Created **entities and relations** workspace:
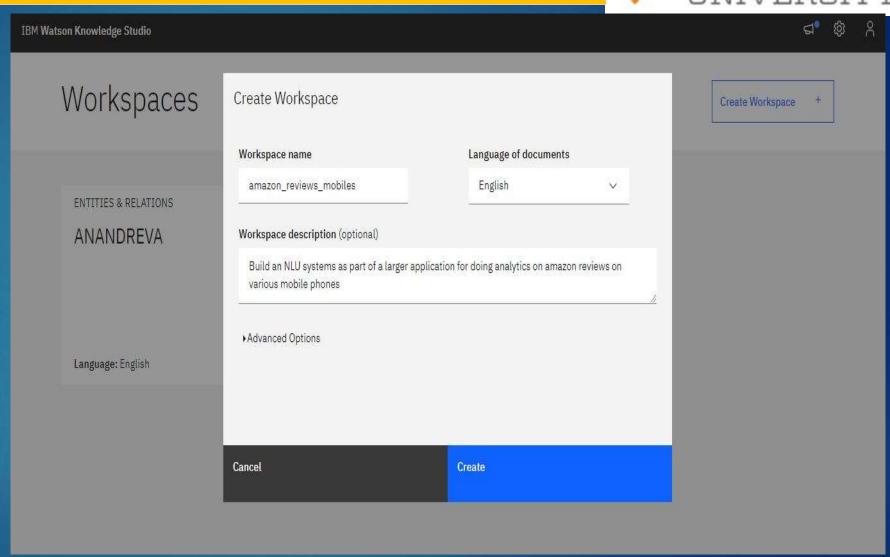
1. **Logged in to IBM Cloud.**
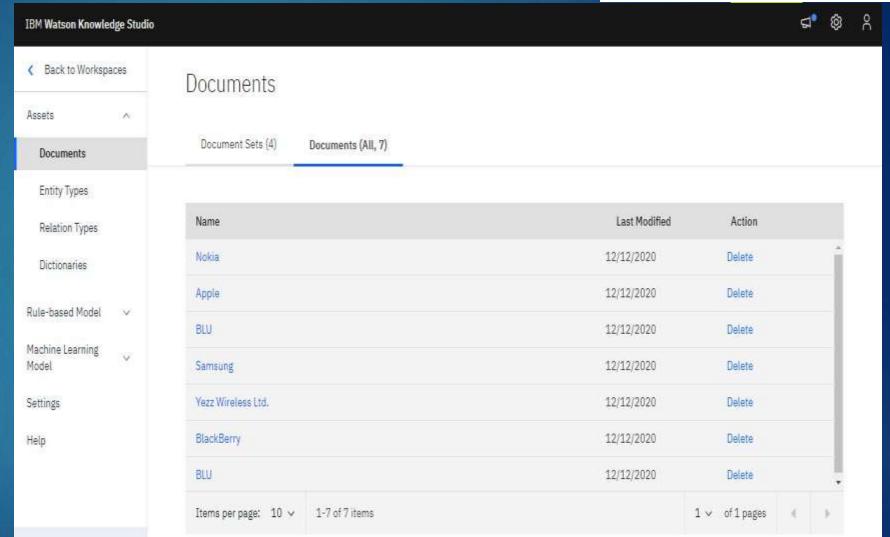
2. **Created a Watson Knowledge Studio instance.**

3. **Created entities and relations workspace" to create my new workspace.**

4. **Also filled the required details meant to create my workspace.**

IBM Watson Knowledge Studio

Workspaces

Create Workspace

ENTITIES & RELATIONS

ANANDREVA

Language: English

Create Workspace +

Workspace name
amazon_reviews_mobiles

Language of documents
English

Workspace description (optional)
Build an NLU systems as part of a larger application for doing analytics on amazon reviews on various mobile phones

▶ Advanced Options

Cancel

Create

Getting started with training corpus:

we had to Build an NLU systems as part of a larger application for doing analytics on amazon reviews on various mobile phones. obtained one dataset from amazon reviews on Amazon_Unlocked_Mobile and Build my own corpus accordingly.

IBM Watson Knowledge Studio

< Back to Workspaces

Assets ∧

Documents

Entity Types

Relation Types

Dictionaries

Rule-based Model ∨

Machine Learning Model ∨

Settings

Help

Documents

Document Sets (4)   Documents (All, 7)

Name	Last Modified	Action
Nokia	12/12/2020	Delete
Apple	12/12/2020	Delete
BLU	12/12/2020	Delete
Samsung	12/12/2020	Delete
Yezz Wireless Ltd.	12/12/2020	Delete
BlackBerry	12/12/2020	Delete
BLU	12/12/2020	Delete

Items per page: 10 ∨   1-7 of 7 items   1 ∨ of 1 pages < >
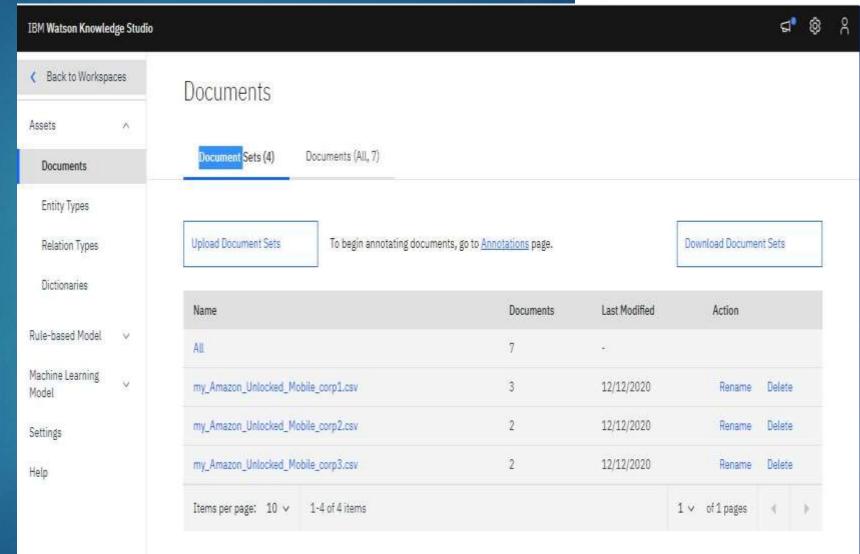
Building my own corpus:

Almost all NLP systems which are adapted to a specific domain are data driven and hence dependent on a set of documents which can be used for training, testing and evaluating the models. Uploaded documents to do exploratory annotation before finalizing type system.

Ideally, documents should be representative of the document source that we wish to analyze.

However for the sake of demonstration as part of our lab assignment I kept my corpus a bit limited only to explain the basic steps that I may need to Capture all the steps as screenshot and demonstrate creating a type system and Building model by doing human annotation in WKS.

REVA
UNIVERSITY

A type system defines the entities and relationships between entities that matter to us. The type system controls how content can be annotated by defining the types of entities that can be labeled and how relationships among different entities can be labeled.
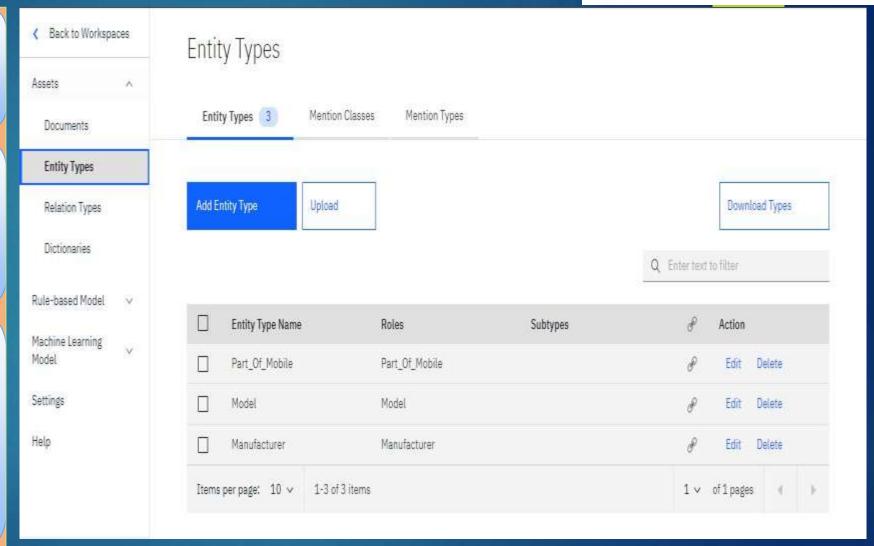
Created the Type System.WKS helps to identify and extract the entities and the relations among the entities. The set of entity types and relation types is called a Type System.

Adding New Entity types:

To add new entity types, clicked on "Entity Types" link under "Assets" menu on the left. Clicked on "Add Entity Type" on the Entity Types tab.

2. Added the entity type "Part_Of_Mobile" and hit the return key. we will see a new entry for Part_Of_Mobile. Similarly added the remaining entity types Model and Manufacturer.

3. Edited the entity types Model,Manufacturer,Part_Of_Mobile and added their respective roles. E.g. to edit Model entity type, I checked the entity type and clicked on the "Edit" link. In the "Select a role" drop down menu, picked "Model" in order to add a role to Model entity type. Similarly added roles for Manufacturer and Part_Of_Mobile.
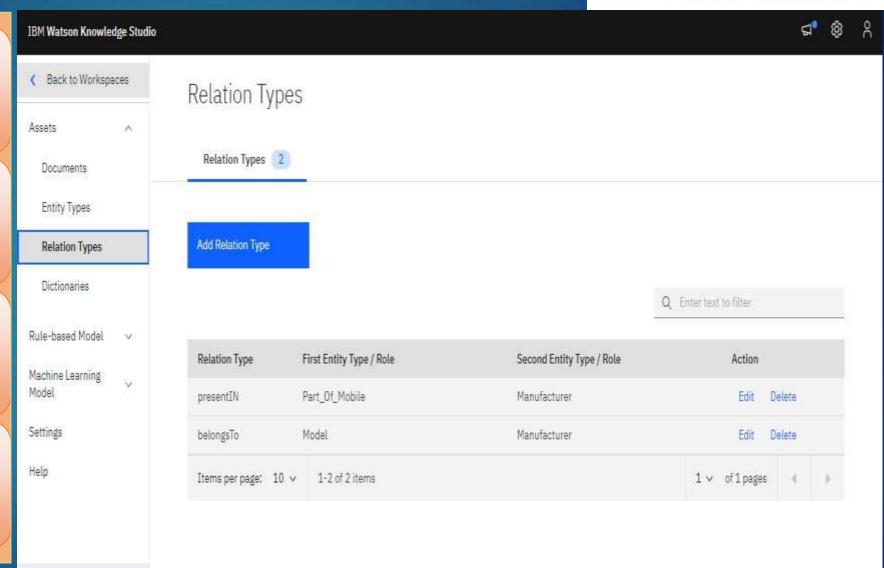
Adding New Relation types:

To add new relation types, click on "Relation Types" link under "Assets" menu on the left. Click on "Add Relation Type" on the Relation Types tab.

Typed the relation type name as "presentIN" and chose the first entity type as Part_Of_Mobile and the second entity type as Manufacturer .

Similarly, added a relation type "belongsTo" and chose the first entity type as Model and the second entity type Manufacturer.

Once the relation types have been added, the Relation Types tab should look like as shown in the snapshots.

IBM Watson Knowledge Studio

< Back to Workspaces

Relation Types

Assets

Documents

Entity Types

Relation Types

Dictionaries

Rule-based Model

Machine Learning Model

Settings

Help

Relation Types 2

Add Relation Type

Q Enter text to filter

Relation Type	First Entity Type / Role	Second Entity Type / Role	Action
presentIN	Part_Of_Mobile	Manufacturer	Edit Delete
belongsTo	Model	Manufacturer	Edit Delete

Items per page: 10 ∨   1-2 of 2 items   1 ∨ of 1 pages   ◄   ►

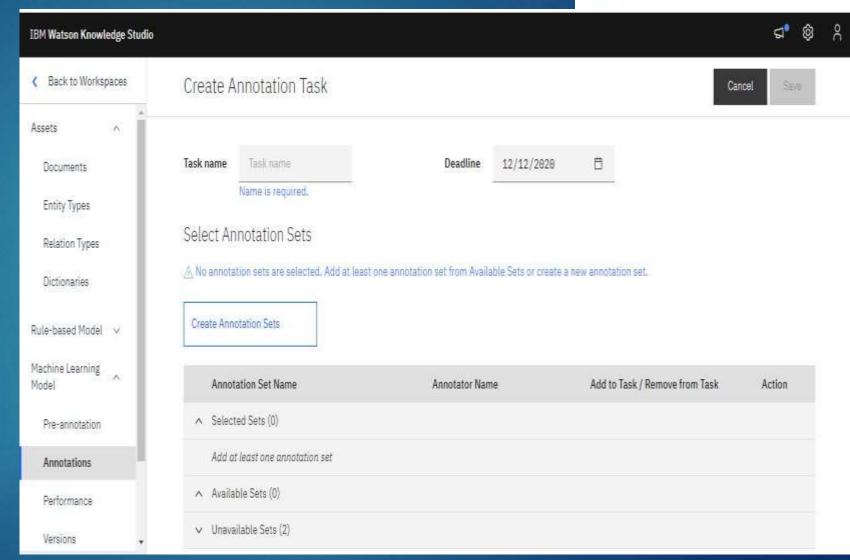# Create Human Annotation Task and Annotation Sets:

REVA UNIVERSITY

In this exercise I will Create annotation task and annotation sets and Annotate mentions, co-references and relations for the documents in the document set assigned to me.

Created Human Annotation Task and Annotation Sets.

Once the documents have been uploaded, annotation sets need to be created so that they can be annotated by multiple human annotators.

1. Clicked on "Machine Learning Model à Annotations" link under "Assets" menu on the left.

2. Clicked the Annotation Tasks tab, then click "Add Task".

3.Specified Task name as "TestTask" and selected a future data for Deadline, then clicked on "Create Annotation Sets".

IBM Watson Knowledge Studio

< Back to Workspaces

Create Annotation Task

Cancel   Save

Assets ∧

Documents

Entity Types

Relation Types

Dictionaries

Rule-based Model ∨

Machine Learning Model ∧

Pre-annotation

Annotations

Performance

Versions

Task name   [Task name]   Deadline   12/12/2020

Name is required.

Select Annotation Sets

⚠ No annotation sets are selected. Add at least one annotation set from Available Sets or create a new annotation set.

[Create Annotation Sets]

Annotation Set Name	Annotator Name	Add to Task / Remove from Task	Action
∧ Selected Sets (0)			
Add at least one annotation set			
∧ Available Sets (0)			
∨ Unavailable Sets (2)			

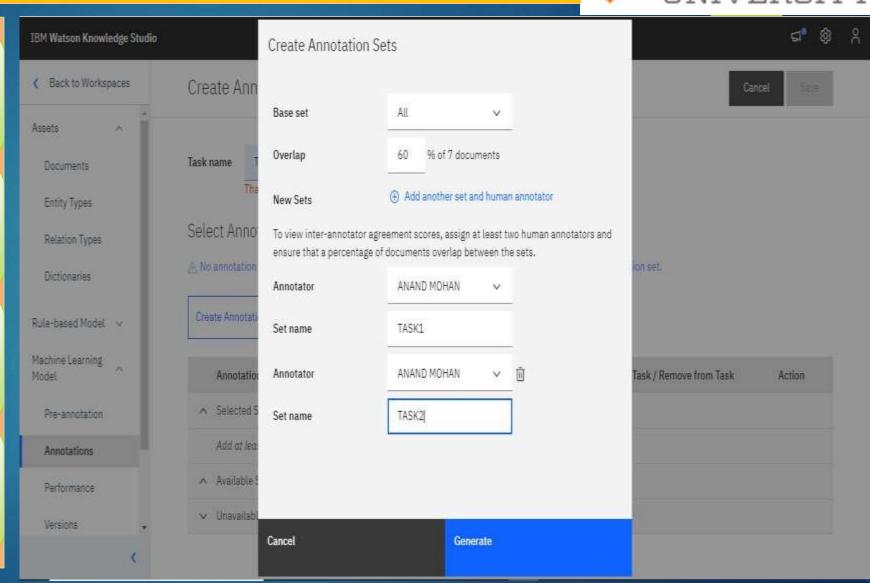# Create Human Annotation Task and Annotation Sets:

In the "Create Annotation Sets" pop up window, chose Base set as Corpus.csv (the uploaded document set), Overlap as 60% and created 2 sets:

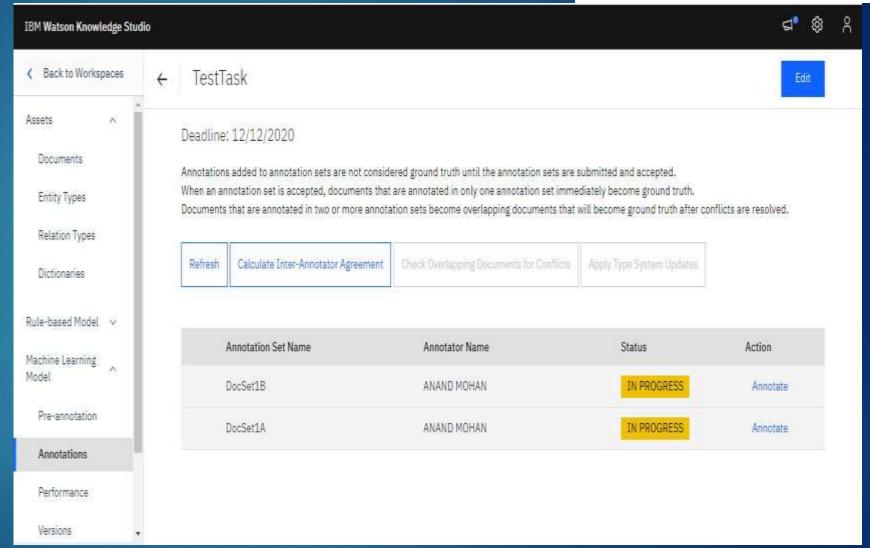- DocSet1A assigned to myself

- DocSet1B assigned to myself

we can add two sets by clicking on the '+' icon.

IBM Watson Knowledge Studio

‹ Back to Workspaces

Create Ann

Assets

Documents

Entity Types

Relation Types

Dictionaries

Rule-based Model ⌄

Machine Learning Model ⌄

Pre-annotation

Annotations

Performance

Versions

Task name

Tha

Select Anno

⚠ No annotation

Create Annotati

Annotation

⌃ Selected S

Add at lea

⌃ Available

⌄ Unavailabl

Cancel   Save

ion set.

Task / Remove from Task   Action

Create Annotation Sets

Base set   All ⌄

Overlap   60   % of 7 documents

New Sets   ⊕ Add another set and human annotator

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

Annotator   ANAND MOHAN ⌄

Set name   TASK1

Annotator   ANAND MOHAN ⌄ 🗑

Set name   TASK2|

Cancel   Generate

REVA UNIVERSITY

Clicked on "Generate". Saved the task by clicking on "Save", the annotation task will be created with 2 annotation sets as shown in the snapshots.

IBM Watson Knowledge Studio

< Back to Workspaces

← TestTask

Edit

Assets ⌄

Documents

Entity Types

Relation Types

Dictionaries

Rule-based Model ⌄

Machine Learning Model ⌄

Pre-annotation

Annotations

Performance

Versions

Deadline: 12/12/2020

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 |

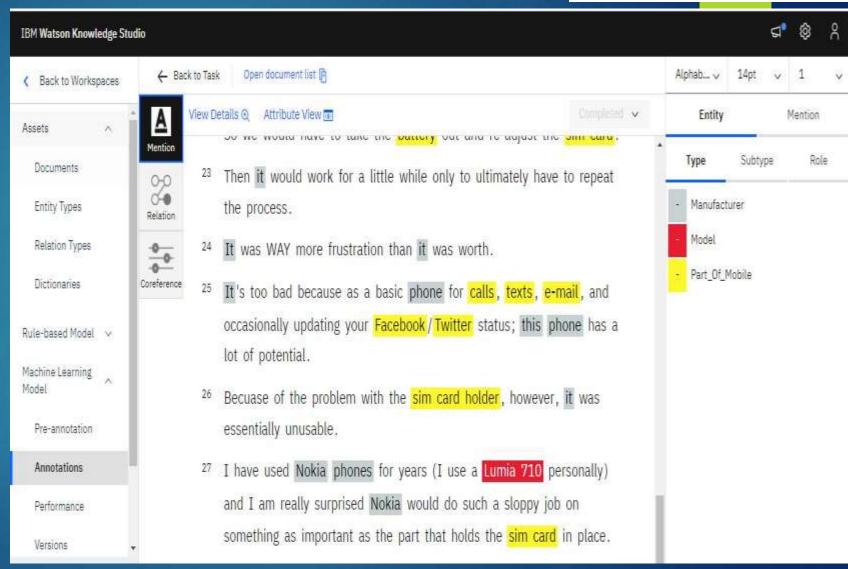| Annotation Set Name	Annotator Name	Status	Action
DocSet1B	ANAND MOHAN	IN PROGRESS	Annotate
DocSet1A	ANAND MOHAN	IN PROGRESS	Annotate

Annotate Mentions:

Clicked on "TestTask" to begin annotation. I will see the document sets assigned for me to annotate. In this case, since I am the only one working on the project, I will see two document sets both assigned to me. Clicked on the "Annotate" for one of the document sets.

clicked the first document in the annotation set. The default section that comes up is 'Mentions'. For annotating mentions, clicked the word or phrase that I would like to annotate and chose the entity type (which is color coded) on the right.

As and when I annotate, I can save the annotations by clicking on the 'Save' at the top right corner.

To annotate a role for a mention, selected the mention, go to the 'Role' tab on the right. I will see the list of roles that are applicable to the entity type that has been chosen for the mention. Selected the role I would like.
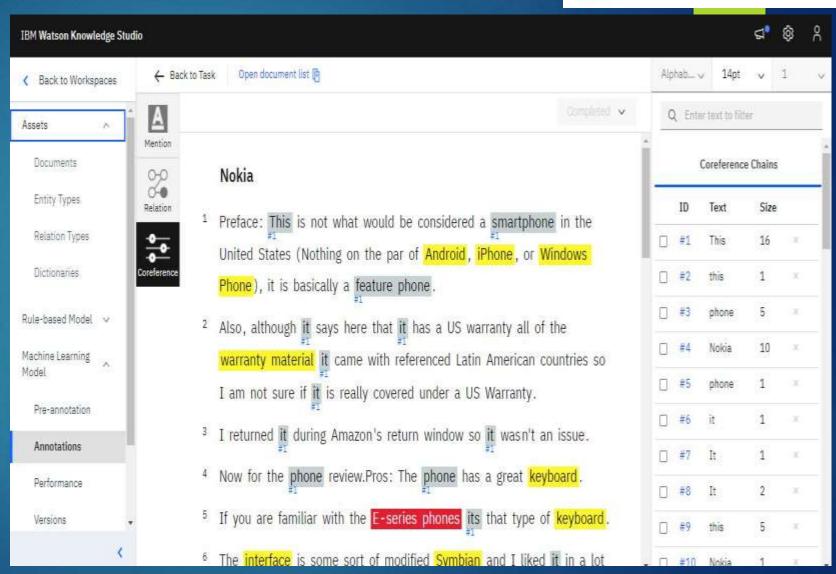
IBM Watson Knowledge Studio

< Back to Workspaces   ← Back to Task   Open document list

Assets ∧   View Details ⊕   Attribute View   Completed ∨

Documents   So we would have to take the battery out and re adjust the sim card.

Entity Types   23   Then it would work for a little while only to ultimately have to repeat
                                    the process.
Relation Types
                              24   It was WAY more frustration than it was worth.
Dictionaries
                              25   It's too bad because as a basic phone for calls, texts, e-mail, and
Rule-based Model ∨   occasionally updating your Facebook / Twitter status; this phone has a
                                   lot of potential.
Machine Learning
Model ∧   26   Becuase of the problem with the sim card holder, however, it was
                                   essentially unusable.
Pre-annotation
                              27   I have used Nokia phones for years (I use a Lumia 710 personally)
Annotations   and I am really surprised Nokia would do such a sloppy job on
Performance   something as important as the part that holds the sim card in place.

Versions

Alphab... ∨   14pt ∨   1 ∨

Entity   Mention

Type   Subtype   Role

- Manufacturer

- Model

- Part_Of_Mobile

Annotate References:

Note: Co-references should be annotated only after mention annotations have been completed.

To annotate co-references, I went to the co-reference page by clicking 'Coreference' on the left.

Chose the mentions that belong to the same entity and double clicked the last mention. I will see the number key on the right with co-referenced mentions having the same number.
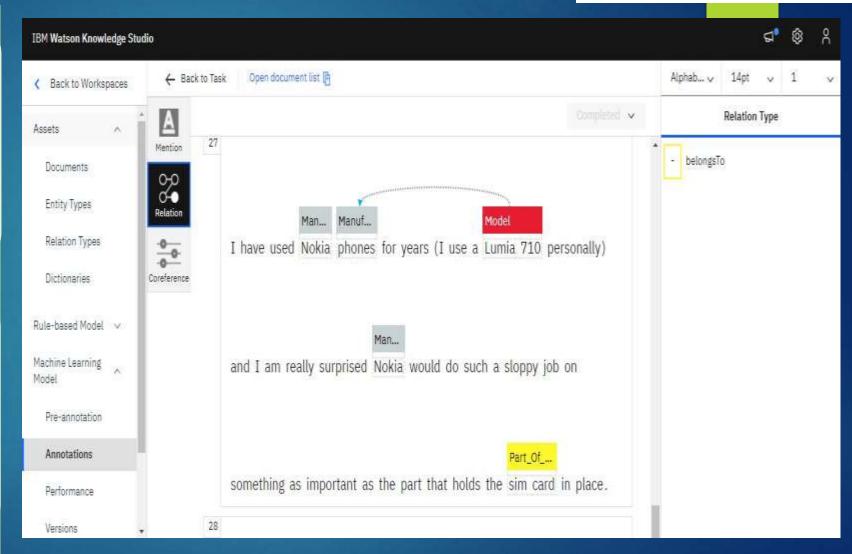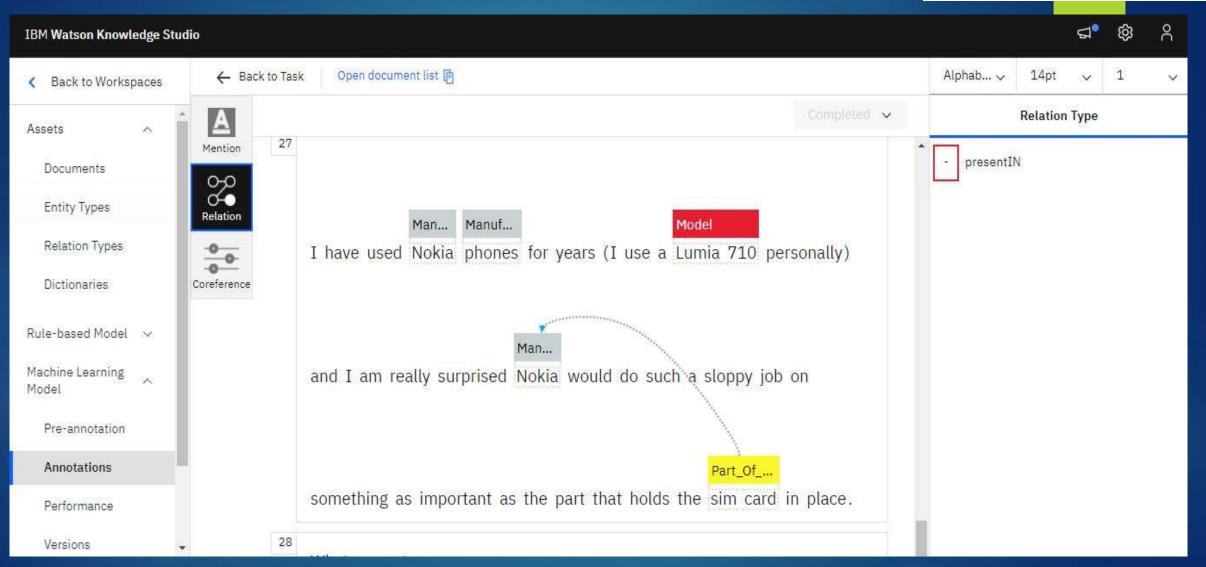
Annotate Relations:

To annotate relations, I went to the relations page by clicking 'Relation' on the left.

Chose the mention which belongs to the first entity type and then the mention that belongs to the second entity type in a relation.Selected the relation on the right to annotate the relation.

Annotate Relations:
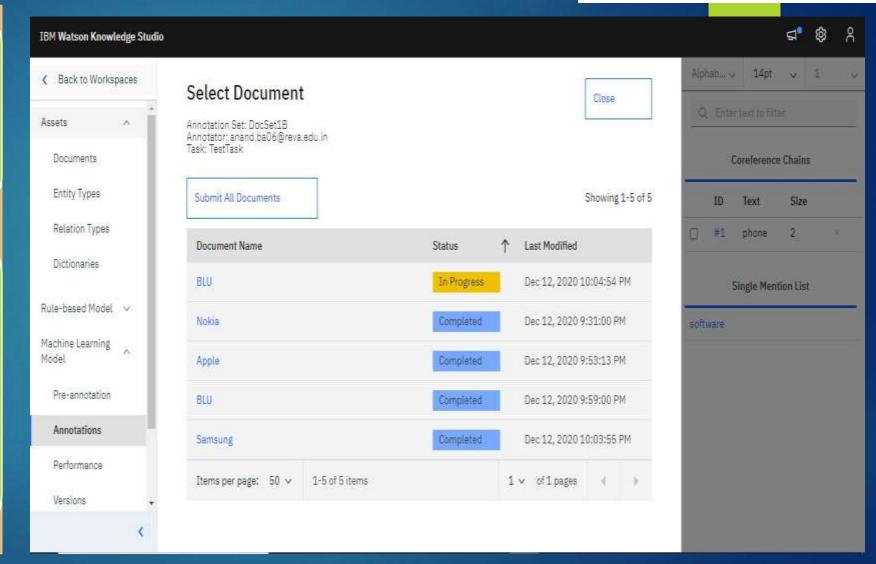
Completing Human Annotations:

Saving and submitting my annotations:

Once I completed annotating mentions, co-references and relations on all the documents, marked each document as "Completed" and Save.
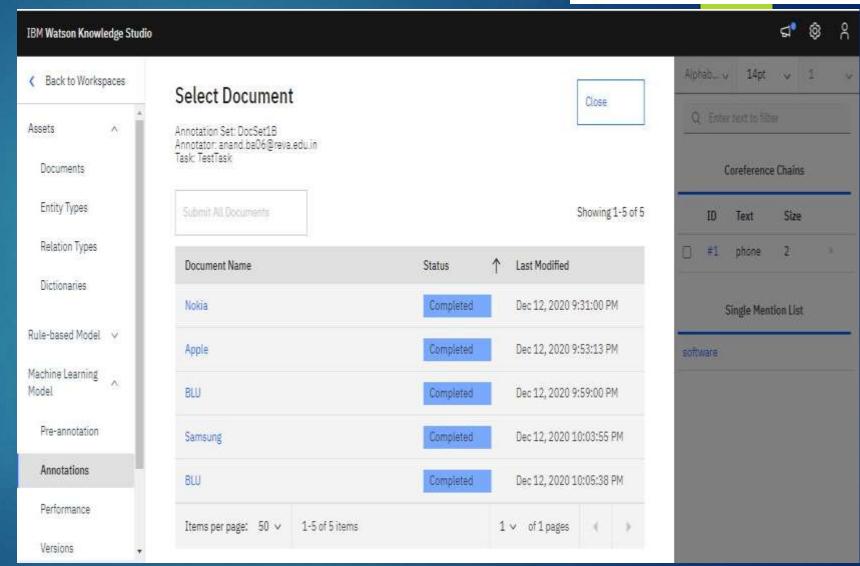
Completing Human Annotations:

went back to annotate the remaining documents by clicking on the "Open document list" on top.

Once I have annotated all the documents in a document set and marked them as complete, the document set status changes to "Submitted" from "In Progress". Once submitted, human annotators cannot edit the annotations. It is now up to the Project Manager to review the annotations, adjudicate and accept or reject the annotations. If accepted, they are promoted to ground truth.
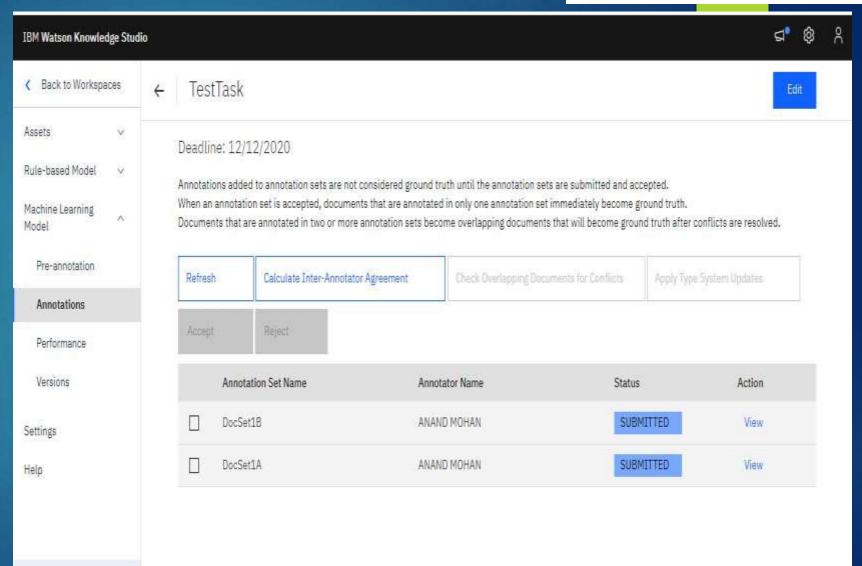
Completing Human Annotations:

Carried out annotation tasks on the second document set also that has been assigned to me.

Note: In order to be able to adjudicate documents, there must be some differences in the overlapping documents so that conflicts can be resolved. So, I wanted to purposely annotate the overlapping documents slightly differently just to be able to go through the process of adjudicating.
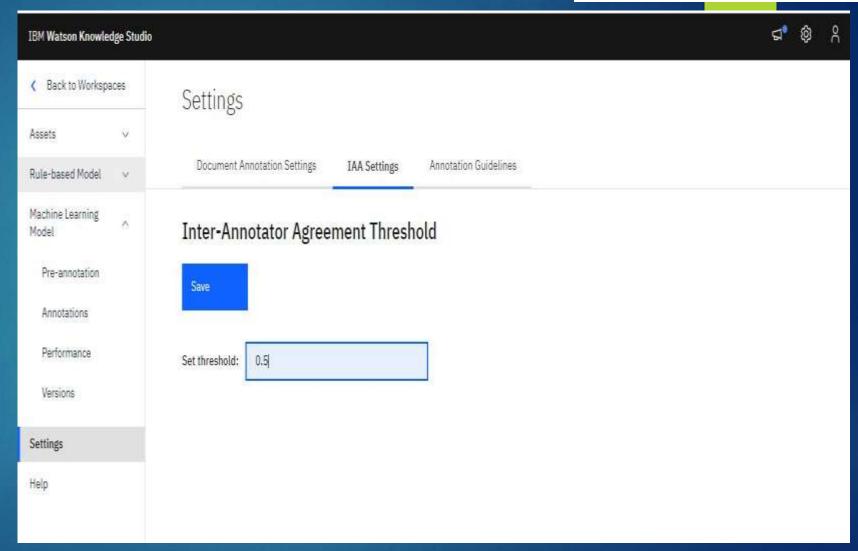

IBM Watson Knowledge Studio

< Back to Workspaces

← TestTask

Edit

Assets ∨

Rule-based Model ∨

Machine Learning Model ∧

Pre-annotation

Annotations

Performance

Versions

Settings

Help

Deadline: 12/12/2020

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 |

| Accept | Reject |

	Annotation Set Name	Annotator Name	Status	Action
☐	DocSet1B	ANAND MOHAN	SUBMITTED	View
☐	DocSet1A	ANAND MOHAN	SUBMITTED	View

Adjudication:

Completed annotation and submitted the documents from both document sets that have been assigned to me.

Clicked the "Settings" menu item on the left. Went to "IAA Settings" tab. Set the threshold IAA value as 0.5 and saved.
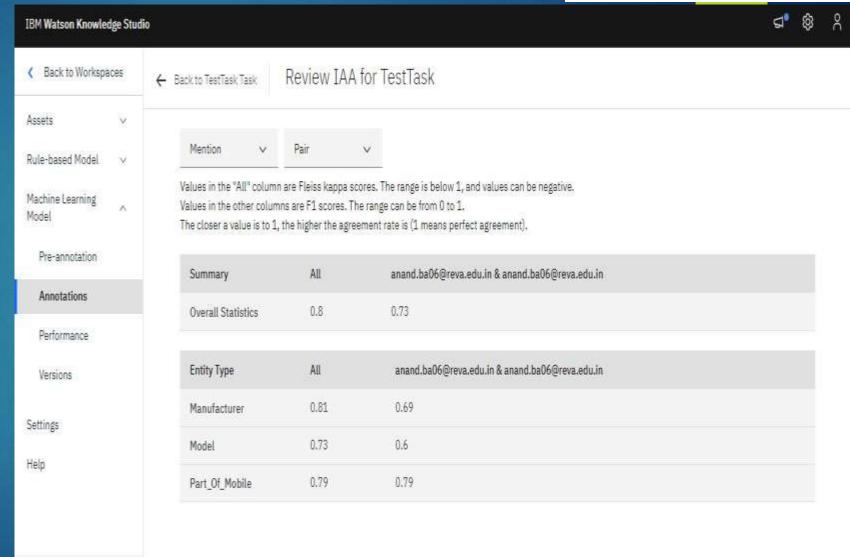
Adjudication

went back to the "Annotation Tasks" link from "Machine Learning Model" menu. Clicked "TestTask".Clicked "Calculate Inter-Annotator Agreement". The IAA value is shown by entity and relation types. The entity types and relation types for which the IAA value is below the threshold are highlighted in red. If the IAA value is very low, the documents can be rejected and returned to the human annotators to annotate the documents again. For the purposes of this exercise, assume that the IAA value is acceptable
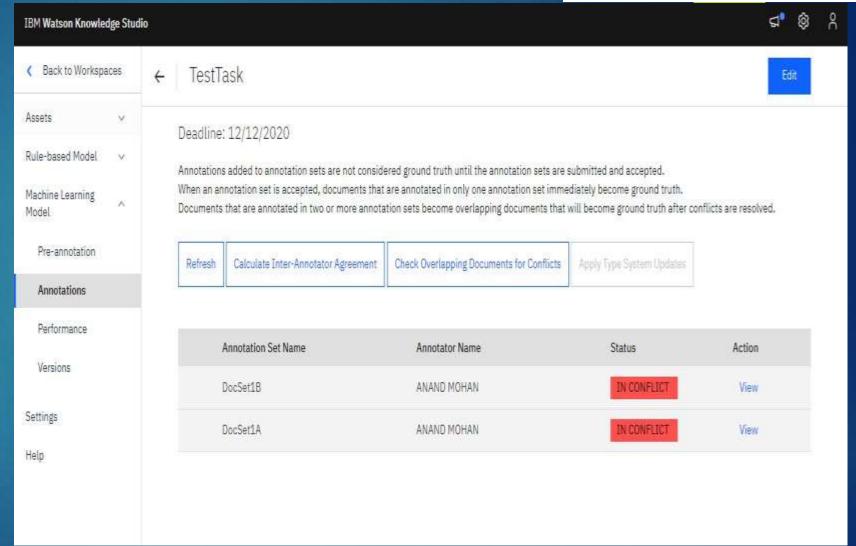
Clicked the "Back to TestTask Task" link. Since the IAA is acceptable, I am now ready to create ground truth that will be used in training the machine learning annotator. Checked DocSet1A and DocSet1B and clicked "Accept".
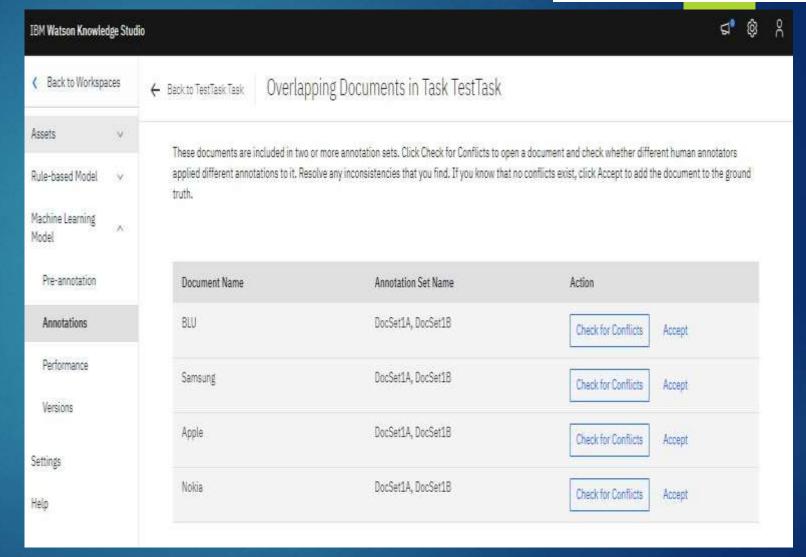
Adjudication:

Once accept is clicked, the documents that do not overlap are automatically promoted to ground truth. The documents that do overlap need to be adjudicated. To adjudicate conflicts, clicked on "Check Overlapping Documents for Conflicts".

IBM Watson Knowledge Studio

← Back to Workspaces

← TestTask

Edit

Assets ⌄

Rule-based Model ⌄

Machine Learning Model ⌃

Pre-annotation

Annotations

Performance

Versions

Settings

Help

Deadline: 12/12/2020

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
DocSet1B	ANAND MOHAN	IN CONFLICT	View
DocSet1A	ANAND MOHAN	IN CONFLICT	View

To start resolving conflicts, clicked "Check for Conflicts". I will see the annotation sets. Checked DocSet1A and DocSet1B and clicked "Check for Conflicts".
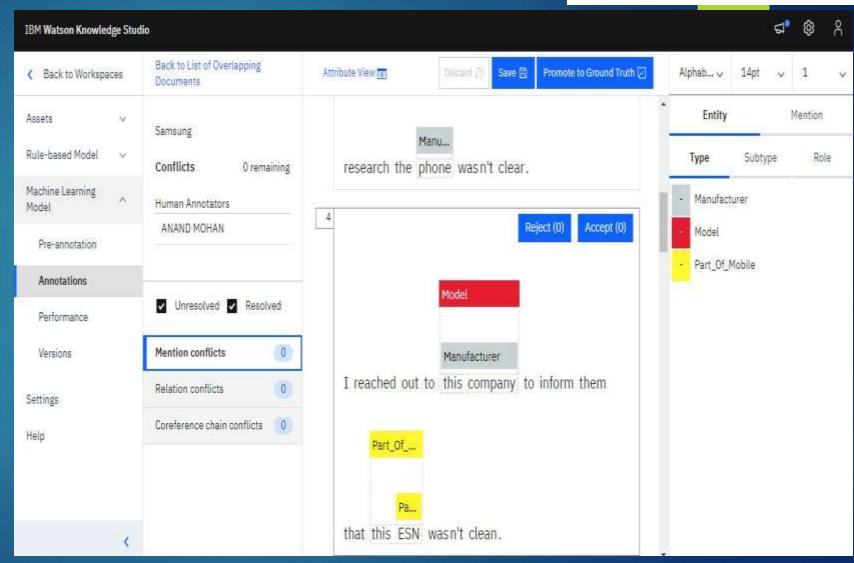
IBM Watson Knowledge Studio

‹ Back to Workspaces

← Back to TestTask Task   Overlapping Documents in Task TestTask

Assets ⌄

Rule-based Model ⌄

Machine Learning Model ⌃

Pre-annotation

Annotations

Performance

Versions

Settings

Help

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	
BLU	DocSet1A, DocSet1B	Check for Conflicts	Accept
Samsung	DocSet1A, DocSet1B	Check for Conflicts	Accept
Apple	DocSet1A, DocSet1B	Check for Conflicts	Accept
Nokia	DocSet1A, DocSet1B	Check for Conflicts	Accept

Adjudication:

The adjudication tool shows how many mention, relation, and co-reference chain conflicts exist. Clicked "Accept" or "Reject" for each individual annotation.
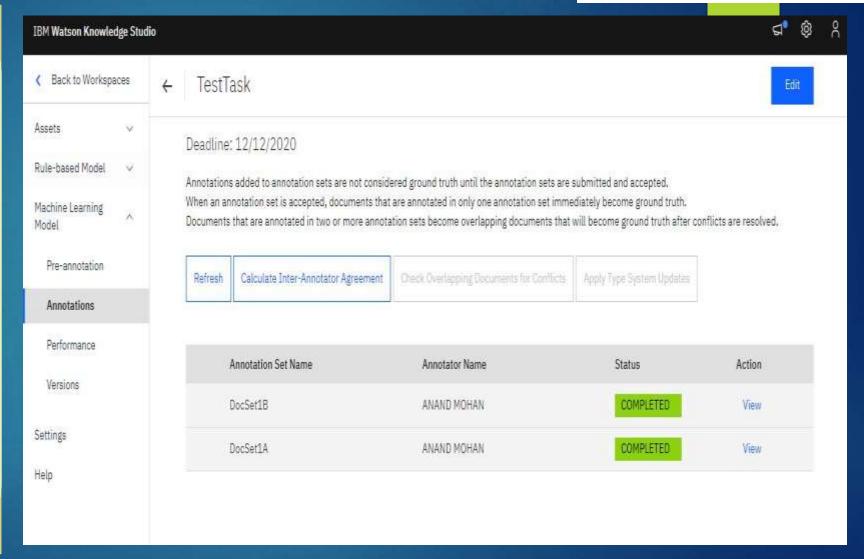
Continued until all mention, relations and co-reference conflicts are resolved. Once conflicts are resolved, the documents can be promoted to ground truth by clicking on the "Promote to ground truth" link on the top right corner.

Once all the conflicts are resolved and promoted to ground truth, the document sets are moved to "Completed" status.

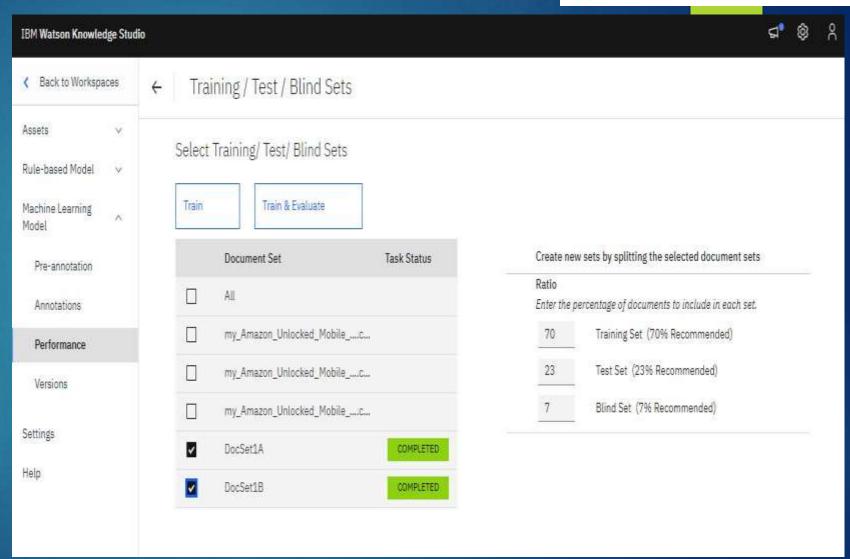IBM Watson Knowledge Studio

< Back to Workspaces

← TestTask

Edit

Assets ˅

Rule-based Model ˅

Machine Learning Model ˄

Pre-annotation

Annotations

Performance

Versions

Settings

Help

Deadline: 12/12/2020

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
DocSet1B	ANAND MOHAN	COMPLETED	View
DocSet1A	ANAND MOHAN	COMPLETED	View

Creating a machine learning model:

To create a machine learning model, I selected the document sets that I wanted to use to train the model and specified the percentage of documents that are to be used as training data, test data, and blind data.

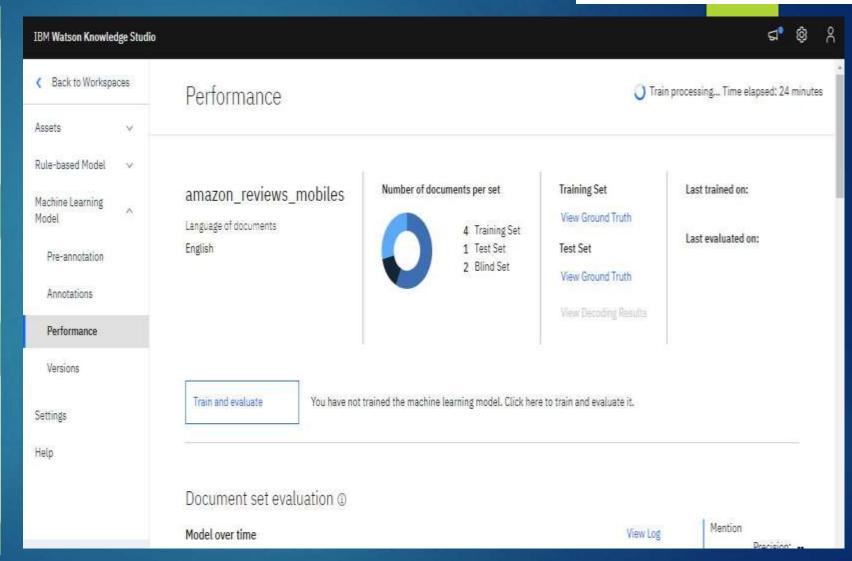**Clicked on "Machine Learning Model à Performance" link on the left.**

IBM Watson Knowledge Studio

< Back to Workspaces

← Training / Test / Blind Sets

Assets ⌄

Rule-based Model ⌄

Machine Learning Model ⌃

Pre-annotation

Annotations

Performance

Versions

Settings

Help

Select Training/ Test/ Blind Sets

| Train | Train & Evaluate |

Document Set	Task Status
☐ All	
☐ my_Amazon_Unlocked_Mobile_....c...	
☐ my_Amazon_Unlocked_Mobile_....c...	
☐ my_Amazon_Unlocked_Mobile_....c...	
☑ DocSet1A	COMPLETED
☑ DocSet1B	COMPLETED

Create new sets by splitting the selected document sets

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

70	Training Set (70% Recommended)
23	Test Set (23% Recommended)
7	Blind Set (7% Recommended)

REVA UNIVERSITY

Creating a machine learning model:

Verified that all of the document sets have been approved and all annotation conflicts have been resolved through adjudication. Only documents that have become ground truth through adjudication or approval can be used to train the model.

Clicked on "Train and evaluate".
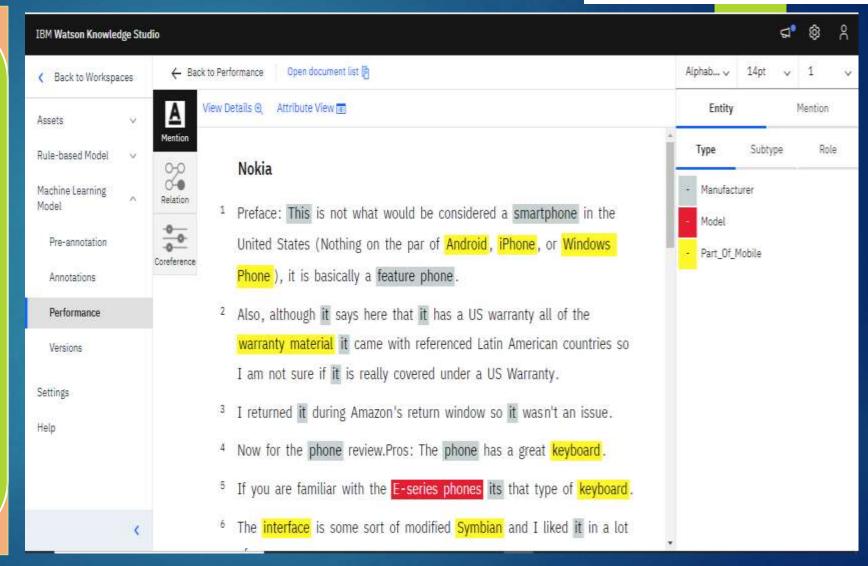
Checked "DocSet1A" and "DocSet1B" and clicked "Train & Evaluate" to train the model, evaluated annotations added by the machine learning model, and analyzed the performance statistics.

Evaluate annotations added by the model:
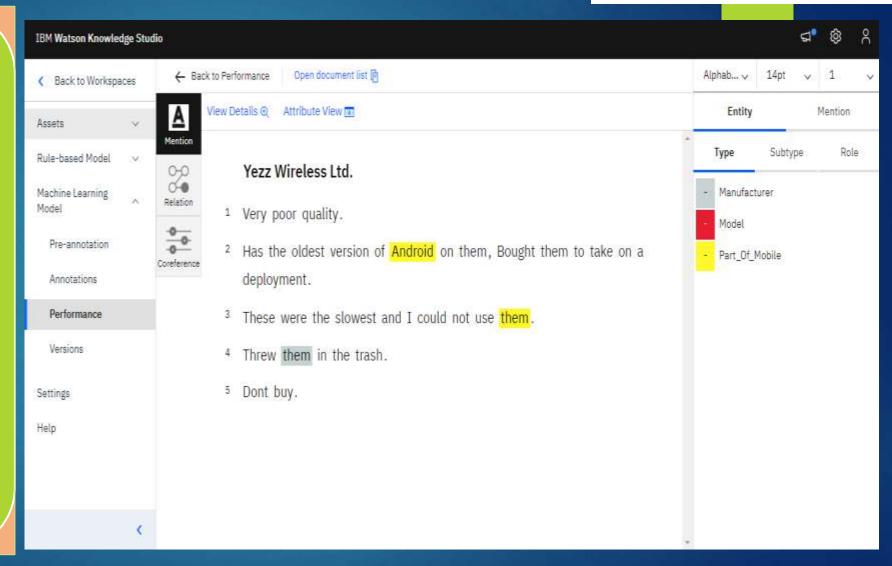
Clicked "View Ground Truth" link for the training set or test set to see the annotations that were added through pre-annotation and by human annotators. The ground truth editor opens. Click to open individual documents and saw how the mentions, relations, and referenced mentions were annotated.

Evaluate annotations added by the model:

REVA UNIVERSITY

On Performance page, clicked "View Decoding Results" link to see the annotations that the machine learning model added to documents in the test set. You can see how well the machine learning model labeled mentions, relations, and referenced mentions in the test data.
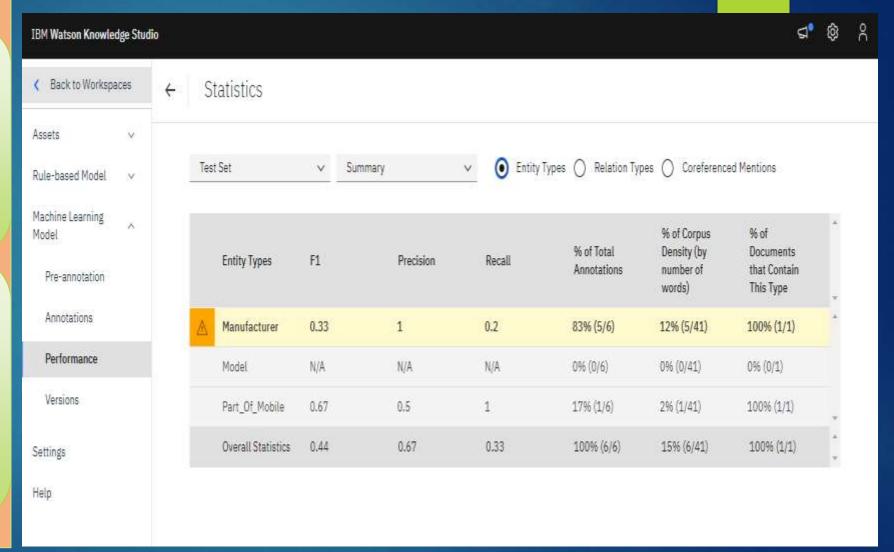
IBM Watson Knowledge Studio

< Back to Workspaces

← Back to Performance | Open document list

Assets

View Details ⊕   Attribute View

Mention

Rule-based Model

Relation

Machine Learning Model

Pre-annotation

Coreference

Annotations

Performance

Versions

Settings

Help

Alphab... ∨   14pt ∨   1 ∨

Entity   Mention

Type   Subtype   Role

- Manufacturer

- Model

- Part_Of_Mobile

Yezz Wireless Ltd.

1 Very poor quality.

2 Has the oldest version of Android on them, Bought them to take on a deployment.

3 These were the slowest and I could not use them.

4 Threw them in the trash.

5 Dont buy.

Analyze performance of the model:

To view performance statistics for how well the model was trained, Clicked on "Machine Learning Model à Performance" link on the left. For the mentions, relations, or coreferences, selected the "Detailed Statistics" link.

In the "Summary" view, specified whether I want to evaluate test data or training data, and then specified the type of annotations I want to see statistics for: entity types, relation types, or coreferenced mentions. As I scroll through the data, items that have low scores are flagged and highlighted to indicate that they require investigation and improvement. The triangle warning icon indicates that the F1 value is less than the fixed value, 0.5.

IBM Watson Knowledge Studio

< Back to Workspaces

← Statistics

Assets ∨

Rule-based Model ∨

Machine Learning Model ∧

Pre-annotation

Annotations

Performance

Versions

Settings

Help

Test Set ∨   Summary ∨   ⦿ Entity Types ◯ Relation Types ◯ Coreferenced Mentions

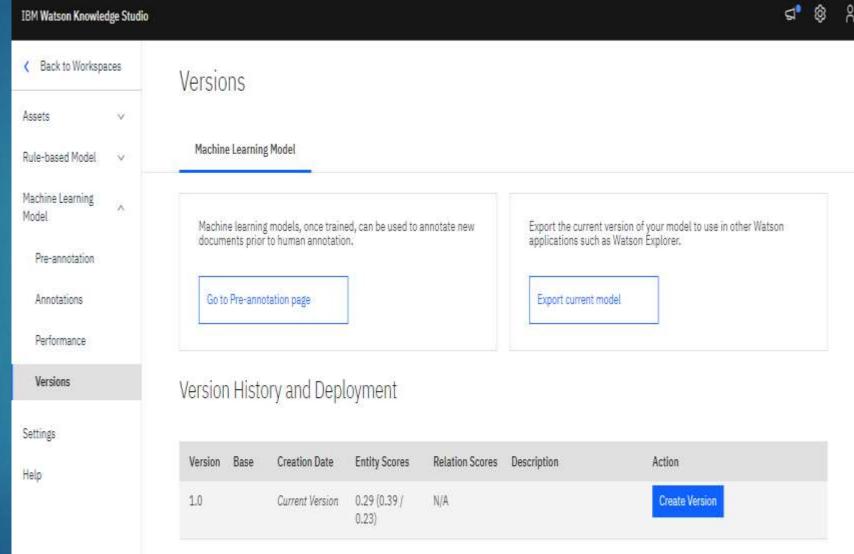|	Entity Types	F1	Precision	Recall	% of Total Annotations	% of Corpus Density (by number of words)	% of Documents that Contain This Type
⚠	Manufacturer	0.33	1	0.2	83% (5/6)	12% (5/41)	100% (1/1)
	Model	N/A	N/A	N/A	0% (0/6)	0% (0/41)	0% (0/1)
	Part_Of_Mobile	0.67	0.5	1	17% (1/6)	2% (1/41)	100% (1/1)
	Overall Statistics	0.44	0.67	0.33	100% (6/6)	15% (6/41)	100% (1/1)

Analyze performance of the model:

In the "Confusion Matrix" view for test data, specified the type of annotations that I want to see statistics for entity types or relation types. For each entity type or relation type: Each row in the matrix shows ground truth

Each column in the matrix shows the decoding results. The numbers in each cell represent the number of annotation tokens per mention. The column labeled O identifies tokens that should have been annotated as the type that is identified in the intersecting row (the human annotation), but they are not annotated as any type. The value N/A is shown when there are no annotations for a given type in the document sets.

IBM Watson Knowledge Studio

< Back to Workspaces

← Statistics

Assets ⌄

Rule-based Model ⌄

Machine Learning Model ⌃

Pre-annotation

Annotations

Performance

Versions

Settings

Help

| Test Set ⌄ | Confusion Matrix ⌄ | ⦿ Entity Types ◯ Relation Types |

Entity Types	Manufacturer	0	Part_Of_Mobile	Total
Manufacturer	1	3	1	5
0	0	35	0	35
Part_Of_Mobile	0	0	1	1
Total	1	38	2	41
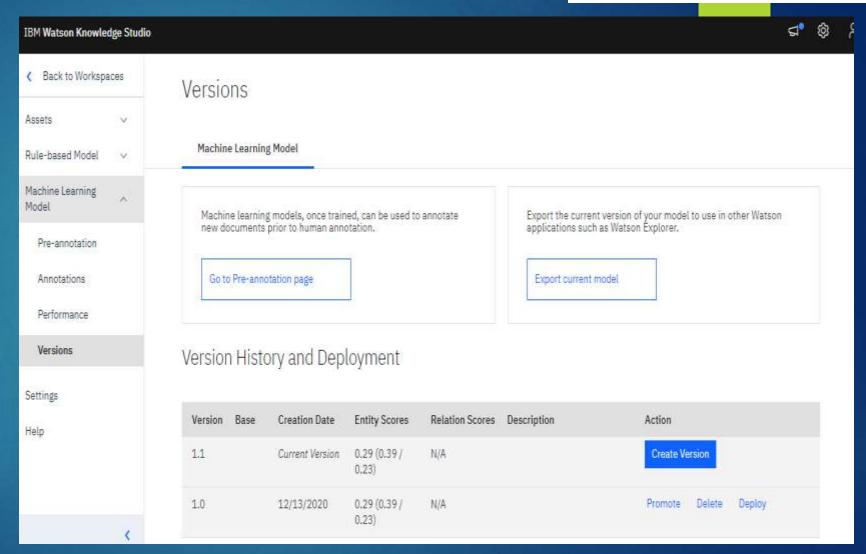
Clicked on "Machine Learning Model à Versions" link on the left. Chose the version of the model that I want to deploy. clicked on "Create Version" to take a snapshot of the current model.
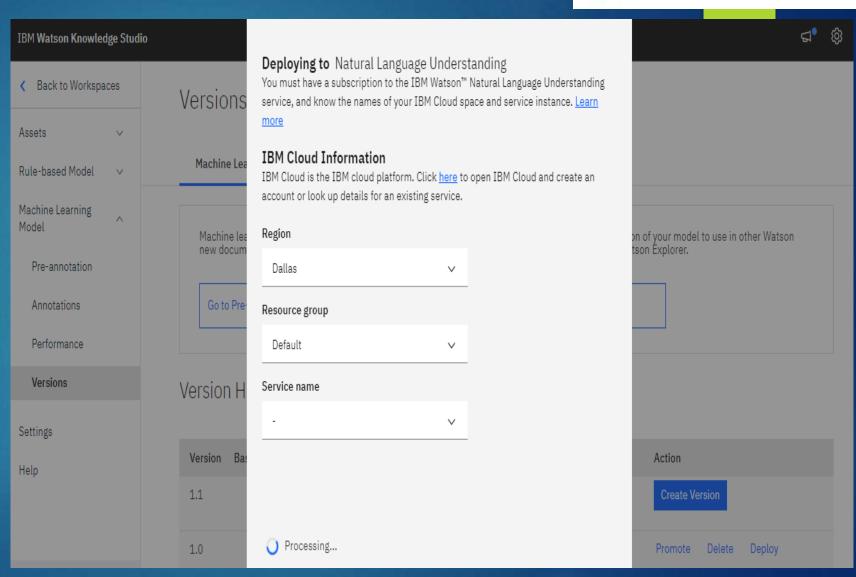
Clicked "Deploy", choose to deploy it to Natural Language Understanding, and then clicked Next.
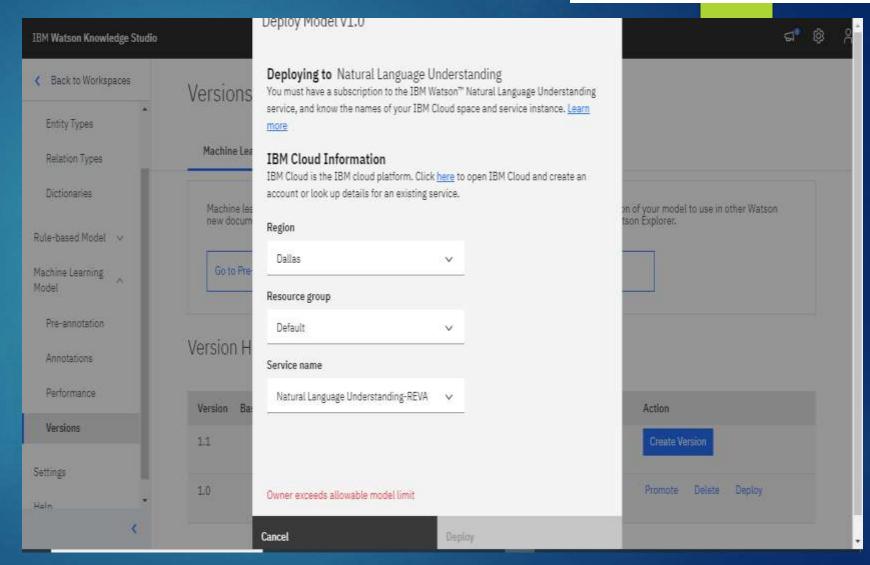
Deploy model to IBM Watson Natural Language Understanding (NLU):

Selected the IBM Cloud space/ resource group and instance. Clicked "Deploy". The deployment process took a few minutes. To check the status of the deployment, clicked "Status" on the Versions tab next to the version that I deployed.

Once deployed and available, made a note of the model ID (model_id).I will provide this ID to the Natural Language Understanding service to enable the service to use my custom model.

Got an error as shown in snapshot.so Demonstrating using appropriate NLU Query, the entities created in WKS and their targeted sentiments for the earlier lab project done in the previous class.

The command used on curl is as follows:

```
curl -X POST -u
"apikey:GBtKeOGto9F3yRzwVtVVz02Ns6LF
oklyvIoEAwUsCwbd" --header "Content-
Type: application/json" --data
"{\"text\":\'This on-site investigative effort
focused on the side impact inflatable
occupant protection system in a 2007
Hyundai Sonata and the injury sources of
the restrained 31-year-old female
driver.\",\"features\":{\"sentiment\":{},\"en
tities\":{\"model\": \"43f02987-2e31-4eb8-
a56d-aa464362890b\"}}}" "https://api.us-
south.natural-language-
understanding.watson.cloud.ibm.com/inst
ances/a03ff64e-e9cc-4a69-89d1-
0b9e727d9527/v1/analyze?version=2019-
07-12"
```

```
4eb8-a56d-aa464362890b\"}}}" "https://api.us-south.natural-language-understanding.watson.cloud.ibm.com/instances/a03ff64e-e9cc-4a69-89d1-0b9e727d9527/v1/analyze?version
=2019-07-12"
{
  "usage": {
    "text_units": 1,
    "text_characters": 191,
    "features": 2
  },
  "sentiment": {
    "document": {
      "score": 0.908132,
      "label": "positive"
    }
  },
  "language": "en",
  "entities": [
    {
      "type": "Manufacturer",
      "text": "Hyundai",
      "disambiguation": {
        "subtype": [
          "NONE"
        ]
      },
      "count": 1,
      "confidence": 0.997192
    },
    {
      "type": "Person",
      "text": "31-year-old female driver",
      "disambiguation": {
        "subtype": [
          "NONE"
        ]
      },
      "count": 1,
      "confidence": 0.985835
    },
    {
      "type": "Model_Year",
      "text": "2007",
      "disambiguation": {
        "subtype": [
          "NONE"
```