

# **IBM Training**

## **Student Exercises**

**Lab-2: Create a knowledge management system and develop a COVID-19 vulnerability index**

**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>Prerequisites .....</b>	<b>3</b>
<b>Introduction .....</b>	<b>5</b>
<b>Objectives.....</b>	<b>6</b>
<b>Exercise 1: Create a Discovery collection .....</b>	<b>6</b>
<b>Exercise 2: Upload the documents .....</b>	<b>9</b>
<b>Exercise 3: Add the entity model from Knowledge Studio.....</b>	<b>12</b>
<b>Exercise 4: Perform Custom Entity Extraction .....</b>	<b>16</b>
<b>Exercise 5: Retrieve the analyzed files using the Discovery API .....</b>	<b>18</b>
<b>Exercise 6: Calculate the COVID-19 vulnerability index.....</b>	<b>28</b>
<b>Exercise 7: Create a collection for a COVID-19 publication.....</b>	<b>31</b>
<b>Exercise 8: Perform Smart Document Understanding .....</b>	<b>35</b>
<b>Exercise 9: Create and run Natural Language Queries.....</b>	<b>44</b>
<b>Exercise 10: Improve accuracy with Relevancy Training .....</b>	<b>51</b>

# Prerequisites

## 1. Save all lab files to Desktop

Before starting Lab 2, please download all files from the Lab 2 Folder on GitHub ([https://github.com/bleonardb3/AI\\_POT\\_06-11-2020/tree/master/Lab-2](https://github.com/bleonardb3/AI_POT_06-11-2020/tree/master/Lab-2)) and save them to a new folder on the Desktop of your computer.

- Create a new folder on your Desktop called Watson-Lab2
- Transfer all Lab 2 files and folders to the Watson-Lab2 folder

## 2. Download Node.js

In order to call the Discovery API to retrieve the analyzed files and later use them to calculate the social vulnerability index (SVI) for each city, you will need to make sure that you have Node.js installed on your computer.

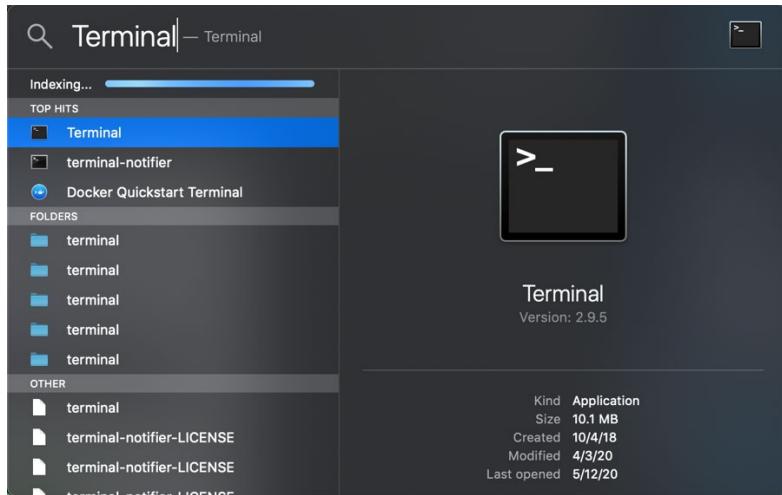
If you don't have Node.js already installed, please go to <https://nodejs.org/en/> and download the most current version of Node for your operating system.

The screenshot shows the official Node.js website. At the top, there is a dark navigation bar with links for HOME, ABOUT, DOWNLOADS, DOCS, GET INVOLVED, SECURITY, and NEWS. Below the navigation bar, a green banner states "Node.js® is a JavaScript runtime built on Chrome's V8 JavaScript engine." In the center, there is a heading "Download for macOS (x64)". Below it, two download buttons are visible: one for "12.16.3 LTS" (labeled "Recommended For Most Users") and another for "14.3.0 Current" (labeled "Latest Features"). A red oval highlights the "14.3.0 Current" button. At the bottom of the page, there are links for "Other Downloads | Changelog | API Docs" and "Other Downloads | Changelog | API Docs".

After downloading the latest version, double-click the file to install Node.js on your computer.

Once you have successfully installed it, please verify that install was successful by running a command from a command line interface.

- Open up a Terminal or command line window and type node -v



```
[Asads-MacBook-Pro:~ asadmahmood$ node -v  
v8.12.0  
Asads-MacBook-Pro:~ asadmahmood$ ]
```

If you are able to see a version number, then Node was successfully installed.

### 3. Install latest version of npm

Now that you have Node.js installed on your computer, you should automatically have npm on your system. We will be using npm (Node Package Manager) in order to run Node.js programs from the command line.

Let's make sure that we have the latest version of npm by running the following command from the Terminal or command line prompt: **sudo npm install -g npm**

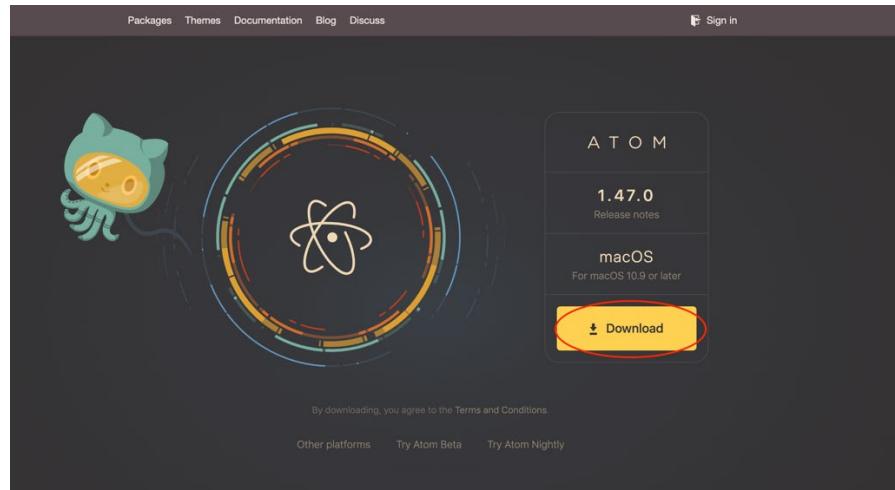
After entering the password for your computer, you'll get the latest version of npm.

```
[Asads-MacBook-Pro:~ asadmahmood$ node -v  
v8.12.0  
[Asads-MacBook-Pro:~ asadmahmood$ sudo npm install -g npm  
/usr/local/bin/npx -> /usr/local/lib/node_modules/npm/bin/npx-cli.js  
/usr/local/bin/npm -> /usr/local/lib/node_modules/npm/bin/npm-cli.js  
+ npm@6.14.5  
updated 1 package in 6.313s  
Asads-MacBook-Pro:~ asadmahmood$ ]
```

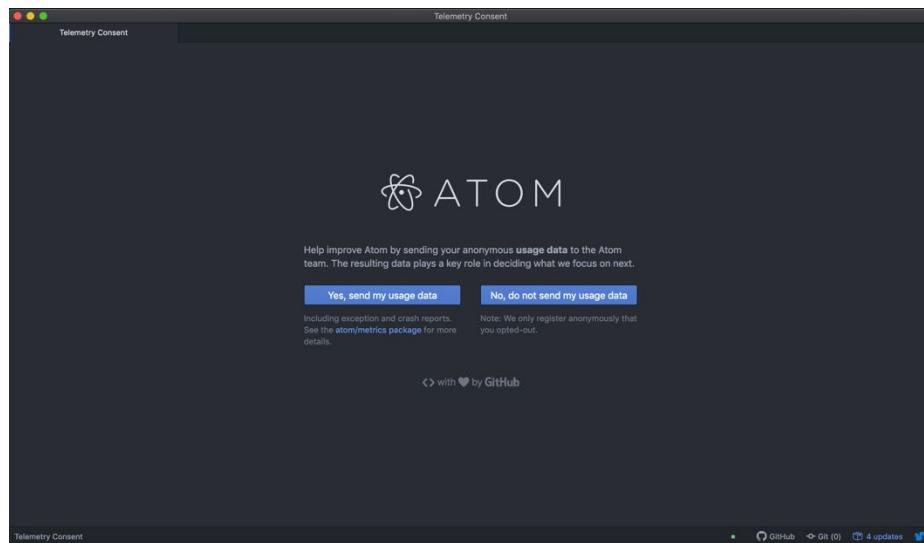
### 4. Download the Atom code editor

If you don't already have a code editor on your computer, let's download and install the Atom code editor, which we will use to edit a few Node.js programs.

- Go to atom.io and download the latest version for your operating system



- Double-click the installer file to complete installation of the Atom code editor.
- Once Atom is successfully installed, double-click the Atom icon to launch the code editor



You can use the Atom editor to open folders with specific program files by clicking **File->Add Project Folder**. We will access the files in our Watson-Lab2 folder using the Atom code editor later in the lab.

## Introduction

In this lab you will create a knowledge management system (KMS), train the KMS to generate knowledge and analyze information to create a COVID-19 vulnerability index. IBM Watson Discovery will be used to develop and train the KMS.

# Objectives

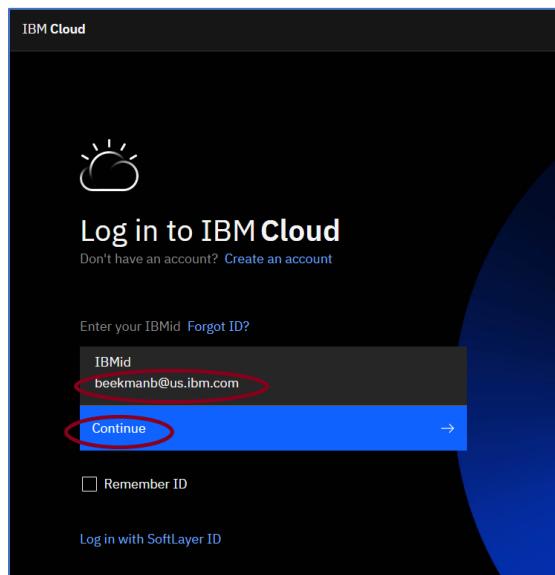
The goal of this lab is to familiarize the user with the Watson Discovery service. Watson Discovery is an enterprise AI search technology that leverages machine learning, including natural language processing (NLP), to retrieve specific answers to your questions and analyze trends and relationships buried in enterprise data. By integrating a machine learning annotator from Watson Knowledge Studio (which we created and deployed in Lab 1), Watson Discovery can be trained on the language of your domain to perform customized NLP. The Watson Discovery service can be deployed on any cloud or on-premises environment.

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

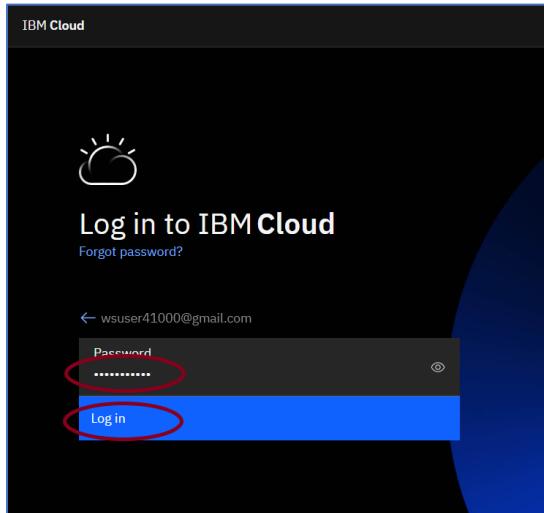
1. Create a Discovery collection
2. Upload the documents
3. Add the entity model from Knowledge Studio
4. Perform custom entity extraction
5. Retrieve the analyzed files using the Discovery API
6. Calculate the COVID-19 vulnerability index
7. Create a collection for a COVID-19 publication
8. Perform Smart Document Understanding
9. Create and run Natural Language Queries
10. Improve accuracy with Relevancy Training

## Exercise 1: Create a Discovery collection

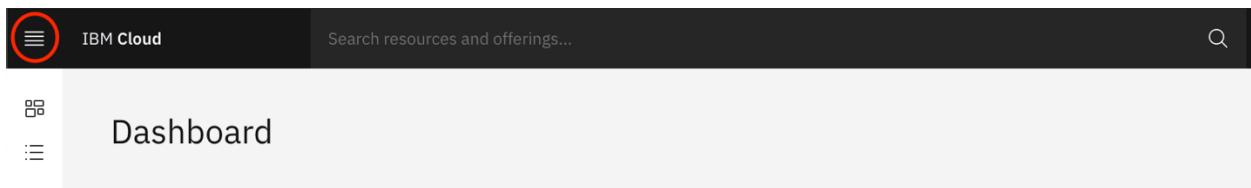
1. Log into your IBM Cloud account by typing **cloud.ibm.com** into the URL address bar of your Firefox or Chrome browser.
2. If you have been logged out, enter your **IBMid** and click **Continue**.



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



4. On the Dashboard screen, click the **hamburger icon**.



5. Select **Resource List** from the drop-down menu.



6. Under services, click on the **name of the Discovery instance** that you created in Lab 1.

A screenshot of the IBM Cloud Resource list. The top navigation bar has "Resource list" selected. The main area shows a table with columns: Name, Group, Location, Offering, Status, and Tags. There is a filter bar above the table. In the "Name" column, there is a row for "Discovery-kf" which is circled in red.

7. On the Manage screen, click on **Show Credentials** and copy the **API key** and **URL** for your Discovery instance. We will be using these credentials when calling the Discovery API later in the lab.

Start by launching the tool

The screenshot shows the 'Credentials' section of the IBM Watson Discovery service. It displays two main fields: 'API key:' and 'URL:'. The 'API key:' field contains the value 'JfXr729jMa\_YBYKGpsxf4awFhfUP-DyBEwpjGDZ5mqcy', and the 'URL:' field contains the value 'https://gateway.watsonplatform.net/discovery/api'. Both fields have a red oval drawn around them. Above the 'API key:' field is a 'Download' button with a downward arrow and a 'Hide credentials' button with a circular arrow icon.

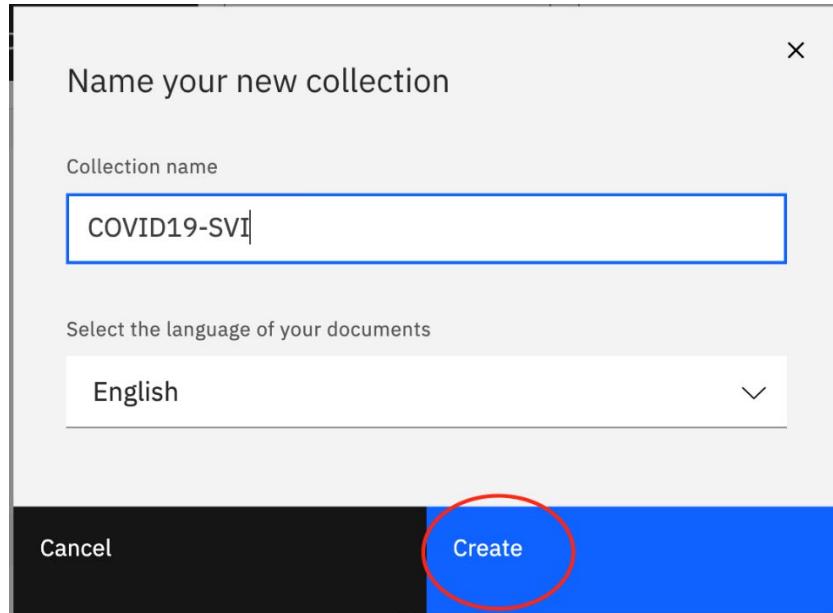
#### 8. Click **Launch Watson Discovery** in order to start your Discovery instance.

The screenshot shows the 'Manage' section of the IBM Watson Discovery service. It features a 'Launch Watson Discovery' button, which is highlighted with a red oval. To the right, there is a 'Plan' section labeled 'Lite' and an 'Upgrade' button. Below the 'Launch Watson Discovery' button is a 'Credentials' section with 'API key:' and 'URL:' fields, both of which are also circled with red ovals. The URL field contains 'https://api.us-south.discovery.watson.cloud.ibm.com/instances/ffe3214a-5b4b-4d1d-8529-7ac4a...'. On the far right, there is a 'FEEDBACK' link.

#### 9. On the Manage data screen, click **Upload your own data** to create a new collection.

The screenshot shows the 'Manage data' screen of the IBM Watson Discovery service. At the top, there is a search bar with 'Create COVID-19 Kit' and a button labeled 'Upload your own data' with a red oval around it. Below the search bar, there is a 'Create a new data collection' button and a 'Connect a data source' button.

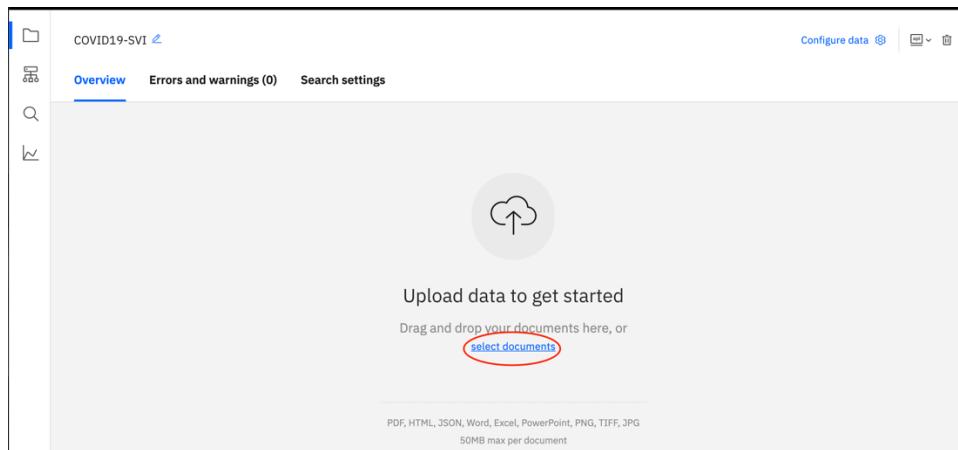
#### 10. Give your collection a name of **COVID19-SVI** and click **Create**.



## Exercise 2: Upload the documents

Now that we have created a collection, we can upload all of our social media posts to this collection. In Watson Discovery, a collection stores all of the relevant documents (preferably in the same file format) and is subsequently used to perform content mining and passage retrieval and query the analyzed document set.

1. Click **select documents**.

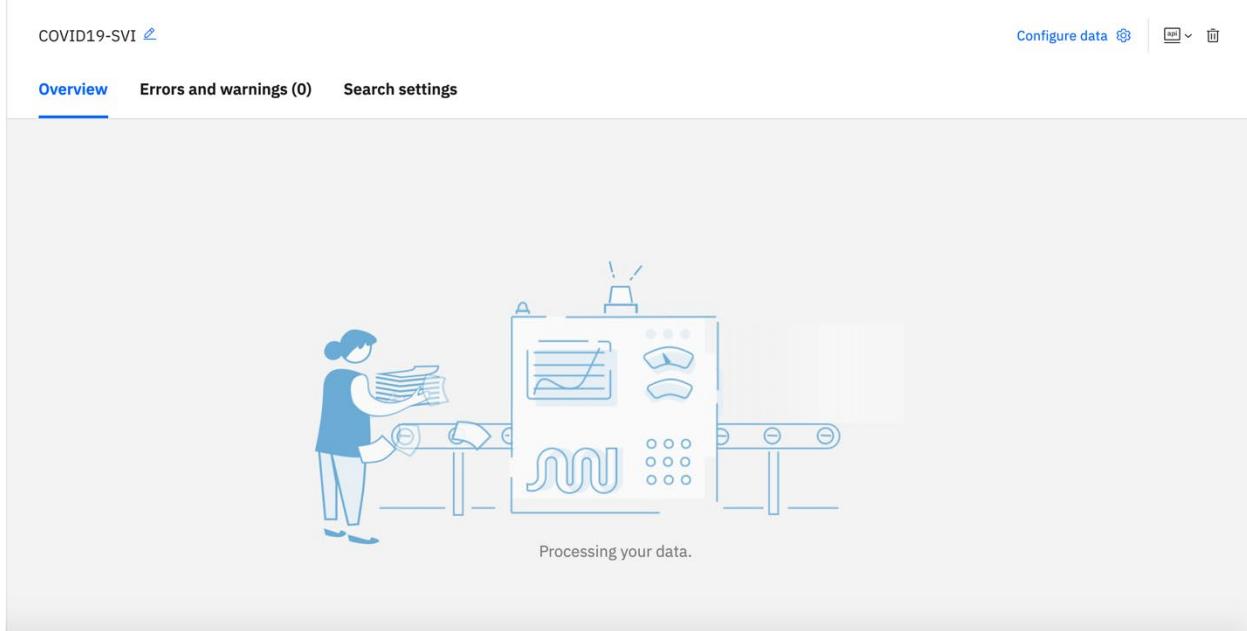


2. Navigate to the Posts folder, **shift select all of the files** in the folder and click **Open**.

Name	Date Modified	Size	Kind
relevant-seattle-292.json	Apr 28, 2020 at 2:54 AM	725 bytes	JSON Document
relevant-seattle-267.json	Apr 28, 2020 at 2:54 AM	633 bytes	JSON Document
relevant-seattle-285.json	Apr 28, 2020 at 2:54 AM	568 bytes	JSON Document
relevant-seattle-221.json	Apr 28, 2020 at 2:54 AM	476 bytes	JSON Document
relevant-seattle-139.json	Apr 28, 2020 at 2:53 AM	675 bytes	JSON Document
relevant-seattle-172.json	Apr 28, 2020 at 2:53 AM	1 KB	JSON Document
relevant-seattle-114.json	Apr 28, 2020 at 2:53 AM	720 bytes	JSON Document
relevant-seattle-199.json	Apr 28, 2020 at 2:53 AM	478 bytes	JSON Document
relevant-seattle-126.json	Apr 28, 2020 at 2:53 AM	472 bytes	JSON Document
relevant-seattle-134.json	Apr 28, 2020 at 2:53 AM	528 bytes	JSON Document
relevant-seattle-175.json	Apr 28, 2020 at 2:53 AM	440 bytes	JSON Document
relevant-seattle-151.json	Apr 28, 2020 at 2:53 AM	783 bytes	JSON Document
relevant-seattle-158.json	Apr 28, 2020 at 2:53 AM	635 bytes	JSON Document
relevant-seattle-137.json	Apr 28, 2020 at 2:53 AM	295 bytes	JSON Document
relevant-seattle-36.json	Apr 28, 2020 at 2:52 AM	610 bytes	JSON Document
relevant-seattle-14.json	Apr 28, 2020 at 2:52 AM	542 bytes	JSON Document
relevant-seattle-13.json	Apr 28, 2020 at 2:52 AM	3 KB	JSON Document
relevant-seattle-12.json	Apr 28, 2020 at 2:52 AM	775 bytes	JSON Document
relevant-seattle-65.json	Apr 28, 2020 at 2:52 AM	2 KB	JSON Document
relevant-seattle-94.json	Apr 28, 2020 at 2:52 AM	896 bytes	JSON Document
relevant-seattle-42.json	Apr 28, 2020 at 2:52 AM	6 KB	JSON Document
relevant-seattle-64.json	Apr 28, 2020 at 2:52 AM	955 bytes	JSON Document
relevant-seattle-5.json	Apr 28, 2020 at 2:52 AM	640 bytes	JSON Document
relevant-seattle-86.json	Apr 28, 2020 at 2:52 AM	790 bytes	JSON Document
relevant-seattle-8.json	Apr 28, 2020 at 2:52 AM	714 bytes	JSON Document
relevant-seattle-32.json	Apr 28, 2020 at 2:52 AM	856 bytes	JSON Document
relevant-seattle-24.json	Apr 28, 2020 at 2:52 AM	653 bytes	JSON Document
relevant-seattle-97.json	Apr 28, 2020 at 2:52 AM	615 bytes	JSON Document
relevant-seattle-44.json	Apr 28, 2020 at 2:52 AM	1 KB	JSON Document
relevant-seattle-25.json	Apr 28, 2020 at 2:52 AM	350 bytes	JSON Document
relevant-seattle-79.json	Apr 28, 2020 at 2:52 AM	924 bytes	JSON Document
relevant-seattle-51.json	Apr 28, 2020 at 2:52 AM	303 bytes	JSON Document

Cancel **Open**

3. It will take approximately 6 minutes to ingest all of the social media posts into the collection. During this ingestion process, Discovery additionally applies out-of-box enrichments such as entity extraction, keyword extraction, concept tagging and sentiment analysis to all of the documents.



4. As soon as all of the documents have been successfully ingested, you should see the following screen:

COVID19-SVI

**Overview** Errors and warnings (0) Search settings

364 documents

0 documents failed View details

Created on 5/18/2020 12:32:41 pm EDT Last updated 5/18/2020 12:32:41 pm EDT

Upload documents

Identified 5 fields from your data

- text
- author\_fullname
- extracted\_metadata
- id
- title

Added 4 enrichments to your data

Entity Extraction

Seattle (36) | DC (21) | \$600 (10) | 2 weeks (10) | \$0 (9)

Sentiment Analysis

positive	28%	neutral	1%	negative	72%
----------	-----	---------	----	----------	-----

Now you're ready to query!

Entities of type **Quantity** which have negative sentiment

Run

Documents that contain English-language films, but not Landlord

Run

Top entities with their average, min, max sentiment score

On this overview screen, you should be able to see the total number of posts in the collection, the number of fields identified per post, the top entities extracted, the overall sentiment of the documents and some sample queries that can be applied to the collection.

We will now configure the dataset to only apply entity extraction with the machine learning annotator from Watson Knowledge Studio. Hopefully, you saved the Model ID at the conclusion of Lab 1.

*If you didn't copy it or misplaced the Model ID number,* this would be a good opportunity to revisit your COVID19-Vulnerability workspace in Watson Knowledge Studio and copy the Model ID underneath Deployed Models on the Versions page.

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 current model

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

## Exercise 3: Add the entity model from Knowledge Studio

1. Click **Configure Data**.

The screenshot shows the 'COVID19-SVI' project overview page. At the top right, there is a 'Configure data' button with a gear icon, which is highlighted with a red circle. Below the header, there are tabs for 'Overview', 'Errors and warnings (0)', and 'Search settings'. The 'Overview' tab is selected. In the center, it displays '364 documents'. To the right, there are sections for '0 documents failed' (with a 'View details' link), 'Created on 5/18/2020 12:32:41 pm EDT', 'Last updated 5/18/2020 12:32:41 pm EDT', and a 'Upload documents' button.

2. On the Configure Data screen, click the **Enrich fields** tab.

Here we can specify the sections of our files that will be subjected to the NLP enrichments (in our case, we will only be selecting entity extraction). Click the drop-down menu next to Add a field to enrich and select **title**.

The screenshot shows the 'Configure data' screen with the 'Enrich fields' tab selected. The title 'Enrich your data with additional Watson insights' is displayed, along with a note to 'Set up rules for which fields you want to apply enrichments to.' A 'Learn more' link is provided. Below this, there is a section titled 'Add a field to enrich' with a 'Choose a field' dropdown containing 'author\_fullname' and 'title'. The 'title' option is highlighted and circled in red. A 'Fields to be enriched' dropdown also contains 'title'. At the bottom, there are four enrichment categories: 'categories', 'concepts', 'entities', and 'sentiment', each with a remove ('X') button. A '+ Add enrichments' button is located at the bottom right.

3. To the title field of each post, we will apply entity extraction. Click **+Add enrichments** on the title row.

Add a field to enrich Choose a field

Fields to be enriched	Enrichments
title	No enrichments applied yet + Add enrichments
text	categories × concepts × entities × sentiment × + Add enrichments

- On the Add Enrichments pop-up screen, click the **Add** button inside of the **Entity Extraction** card.

Add Enrichments

title:	
<b>Keyword Extraction</b> Determines important keywords in this field, ranks them, and optionally detects the sentiment. <a href="#">Learn more</a> <b>Add</b>	<b>Sentiment Analysis</b> Identifies the overall positive or negative sentiment within this field. <a href="#">Learn more</a> <b>Add</b>
<b>Concept Tagging</b> Identifies general concepts that aren't necessarily directly referenced in this field. <a href="#">Learn more</a> <b>Add</b>	<b>Category Classification</b> Classifies this field into a hierarchy of categories that's five levels deep. <a href="#">Learn more</a> <b>Add</b>
<b>Semantic Role Extraction</b> Parses sentences into subject, action, and object form and returns additional semantic information. <a href="#">Learn more</a> <b>Add</b>	<b>Emotion Analysis</b> Analyzes the emotions (anger, disgust, fear, joy, and sadness) in this field. <a href="#">Learn more</a> <b>Add</b>
<b>Entity Extraction</b> Extracts people, companies, organizations, cities, geographic features, and more from this field. You can also create and add custom entity models with Watson Knowledge Studio. <a href="#">Learn more</a> <b>Add</b>	<b>Relation Extraction</b> Recognizes when two entities are related and identifies the type of relation. You can also create and add custom relation models with Watson Knowledge Studio. <a href="#">Learn more</a> <b>Add</b>
Custom Model ID ⓘ enter model ID <input type="text"/> Apply	Custom Model ID ⓘ enter model ID <input type="text"/> Apply

5. Paste the Model ID number for the machine learning annotator from Lab 1 underneath **Custom Model ID** and click **Apply**. Then click the **x** in the top right corner of the pop-up screen to save your changes.

### Add Enrichments

**X**

**title:** entities

**Keyword Extraction**  
Determines important keywords in this field, ranks them, and optionally detects the sentiment.

[Learn more](#) [Add](#)

**Sentiment Analysis**  
Identifies the overall positive or negative sentiment within this field.

[Learn more](#) [Add](#)

**Concept Tagging**  
Identifies general concepts that aren't necessarily directly referenced in this field.

[Learn more](#) [Add](#)

**Category Classification**  
Classifies this field into a hierarchy of categories that's five levels deep.

[Learn more](#) [Add](#)

**Semantic Role Extraction**  
Parses sentences into subject, action, and object form and returns additional semantic information.

[Learn more](#) [Add](#)

**Emotion Analysis**  
Analyzes the emotions (anger, disgust, fear, joy, and sadness) in this field.

[Learn more](#) [Add](#)

**Entity Extraction**  
Extracts people, companies, organizations, cities, geographic features, and more from this field. You can also create and add custom entity models with Watson Knowledge Studio

[Learn more](#) Added!

**Relation Extraction**  
Recognizes when two entities are related and identifies the type of relation. You can also create and add custom relation models with Watson Knowledge Studio.

[Learn more](#) [Add](#)

**Custom Model ID** [?](#)

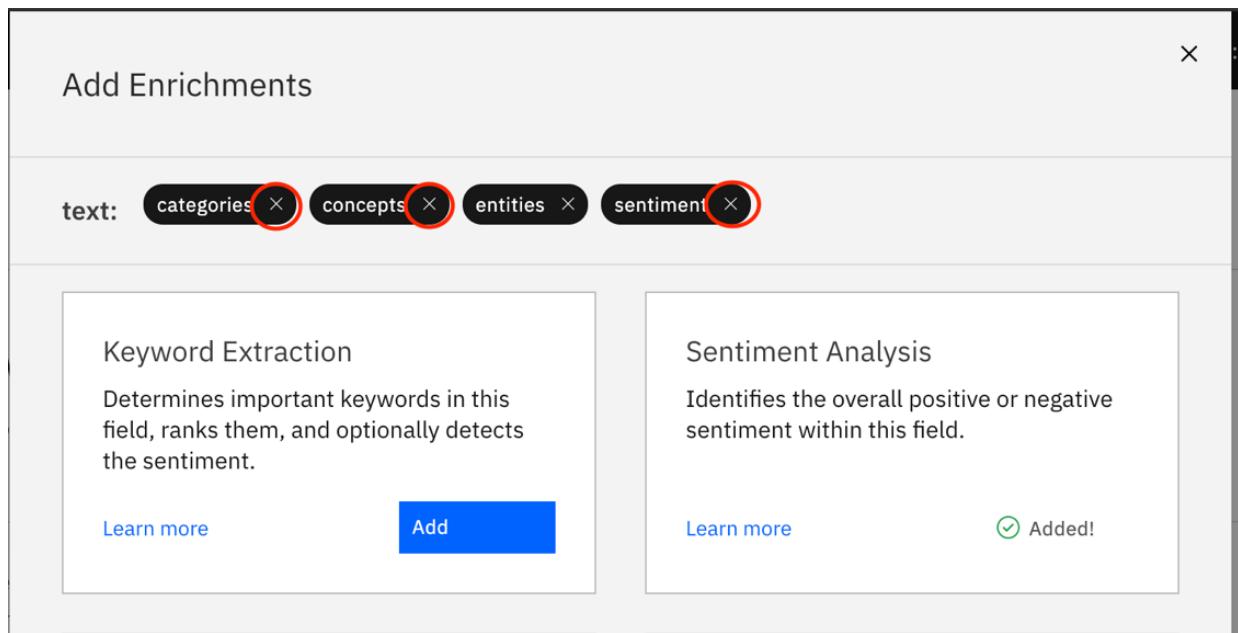
63d1efc3-6d00-4273-a034-Apply

enter model ID [Apply](#)

6. Now let's do the same thing for the text field of each post.  
Click on **+ Add enrichments** on the text row.

Fields to be enriched	Enrichments
title	<b>entities</b> <span style="color: red;">×</span>
text	<b>categories</b> <span style="color: red;">×</span> <b>concepts</b> <span style="color: red;">×</span> <b>entities</b> <span style="color: red;">×</span> <b>sentiment</b> <span style="color: red;">×</span>  <b>+ Add enrichments</b> <span style="color: red; border: 1px solid red; border-radius: 50%; padding: 2px;"> </span>

7. Remove the NLP enrichments of categories, concepts and sentiment by clicking on each **x** next to **categories**, **concepts** and **sentiment**.



8. Scroll down to the bottom of the Add Enrichments pop-up screen and paste the Model ID number for the ML annotator underneath **Custom Model ID**, click **Apply** and then click **X** to exit this screen.

Add Enrichments

text: entities

**Keyword Extraction**  
Determines important keywords in this field, ranks them, and optionally detects the sentiment.

**Sentiment Analysis**  
Identifies the overall positive or negative sentiment within this field.

**Concept Tagging**  
Identifies general concepts that aren't necessarily directly referenced in this field.

**Category Classification**  
Classifies this field into a hierarchy of categories that's five levels deep.

**Semantic Role Extraction**  
Parses sentences into subject, action, and object form and returns additional semantic information.

**Emotion Analysis**  
Analyzes the emotions (anger, disgust, fear, joy, and sadness) in this field.

**Entity Extraction**  
Extracts people, companies, organizations, cities, geographic features, and more from this field. You can also create and add custom entity models with Watson Knowledge Studio.

**Relation Extraction**  
Recognizes when two entities are related and identifies the type of relation. You can also create and add custom relation models with Watson Knowledge Studio.

**Custom Model ID** (?)  
63d1efc3-6d00-4273-a034 **Apply**

## Exercise 4: Perform Custom Entity Extraction

- Now that we have specified that entity extraction will occur on the title and text fields of each document using our ML annotator, click **Apply changes to collection**.

COVID19-SVI / Configure data

Identify fields Manage fields **Enrich fields**

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

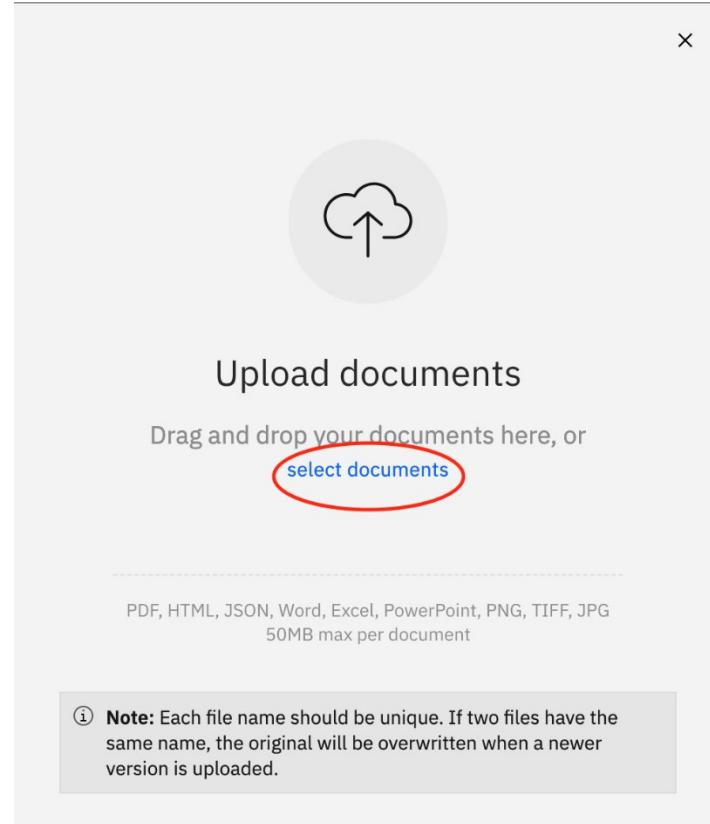
Add a field to enrich Choose a field

Fields to be enriched	Enrichments
title	entities <small>(?)</small>
	+ <a href="#">Add enrichments</a>
text	entities <small>(?)</small>
	+ <a href="#">Add enrichments</a>

**Apply changes to collection** (?)

2. We will need to tell Discovery that we are extracting entities from all the documents by selecting our documents again.

Click **select documents**.



3. Shift select all of the documents in the Posts folder and click **Open**.

Name	Date Modified	Size	Kind
relevant-seattle-292.json	Apr 28, 2020 at 2:54 AM	725 bytes	JSON Document
relevant-seattle-267.json	Apr 28, 2020 at 2:54 AM	633 bytes	JSON Document
relevant-seattle-285.json	Apr 28, 2020 at 2:54 AM	568 bytes	JSON Document
relevant-seattle-221.json	Apr 28, 2020 at 2:54 AM	476 bytes	JSON Document
relevant-seattle-139.json	Apr 28, 2020 at 2:53 AM	675 bytes	JSON Document
relevant-seattle-172.json	Apr 28, 2020 at 2:53 AM	1 KB	JSON Document
relevant-seattle-114.json	Apr 28, 2020 at 2:53 AM	720 bytes	JSON Document
relevant-seattle-199.json	Apr 28, 2020 at 2:53 AM	478 bytes	JSON Document
relevant-seattle-126.json	Apr 28, 2020 at 2:53 AM	472 bytes	JSON Document
relevant-seattle-134.json	Apr 28, 2020 at 2:53 AM	528 bytes	JSON Document
relevant-seattle-175.json	Apr 28, 2020 at 2:53 AM	440 bytes	JSON Document
relevant-seattle-161.json	Apr 28, 2020 at 2:53 AM	783 bytes	JSON Document
relevant-seattle-158.json	Apr 28, 2020 at 2:53 AM	635 bytes	JSON Document
relevant-seattle-137.json	Apr 28, 2020 at 2:53 AM	295 bytes	JSON Document
relevant-seattle-36.json	Apr 28, 2020 at 2:52 AM	510 bytes	JSON Document
relevant-seattle-14.json	Apr 28, 2020 at 2:52 AM	542 bytes	JSON Document
relevant-seattle-13.json	Apr 28, 2020 at 2:52 AM	3 KB	JSON Document
relevant-seattle-12.json	Apr 28, 2020 at 2:52 AM	775 bytes	JSON Document
relevant-seattle-65.json	Apr 28, 2020 at 2:52 AM	2 KB	JSON Document
relevant-seattle-94.json	Apr 28, 2020 at 2:52 AM	896 bytes	JSON Document
relevant-seattle-42.json	Apr 28, 2020 at 2:52 AM	6 KB	JSON Document
relevant-seattle-64.json	Apr 28, 2020 at 2:52 AM	955 bytes	JSON Document
relevant-seattle-5.json	Apr 28, 2020 at 2:52 AM	640 bytes	JSON Document
relevant-seattle-86.json	Apr 28, 2020 at 2:52 AM	790 bytes	JSON Document
relevant-seattle-8.json	Apr 28, 2020 at 2:52 AM	714 bytes	JSON Document
relevant-seattle-32.json	Apr 28, 2020 at 2:52 AM	856 bytes	JSON Document
relevant-seattle-24.json	Apr 28, 2020 at 2:52 AM	653 bytes	JSON Document
relevant-seattle-97.json	Apr 28, 2020 at 2:52 AM	615 bytes	JSON Document
relevant-seattle-44.json	Apr 28, 2020 at 2:52 AM	1 KB	JSON Document
relevant-seattle-25.json	Apr 28, 2020 at 2:52 AM	350 bytes	JSON Document
relevant-seattle-79.json	Apr 28, 2020 at 2:52 AM	924 bytes	JSON Document
relevant-seattle-51.json	Apr 28, 2020 at 2:52 AM	303 bytes	JSON Document

It will take approximately 6 minutes for entity extraction to occur on all the documents in the collection. When it is complete, you should see the following screen:

The screenshot shows the IBM Watson Discovery UI. On the left, there are navigation icons for folder, database, search, and list. The main title is "COVID19-SVI". Below the title, there are three tabs: "Overview" (selected), "Errors and warnings (0)", and "Search settings". The "Overview" section displays the following data:

- 364 documents
- 0 documents failed (View details)
- Created on: 5/18/2020 12:32:41 pm EDT
- Last updated: 5/18/2020 12:32:41 pm EDT
- Upload documents

Below this, there are three columns of information:

- Identified 5 fields from your data: text, author\_fullname, extracted\_metadata, id, title.
- Added 1 enrichment to your data: Entity Extraction (unemployment (61) | rent (27) | apartment (26) | lease (17) | apartments (10)).
- Now you're ready to query! (with three examples: Top people related to /society/work/unemployment, Entities of type Unemployed which have negative sentiment, and Top entities with their average, min, max sentiment score).

While we are able to see the same number of documents and fields per post, we are now only seeing one enrichment applied to our dataset. Entity extraction has been successfully performed on our dataset and we can now retrieve the output of this process.

## Exercise 5: Retrieve the analyzed files using the Discovery API

Let's start by viewing the entity extraction output inside the Discovery UI.

1. Click on the magnifying glass icon on the left-hand side of the screen.

The screenshot shows the IBM Watson Discovery UI. The magnifying glass icon on the left is highlighted with a red circle. The rest of the interface is identical to the previous screenshot, showing the "Overview" tab selected with 364 documents, 0 errors, and the entity extraction enrichment results.

2. This will take us to the Build Queries page where we can query a data using structured Discovery Language Queries or Natural Language Queries.  
Since we are only interested in viewing the output of the entity extraction process, click **Run Query**.

COVID19-SVI / Build queries

Build a query using one or more of these components. [Learn more.](#)

+ Search for documents

+ Include analysis of your results

+ Filter which documents you query

> More options

**Run query** **Close**

3. This will return a massive JSON file on the right-hand side of the screen consisting of all the documents in our collection after entity extraction.

Train Watson to improve results

**Summary** **JSON**

Query URL <https://gateway.watsonplatform.net/discovery/api/v1/environment>

```
{
  "matching_results": 364,
  "session_token": "1_0SECbKHa0o613VG4_xrczEdrmX",
  "passages": [],
  "results": [
    {
      "id": "f7181e61aecd2d2de8e67f34ff137ca07",
      "result_metadata": {...},
      "author_fullname": "t2_mguoa",
      "enriched_title": {...},
      "text": "I have reached a point of not knowing what else to do, other than reach out to news agencies to possibly shed light on the situation of Unemployment Insurance here in NYC. \\n\\nI, along with thousands of others, filed for UIB over a month ago and have yet to receive any word about whether my claim is accepted or not. Getting in touch with the Dol is impossible, their lines don't even have a queue option. After spending over 5 minutes in a touch-tone operating system, you are advised that they are 'experiencing a high volume of callers' and to 'try back later'. After calling everyday for weeks now, I am no closer to getting an answer than I was when I started this. \\n\\nGroups on Facebook like Restaurant Worker Solidarity NYC will show hundreds of posts of people in the same position. None of us have any idea about what to do, and there's no money coming in. I haven't received any income for a month and a half, have not paid rent, am depending on my sweetheart to help pay for groceries, and still have tuition pay...",
      "enriched_text": {
        "entities": [...]
      }
    },
    ...
  ]
}
```

**Run query** **Close**

Scrolling down this right panel, we can see that the output consists of all the documents with their original text as well as their extracted entities. We will be using the extracted entities for each document in order to compute the social vulnerability index for COVID-19.

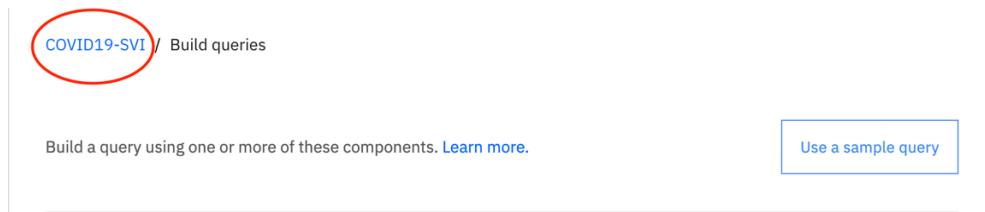
Unfortunately, the output is only presented in a condensed format inside the Discovery UI. We need access to the raw output and thus require the Discovery API. The Discovery API will allow us to use a NodeJS function to retrieve the entity extracted output file as a JSON file, which we can save on our local filesystem and subsequently split so that we have an entity extracted file for each document.

In order to use the Discovery API, we need several pieces of information: Discovery API Key, Discovery URL, Collection ID and Environment ID.

We have already retrieved the Discovery API Key and Discovery URL in Exercise 1 step 7 (pp. 7-8).

Let's get the Discovery Collection ID and Environment ID.

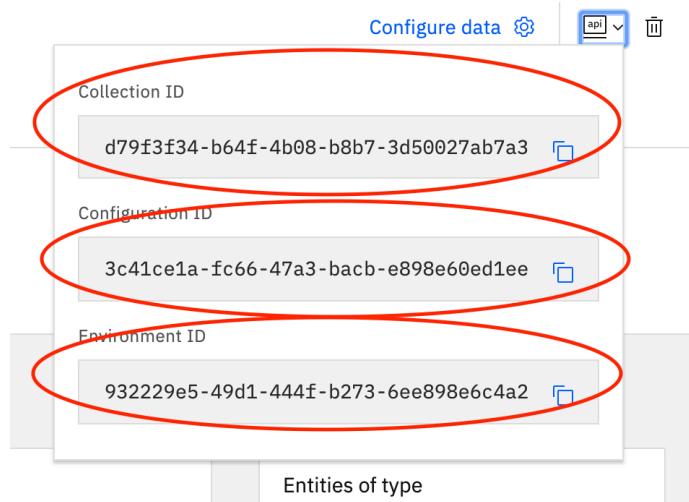
4. Click on the **COVID19-SVI** link on the top left corner of the screen.



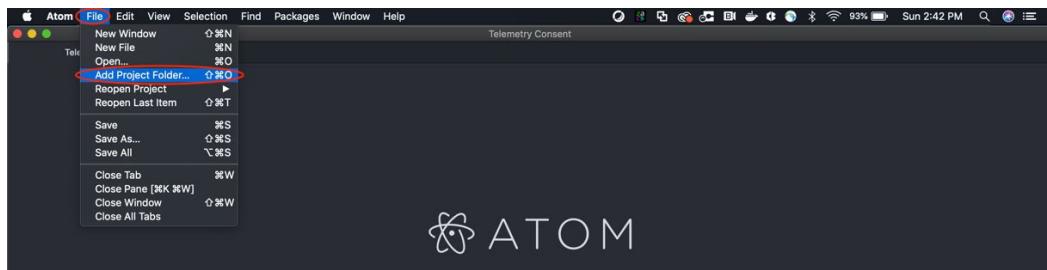
5. On the COVID-SVI screen, click on the **tiny API icon** in the top right corner.



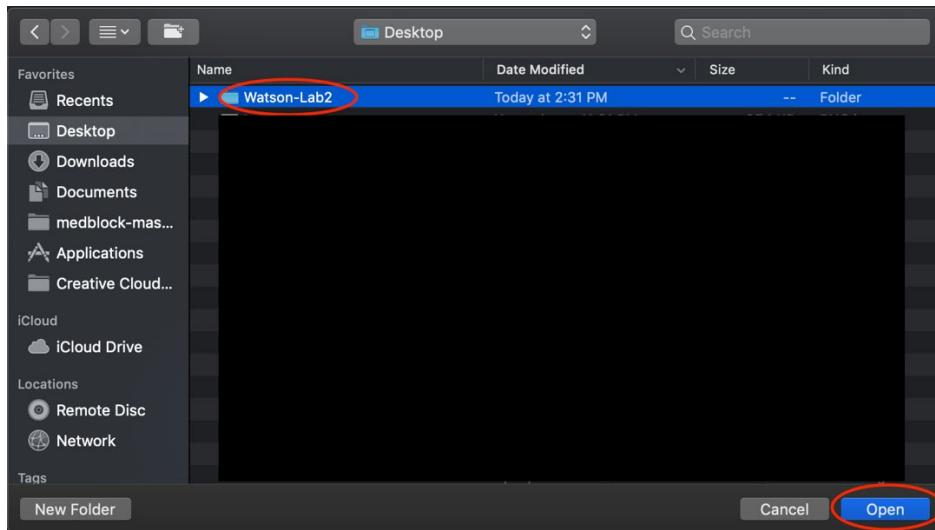
6. Copy the **Collection ID** and **Environment ID** (the Configuration ID is not needed for the NodeJS function)



7. Open the Atom code editor. Go to **File->Add Project Folder**.



8. Select the **Watson-Lab2** folder on your Desktop and click **Open**.



9. You should be able to see all of the files in the Watson-Lab2 folder on the left-hand side, including Discovery.js. Click the **Discovery.js** file to view it on the right.

```

Project              Telemetry Consent          Discovery.js
Watson-Lab2
  - Chicago-Discovery
  - DC-Discovery
  - LA-Discovery
  - NY-Discovery
  - Posts
  - Seattle-Discovery
  - .DS_Store
  - discovery-results-splitter.sh
  - Discovery.js (highlighted with a red circle)

```

```

const fs = require('fs');
const DiscoveryV1 = require('ibm-watson/discovery/v1');
const { IamAuthenticator } = require('ibm-watson/auth');

require("dotenv").config();

// Initialize Discovery and set default query parameters
const discovery = new DiscoveryV1({
  version: process.env.DISCOVERY_VERSION,
  authenticator: new IamAuthenticator({
    apikey: process.env.DISCOVERY_API_KEY,
  }),
  url: process.env.DISCOVERY_URL,
});

const queryParams = {
  environmentId: process.env.DISCOVERY_ENVIRON,
  collectionId: process.env.DISCOVERY_COLLECTION,
  count: 364,
};

let resultsPath = __dirname + "/DiscoveryOutput.json";

// Call Discovery API in order to perform NLP enrichments and save results array as a JSON file
discovery.query(queryParams).then(queryResponse => {
  var jsonData = JSON.stringify(queryResponse, null, 4);
  var data = JSON.parse(jsonData);
  var results = data.result.results;
  var json = JSON.stringify(results, null, 2);
  fs.writeFile(resultsPath, json, (err) =>{
    if(err){
      console.error(err);
    } else {
      console.log('Wrote successfully');
    }
  });
});

```

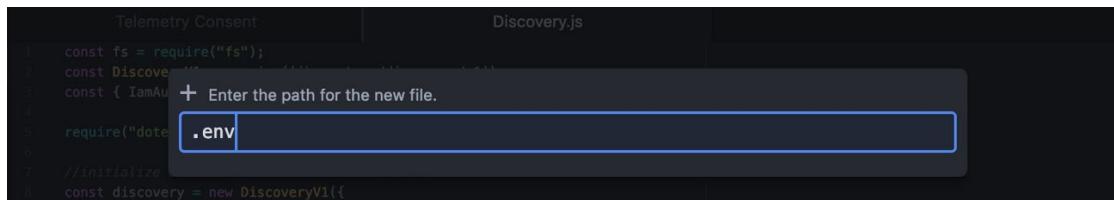
This Discovery.js file is a Node.js program which calls Watson Discovery API to perform entity extraction on all 364 social media posts in our Discovery collection. Upon completion of the entity extraction, the output will be saved as DiscoveryOutput.json in our Watson-Lab2 folder.

We need to add the credentials for our Discovery instance, collection and environment, which we will do right now. Instead of adding them directly to this file, we will add a separate file called .env to save our credentials.

10. Right click the Watson-Lab2 folder on the left side of the screen and click **New file**.



11. Type in a name of .env for this new file and click **ENTER** on your keyboard.



12. Add the following fields to this .env file:

```

Telemetry Consent           Discovery.js          .env
1  DISCOVERY_VERSION=
2  DISCOVERY_API_KEY=
3  DISCOVERY_URL=
4  DISCOVERY_ENVIRON=
5  DISCOVERY_COLLECTION=
6

```

13. Type **2019-04-02** next to DISCOVERY\_VERSION

```

Project              Telemetry Consent          .env — Desktop/Watson-Lab2 — ~/Desktop/Watson-Lab2
Watson-Lab2
  Chicago-Discovery
  DC-Discovery
  LA-Discovery
  NY-Discovery
  Posts
  Seattle-Discovery
  .DS_Store
  .env

```

14. Use the Discovery API Key, Discovery URL, Environment ID and Collection ID that you retrieved earlier in the lab to fill in the rest of the fields. Please refer to Exercise 1 Step 7 to get the Discovery API Key and Discovery URL and Exercise 5 Step 6 to get the Environment ID and Collection ID.

Save your .env file after entering in all the credentials by clicking **File->Save** or pressing **Command/CTRL + S** on your keyboard.

15. In order to run our Node.js program, we need to create a package.json file in our Watson-Lab2 folder.

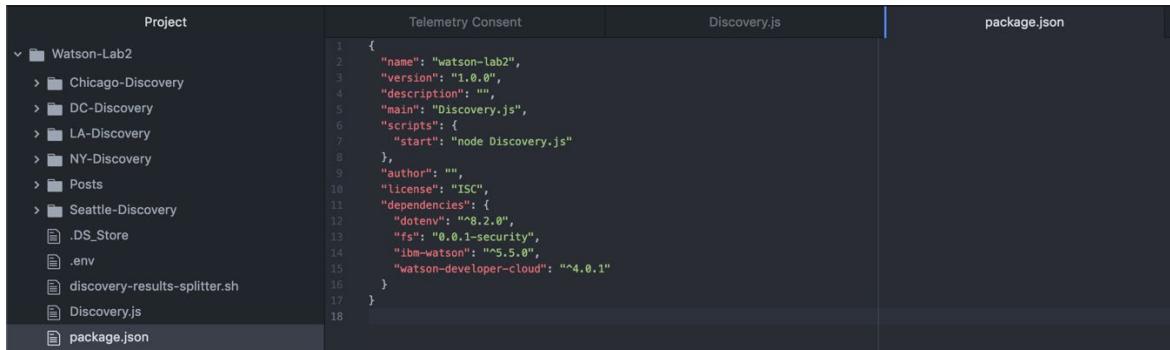
Right-click the Watson-Lab2 folder and click **New File**.

16. Give the new file a name of **package.json** and click **ENTER** on your keyboard.

17. Copy and paste the following code into this package.json file:

```
{  
  "name": "watson-lab2",  
  "version": "1.0.0",  
  "description": "",  
  "main": "Discovery.js",  
  "scripts": {  
    "start": "node Discovery.js"  
  },  
  "author": "",  
  "license": "ISC",  
  "dependencies": {  
    "dotenv": "^8.2.0",  
    "fs": "0.0.1-security",  
    "ibm-watson": "^5.5.0",  
    "watson-developer-cloud": "^4.0.1"  
  }  
}
```

You should now have a package.json file that looks like this:

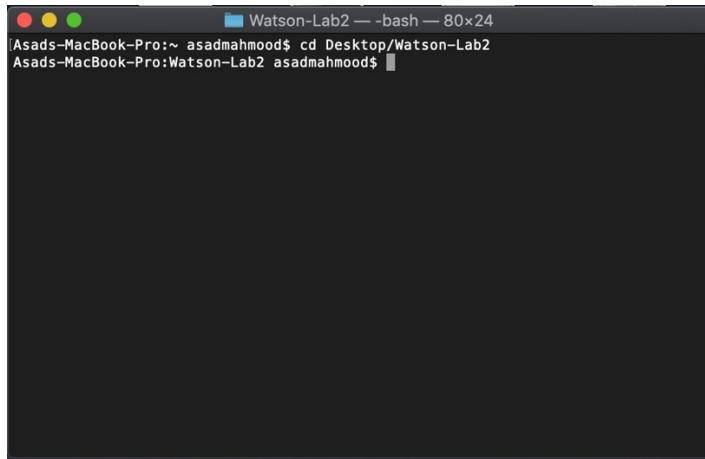


```
1  {  
2   "name": "watson-lab2",  
3   "version": "1.0.0",  
4   "description": "",  
5   "main": "Discovery.js",  
6   "scripts": {  
7     "start": "node Discovery.js"  
8   },  
9   "author": "",  
10  "license": "ISC",  
11  "dependencies": {  
12    "dotenv": "^8.2.0",  
13    "fs": "0.0.1-security",  
14    "ibm-watson": "^5.5.0",  
15    "watson-developer-cloud": "^4.0.1"  
16  }  
17 }
```

Don't forget to save your package.json file with **Command/CTRL + S** or by selecting **File->Save**.

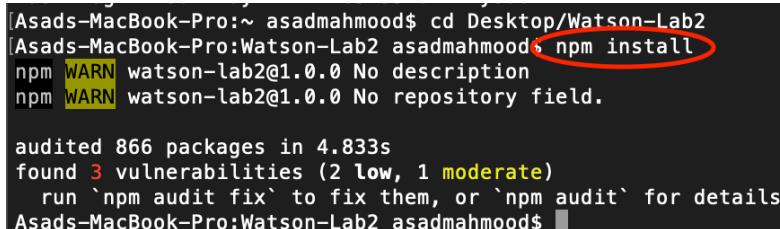
Now that we have a program to call the Discovery API and an associated credentials file, we can capture the results of the entity extraction process in a separate output file, which will be saved to the Watson-Lab2 folder.

18. **Open up the Terminal window** and navigate to the Watson-Lab2 folder on your Desktop with the following command: **cd Desktop/Watson-Lab2**

A screenshot of a terminal window titled "Watson-Lab2 — -bash — 80x24". The window has three colored window control buttons (red, yellow, green) at the top left. The main area is black and contains no text or other content.

19. Since we used a few node libraries to build our program, let's install them before running the program.

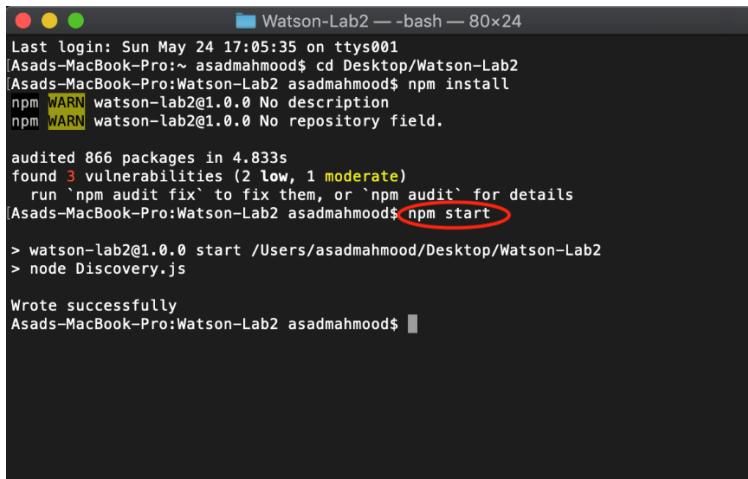
Type **npm install**

A screenshot of a terminal window showing the output of the "npm install" command. The command is highlighted with a red oval. The output shows two "WARN" messages about missing description and repository fields, followed by audit results and a note to fix vulnerabilities. The command "Asads-MacBook-Pro:Watson-Lab2 asadmahmood\$" is also circled in red.

Wait a few minutes for the command to complete and ignore any vulnerabilities.

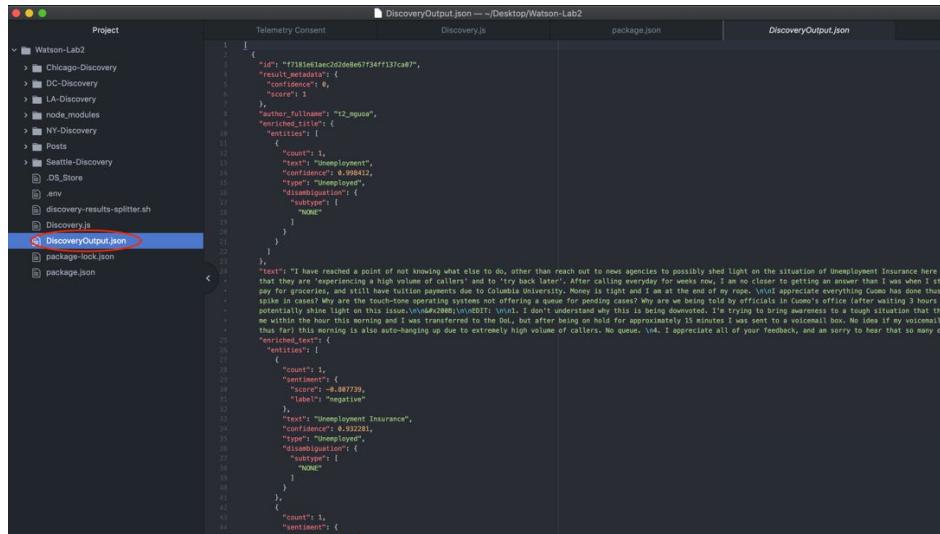
20. We can run the Discovery.js to perform entity extraction and capture the output in a separate file.

Type **npm start**

A screenshot of a terminal window showing the output of the "npm start" command. The command is highlighted with a red oval. The output shows the command being run, followed by a note to write to a file, and the message "Wrote successfully". The command "Asads-MacBook-Pro:Watson-Lab2 asadmahmood\$" is also circled in red.

You should be able to see “Wrote successfully” on your Terminal window if the program is run successfully.

21. Check the output file by opening up the Atom code editor and clicking **DiscoveryOutput.json** on the left-hand side of the screen.



```
Project Telemetry Consent Discovery.js package.json DiscoveryOutput.json
Watson-Lab2
  > Chicago-Discovery
  > DC-Discovery
  > LA-Discovery
  > node_modules
  > NY-Discovery
  > Posts
  > Seattle-Discovery
    .DS_Store
    .env
    discovery-results-splitter.sh
    Discovery.js
    DiscoveryOutput.json
    package-lock.json
    package.json

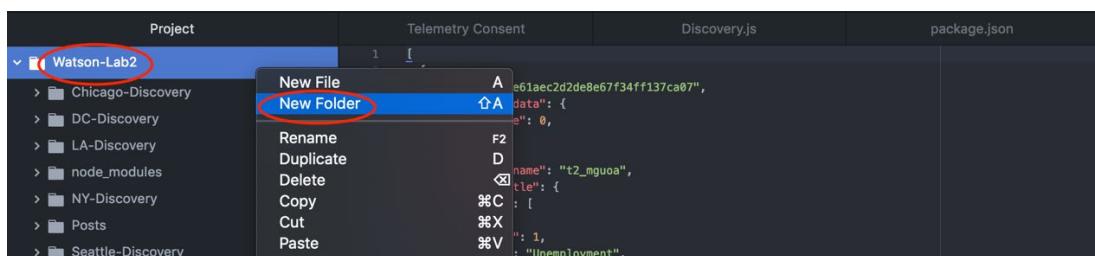
DiscoveryOutput.json
```

```
{
  "id": "f7818e61ec2d2de8e67f34ff137ca07",
  "result_metadata": {
    "confidence": 0,
    "score": 1
  },
  "author_fullname": "t2_egusa",
  "enriched_title": {
    "entities": [
      {
        "count": 1,
        "text": "Unemployment",
        "confidence": 0.99842,
        "type": "disambiguated",
        "disambiguation": {
          "subtype": "NONE"
        }
      }
    ],
    "sentiment": {
      "count": 1,
      "sentiment": {
        "score": -0.88739,
        "label": "negative"
      }
    },
    "text": "Unemployment Insurance",
    "confidence": 0.93228,
    "type": "disambiguated",
    "disambiguation": {
      "subtype": "NONE"
    }
  },
  "count": 1,
  "sentiment": {
```

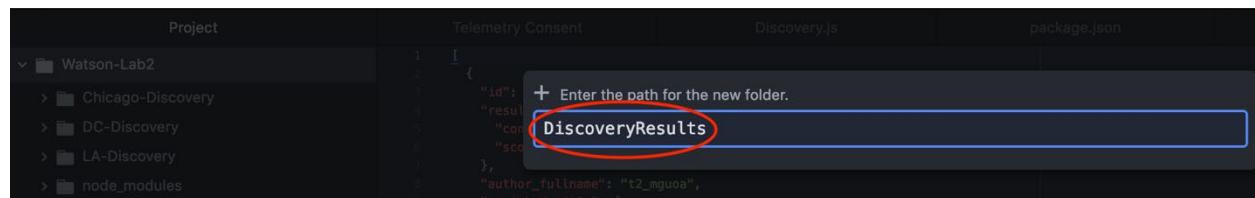
Although we have the raw output file from Discovery now, you can see that it is a nested JSON file containing the combined output for all posts. Let's break this file up into individual output files for each of our posts. We can do this by running a bash script from the command line in the next few steps.

22. First, create a new folder inside your Watson-Lab2 folder.

Inside of the Atom code editor, right-click the Watson-Lab2 folder and click **New Folder**.



23. Give the new folder a name of **DiscoveryResults** and click **ENTER** on your keyboard.



24. Click the **discovery-results-splitter.sh** file on the left side to view it. This will be the bash script that we will use to split the DiscoveryOutput.json into individual output files.

```

Project              Telemetry Consent      Discovery.js      package.json      discovery-results-splitter.sh
Watson-Lab2
  Chicago-Discovery
  DC-Discovery
  DiscoveryResults
  LA-Discovery
  node_modules
  NY-Discovery
  Posts
  Seattle-Discovery
    .DS_Store
    .env
discovery-results-splitter.sh

```

```

1 #!/bin/bash
2 N=0;
3 jq -c '.[]' DiscoveryOutput.json |
4   while read -r json ; do
5     N=$((N+1))
6     jq . <<< "json" > "/Users/asadmahmood/Desktop/Watson-Lab2/DiscoveryResults/discovery-$N.json"
7   done
8

```

The bash script is quite straightforward and easy to understand: it takes the DiscoveryOutput.json, grabs each object in the JSON array and saves each as an individual JSON file in the DiscoveryResults folder.

Let's change the file path of the Discovery output files.

25. If you have a Mac, only change the username in the file path to the username for your computer:

"`/Users/username/Desktop/Watson-Lab2/DiscoveryResults/discovery-$N.json`"

If you have a Windows, change the username and add a C: to the beginning of the path:  
"`C:/Users/username/Desktop/Watson-Lab2/DiscoveryResults/discovery-$N.json`"

If you are running the Linux operating system, make the following change:

"`/home/username/Desktop/Watson-Lab2/DiscoveryResults/discovery-$N.json`"

26. Save the **discovery-results-splitter.sh** file with **Command/CTRL + S** or **File->Save**.

27. Open up the Terminal window and run the following command in order to create

Discovery output files for all posts:

**sh discovery-results-splitter.sh**

```

Watson-Lab2 -- bash -- 131x24
[Asads-MacBook-Pro:Watson-Lab2 asadmahmood$ sh discovery-results-splitter.sh
Asads-MacBook-Pro:Watson-Lab2 asadmahmood$ ]

```

After a pause of a few seconds, the bash script should complete and you can now check the DiscoveryResults folder inside Watson-Lab2 to see the separated Discovery output files.

Name	Date Modified	Size	Kind
discovery-1.json	Today at 7:35 PM	5 KB	JSON Document
discovery-2.json	Today at 7:35 PM	1 KB	JSON Document
discovery-3.json	Today at 7:35 PM	5 KB	JSON Document
discovery-4.json	Today at 7:35 PM	3 KB	JSON Document
discovery-5.json	Today at 7:35 PM	2 KB	JSON Document
discovery-6.json	Today at 7:35 PM	3 KB	JSON Document
discovery-7.json	Today at 7:35 PM	2 KB	JSON Document
discovery-8.json	Today at 7:35 PM	2 KB	JSON Document
discovery-9.json	Today at 7:35 PM	2 KB	JSON Document
discovery-10.json	Today at 7:35 PM	2 KB	JSON Document
discovery-11.json	Today at 7:35 PM	1 KB	JSON Document
discovery-12.json	Today at 7:35 PM	1 KB	JSON Document
discovery-13.json	Today at 7:35 PM	4 KB	JSON Document
discovery-14.json	Today at 7:35 PM	3 KB	JSON Document
discovery-15.json	Today at 7:35 PM	2 KB	JSON Document
discovery-16.json	Today at 7:35 PM	2 KB	JSON Document
discovery-17.json	Today at 7:35 PM	2 KB	JSON Document
discovery-18.json	Today at 7:35 PM	973 bytes	JSON Document
discovery-19.json	Today at 7:35 PM	1 KB	JSON Document
discovery-20.json	Today at 7:35 PM	1 KB	JSON Document
discovery-21.json	Today at 7:35 PM	869 bytes	JSON Document
discovery-22.json	Today at 7:35 PM	2 KB	JSON Document
discovery-23.json	Today at 7:35 PM	3 KB	JSON Document
discovery-24.json	Today at 7:35 PM	1 KB	JSON Document
discovery-25.json	Today at 7:35 PM	2 KB	JSON Document
discovery-26.json	Today at 7:35 PM	2 KB	JSON Document
discovery-27.json	Today at 7:35 PM	2 KB	JSON Document
discovery-28.json	Today at 7:35 PM	2 KB	JSON Document
discovery-29.json	Today at 7:35 PM	3 KB	JSON Document
discovery-30.json	Today at 7:35 PM	1 KB	JSON Document
discovery-31.json	Today at 7:35 PM	2 KB	JSON Document
discovery-32.json	Today at 7:35 PM	935 bytes	JSON Document
discovery-33.json	Today at 7:35 PM	4 KB	JSON Document
discovery-34.json	Today at 7:35 PM	1 KB	JSON Document
discovery-35.json	Today at 7:35 PM	2 KB	JSON Document
discovery-36.json	Today at 7:35 PM	4 KB	JSON Document
discovery-37.json	Today at 7:35 PM	2 KB	JSON Document
discovery-38.json	Today at 7:35 PM	1 KB	JSON Document
discovery-39.json	Today at 7:35 PM	1 KB	JSON Document
discovery-40.json	Today at 7:35 PM	1 KB	JSON Document

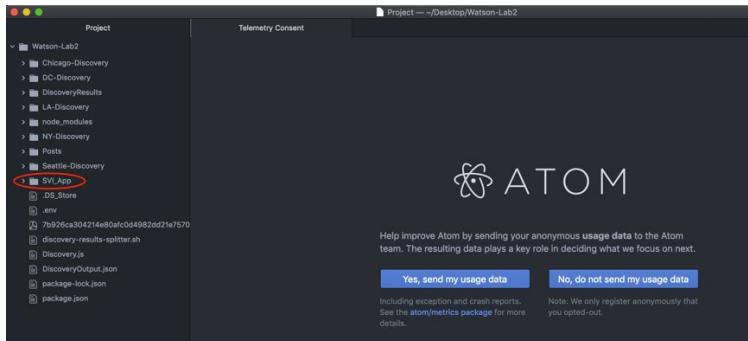
We have just successfully retrieved the entity extracted output files for every social media post uploaded to Watson Discovery. The next step is to use these files to compute a Social Vulnerability Index (SVI) for each of the five cities.

## Exercise 6: Calculate the COVID-19 vulnerability index

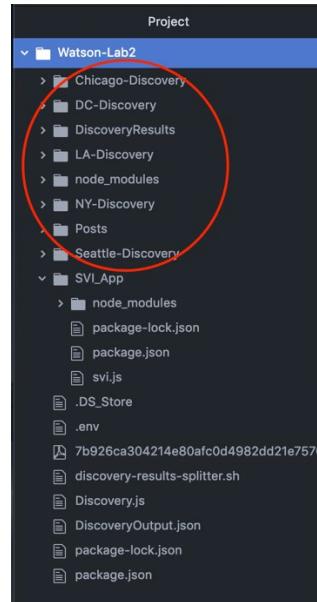
Although we have just saved the entity extracted output files to a separate folder inside our Watson-Lab2 folder called DiscoveryResults, it would be easier for us to work with the city separated folders instead. We will be using the city separated folders – Chicago-Discovery, Seattle-Discovery, LA-Discovery, DC-Discovery and NY-Discovery – to compute a social vulnerability index (SVI) for each of these 5 cities.

Essentially, we will use a NodeJS program called svi.js to compute each city SVI. This program is able to take all five folders simultaneously and use the identified entities in each file to compute an SVI for each city. The greater the computed SVI value, the more vulnerable to COVID-19 the city is estimated to be.

1. Open up the Atom code editor and make sure that the Watson-Lab2 folder contains a folder called SVI\_App.



2. Click on the SVI\_App folder to check if it contains package-lock.json, package.json and svi.js.
3. Make sure that the 5 city folders are inside the Watson-Lab2 folder but are outside the SVI\_App folder.

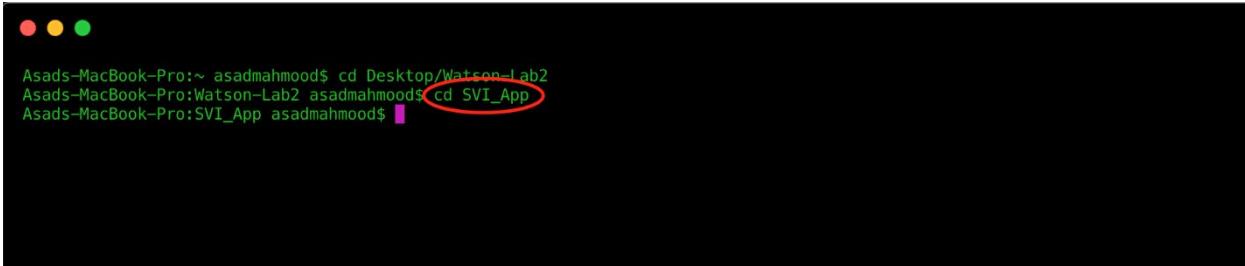


4. Minimize the Atom code editor and open up the Terminal window. Let's run the svi.js program to compute an SVI for each of the 5 cities.

If you closed down your Terminal window earlier and find yourself in your user folder again, return to the Watson-Lab2 folder by typing in **cd Desktop/Watson-Lab2**

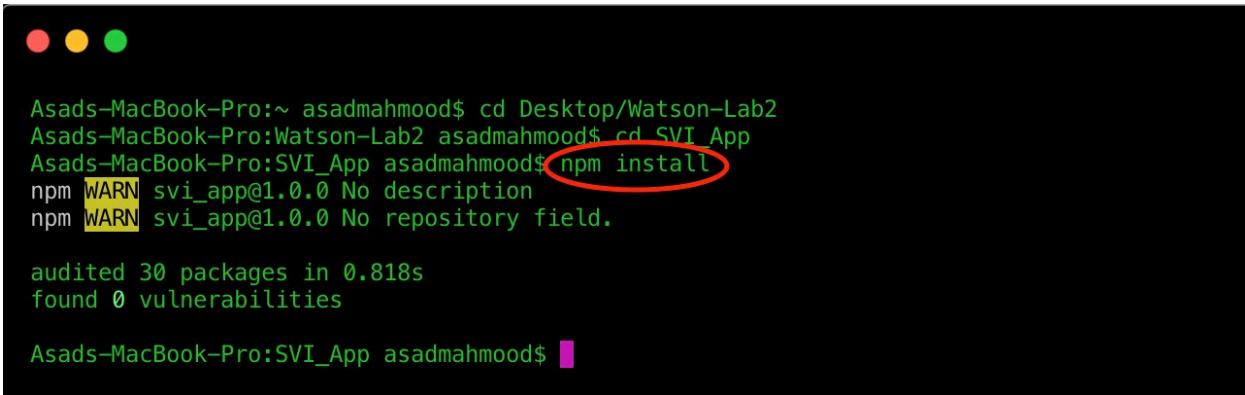
```
Asads-MacBook-Pro:~ asadmahmood$ cd Desktop/Watson-Lab2
Asads-MacBook-Pro:Watson-Lab2 asadmahmood$
```

5. Navigate to the SVI\_App folder by typing **cd SVI\_App**



```
Asads-MacBook-Pro:~ asadmahmood$ cd Desktop/Watson-Lab2
Asads-MacBook-Pro:Watson-Lab2 asadmahmood$ cd SVI_App
Asads-MacBook-Pro:SVI_App asadmahmood$
```

6. Install the node libraries in the program by typing **npm install**

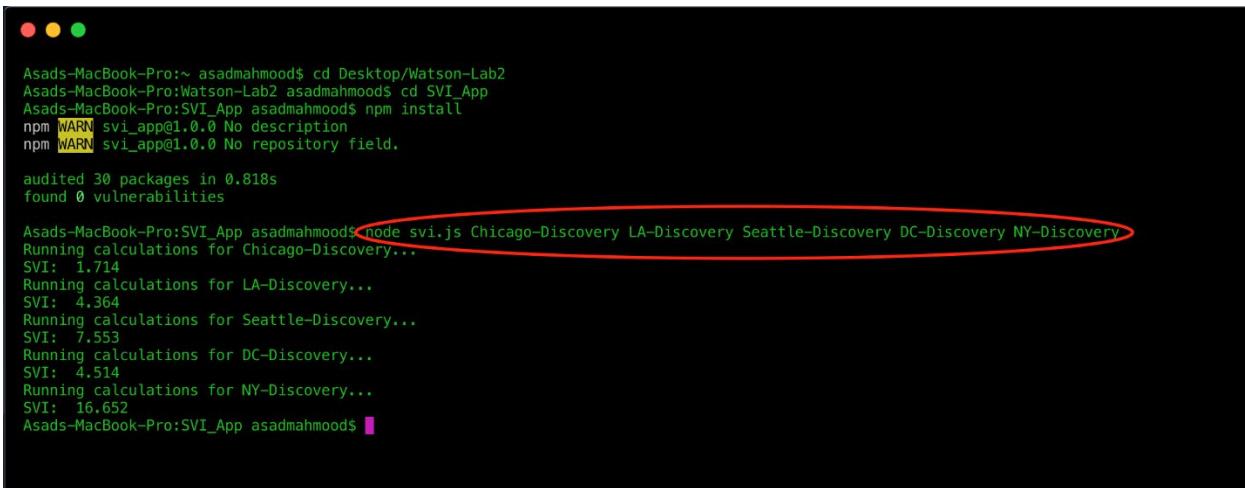


```
Asads-MacBook-Pro:~ asadmahmood$ cd Desktop/Watson-Lab2
Asads-MacBook-Pro:Watson-Lab2 asadmahmood$ cd SVI_App
Asads-MacBook-Pro:SVI_App asadmahmood$ npm install
npm WARN svi_app@1.0.0 No description
npm WARN svi_app@1.0.0 No repository field.

audited 30 packages in 0.818s
found 0 vulnerabilities

Asads-MacBook-Pro:SVI_App asadmahmood$
```

7. Run the svi.js program by typing the following command on a single line:  
**node svi.js Chicago-Discovery LA-Discovery Seattle-Discovery DC-Discovery NY-Discovery**



```
Asads-MacBook-Pro:~ asadmahmood$ cd Desktop/Watson-Lab2
Asads-MacBook-Pro:Watson-Lab2 asadmahmood$ cd SVI_App
Asads-MacBook-Pro:SVI_App asadmahmood$ npm install
npm WARN svi_app@1.0.0 No description
npm WARN svi_app@1.0.0 No repository field.

audited 30 packages in 0.818s
found 0 vulnerabilities

Asads-MacBook-Pro:SVI_App asadmahmood$ node svi.js Chicago-Discovery LA-Discovery Seattle-Discovery DC-Discovery NY-Discovery
Running calculations for Chicago-Discovery...
SVI: 1.714
Running calculations for LA-Discovery...
SVI: 4.364
Running calculations for Seattle-Discovery...
SVI: 7.553
Running calculations for DC-Discovery...
SVI: 4.514
Running calculations for NY-Discovery...
SVI: 16.652
Asads-MacBook-Pro:SVI_App asadmahmood$
```

The svi.js program should be able to automatically calculate and display the SVI for each city in the order we provide each city folder.

As we can see, since NY-Discovery has the largest SVI, New York City was calculated to be the city most socially vulnerable to COVID-19. On the other hand, since Chicago-Discovery has the smallest SVI, Chicago was calculated to be the city least socially vulnerable to COVID-19.

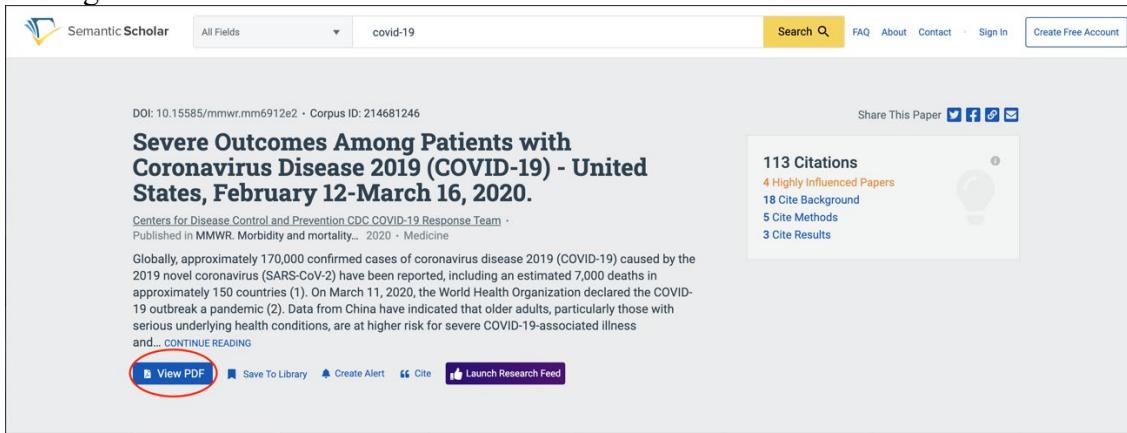
In order of descending social vulnerability, the 5 cities can be ranked as follows: New York City > Seattle > DC > LA > Chicago.

These SVIs can of course be modified by adding more posts to each city folder. In addition, the svi.js program can support even more folders and thus allows us to compute SVI for other U.S. cities. Now that you have access to this SVI\_App folder, please feel free to extend the analysis by adding more posts and more cities!

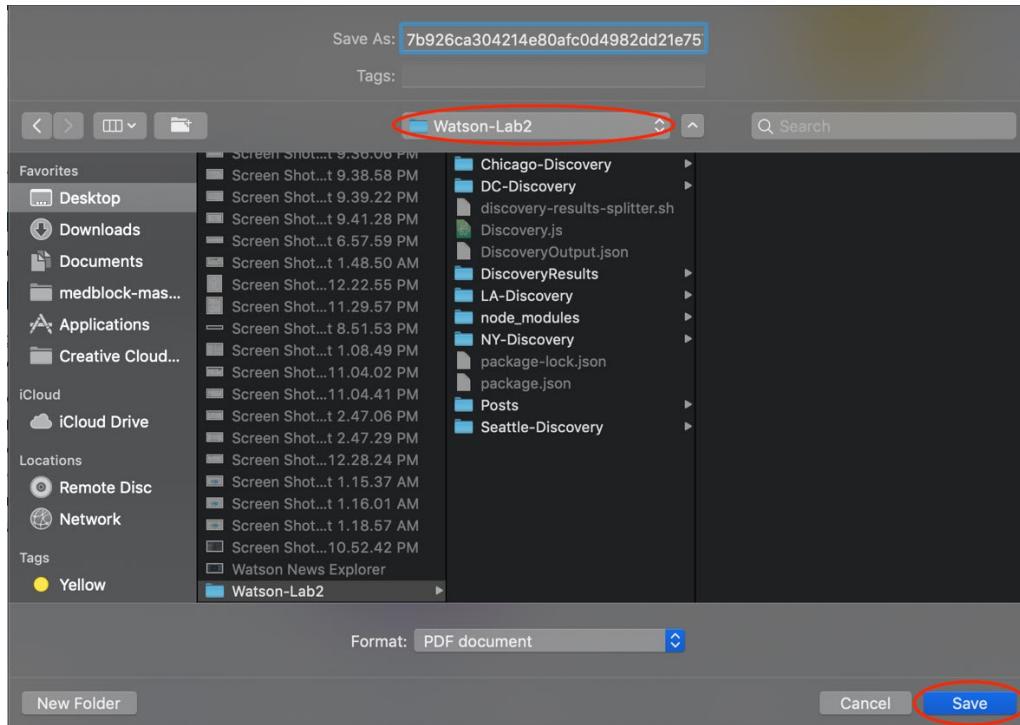
## Exercise 7: Create a collection for a COVID-19 publication

Although we have just used Watson Discovery to ingest a collection of social media data, perform custom entity extraction and use the analyzed files to compute a vulnerability index for different U.S. cities, we have not shown you how to search through a journal publication to answer natural language questions and retrieve relevant passages...yet. In this exercise, we're going to first create a brand new collection for a COVID-19 related journal publication.

1. Let's find our COVID-19 journal article by visiting  
<https://www.semanticscholar.org/paper/Severe-Outcomes-Among-Patients-with-Coronavirus-16%2C-Team/0813446cd9dcf350a7212280b8db7ee3fd05b970>
2. Right-click the **View PDF** button.



3. Click **Save Link As** and save the pdf file to the **Watson-Lab2** folder by clicking **Save**.



- Open up the downloaded pdf file in your Watson-Lab2 folder to verify that you have the complete article, which spans 4 pages. We will be uploaded this article to a brand new collection in Watson Discovery.

**Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020**

**Morbidity and Mortality Weekly Report**

**MMWR**

**Centers for Disease Control and Prevention**

**Early Release / Vol. 69**

**March 18, 2020**

**Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020**

**CDC COVID-19 Response Team**

Please note: This report has been corrected.

Globally, approximately 170,000 confirmed cases of coronavirus disease 2019 (COVID-19) caused by the 2019 novel coronavirus (SARS-CoV-2) have been reported, including an estimated 7,000 deaths in approximately 150 countries (1). On March 11, 2020, the World Health Organization declared the COVID-19 outbreak a pandemic (2). Data from China have indicated that older adults, particularly those with serious underlying health conditions, are at higher risk for severe COVID-19-associated illness and death than are younger persons (3). Although the majority of reported COVID-19 cases in China were mild (81%), approximately 80% of deaths occurred among adults aged ≥60 years; only one (0.1%) death occurred in a person aged ≤19 years (3). In this report, COVID-19 cases in the United States that occurred during February 12–March 16, 2020 and severity of disease (hospitalization, admission to intensive care unit [ICU], and death) were analyzed by age group. As of March 16, a total of 4,226 COVID-19 cases in the United States had been reported to CDC.

developed data collection forms (6). The cases described in this report include both COVID-19 cases confirmed by state or local public health laboratories as well as those with a positive test at the state or local public health laboratories and confirmation at CDC. No data on serious underlying health conditions were available. Data on these cases are preliminary and are missing for some key characteristics of interest, including hospitalization status (1,5,14), ICU admission (2,253), death (2,001), and age (386). Because of these missing data, the percentages of hospitalizations, ICU admissions, and deaths (case-fatality percentages) were estimated as a range. The lower bound of these percentages was estimated by using all cases within each age group as denominators. The corresponding upper bound of these percentages was estimated by using only cases with known information on each outcome as denominators.

As of March 16, a total of 4,226 COVID-19 cases had been reported in the United States, with reports increasing to 500 or more cases per day beginning March 14 (Figure 1). Among

- Open up your Discovery instance (you may have to log in again) and click **Upload your own data**.

The screenshot shows the 'Manage data' section of the Watson Studio interface. At the top, there's a header with icons for folder, file, and search. Below it, a sub-header says 'Manage data' and describes it as 'Collections of your private data and pre-enriched data to configure and query against.' There are three main buttons: 'Create a new data collection' (disabled), 'Create COVID-19 Kit' (disabled), and 'Upload your own data' (highlighted with a red circle). To the right of these is a 'Connect a data source' button.

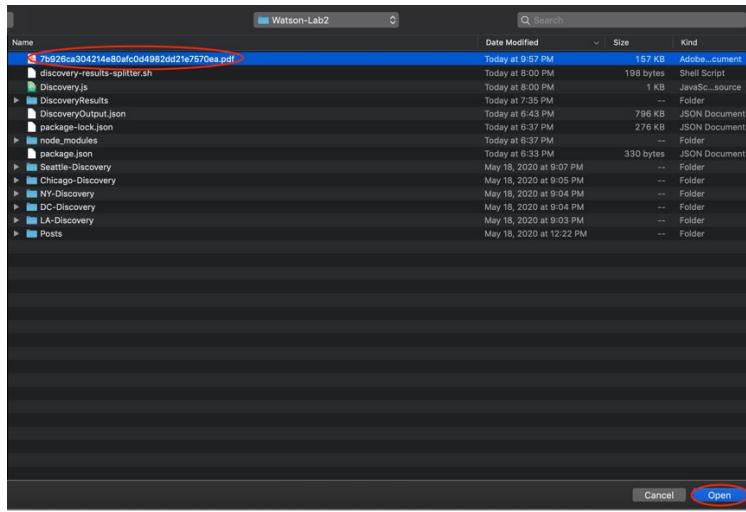
6. Give the new collection a name of **COVID-19 Article** and click **Create**.

This is a modal dialog titled 'Name your new collection'. It contains a 'Collection name' input field where 'COVID-19Article' is typed and highlighted with a red circle. Below it is a dropdown menu for selecting the language of documents, set to 'English'. At the bottom are 'Cancel' and 'Create' buttons, with 'Create' highlighted with a red circle.

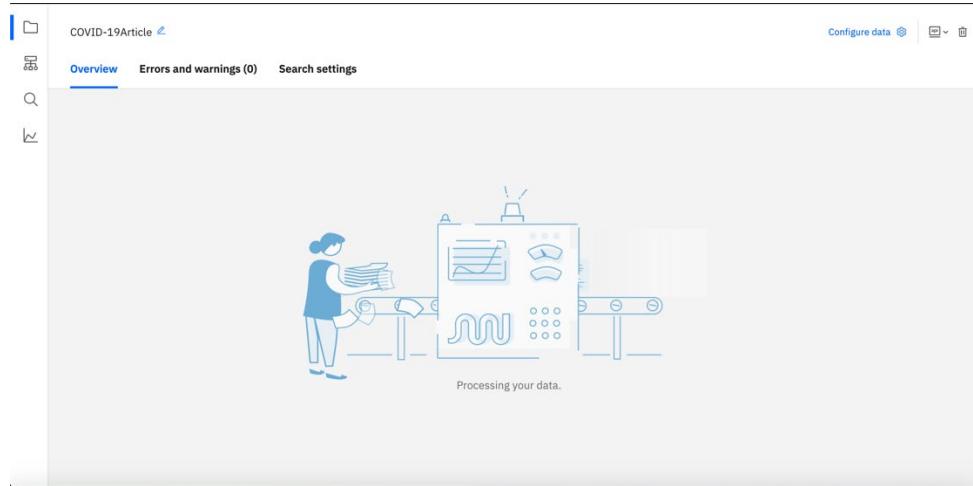
7. Inside the new collection, click **select documents**.

This screenshot shows the 'COVID-19Article' collection overview page. The left sidebar has icons for folder, file, and search. The main area has tabs for 'Overview' (which is selected and underlined), 'Errors and warnings (0)', and 'Search settings'. A central panel has a cloud icon with an upward arrow and the text 'Upload data to get started'. Below it is a placeholder: 'Drag and drop your documents here, or [select documents](#)' (with 'select documents' highlighted with a red circle). At the bottom, it specifies supported file types: 'PDF, HTML, JSON, Word, Excel, PowerPoint, PNG, TIFF, JPG' and '50MB max per document'. Top right of the main area includes 'Configure data' and other small icons.

8. Select the pdf article inside your Watson-Lab2 folder and click **Open**.



9. Wait approximately 13 minutes for the document to be completely uploaded to the new collection.



10. After the document is successfully uploaded, you should see the following screen:

The screenshot shows the 'Overview' tab of the IBM Watson Discovery interface. At the top, it displays 'COVID-19Article' with a document count of '1 document'. It indicates '0 documents failed' and provides 'View details', 'Created on 5/24/2020 10:15:40 pm EDT', and 'Last updated 5/24/2020 10:15:40 pm EDT'. A blue button 'Upload documents' is visible.

Below this, there are three main sections: 'Identified 1 field from your data' (text), 'Added 4 enrichments to your data' (Entity Extraction, Sentiment Analysis, Concept Tagging, Category Classification), and 'Now you're ready to query!' (three cards: JobTitle entities with positive sentiment, documents containing Ageing but not Death, and documents about Washington as a location with negative sentiment).

The default NLP enrichments of entity extraction, concept tagging, sentiment analysis and category classification have already been applied to this journal article. We will not be making any changes to these enrichments and we will instead teach Watson to understand the underlying structure of the article in the next section.

## Exercise 8: Perform Smart Document Understanding

Before we can search through the pages of our document for answers to specific questions, we must first train Watson to understand the underlying structure and format of our entire document. This is made possible through a capability known as Smart Document Understanding (SDU). We can access SDU within our Discovery instance.

1. Click **Configure Data**.

This screenshot is identical to the one above, showing the 'Overview' tab of the IBM Watson Discovery interface. The 'Configure data' button at the top right is highlighted with a red oval.

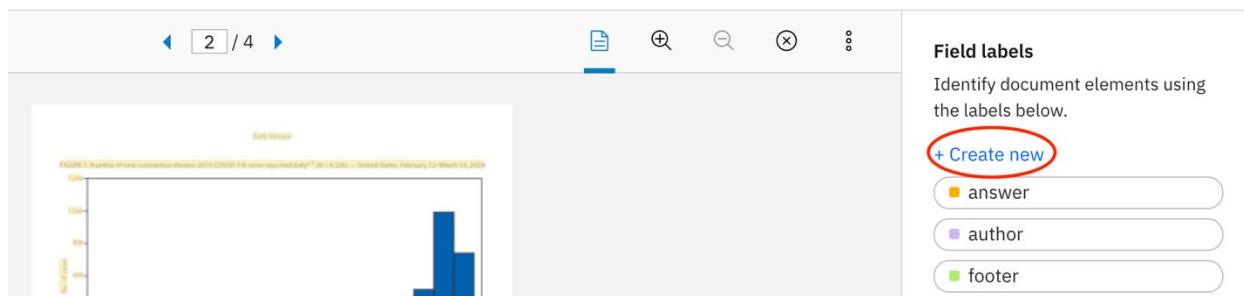
2. On the **Identify fields** page, you should be able to see a page by page preview of the uploaded article. This is where we can access SDU.

We are now going to use SDU on each page of our document and mark each section with the appropriate field on the right hand side in order to train Watson to understand the structure and format.

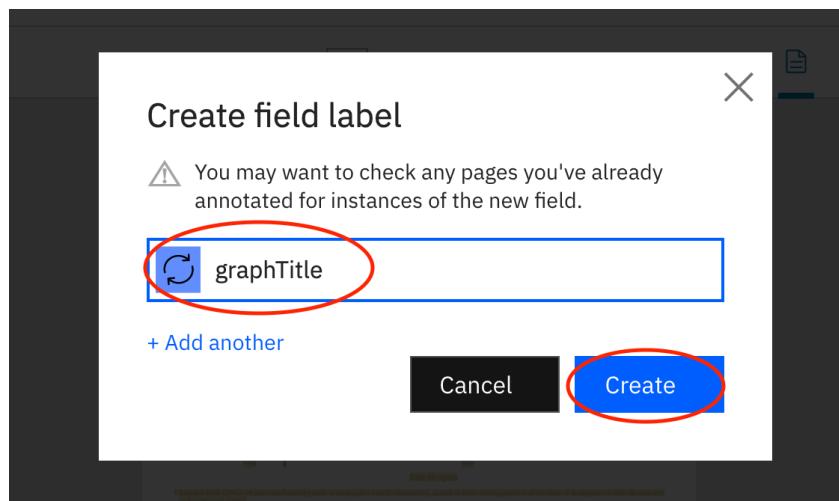
- Click on the **small book icon** near the center of the screen to revert to a single page view.

- On the first page of the document, we are going to click on title underneath Field labels on the right side of the screen and then use it to highlight “Morbidity and Mortality Weekly Report” and “Severe Outcomes Among Patients with Coronavirus Disease....” To highlight a phrase with a field label, simply select the field label first and then click and drag it over the desired phrase. You should be able to see the following labeled page:

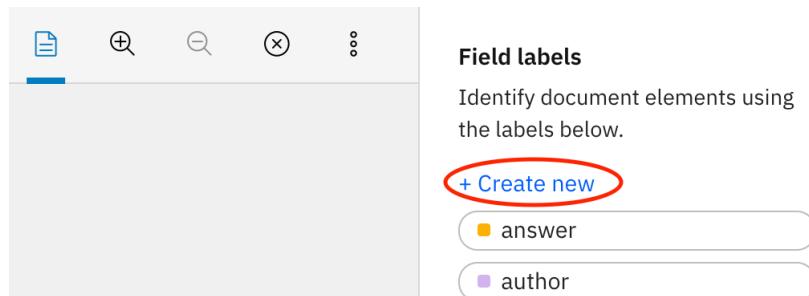
- Click the **Submit page** button to move on to the next page.
  - On the second page of the document, we're going to create two new field labels - **graphTitle** and **caption**.
- Click **+ Create new**

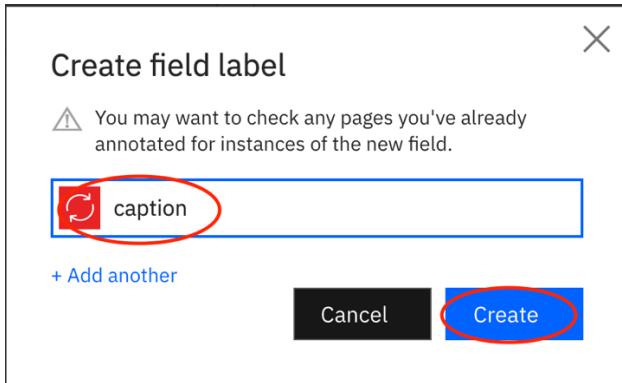


- Give the new field label a name of **graphTitle** and click **Create**.



- Click **+ Create new** again and this time give the new field label a name of **caption** before clicking **Create**.





9. Return to the second page of the document by clicking **Submit page**.
10. Select the **graphTitle** field and then click the title above the graph so that it turns into the same color as the **graphTitle**.  
Then, select the **caption** field and click each of the 3 lines below the graph.  
You should be able to see the following labeled page:

- Click the **Submit page** button to move to the next page.
11. Label the third page using the **graphTitle**, **caption** and **footer** fields so that your page resembles the following labeled page:

Viewing: Your training

Submit page

Learn more about how to use.

Click the **Submit page** button to move on to the last page.

12. Label the fourth and last page so that it resembles the following labeled page:

Submit page

Learn more about how to use.

Click the **Submit page** button to save these changes.

13. Click the **Apply changes to collection** button on the top right corner of the screen to save the labelled pages.

Apply changes to collection

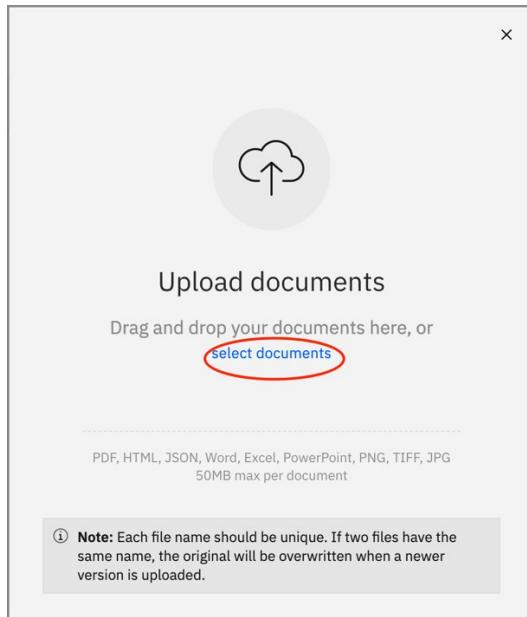
Field labels

Identify document elements using the labels below.

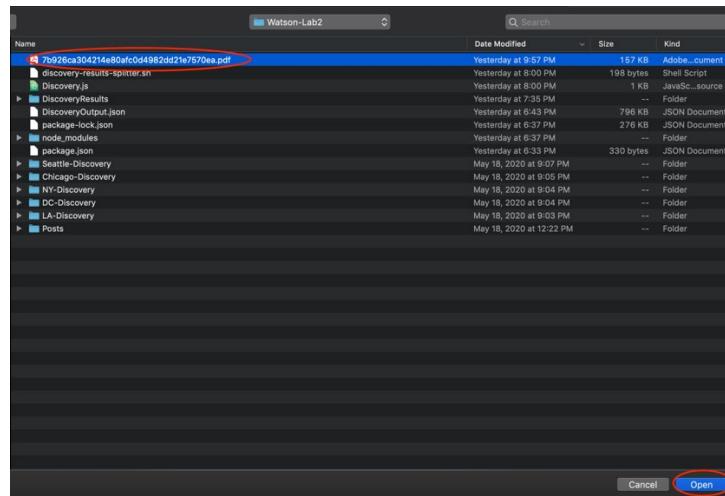
+ Create new

answer

14. You will be asked to select your document again in the Upload documents screen.  
Click **select documents**.



15. Select the pdf article in the Watson-Lab2 folder and click **Open**.



16. In a few seconds, the changes will be applied to the collection and you will be taken back to the Collection overview page.

COVID-19Article

Overview Errors and warnings (0) Search settings

1 document

0 documents failed View details

Created on Last updated 5/24/2020 10:15:40 pm EDT 5/24/2020 10:15:40 pm EDT

Upload documents

Identified 1 field from your data Added 4 enrichments to your data Now you're ready to query!

We have successfully used SDU to train Watson on the structure and format of each page of our PDF document. This will enable us to create queries to search through our document in the next exercise and retrieve relevant passages in order to answer natural language questions.

A final note about SDU: when we have multiple documents in a collection to train, we can label one document and save its corresponding SDU model to use on other similarly-formatted documents.

Before we start creating queries, let's make one change to the configuration of the document.

### 17. Click on **Configure data**.

COVID-19Article

Overview Errors and warnings (0) Search settings

1 document

0 documents failed View details

Created on Last updated 5/24/2020 10:15:40 pm EDT 5/24/2020 10:15:40 pm EDT

Upload documents

### 18. Click on **Manage fields**.

COVID-19Article / Configure data

Identify fields Manage Fields Enrich fields

Identify fields to index

All fields are indexed by default. Switch off any fields you do not want to be indexed. [Learn more](#).

answer	On
author	On
caption	On
footer	On
graphTitle	On
header	On
image	On

Improve query results by splitting your documents

You can split your documents into segments based on fields. Once split, each segment is a separate document that will be enriched, indexed, and returned as a query separately. [Learn more](#).

+ Split document

Apply changes to collection

### 19. In order to improve the results of the queries that we will be creating in the next exercise, we can split the journal article on each occurrence of a specific field. In our case, we will split the article on each occurrence of the text field, which approximately represents each paragraph in the document. Once the split is complete, each paragraph in the article will now be a separate document that will be enriched, indexed and returned as a query result. Click **+ Split document**.

---

**Identify fields to index**

All fields are indexed by default. Switch off any fields you do not want to be indexed. [Learn more.](#)

answer	<input checked="" type="checkbox"/> On
author	<input checked="" type="checkbox"/> On
caption	<input checked="" type="checkbox"/> On

**Improve query results by splitting your documents**

You can split your documents into segments based on fields. Once split, each segment is a separate document that will be enriched, indexed, and returned as a query separately. [Learn more.](#)

**+ Split document**

## 20. Click **Select field** and choose **text**.

**Improve query results by splitting your documents**

You can split your documents into segments based on fields. Once split, each segment is a separate document that will be enriched, indexed, and returned as a query separately. [Learn more.](#)

Split document on each occurrence of

**Select field**

subtitle

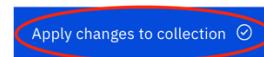
table

table\_of\_contents

**text**

title

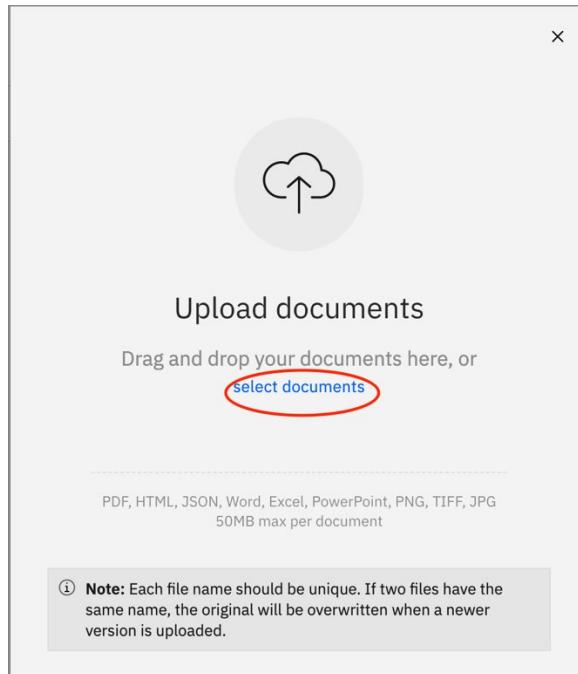
## 21. Click **Apply changes to collection** for the journal article to be split on each occurrence of the **text** field.



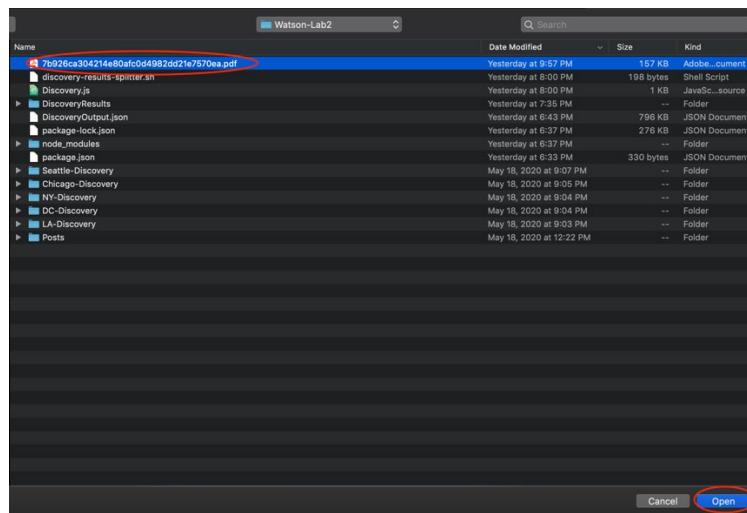
### Improve query results by splitting your documents

You can split your documents into segments based on fields. Once split, each segment is a separate document that will be enriched, indexed, and returned as a query separately. [Learn more.](#)

## 22. You will be prompted to select the journal article. Click **select documents**.



23. Select the journal article and click **Open**.



24. You will be taken back to the overview page where you'll see that the original document has been split into 52 documents. Do not proceed to the next exercise until you see the additional documents (wait approximately a minute and refresh the page if needed).

The screenshot shows the Watson Discovery Overview page for a dataset named 'COVID-19Article'. Key details include:

- Overview:** Shows 52 documents, 0 failed documents, and was created on 5/24/2020 at 10:15:40 pm EDT.
- Identified fields:** caption, text, title.
- Added enrichments:**
  - Entity Extraction: CDC (15), 19 years (7), United States (7), 84 years (6), Atlanta (6).
  - Sentiment Analysis: 6% positive, 67% neutral, 27% negative.
  - Concept Tagging: Medicine (9), Epidemiology (8), United States (7), Ageing (4), Death (6).
  - Category Classification: health and fitness → disease → epidemic.
- Query suggestions:**
  - Documents about CDC as an Organization with a very negative sentiment (Run).
  - Entities of type Organization which have negative sentiment (Run).
  - Documents that contain Medicine, but not Epidemiology (Run).

## Exercise 9: Create and run Natural Language Queries

There are two types of queries that we can create inside of Watson Discovery in order to search through documents – structured queries and natural language queries. Let's start by running a few structured queries on our uploaded document.

Given the NLP enrichments that were applied to our document by default – entity extraction, sentiment analysis, concept tagging and category classification – we can use any combination of these enrichments to produce structured queries. Let's start by creating a structured query using the entity enrichment.

1. Let's navigate to the Query page by clicking on the **magnifying glass icon** on the left-hand side of the screen.

The screenshot shows the Watson Discovery Overview page for the 'COVID-19Article' dataset. The magnifying glass icon on the left sidebar is circled in red.

2. We will be running a sample query to retrieve the most common entity types identified in our document and their top entities.  
Click on **Use a sample query** and select **Most common entity types and their top entities**.

The screenshot shows the Watson Discovery Query builder interface. A dropdown menu is open, listing several query components. The option 'Most common entity types and their top entities' is circled in red.

3. You should immediately see the results of this query on the right-hand side of the screen:

The screenshot shows the Watson Discovery Query results page. The 'Summary' tab is selected. The results section displays a list of aggregated entity types and their counts. The most common entity types are:

- Quantity (82)
  - term(enriched\_text.entities.type) 19 years (7)
  - term(enriched\_text.entities.type) 84 years (6)
  - term(enriched\_text.entities.type) 64 years (5)
  - term(enriched\_text.entities.type) 85 years (5)
  - term(enriched\_text.entities.type) 31% (4)
  - term(enriched\_text.entities.type) 44 years (4)
  - term(enriched\_text.entities.type) 53% (3)
  - term(enriched\_text.entities.type) 54 years (3)
  - term(enriched\_text.entities.type) 60 years (3)
  - term(enriched\_text.entities.type) 65 years (3)
- Location (32)
  - term(enriched\_text.entities.type) United States (8)
  - term(enriched\_text.entities.type) Atlanta (6)
  - term(enriched\_text.entities.type) China (5)
  - term(enriched\_text.entities.type) Washington (3)
  - term(enriched\_text.entities.type) GA (2)
  - term(enriched\_text.entities.type) Geneva (2)
  - term(enriched\_text.entities.type) Switzerland (2)
  - term(enriched\_text.entities.type) U.S. (2)
  - term(enriched\_text.entities.type) Japan (1)
  - term(enriched\_text.entities.type) Wuhan (1)
- Organization (32)
  - term(enriched\_text.entities.type) CDC (15)
  - term(enriched\_text.entities.type) US Department of Health and Human Services

As you can see, the most entity types in our PDF article are Quantity, Location, Organization, Person and EmailAddress and we can also see the top entities pertaining to each entity type.

4. Now instead of using a sample query, let's build our own structured query to determine the top entities with their average sentiment score.  
Click the **trash icon** next to Include analysis of your results

The screenshot shows the Watson Discovery Query editor interface. The 'Include analysis of your results' checkbox is checked. The 'Build in visual mode' button is circled in red. The query code in the editor is:

```
nested(enriched_text.entities).term(enriched_text.entities.type,count:5).term(enriched_text.entities.text)
```

## 5. Click on + Include analysis of your results

The screenshot shows the 'COVID-19Article / Build queries' interface. At the top, there is a header with the title and a 'Build queries' button. Below the header, there is a message: 'Build a query using one or more of these components. [Learn more.](#)' and a 'Use a sample query' button. There are three main options listed: '+ Search for documents', '+ Include analysis of your results' (which is circled in red), and '+ Filter which documents you query'. At the bottom left, there is a 'More options' link, and at the bottom right, there are 'Run query' and 'Close' buttons.

## 6. From the **Field** drop-down, select **enriched\_text.entities.text**

The screenshot shows the 'Include analysis of your results' configuration screen. The 'Output' dropdown is set to 'Top values' and the 'Count' input field is set to '10'. The 'Field' dropdown has 'enriched\_text.entities.text' selected. A red circle highlights the 'enriched\_text.entities.text' option in the dropdown list. Other available fields include 'enriched\_text.entities.disambiguation.name', 'enriched\_text.entities.disambiguation.subtype', 'enriched\_text.entities.sentiment.label', and 'enriched\_text.entities.type'. At the bottom right, there are 'Run query' and 'Close' buttons.

## 7. Click + Add child aggregation

The screenshot shows the 'Include analysis of your results' configuration screen. The 'Output' dropdown is set to 'Top values' and the 'Count' input field is set to '10'. The 'Field' dropdown has 'enriched\_text.entities.text' selected. A red circle highlights the '+ Add child aggregation' button. Other available buttons include '+ Add condition' and '+ Add top-level aggregation'. Below the configuration area, there is a preview of the query: 'term(enriched\_text.entities.text,count:10)'. At the bottom right, there are 'Run query' and 'Close' buttons.

8. Under **Output**, select **Average** and under **Field**, select **enriched\_text.entities.sentiment.score**. Click **Run query** to view the results.

The screenshot shows the 'Include analysis of your results' interface. At the top right are 'Edit in query language' and a trash icon. Below is a table for aggregations:

Output	Field	Count
Top values	enriched_text.entities.text	10
Average	enriched_text.entities.sentiment.score	

Buttons include '+ Add condition', '+ Add top-level aggregation', 'Run query' (circled in red), and 'Close'.

9. You should now be able to see the top entities in the document with their associated average sentiment scores.

The screenshot shows the 'Summary' tab of the results page. At the top right is a 'Train Watson to improve results' link. Below is a 'Query URL' field with the value <https://gateway.watsonplatform.net/discovery/api/v1/environments/9>. The main area is titled 'Aggregations' and lists various entities and their average sentiment scores:

- term(enriched\_text.entities.text) **CDC** (15)
  - **average** (enriched\_text.entities.sentiment.score): 0.08059964556962025
- term(enriched\_text.entities.text) **United States** (8)
  - **average** (enriched\_text.entities.sentiment.score): -0.04446059322033898
- term(enriched\_text.entities.text) **19 years** (7)
  - **average** (enriched\_text.entities.sentiment.score): -0.0229655641025641
- term(enriched\_text.entities.text) **84 years** (6)
  - **average** (enriched\_text.entities.sentiment.score): -0.007347934426229509
- term(enriched\_text.entities.text) **Atlanta** (6)
  - **average** (enriched\_text.entities.sentiment.score): 0.13650737500000001
- term(enriched\_text.entities.text) **US Department of Health and Human Services** (6)
  - **average** (enriched\_text.entities.sentiment.score): 0.14098004761904762
- term(enriched\_text.entities.text) **64 years** (5)
  - **average** (enriched\_text.entities.sentiment.score): -0.008300444444444445
- term(enriched\_text.entities.text) **85 years** (5)
  - **average** (enriched\_text.entities.sentiment.score): -0.02315672413793103
- term(enriched\_text.entities.text) **China** (5)
  - **average** (enriched\_text.entities.sentiment.score): -0.06504192105263158
- term(enriched\_text.entities.text) **31%** (4)
  - **average** (enriched\_text.entities.sentiment.score): -0.027981041666666664

While it is certainly useful to run queries to learn more about the NLP enrichments that were identified in the document, we haven't actually performed a detailed search through the content of the document. Moreover, it would be convenient if we could ask colloquial questions about our document without using the Discovery Query Language. Fortunately, we are able to do this with Natural Language Queries.

10. Click on the trash icon next to **Include analysis of your results** to clear the previous query.

The screenshot shows the IBM Watson Discovery interface. At the top, there's a header with the text "Include analysis of your results" and a "Build in visual mode" button, which is circled in red. Below the header, there's a section titled "Write an aggregation query using the Discovery Query Language". A code input field contains the query: `term(enriched_text.entities.text,count:10).average(enriched_text.entities.sentiment.score)`. To the right of the code field is a blue question mark icon.

## 11. Click on + Search for documents.

The screenshot shows a list of search-related components. At the top, it says "Build a query using one or more of these components. [Learn more.](#)" and "Use a sample query". Below are three items: "+ Search for documents" (circled in red), "+ Include analysis of your results", and "+ Filter which documents you query".

Here we can create natural language queries to ask specific questions about our document. After a quick glance at our article, we can ask the following questions:

- When was COVID-19 declared as a pandemic?
- Which age group in the United States has suffered the highest percentage of severe outcomes?
- How can we protect older adults from COVID-19?
- What was the percentage of fatalities among people that are less than 19 years old?

Let's create natural language queries with each of these questions.

12. Creating a new natural language query is as simple as typing in a question you would like to ask about the document. As soon as the natural language query is created, Watson uses this query to retrieve relevant passages from the document in an attempt to answer the question. Since this Discovery collection has applied NLP enrichments to only the text field of the document, the passages retrieved to answer each query will originate from the body text of the document (and not from the title, graphTitle, caption or footer fields).

Underneath **Use natural language**, type in **When was COVID-19 declared as a pandemic?** Then click the **Run query** button.

Search for documents

Use natural language   Use the Discovery Query Language

When was COVID-19 declared as a pandemic?

+ Include analysis of your results

+ Filter which documents you query

> More options

Run query   Close

13. You should be able to see 5 passages on the right side of the screen that were retrieved in order to answer this question. We will be improving the accuracy of this query in the next exercise.

Train Watson to improve results

**Summary**   **JSON**

Query URL <https://gateway.watsonplatform.net/discovery/api/v1/environments/9>

**Passages**

"In contrast, persons aged ≤19 years appear to have milder COVID-19 illness, with almost no hospitalizations or deaths reported to date in the United States in this age group. Given the spread of COVID-19 in many U.S"

"Although the majority of reported COVID-19 cases in China were mild (81%), approximately 80% of deaths occurred among adults aged ≥60 years; only one (0.1%) death occurred in a person aged ≤19 years (3). In this report, COVID-19 cases in the United States that occurred during February"

"Data from China have indicated that older adults, particularly those with serious underlying health conditions, are at higher risk for severe COVID-19-associated illness and death than are younger persons (3). Although the majority of reported COVID-19 cases in China were mild (81%), approximately"

"CDC COVID-19 Response Team Globally, approximately 170,000 confirmed cases of coronavirus disease 2019 (COVID-19) caused by the 2019 novel coronavirus (SARS-CoV-2) have been reported, including an estimated 7,000 deaths in approximately 150 countries (1)."

"World Health Organization. Coronavirus disease 2019 (COVID-19) situation report-57. Geneva, Switzerland: World Health Organization; 2020.  
[https://www.who.int/docs/default-source/coronavirus/situation-reports/20200317-sitrep-57-covid-19.pdf?sfvrsn=a26922f2\\_2](https://www.who.int/docs/default-source/coronavirus/situation-reports/20200317-sitrep-57-covid-19.pdf?sfvrsn=a26922f2_2)"

14. Repeat steps 13 and 14 in order to generate and run the following natural language queries:

**Which age group in the United States has suffered the highest percentage of severe outcomes?**

**How can we protect older adults from COVID-19?**

**What was the percentage of fatalities among people that are less than 19 years old?**

15. You should see the following passages returned for the aforementioned queries:

Summary	JSON	Train Watson to improve results
Query URL <a href="https://gateway.watsonplatform.net/discovery/api/v1/environments/9">https://gateway.watsonplatform.net/discovery/api/v1/environments/9</a>		
<b>Passages</b>		
<p>"Among 2,449 patients with known age, 6% were aged ≥85, 25% were aged 65–84 years, 18% each were aged 55–64 years and 45–54 years, and 29% were aged 20–44 years (Figure 2). Only 5% of cases occurred in persons aged 0–19 years."</p> <p>"This first preliminary description of outcomes among patients with COVID-19 in the United States indicates that fatality was highest in persons aged ≥85, ranging from 10% to 27%, followed by 3% to 11% among persons aged 65–84 years, 1% to 3% among persons aged 55–64 years, &lt;1% among persons aged 20–54 years, and no fatalities among persons aged ≤19 years."</p> <p>"Among 508 (12%) patients known to have been hospitalized, 9% were aged ≥85 years, 36% were aged 65–84 years, 17% were aged 55–64 years, 18% were 45–54 years, and 20% were aged 20–44 years. Less than 1% of hospitalizations were among persons aged ≤19 years (Figure 2)."</p> <p>"Less than 1% of hospitalizations were among persons aged ≤19 years (Figure 2). The percentage of persons hospitalized increased with age, from 2%–3% among persons aged ≤19 years, to ≥31% among adults aged ≥85 years. (Table)."</p> <p>"aged ≤19 years. Percentages of ICU admissions were lowest among adults aged 20–44 years (2%–4%) and highest among adults aged 75–84 years (11%–31%) (Table)."</p>		

Summary	JSON	Train Watson to improve results
Query URL <a href="https://gateway.watsonplatform.net/discovery/api/v1/environments/9">https://gateway.watsonplatform.net/discovery/api/v1/environments/9</a>		
<b>Passages</b>		
<p>"COVID-19 can result in severe disease, including hospitalization, admission to an intensive care unit, and death, especially among older adults. Everyone can take actions, such as social distancing, to help slow the spread of COVID-19 and protect older adults from severe illness, nonessential air travel, and stay home as much as possible to further reduce the risk of being exposed (7)."</p> <p>". 69 Summary What is already known about this topic? Early data from China suggest that a majority of coronavirus disease 2019 (COVID-19) deaths have occurred among adults aged ≥60 years and among persons with serious underlying health conditions."</p> <p>"CDC COVID-19 Response Team Globally, approximately 170,000 confirmed cases of coronavirus disease 2019 (COVID-19) caused by the 2019 novel coronavirus (SARS-CoV-2) have been reported, including an estimated 7,000 deaths in approximately 150 countries (1)."</p> <p>"CDC. Coronavirus disease 2019 (COVID-19): if you are at higher risk."</p> <p>"What are the implications for public health practice? COVID-19 can result in severe disease, including hospitalization, admission to an intensive care unit, and death, especially among older adults."</p>		

Summary	JSON	Train Watson to improve results
Query URL <a href="https://gateway.watsonplatform.net/discovery/api/v1/environments/9">https://gateway.watsonplatform.net/discovery/api/v1/environments/9</a>		
<b>Passages</b>		
<p>"Less than 1% of hospitalizations were among persons aged ≤19 years (Figure 2). The percentage of persons hospitalized increased with age, from 2%–3% among persons aged ≤19 years, to ≥31% among adults aged ≥85 years. (Table)."</p> <p>"This first preliminary description of outcomes among patients with COVID-19 in the United States indicates that fatality was highest in persons aged ≥85, ranging from 10% to 27%, followed by 3% to 11% among persons aged 65–84 years, 1% to 3% among persons aged 55–64 years, &lt;1% among persons aged 20–54 years, and no fatalities among persons aged ≤19 years."</p> <p>". 69 Summary What is already known about this topic? Early data from China suggest that a majority of coronavirus disease 2019 (COVID-19) deaths have occurred among adults aged ≥60 years and among persons with serious underlying health conditions."</p> <p>"aged ≤19 years. Percentages of ICU admissions were lowest among adults aged 20–44 years (2%–4%) and highest among adults aged 75–84 years (11%–31%) (Table)."</p> <p>"Case-fatality percentages increased with increasing age, from no deaths reported among persons aged ≤19 years to highest percentages (10%–27%) among adults aged ≥85 years (Table) (Figure 2)."</p>		

Don't worry if you are seeing passages that don't quite answer each question – we will be training Watson to return more relevant passages using Relevancy Training in the next exercise.

## Exercise 10: Improve accuracy with Relevancy Training

Now that we have been able to search through our document by answering specific content-related questions, we can proceed to improving the accuracy of the responses retrieved by Watson using another capability of Watson Discovery known as Relevancy Training. Relevancy Training allows us to train Watson to improve passage retrieval results. Let's do this for our four natural language queries from the previous exercise.

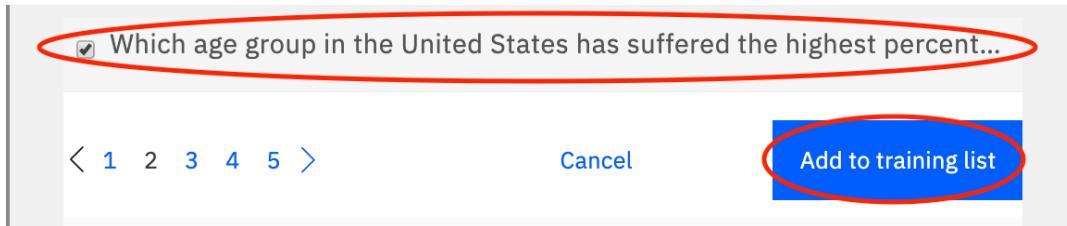
1. Click on **Train Watson to improve results** in the top right corner of the screen.

The screenshot shows the Watson Discovery interface with the 'Summary' tab selected. At the top right, there is a blue button labeled 'Train Watson to improve results' with a red oval highlighting it. Below the tabs, a query URL is displayed: 'Query URL https://gateway.watsonplatform.net/discovery/api/v1/environments/9'. Under the 'Passages' section, there is some placeholder text.

2. Now let's add all of the natural language queries that we created in the previous exercise. Click **+ Add recent queries from Watson Discovery to COVID-19Article**

The screenshot shows the 'Train Watson' interface for the 'COVID-19Article' environment. It displays a sidebar with icons for folder, database, search, and list. The main area shows instructions for rating queries and three buttons: 'Add more queries', 'Rate more results', and 'Add more variety to your ratings'. Below this, there is a section for 'Queries (0)' with a note to 'Train Watson by adding natural language queries and rating the results.' A red oval highlights the '+ Add recent queries from Watson Discovery to COVID-19Article' button. Another button '+ Add a natural language query' is also visible.

3. Search through the list of recent natural language queries and select each of the four queries from the previous exercise. After selecting all 4 queries, click **Add to training list**.



4. After adding all 4 queries to the training list, close the pop-up screen by clicking the X in the right corner of the screen.



5. You should now be able to see all 4 queries listed on the screen:

Watson will learn which are the best results for your queries after you've rated enough.	<a href="#">Add more queries</a>	<a href="#">Rate more results</a>	<a href="#">Add more variety to your ratings</a>
+ Add recent queries from Watson Discovery to COVID-19Article			
+ Add a natural language query			
How can we protect older adults from COVID-19?	<a href="#">Rate results</a>	Not rated yet	
What was the percentage of fatalities among people that are less than 19 years old?	<a href="#">Rate results</a>	Not rated yet	
When was COVID-19 declared as a pandemic?	<a href="#">Rate results</a>	Not rated yet	
Which age group in the United States has suffered the highest percentage of severe outcomes?	<a href="#">Rate results</a>	Not rated yet	

6. Let's work with the first query (How can we protect older adults from COVID-19?) by clicking on the **Rate Results** button on the same row.

+ Add recent queries from Watson Discovery to COVID-19Article			
+ Add a natural language query			
How can we protect older adults from COVID-19?	<a href="#">Rate results</a>	Not rated yet	

7. We can go through all of the documents returned for this query and mark each as Relevant or Not relevant. For this question, we are looking for passages that can answer the question of how to protect older adults from COVID-19. Any passage that answers this question should be marked as **Relevant**; any other passage should be marked as **Not Relevant**.

**relevant.** Make sure that you review all the passages for this query by clicking through the pages of results.

The screenshot shows the Watson interface with the following details:

- Header:** COVID-19Article / Train Watson
- Text:** Watson will learn which are the best results for your queries after you've rated enough.
- Buttons:** Add more queries, Rate more results, Add more variety to your ratings.
- Query:** How can we protect older adults from COVID-19?
- Rating Instructions:** Rate some documents as relevant or not relevant results for this query. Learn more.
- Results:** Three search results are displayed, each with a "View document" link and a brief summary:
  - Severe Outcomes Among Patients with Coronavirus Disease ...**  
Summary: COVID-19 can result in severe disease, including hospitalization, admission to an intensive care unit, and death, especially among older adults. ...  
Action: To help slow the spread of COVID-19 and protect older adults from severe illness, reduce the risk of being exposed to COVID-19, ...  
Rating buttons: Relevant (blue), Not relevant (grey).
  - Severe Outcomes Among Patients with Coronavirus Disease ...**  
Summary: Social distancing is recommended for all ages to slow the spread of the virus, protect the health care system, and help protect vulnerable older adults. ...  
Rating buttons: Relevant (blue), Not relevant (grey).
  - Severe Outcomes Among Patients with Coronavirus Disease ...**  
Summary: Administration for Community Living, 2017 profile of older Americans. ...  
Rating buttons: Relevant (blue), Not relevant (grey).
- Pagination:** 1 2 3 4 5 (with the number 5 circled in red)

- When you are done reviewing all the passages, click on **Back to queries** to return to the list of our natural language queries.

The screenshot shows the Watson interface with the following details:

- Text:** Watson will learn which are the best results for your queries after you've rated enough.
- Buttons:** Add more queries, Rate more results, Add more variety to your ratings.
- Text:** How can we protect older adults from COVID-19?
- Rating Instructions:** Rate some documents as relevant or not relevant results for this query. Learn more.

A red circle highlights the "Back to queries" button.

- You should have about 3 relevant results and 49 not relevant results for the first query. Click on the **Rate Results** button next to What was the percentage of fatalities among people that are less than 19 years old?

The screenshot shows the Watson interface with the following details:

- Section:** Queries (4)
- Text:** Train Watson by adding natural language queries and rating the results. Learn more.
- Buttons:** Add recent queries from Watson Discovery to COVID-19Article, Add a natural language query.
- Query:** How can we protect older adults from COVID-19?
- Rating Buttons:** Rate results (blue), 3 relevant, 49 not relevant.
- Query:** What was the percentage of fatalities among people that are less than 19 years old?
- Rating Buttons:** Rate results (blue), Not rated yet.

10. Review all of the passages for this second query and tag all of the passages that mention the fatality rate for people  $\leq 19$  years old as **Relevant** and anything else as **Not relevant**.

COVID-19Article / Train Watson

Watson will learn which are the best results for your queries after you've rated enough.

Add more queries    Rate more results    Add more variety to your ratings.

[Back to queries](#)

What was the percentage of fatalities among people that are less than 19 years old?

Rate some documents as relevant or not relevant results for this query. [Learn more.](#)

**Severe Outcomes Among Patients with Coronavirus Disease ...**

View document

... Less than 1% of hospitalizations were among persons aged  $\leq 19$  years (Figure 2). The percentage of persons hospitalized increased with age, from 2%-3% among persons aged  $\leq 19$  years, to 13% among adults aged  $\geq 65$  years. ...

... Among 508 (2%) patients known to have been hospitalized, 9% were aged 85 years, 36% were aged 65-84 years, 17% were aged 55-64 years, 18% were 45-54 years, and 20% were aged 20-44 years. Less than 1% of hospitalizations were among persons aged  $\leq 19$  years (Figure 2). ...

Show less

Relevant    Not relevant

**Severe Outcomes Among Patients with Coronavirus Disease ...**

View document

... This first preliminary description of outcomes among patients with COVID-19 in the United States indicates that fatality was highest in persons aged  $\geq 65$ , ranging from 1.0% to 27%, followed by 3% to 11% among persons aged 45-64 years, 1% to 3% among persons aged 55-64 years, <1% among persons aged 20-54 years, and no fatalities among persons aged  $\leq 19$  years. ...

Relevant    Not relevant

**Severe Outcomes Among Patients with Coronavirus Disease ...**

View document

... Case-fatality percentages increased with increasing age, from no deaths reported among persons aged  $\leq 19$  years to highest percentages (10%-27%) among adults aged  $\geq 85$  years (Table) (Figure 2). ...

Relevant    Not relevant

11. When you are done reviewing all the passages, click on **Back to queries** to return to the list of our natural language queries.

COVID-19Article / Train Watson

Watson will learn which are the best results for your queries after you've rated enough.

Add more queries    Rate more results    Add more variety to your ratings.

[Back to queries](#)

12. Click on the **Rate Results** button next to When was COVID-19 declared as a pandemic?

Queries (4)

Train Watson by adding natural language queries and rating the results. [Learn more.](#)

+ Add recent queries from Watson Discovery to COVID-19Article

+ Add a natural language query

---

How can we protect older adults from COVID-19?  Rate results   3 relevant 49 not relevant

---

What was the percentage of fatalities among people that are less than 19 years old?  Rate results   4 relevant 48 not relevant

---

When was COVID-19 declared as a pandemic?  Rate results   Not rated yet

13. Review all of the passages for this third query and tag all of the passages that mention exactly when COVID-19 was declared as a pandemic as **Relevant** and anything else as **Not relevant**.

Watson will learn which are the best results for your queries after you've rated enough.

[Add more queries](#) [Rate more results](#) [Add more variety to your ratings](#)

[Back to queries](#)

When was COVID-19 declared as a pandemic?

Rate some documents as relevant or not relevant results for this query. [Learn more](#).

**Severe Outcomes Among Patients with Coronavirus Disease 2019 (CO...**

[View document](#)

.. On March 11, 2020, the World Health Organization declared the COVID-19 outbreak a pandemic (2). Data from China have indicated that most adults, particularly those with serious underlying health conditions, are at higher risk for severe COVID-19-associated illness and death than are younger persons (3)...

.. Data from China have indicated that older adults, particularly those with serious underlying health conditions, are at higher risk for severe COVID-19-associated illness and death than are younger persons (3). Although the majority of reported COVID-19 cases in China were among adults aged <60 years (81%),...

.. Although the majority of reported COVID-19 cases in China were among adults aged <60 years (81%), approximately 80% of deaths occurred among adults aged ≥60 years; only one (0.1%) death occurred in a person aged ≤19 years (3). In this report, COVID-19 cases in the United States that occurred during February ...

Show less

Relevant  Not relevant

**Severe Outcomes Among Patients with Coronavirus Disease 2019 (CO...**

[View document](#)

.. Approximately 49 million U.S. persons are aged ≥65 years (9), and many of these adults, who are at risk for severe COVID-19-associated illness, might depend on services and support to maintain their health and independence. To prepare for potential COVID-19 illness among persons at high risk, family members and caregivers of older adults should...

.. Approximately 49 million U.S. persons are aged ≥65 years (9), and many of these adults, who are at risk for severe COVID-19-associated illness, might depend on services and support to maintain their health and independence. To prepare for potential COVID-19 illness among persons at high risk, family members and caregivers of older adults should...

.. To prepare for potential COVID-19 illness among persons at high risk, family members and caregivers of older adults should know what medications they are taking and ensure that food and required medical supplies are available. Long-term care facilities should be particularly vigilant to prevent the introduction and spread of COVID-19 (10). In addition, clinicians who ...

Show less

Relevant  Not relevant

**Severe Outcomes Among Patients with Coronavirus Disease 2019 (CO...**

[View document](#)

.. CDC COVID-19 Response Team Globally, approximately 170,000 confirmed cases of coronavirus disease 2019 (COVID-19) caused by the 2019 novel coronavirus (SARS-CoV-2) have been reported, including an estimated 7,000 deaths in approximately 130 countries (1)...

Relevant  Not relevant

14. When you are done reviewing all the passages, click on **Back to queries** to return to the list of our natural language queries.

[COVID-19Article](#) / Train Watson

Watson will learn which are the best results for your queries after you've rated enough.

[Add more queries](#) [Rate more results](#) [Add more variety to your ratings](#)

[Back to queries](#)

15. Click on the **Rate Results** button next to Which age group in the United States has suffered the highest percentage of severe outcomes?

**Queries (4)**

Train Watson by adding natural language queries and rating the results. [Learn more](#).

+ Add recent queries from Watson Discovery to COVID-19Article  
+ Add a natural language query

How can we protect older adults from COVID-19? [Rate results](#) 3 relevant 49 not relevant

What was the percentage of fatalities among people that are less than 19 years old? [Rate results](#) 4 relevant 48 not relevant

When was COVID-19 declared as a pandemic? [Rate results](#) 1 relevant 51 not relevant

Which age group in the United States has suffered the highest percentage of severe outcomes? [Rate results](#) Not rated yet

16. Review all of the passages for this final query and tag all passages that mention the age group suffering the most from severe outcomes as **Relevant** and anything as **Not relevant**.

Watson will learn which are the best results for your queries after you've rated enough.

[Add more queries](#) [Rate more results](#) [Add more variety to your ratings](#)

[Back to queries](#)

Which age group in the United States has suffered the highest percentage of severe outcomes?

Rate some documents as relevant or not relevant results for this query. [Learn more](#).

**Severe Outcomes Among Patients with Coronavirus Disease 2019 (CO...**

[View document](#)

.. In contrast, persons aged <10 years appear to have milder COVID-19 illness, with almost no hospitalizations or deaths reported to date in the United States in this age group. Given the spread of COVID-19 in many U.S. ....

.. Discussion Since February 12, 4,226 COVID-19 cases were reported in the United States; 31% of cases, 45% of hospitalizations, 53% of ICU admissions, and 80% of deaths occurred among adults aged ≥65 years with the highest percentage of severe outcomes among persons aged ≥85 years. These findings are similar to data from China, which indicated >80% of deaths occurred among persons aged ≥60 years. ....

Show less

[Relevant](#) [Not relevant](#)

**Severe Outcomes Among Patients with Coronavirus Disease 2019 (CO...**

[View document](#)

.. fatality percentages for reported COVID-19 cases, by age group — United States, February 12–March 16, 2020 Age group (yrs) (no. of cases) % Hospitalization ICU admission Case-fatality 0–19 (123) 1.6 ...

.. Overall, 31% of cases, 45% of hospitalizations, 53% of ICU admissions, and 80% of deaths associated with COVID-19 were among adults aged ≥65 years with the highest percentage of severe outcomes among persons aged ≥85 years. In contrast, no ICU admissions or deaths were reported among persons aged ≥19 years. ....

.. In this report, COVID-19 cases in the United States that occurred during February 12–March 16, 2020 and severity of disease (hospitalization, admission to intensive care unit [ICU], and death) were analyzed by age group. As of March 16, a total of 4,226 COVID-19 cases in the United ...

Show more

[Relevant](#) [Not relevant](#)

< 1 2 3 4 5 >

17. When you are done reviewing all the passages, click on **Back to queries** to return to the list of our natural language queries.

[COVID-19Article](#) / Train Watson

Watson will learn which are the best results for your queries after you've rated enough.

[Add more queries](#) [Rate more results](#) [Add more variety to your ratings](#)

[Back to queries](#)

**Queries (4)**  
Train Watson by adding natural language queries and rating the results. [Learn more](#).

- + Add recent queries from Watson Discovery to COVID-19Article
- + Add a natural language query

How can we protect older adults from COVID-19?	<a href="#">Rate results</a> 3 relevant 49 not relevant
What was the percentage of fatalities among people that are less than 19 years old?	<a href="#">Rate results</a> 4 relevant 48 not relevant
When was COVID-19 declared as a pandemic?	<a href="#">Rate results</a> 1 relevant 53 not relevant
Which age group in the United States has suffered the highest percentage of severe outcomes?	<a href="#">Rate results</a> 8 relevant 44 not relevant

Now that we have completed Relevancy Training for our natural language queries, we should be able to get more accurate results when we run one of these trained queries in the future. This is especially useful for conversational applications containing a virtual agent (such as Watson Assistant in Lab 3), which require accurate real-time responses to user inquiries that are often more detailed and long-tail.

**You have completed Lab 2!**