# Institute of Computer Technology
B. Tech Computer Science and Engineering
Sub: Cognitive Computing (2CSE70E23)

PRACTICAL 8: WATSON KNOWLEDGE STUDIO

Aman got a task to extract information from news portals. However, unable to find a specific outcome. So for that, he needs to train a custom ML model to extract meaningful insights from news data. So he will use Watson Knowledge studio.

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

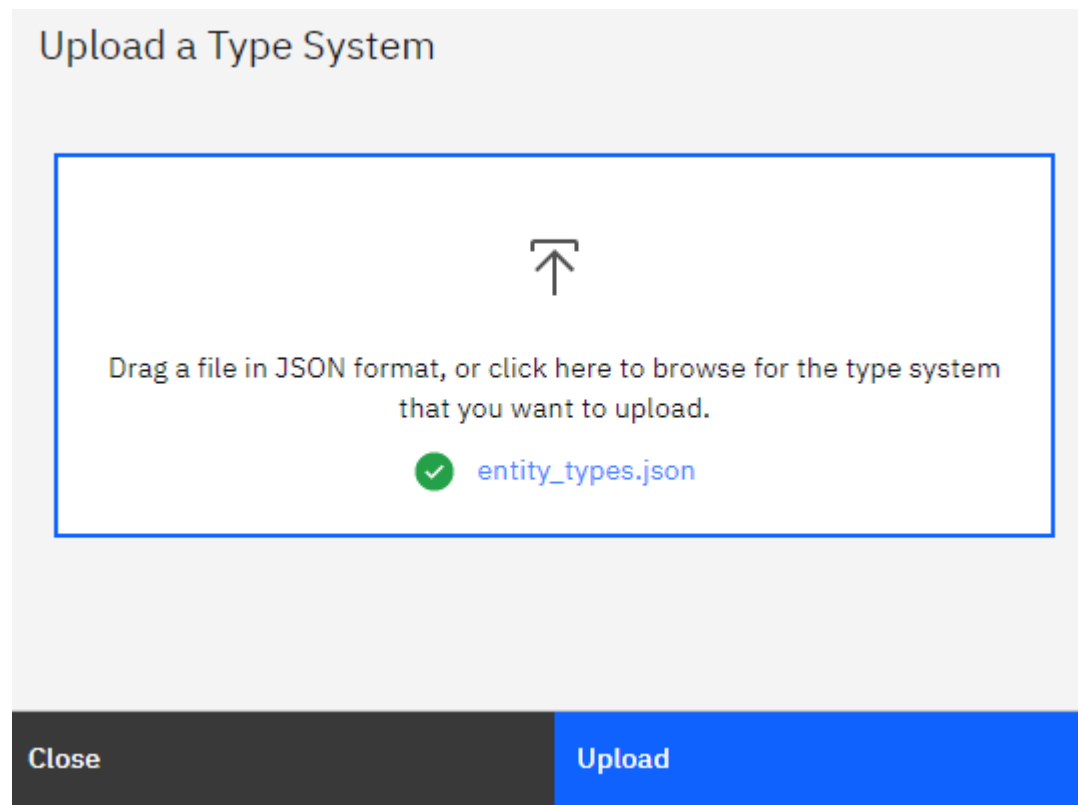Open Knowledge Studio in IBM Cloud in the same region as Watson Discovery.
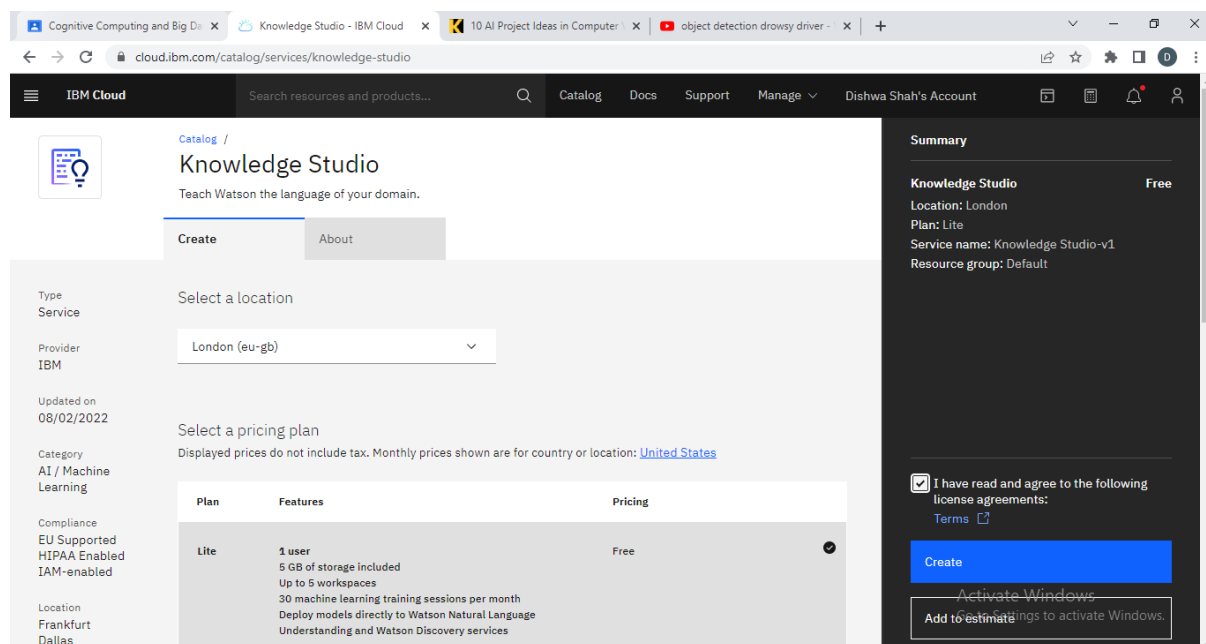
We can't just use NLP to detect entities since we have a lot of different varieties of entities which make it hard to detect them.
Knowledge Studio is used to create a completely customized model for us according to our needs. Then, we will use API of Watson Discovery so that it can work o]as our project.
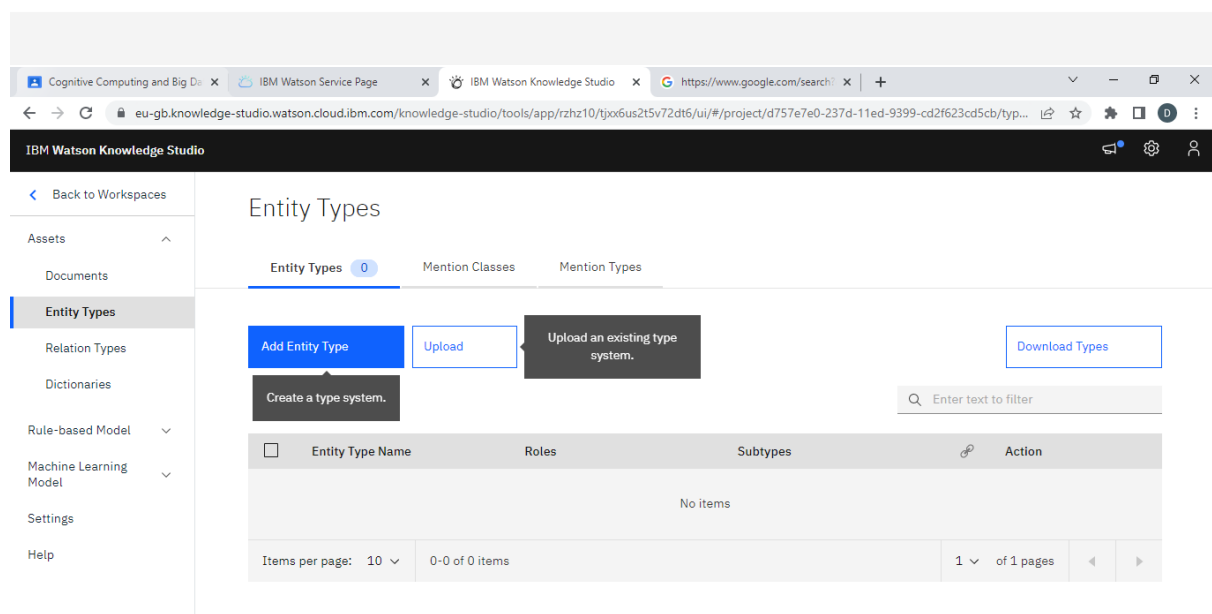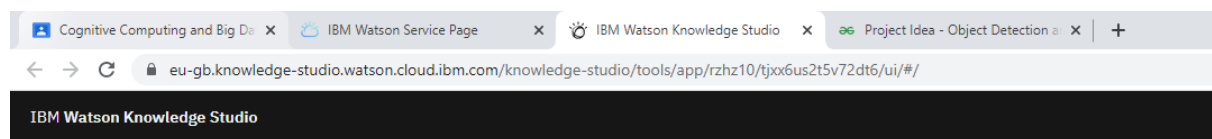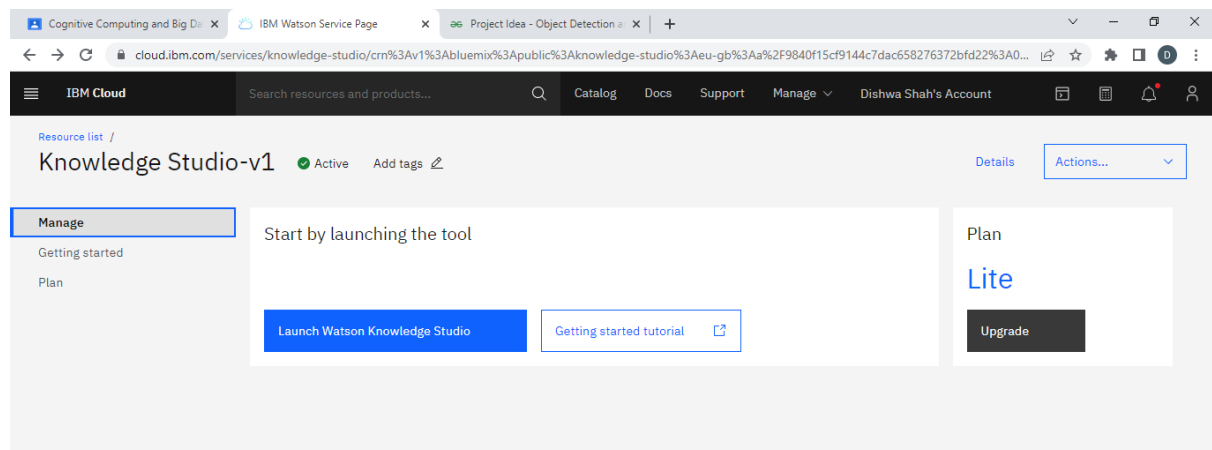
The left portion showing assets and models is to customize your project.

Our file has entities and relations.

Upload a Type System

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

✓ entity_types.json

Close   **Upload**

IBM Watson Knowledge Studio

Back to Workspaces

Assets ∧
  Documents
  Entity Types
  Relation Types
  Dictionaries

Rule-based Model ∨

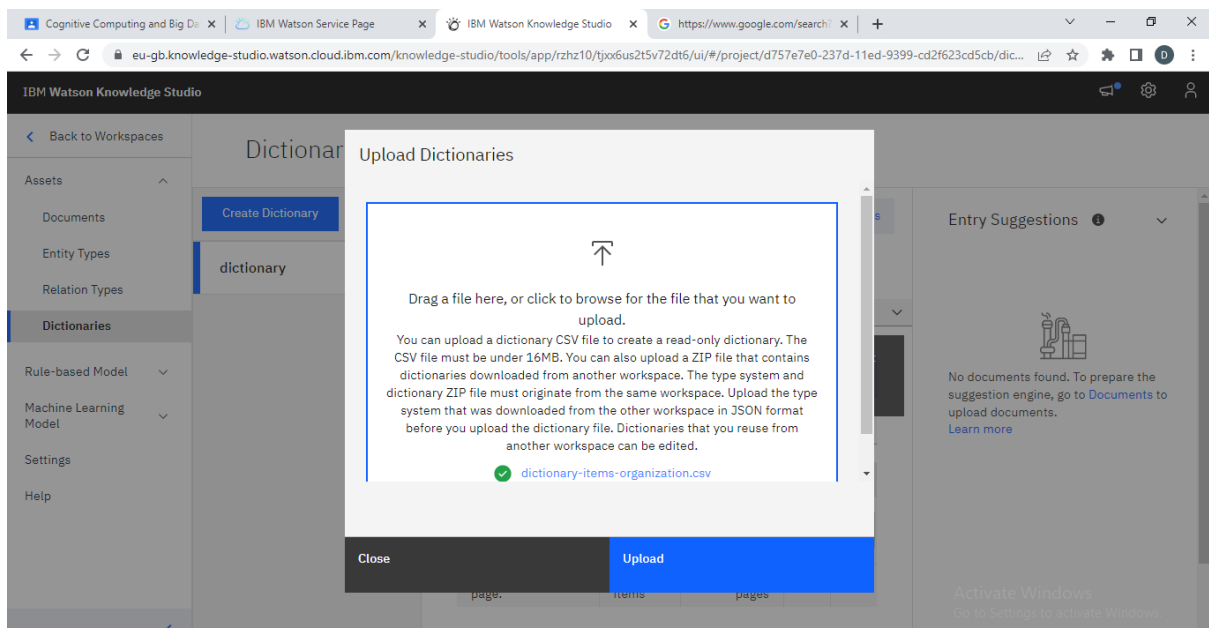Machine Learning Model ∨

Settings

Help

Entity Types

Entity Types 52	Mention Classes	Mention Types

Add Entity Type   Upload   Download Types

🔍 Enter text to filter

	Entity Type Name	Roles	Subtypes	🔗	Action
☐	ORDINAL	ORDINAL ANIMAL Show all...			Edit Delete
☐	MONEY	MONEY AWARD	OTHER UNSPECIFIED	🔗	Edit Delete

Items per page: 10 ∨   1-10 of 52 items   1 of 6 pages   ▶

Entity Types

Entity Types 52	Mention Classes	Mention Types

This is the type of entity (customized) that will not be present by default. This is why we add entities according to our needs.

You can also add more entities by clicking on Add Entity Type.

These are the relations between entities, which we uploaded through JSON file along with the entities.

Create a new dictionary and upload the dictionary file in it.

For entity type ORGANIZATION

Pre-annotation: labelling
These are the 4 ways to do it.
We can only see dictionaries since we only added dictionaries.

Run pre-annotators:

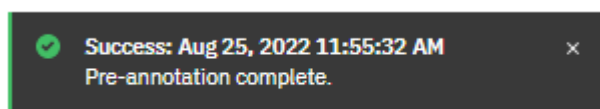## Run Pre-annotators

Back   Run

Select document sets

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

☐ Wipe previous pre-annotation results

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

	Document set	Documents	Pre-annotated documents	Human annotated documents
☑	documents-new.csv	14	📄 0	📄 0

🔵 Running pre-annotators...Time elapsed: 10 seconds

✅ **Success: Aug 25, 2022 11:55:32 AM**   ×
Pre-annotation complete.

d make the necessary changes to enable the pre-annotator

Now, the organizations' names in the documents will be highlighted since the pre-annotation has been completed.

We can see that only 5/14 documents have been annoted because only those had organization entities.

Here, we are not collaborating, so we will perform all the tasks that are usually divided amongst people.

Add task.

Set the task name and the due date.
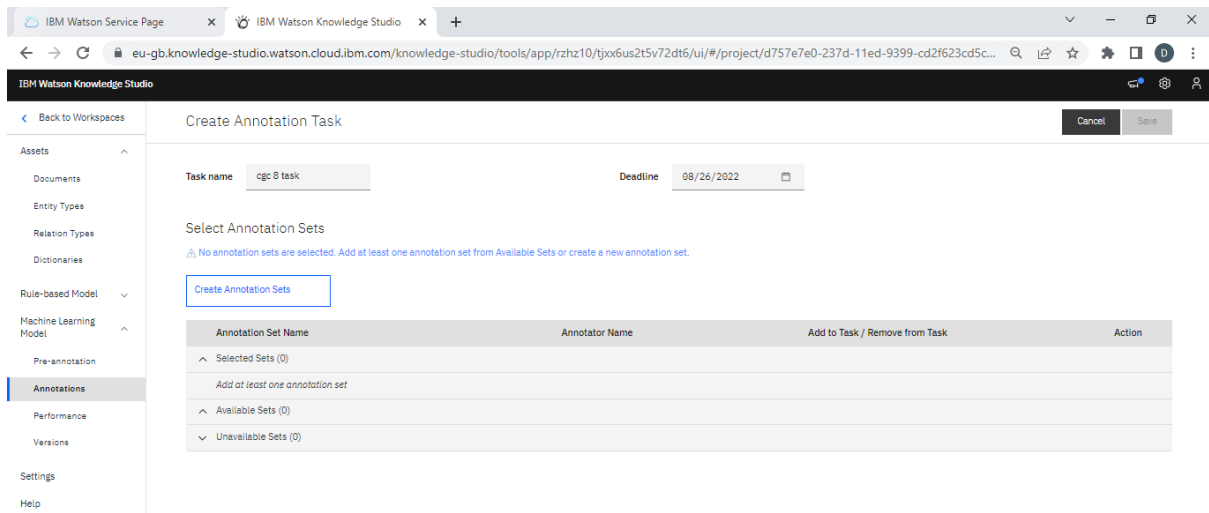
Create annotation set.

Overlapping will be reduced if you want to divide the work between the team members. If we keep 100% overlap, all members will get all the files. Here, we will keep 100% since we are working alone.

We created two sets to see how this works in collaboration.

Click on Save.

Click on Annotate of SET_1.

Click and open the third document.
Now we can see the content of the document and on the right side, we can see the entities we added.
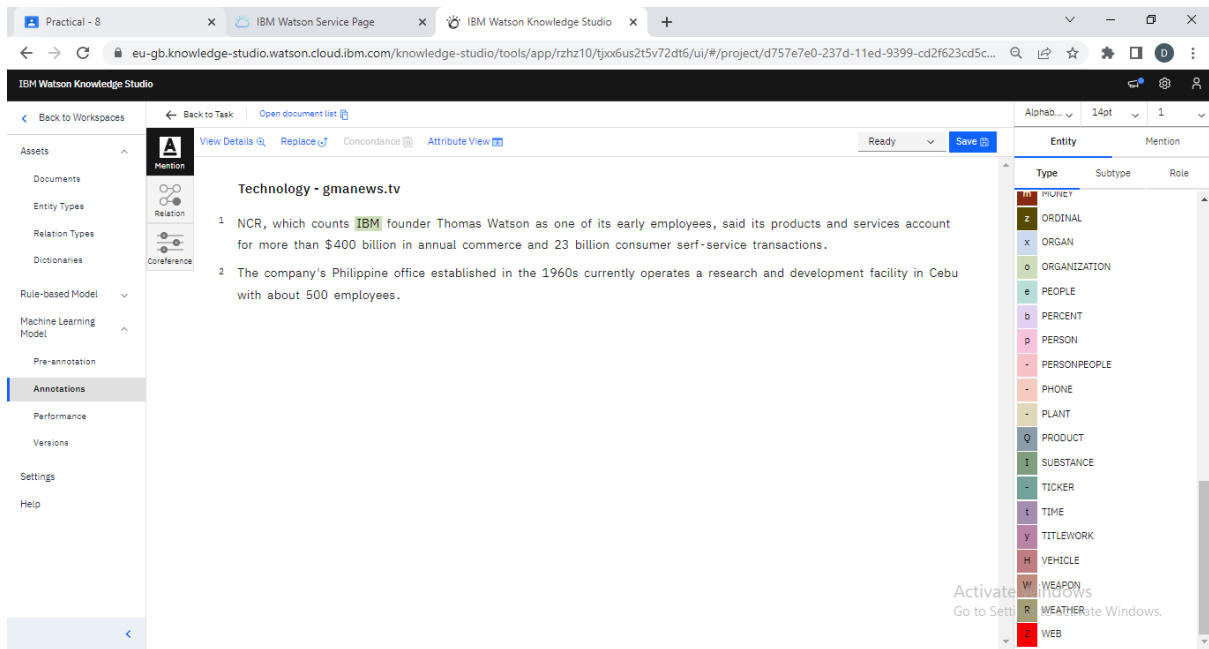
Not all the organizations will be highlighted, only the ones mentioned in our dictionary will (according to the colors on the right side).

List of the documents where the annotations have been done:
Brighthand.com
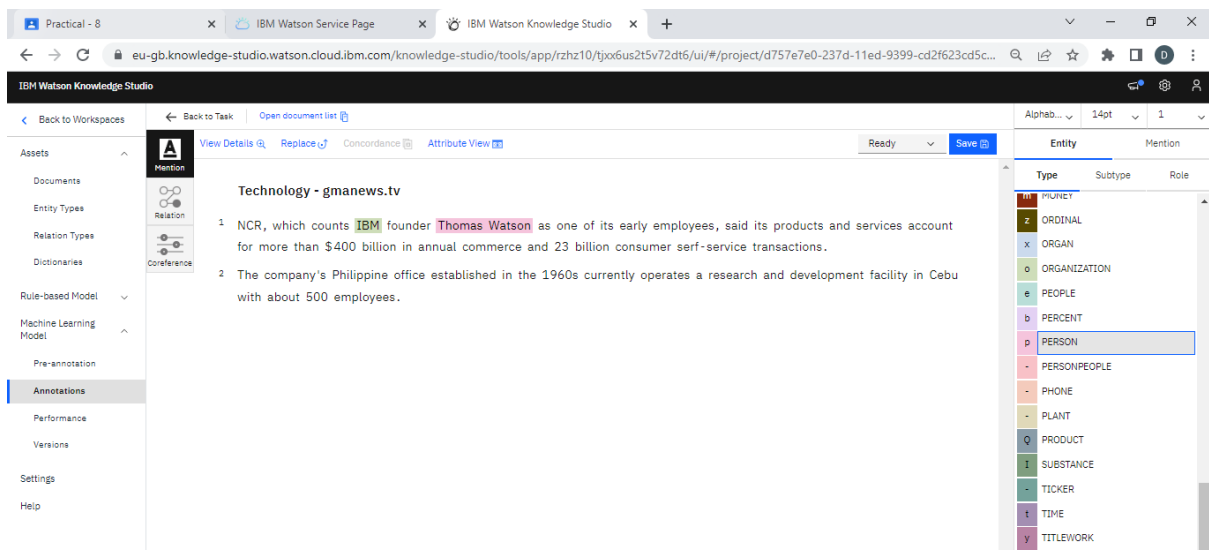Informationweek.com
Computerworld.com
Computerworlds.com
Foxnews.com
Gmanews.tv

If you want to add any annotations, select that word and click on the entity on the right side.
This is called manual annotations.

Use CTRL Z to undo any changes.
Save the work when you are ready. These changes will be applied to all the documents.

Now, select the relations tab. You will see the relations and on the right side you will see all the relation types that we added.

Click on person and then click on organization and then select the relation type. Click on Save once you are ready.

Submit all documents.

Now, we will annotate SET_2.

Add the same entity as in SET_1, but give a different relation.

Save the changes and submit all documents.

You can either Accept or Reject after reviewing. Here, we will accept.

We have a conflict.

← | cgc 8 task   Edit

Deadline: 08/26/2022

Annotations added to annotation sets are not considered ground truth until the annotation sets are submitted and accepted.
When an annotation set is accepted, documents that are annotated in only one annotation set immediately become ground truth.
Documents that are annotated in two or more annotation sets become overlapping documents that will become ground truth after conflicts are resolved.

| Refresh | Calculate Inter-Annotator Agreement | Check Overlapping Documents for Conflicts | Apply Type System Updates |

Annotation Set Name	Annotator Name	Status	Action
SET_2	Dishwa Shah	IN CONFLICT	View
SET_1	Dishwa Shah	IN CONFLICT	View

We can now check for overlapping documents for conflicts.

← Back to cgc 8 task Task | Overlapping Documents in Task cgc 8 task

These documents are included in two or more annotation sets. Click Check for Conflicts to open a document and check whether different human annotators applied different annotations to it. Resolve any inconsistencies that you find. If you know that no conflicts exist, click Accept to add the document to the ground truth.

Document Name	Annotation Set Name	Action
Technology - gmanews.tv	SET_2, SET_1	Check for Conflicts Accept
Technology - gamasutra.com	SET_2, SET_1	Check for Conflicts Accept

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

☑ SET_2(dishwashah19@gnu.ac.in)
☑ SET_1(dishwashah19@gnu.ac.in)
☐ Current Ground Truth

| Cancel | Check for Conflicts |

Now, click on the relation conflicts to resolve them.

Accept the other one.

Save your work. Click on promote to ground truth if you want.

Resolve all such conflicts.

← | cgc 8 task   Edit

Deadline: 08/26/2022

Annotations added to annotation sets are not considered ground truth until the annotation sets are submitted and accepted.
When an annotation set is accepted, documents that are annotated in only one annotation set immediately become ground truth.
Documents that are annotated in two or more annotation sets become overlapping documents that will become ground truth after conflicts are resolved.

Refresh | Calculate Inter-Annotator Agreement | Check Overlapping Documents for Conflicts | Apply Type System Updates

Annotation Set Name	Annotator Name	Status	Action
SET_2	Dishwa Shah	COMPLETED	View
SET_1	Dishwa Shah	COMPLETED	View

The task is now completed.

Now, in the Performance tab, select Train and Evaluate.

Click on Train and evaluate.

Create version and deploy model.

Click on Deploy. We will deploy it to Discovery.

Region London. Service Watson Discovery. Click on Deploy.

You will get a model id (f2f3dfa5-39e4-46e9-a59e-b9a8cd6e2868). Save that model id so that you can use that later in Watson discovery under Enrichments option.

Deployment Started.

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

Model ID: f2f3dfa5-39e4-46e9-a59e-b9a8cd6e2868

You can view documentation to learn how to implement the deployed model into your application.

OK

IBM Watson Knowledge Studio

< Back to Workspaces

Assets ∧
  Documents
  Entity Types
  Relation Types
  Dictionaries
Rule-based Model ∨
Machine Learning Model ∧
  Pre-annotation
  Annotations
  Performance
  Versions
Settings
Help

Versions

Machine Learning Model

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

Go to Pre-annotation page

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

Export current model

Version History and Deployment

Version	Base	Creation Date	Entity Scores	Relation Scores	Description	Action
1.1		Current Version	1.00 (1.00 / 1.00)	N/A		Create Version
1.0		08/25/2022	1.00 (1.00 / 1.00)	N/A	cgc 8	Promote Delete Deploy
	▼ Deployed Models (1)					
	Model ID: f2f3dfa5-39e4-46e9-a59e-b9a8cd6e2868		Service ID: 4c803f0c-9591-4f7b-bede-d57c17937873			Undeploy Status

Go to Watson Discovery and create a new collection by Uploading your own data:

| IBM Watson Service Page ✕ | IBM Watson Discovery - Manage ✕ | IBM Watson Knowledge Studio ✕ | New Tab ✕ | + |

← → C   🔒 eu-gb.discovery.watson.cloud.ibm.com/regions/eu-gb/services/crn%3Av1%3Abluemix%3Apublic%3Adiscovery%3Aeu-gb%3Aa%2F9840f15cf

≡   IBM **Watson Discovery**

Manage data

Collections of your private data and pre-enriched data to configure and query against. Learn more.

Create a new data collection   Upload your own data ⬏   Connect a data source ⚲

PRE-ENRICHED DATA
Watson Discovery News

News sources: English ∨

PRIVATE DATA   🗑
cgc 7

Indexed on 8/18/2022

Give your collection a name and Create:

Now, upload your news article. Click on Add enrichments.

Again, click on Add enrichments:
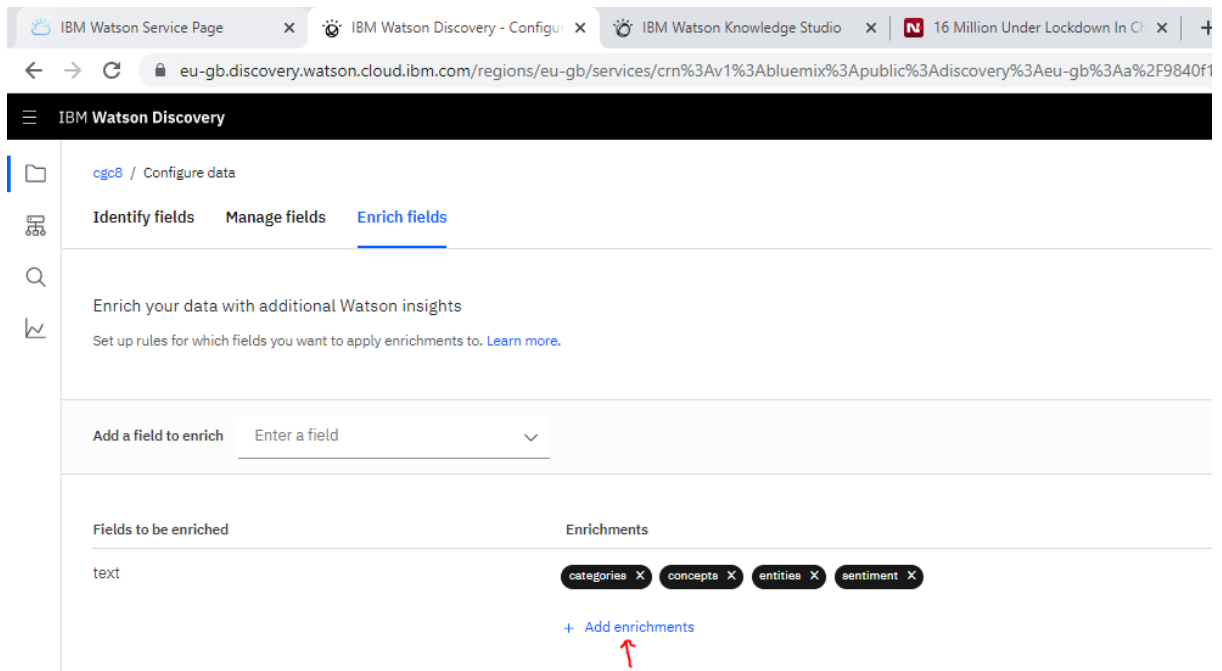
Enter your model id from Knowledge Studio and Apply:
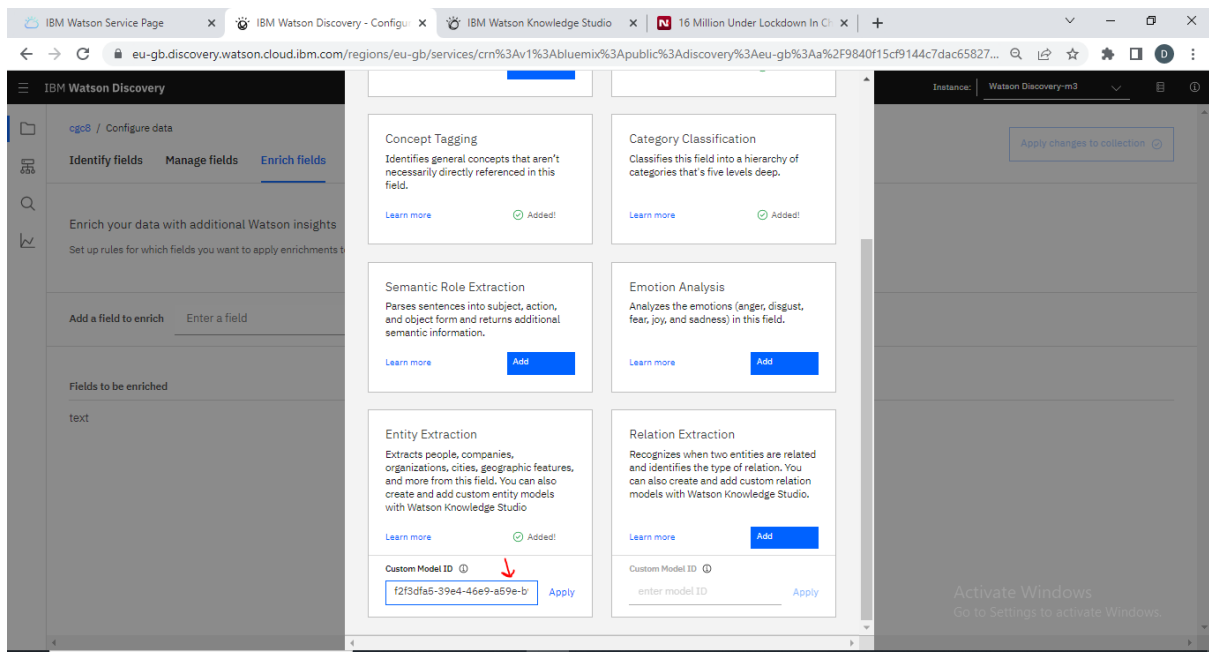
Now, click on Apply changes to the collection. It will again ask you to upload your data. Upload it. Now, our model is deployed and the changes in entities will be updated:

You can now see the entities: