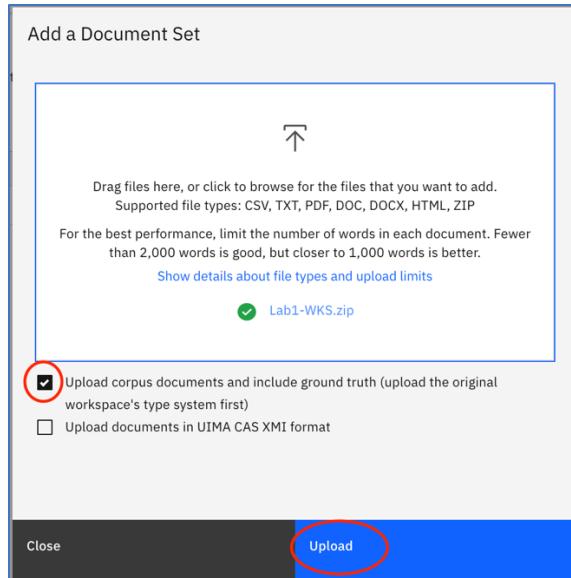
2. On the Documents screen, click **Upload Document Sets**.

The screenshot shows the 'Documents' section of the IBM Watson Knowledge Studio. On the left, there's a sidebar with categories like Assets, Documents (which is selected), Entity Types, Relation Types, Dictionaries, Rule-based Model, and Machine Learning Model. The main area shows 'Document Sets (2)' and 'Documents (All, 5)'. A red circle highlights the 'Upload Document Sets' button at the top left of the document list area. Below it, there's a message: 'To begin annotating documents, go to [Annotations](#) page.' To the right is a 'Download Document Sets' button.

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



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

| 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 | 36 | 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 |

5. Under Machine Learning Model, click on Performance.

The screenshot shows the 'Documents' screen in IBM Watson Knowledge Studio. The left sidebar is open, showing categories such as Entity Types, Relation Types, Dictionaries, Rule-based Model, Machine Learning Model, Pre-annotation, Annotations, and Performance. The 'Performance' category is highlighted and circled in red.

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

The screenshot shows the 'Performance' screen in IBM Watson Knowledge Studio. The left sidebar is open, showing categories such as Entity Types, Relation Types, Dictionaries, Rule-based Model, Machine Learning Model, Pre-annotation, Annotations, and Performance. The 'Train and evaluate' button is highlighted and circled in red.

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

| | |
|--------------------------------------|---|
| <input type="button" value="Train"/> | <input style="outline: 2px solid red; border-radius: 10px; padding: 2px 10px; border: none; color: inherit; background-color: inherit; font-size: inherit; font-weight: inherit; font-family: inherit; font-style: inherit; text-decoration: none; border-radius: 10px; border: 1px solid #ccc; padding: 5px; margin: 5px 10px;" type="button" value="Train & Evaluate"/> |
|--------------------------------------|---|

| 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:

Performance

COVID-19 Vulnerability
Language of documents English

Number of documents per set

| | |
|-----|--------------|
| 312 | Training Set |
| 36 | Test Set |
| 20 | Blind Set |

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

Last trained on: Apr 28, 2021 2:43:49 PM
Last evaluated on: Apr 28, 2021 2:45:50 PM

Low performance? Click here to train.

Document set evaluation ⓘ

Model over time

View Log

Mention
0.65 Precision: 0.69 Recall: 0.62

Relation
-- Precision: -- Recall: --

Coreference
-- Precision: -- Recall: --

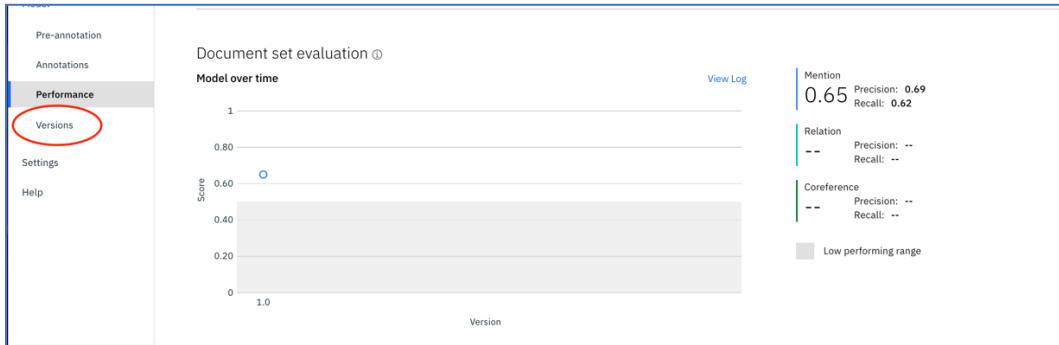
Low performing range

Exercise 8: Save and Deploy the ML Annotator to Discovery

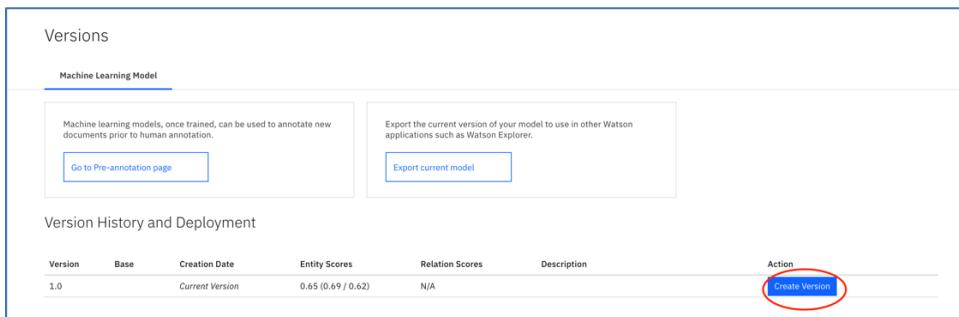
Now that we have a machine learning annotator, we can use to automatically perform entity extraction inside of Watson Discovery. The automated entity extraction of social media posts for all 5 cities will get us closer to determining the social vulnerability index of each city.

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

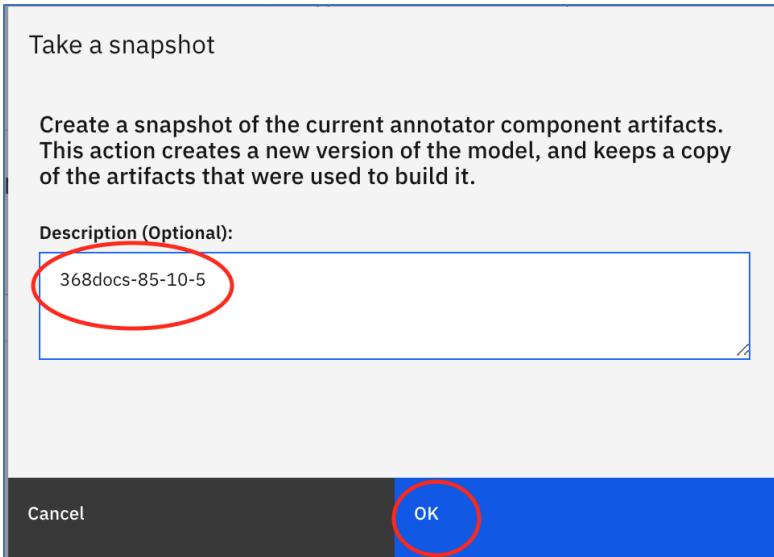
1. Under Machine Learning Model, click on **Versions**.



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



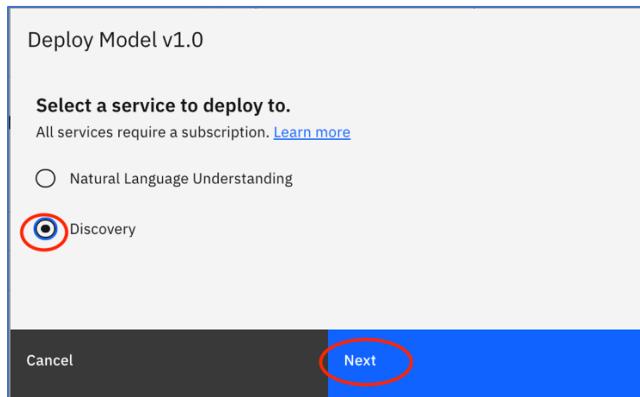
3. Type **368docs-85-10-5** (to distinguish this as an entity model using 368 docs with an 85-10-5 split) under Description and click **OK**.



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 | 05/18/2020 | 0.65 (0.69 / 0.62) | N/A | | <button>Create Version</button> |
| 1.0 | 05/18/2020 | 05/18/2020 | 0.65 (0.69 / 0.62) | N/A | 368docs-85-10-5 | <button>Promote</button> <button>Delete</button> Deploy |

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**.

Deploy Model v1.0

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

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

Region
Dallas

Resource group
default

Service name
Discovery-kf

Cancel Deploy

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

Deployment Started.

Deploying to Discovery
It might take a few minutes for publishing and deployment to complete, and for this model to be available to your applications.
You can view your deployed models, withdraw a model from deployment, or deploy a newer version.

Model ID: 63d1efc3-6d00-4273-a034-7034a996c8f0

You can [view documentation](#) to learn how to implement the deployed model into your application.

OK

8. Click on the right arrow ➡ adjacent to **Deployed Models (1)**. You should 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. | Export the current version of your model to use in other Watson applications such as Watson Explorer. |
| Go to Pre-annotation page | Export current model |

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 | | Create Version |
| 1.0 | | 05/18/2020 | 0.65 (0.69 / 0.62) | N/A | 368docs-85-10-5 | Promote Delete Deploy |
| ▼ Deployed Models (1) | | | | | | |
| Model ID: 63d1efc3-6d00-4273-a034-7034a996c8f0 | | | | Service ID: 03b54347-0aad-4da9-b59a-e1f2df1070cc | | Undeploy Status |

You have completed Lab 1!