IBM Watson – HRM use case model

**INDEX**

1. Create HRM use case questions  
2. Getting started with Knowledge Studio  
 2.1 Create Knowledge Studio resource  
 2.2 Creating a workspace  
 2.3 Creating a type system  
 2.4 Adding a dictionary  
 2.5 Creating a machine learning model  
 2.5.1 Adding documents for annotation:  
 2.5.2 Pre-annotating with a dictionary-based annotator  
 2.5.3 Creating an annotation task  
 2.5.4 Annotating documents  
 2.6 Creating a machine learning model  
 2.6.1 Training the machine learning model  
 2.6.2 Evaluating annotations added by the model  
 2.6.3 Create a machine learning model  
 2.7 Pre-annotating new documents with a machine learning model  
 2.8 Creating a custom categories model  
 2.8.1 Import the training file.  
 2.8.2 Tune your model for each category.

3. Getting started with NLU  
 3.1 Create NLU Service  
 3.2 Deploy your custom model to IBM Watson™ Natural Language Understanding  
 3.3 Test your custom model with Natural Language Understanding or Discovery

1. **Created HRM use case questions**

Creating HRM Use Case model using IBM Watson. To find out the entities and intents for   
 given text using IBM Watson services.

Sample list of questions for HRM use case with Intent and Entities and save the file as .txt format

* + 1. How many jobs are available in your company – (Intent – JobInfo, Entity – None)
    2. How many jobs are available at Hyderabad (Intent – JobInfo, Entity – Hyderabad)
    3. What is the base experience for Software Engineer (Intent - JobInfo, Entity – Experience)
    4. How many years of experience do you expect for a Software Engineer (Intent – JobInfo, Entity = experience, Software Engineer)
    5. What is job duration of the project (Intent- JobInfo, Entity-duration)
    6. How to contact company HR (Intent- HR, Entity-contact)
    7. Should onboarding have been longer, shorter, or was it just about right (Intent-onboarding, Entity-longer, shorter)
    8. How many employees onboarded in your company till now (Intent-onboarding, Entity-none)
    9. How many positions and employees do you plan to onboard in 2019(Intent-onboarding, Entity-positions)
    10. What is the training duration and process in company (Intent- training? Entity-duration)
    11. How does the job role fit within the organization’s broader strategy? (Intent- jobinfo, Entity-organization)

1. **Getting started with Knowledge Studio**

**2.1 Create Knowledge Studio resource**

**Procedure:**

* + Login with your IBM Cloud ID.
  + Click on **“Create resource”** which you can see top right corner of the page.
  + Under catalog search **“knowledge studio”** service.
  + Open Knowledge studio service, under **create -> select region-> choose one region**, here I choose Frankfurt which is Free plan and click on “create”. Make sure that all regions would be the same city

Tips: Creating a real-life model typically involves multiple human annotators in addition to an administrator or project manager. However, for purposes of the tutorial, you can continue with a single user ID.

**2.2 Creating a workspace**

**About this task:**   
A workspace defines all the resources that are required to create a machine learning model, including training documents, the type system, dictionaries, and annotations that are added by human annotators.

**Procedure:**

* + Click Launch tool from the Manage page
  + Click Create Workspace
  + Specify the details for the new workspace:
  + In the Workspace name field, type HRM (workspace name)
  + In the Workspace description field, type Watson Knowledge Studio tutorial workspace
  + In the Language of documents field, use the default value, English. The sample files you will be using for this tutorial are in English
  + Click **Create**

**Results:**  
The workspace is created and opens automatically.

**What to do next:**  
You can now start configuring the workspace resources, such as the type system.

**2.3 Creating a type system**

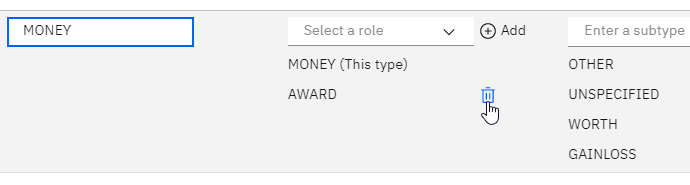
You must create or upload a type system before you begin any annotation tasks.

**Procedure:**

* + Download the en-klue2-types.json External link icon file to your computer. This file contains an example KLUE type system.

Note: For more information about creating a type systems, see (https://cloud.ibm.com/services/knowledge-studio/crn%3Av1%3Abluemix%3Apublic%3Aknowledge-studio%3Aeu-de%3Aa%2Fff04c0afc62b4acab8e02555213a5d84%3Af591d9ba-dea1-4476-a8b7-80302336869b%3A%3A?paneId=gettingStarted)

* + Click **Assets> Entity Types.**
  + On the Entity Types page, click Upload.
  + Upload the en-klue2-types.json file from your computer. The uploaded type system is displayed in the table.
  + Browse the type system so you can see the data that was uploaded.
  + Edit an entity type:
* Locate the MONEY entity type.
* Double-click anywhere in the table row to edit the entity type.
* In the Roles column, click the Delete a role icon The "Delete a role" icon. next to the AWARD role.



* Click Save.

**2.4 Adding a dictionary**

Dictionaries are used for pre-annotating text when creating a machine learning model.

**Procedure**

* + Download the dictionary-items-organization.csv External link icon file to your computer. This file contains dictionary terms in CSV format, suitable for uploading into a Knowledge Studio dictionary.
  + Click Assets > Dictionaries.
  + Click Create Dictionary to add a dictionary.

Important: Do not click Upload dictionary, which is used to upload a dictionary that you   
 want to use as-is. For the tutorial, you will create a new editable dictionary and then upload  
 terms into it.

* + In the Name field, type Test dictionary and click Save to create the (empty) dictionary
  + The new dictionary is created and automatically opened for editing
  + In the dictionary pane, click Upload
  + Upload the dictionary-items-organization.csv file from your computer. The terms in the file are uploaded into the dictionary
  + Click Add Entry to create a new term. An editable row is added at the top of the table
  + In the Surface Forms column, type IBM and International Business Machines Corporation on separate lines. (When you begin to type a new surface form, a space is added below for an additional surface form.) Leave the radio button next to IBM selected, which indicates that IBM is the lemma
  + In the Part of Speech column, select Noun
  + Click **Save**

**2.5 Creating a machine learning model**

Process for building a machine learning model that you can deploy and use with other Watson services.

**2.5.1 Adding documents for annotation:** We will learn how to add documents to a workspace in Knowledge Studio that can be  
 annotated by human annotators.

**Procedure:**

* + Download the documents-new.csv file to your computer. This file contains example documents suitable for uploading
  + Within your workspace, click **Assets > Documents**
  + On the Documents page, click **Upload Document Sets**
  + Upload the documents-new.csv file from your computer. The uploaded file is displayed in the table. Add .txt file (HR questions).

**2.5.2 Pre-annotating with a dictionary-based annotator**

We will learn how to use a dictionary-based annotator to pre-annotate documents in Knowledge Studio.

**Procedure:**

* + Within your workspace, click **Assets > Dictionaries**
  + The Adding a dictionary is same as 2.4 Adding Dictionary
  + From the **Entity type** list, select the **ORGANIZATION** entity type to map it to the **Test dictionary** dictionary
  + The Creating a type system (2.3) of getting started with Knowledge Studio tutorial shows how to create the type system that contains the **ORGANIZATION** entity type
  + On the **Machine Learning Model > Pre-annotation > Dictionaries tab**, click Apply This **Pre-annotator**
  + Click Run

**Results:** The documents in the selected sets are pre-annotated by using the dictionary that you created. If you like, you can use the dictionary to pre-annotate document sets or annotation sets that you add later.

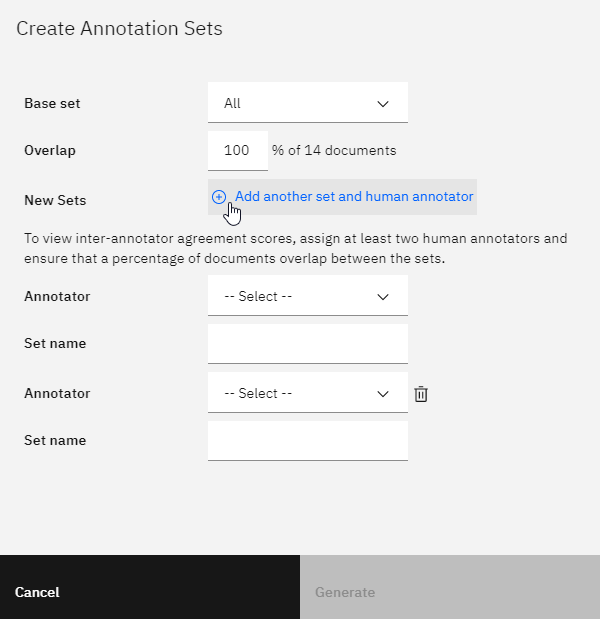
**2.5.3 Creating an annotation task**

We will learn how to create annotation sets and use annotation tasks to track the work of   
 human annotators in Knowledge Studio.

**About this task:** An annotation set is a subset of documents from an uploaded document set   
that you assign to a human annotator. The human annotator annotates the documents in the annotation set. To later use inter-annotator scores to compare the annotations that are added by each human annotator, you must assign at least two human annotators to different annotation sets. You must also specify that some percentage of documents overlap between the sets.

**Procedure:**

* + Within your workspace, click Machine Learning Model > Annotations
  + Click the Annotation Tasks tab, then click Add Task
  + Specify the details for the task:
* In the Task name field, enter HRM (I choose it as task name)
* In the Deadline field, select a date in the future
  + Click Create Annotation Sets
  + The Create Annotation Sets window opens. By default, this window shows the base set, which contains all documents, and fields where you can specify the information for a new annotation set.



* + In the Overlap field, specify 100. This value specifies that you want 100 percent of the documents in the base set to be included in all the new annotation sets so they can be annotated by all human annotators.
  + For each new annotation set, specify the required information.
* In the **Annotator field**, select a human annotator user ID to assign to the new annotation set. In a realistic scenario, each annotation set is assigned to a different human annotator.

**Note:** If you have only a single administrator ID to use for the use case, assign that user to all annotation sets. In a realistic scenario, you would have multiple human annotators, but for the tutorial, the administrator can act as human annotator.

* In the **Set name** field, specify a descriptive name for the annotation set. For this tutorial, you can use the names, Set 1 and Set 2.
  + Click **Generate.**
  + Click **Save.**
  + As human annotators begin annotating documents, you can open tasks to see their progress

**2.5.4 Annotating documents**

We’ll know how to use the **ground truth editor** to annotate documents in Knowledge studio.

* + Open the My workspace workspace and click **Machine Learning Model > Annotations.**
  + Click the Annotation Tasks tab, then open the Test annotation task you created
  + Click Annotate for one of the assigned annotation sets.
  + Depending on how you set up the annotation tasks, you could have one or more annotation tasks assigned to the user ID you logged in with.
  + Notice that the term IBM was already annotated with the ORGANIZATION entity type. This pre-annotation is correct, so it does not need to be modified.
  + Annotate a mention:
* Click the Entity tab.
* In the document body, select the text Thomas Watson.
* In the list of entity types, click **Duration, Contact, Hyderabad, Software Engineer, PERSON**. The entity type **Duration** is applied to the selected mention.
  + From the status menu, select Completed, and then click the Save button
  + Click Open document list to return to the list of documents for this task and click Submit All Documents to submit the documents for approval

**Note:** In a realistic situation, you would create many more annotations and complete all the documents in the set before submitting

* + Close this annotation set, and then open the other annotation set in the Test task
  + Depending on how you set up the annotation tasks and which users you assigned them to, you might need to log in to Knowledge Studio as the user who is assigned to the other annotation set in the annotation task
  + Logging in as another user will help illustrate inter-annotator agreement in the next lesson. But if you have only one user, you can still complete the tutorial to get an understanding of how inter-annotator agreement works
  + After you complete the annotations for the second annotation set, click Submit All Documents

**Note:** Analyzing inter-annotator agreement (2.5.5), to compare the work of multiple human annotators in Knowledge Studio. Adjudicating conflicts in annotated documents (2.5.6) is to cross check that overlap between document sets in Knowledge Studio. This will work for multiple human annotators usage, as of now we used single human annotator.

**2.6 Creating a machine learning model** When you create a machine learning model, you select the document sets that you want to   
 use to train it. You also specify the percentage of documents that are to be used as training   
 data, test data, and blind data. Only documents that became ground truth through approval   
 or adjudication can be used to train the machine learning model.

**2.6.1 Training the machine learning model**

When you create a machine learning model, you select the document sets that you want to use to train the model and specify the percentage of documents that are to be used as training data, test data, and blind data.

**Restriction:** Only three machine learning models can be trained at a time per Knowledge Studio instance. If your instance contains multiple workspaces and the number of machine learning models that are being trained in other workspaces totals 3 already, then your request to train the machine learning model in your workspace will be queued until the other training processes are done.

**Procedure:**

To create a machine learning model:

* + Log in as a Knowledge Studio administrator and select your workspace.
  + Select **Machine Learning Model > Performance.**
  + Verify that all the document sets have been approved and that 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.
  + Click **Train and evaluate**
  + Click **Train** to train the model, or click **Train & Evaluate** to train the model, evaluate annotations added by the machine learning model, and analyze the performance statistics.

**Important:** Training a machine learning model can take several minutes or several hours, depending on the number of human annotations that exist and the total number of words across all documents.

* + Select the document sets that you want to use for training the model.

**Note:** The document sets must contain at least 10 annotated documents.

**2.6.2 Evaluating annotations added by the model**You can compare the ground truth view for annotations added by human annotators to the annotations added by the model.

**Procedure:**

To evaluate the annotations added by the model:

* Select **Machine Learning Model > Performance > Train and evaluate**. The Training/Test/Blind Sets page is displayed.
  + Click **View Ground Truth** 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 see how the mentions, relations, and coreferenced mentions were annotated.
  + On the **Performance page**, click **View Decoding Results** to see the annotations that the machine learning model added to documents in the test set. This button is available only after you evaluate the model. By viewing results, you can see how well the machine learning model labeled mentions, relations, and coreferenced mentions in the test data.
  + If you want to change how the documents are divided between training, test, and blind data sets, click **Performance > Train and evaluate > Edit Settings**. For example, if initial results seem acceptable, you might want to increase the number of documents in the test set to further test the machine learning model's results. You can change the ratio for how documents are automatically divided for different purposes, or you can select specific document sets to use as training data, test data, and blind data.
  + If you made any changes, click **Train & Evaluate** to retrain the model and re-evaluate the annotations

**2.6.3 Create a machine learning model**

When you create a machine learning model, you select the document sets that you want to  
use to train it. You also specify the percentage of documents that are to be used as training data, test data, and blind data. Only documents that became ground truth through approval or adjudication can be used to train the machine learning model.

**Procedure:**

* + Click **Machine Learning Model > Performance > Train and evaluate.**
  + Select **All,** and then click **Train & Evaluate.**

**Note:** Training might take more than ten minutes, or even hours, depending on the number of human annotations and the number of words in all the documents.

* + After the machine learning model is trained, you can export it from the Version page, or you can view detailed information about its performance by clicking the **Detailed Statistics** links that are located above each of the graphs on the Performance page.
  + To view the Training / Test / Blind Sets page, click the **Train and evaluate** button.
  + To see the documents that human annotators worked on, click **View Ground Truth.**
  + To see the annotations that the trained machine learning model created on that   
    same set of documents, click **View Decoding Results**
  + Click **Versions.** On the Versions page, you can take a snapshot of the model and the resources that were used to create it (except for dictionaries and annotation tasks). For example, you might want to take a snapshot before you retrain the model. If the statistics are poorer the next time you train it, you can promote the older version and delete the version that returned poorer results.

**Results**:  
You created a machine learning model, trained it, and evaluated how well it performed when annotating test data and blind data. By exploring the performance metrics, you can identify ways to improve the accuracy of the machine learning model.

**2.7 Pre-annotating new documents with a machine learning model**

After you train a machine learning model, you can use it to pre-annotate new documents that you add to the corpus.

**Attention:** Do not run a pre-annotator on documents that have been annotated by humans, but not been added to the ground truth yet. If you do, all current annotations will be stripped from the documents.

In this tutorial, you can add a second set of documents by using the documents-ml.csv file. Do not re-add the documents-new.csv file, since this addition would result in duplicate documents in the ground truth. Duplication causes the following problems:

* + If annotations on each document do not match, they lower the quality of the machine learning model.
  + If annotations on each document match, they over-train the machine learning model on the duplicated files.

**Procedure:**

* + Create an annotation set that uses the documents-ml.csv file as the base set, and assign it to yourself, the administrator.
  + After you complete the following steps to pre-annotate the new documents, you can view the annotation set to see how the machine learning model annotated the documents. Typically, you assign annotation sets to one or more human annotators.
  + To pre-annotate the new documents, click **Machine Learning Model > Versions**, and then click **Run this model.**
  + Select the document set that you added to the corpus, documents-ml.csv, and click Run.
  + After the pre-annotation is complete, create a human annotation task that includes the annotation set you created.
  + To view the annotations that were applied by the machine learning model to the new documents, open the annotation task.

Because the new documents were pre-annotated with the machine learning model, human annotation requires less time.

**Results:**  
By using your machine learning model to pre-annotate new document sets, you can expedite human annotation tasks for those documents.

**2.8 Creating a custom categories model**

This describes how to train a custom categories model and deploy it to Natural Language Understanding or Discovery.

Custom categories models can be deployed only to Natural Language Understanding or Discovery service instances that are hosted in the Frankfurt location (I used same location for ever service).

**2.8.1 Import the training file.**

* + On the Train your categories model screen, upload your CSV file by dragging and dropping the file to the Drag and drop file here area or by clicking the or browse file link on the screen. This automatically initiates training of the model.
  + If the model training succeeds, you will see the Test Your Model screen. If training fails and the error message suggests that you can fix the issue, edit your CSV file to correct any errors and upload it again.

**2.8.2 Tune your model for each category.**

* + Find a text sample that pertains to a category label from your CSV file and paste it into the Enter input text area.
  + Click Run Test to evaluate the performance of your model against your text sample. A score closer to 1.0 indicates a very high level of certainty that the text passage corresponds to the respective category.
  + If the model does not perform well for the tested category, adjust the key phrases for the category in your CSV file, then click Retrain model to upload the updated file.
  + Repeat steps 2.8.1 until you are satisfied with the model performance.

1. **Getting started with NLU**

**3.1 Create NLU Service**

* + Click on “Create resource” which you can see top right corner of the page.
  + Under catalog search “**Natural Language Understanding**” service.
  + Open **Natural Language Understanding**, under create -> select region-> choose one region, here I choose Frankfurt which is Free plan and click on “create”. Make sure that all regions would be the same city

**Before you begin**

* Copy the credentials to authenticate to your service instance:
* On the Manage page (https://cloud.ibm.com/services/natural-language-understanding/crn%3Av1%3Abluemix%3Apublic%3Anatural-language-understanding%3Aeu-de%3Aa%2Fff04c0afc62b4acab8e02555213a5d84%3A40897cd1-028d-496f-880a-66f1df9d6cb9%3A%3A?paneId=gettingStarted), click Show to view your credentials.
* Copy the API Key and URL values.
* Make sure that you have the curl command:

The examples use curl command to call methods of the HTTP interface. Install the version for your operating system from curl.haxx.se External link icon. Install the version that supports the Secure Sockets Layer (SSL) protocol. Make sure to include the installed binary file on your PATH environment variable.



**3.2 Deploy your custom model to IBM Watson™ Natural Language Understanding**

* + Select Machine Learning Model > Versions.
  + Choose the version of the model that you want to deploy.
  + If there is only one working version of the model, create a snapshot of the current model. This version the model, which enables you to deploy a version, while you continue to improve the current version. The option to deploy does not appear until you create at least one version.

Note: Each version can be deployed to any number of service instances. Each deployed instance of a model version is given a unique Model ID, but is identical in all other ways.

* + In the Create a version dialog, enter a description for the model, and click OK to save a version of the model.
  + Follow the instructions to select a target service instance and click Deploy to execute the model deployment. You will get a model ID which is required to use the custom model with Natural Language Understanding or Discovery.
  + You can manage model versions and deployment from the Versions page in the left navigation menu.

**3.3 Test your custom model with Natural Language Understanding or Discovery**

You can list deployed model in the Natural Language Understanding service instance by calling the following API method.

To use the deployed model, you must specify the model ID of your custom model in the entities.model parameter of an analyze call.

You can use the model with the Natural Language Understanding GET /analyze request to extract the following features:

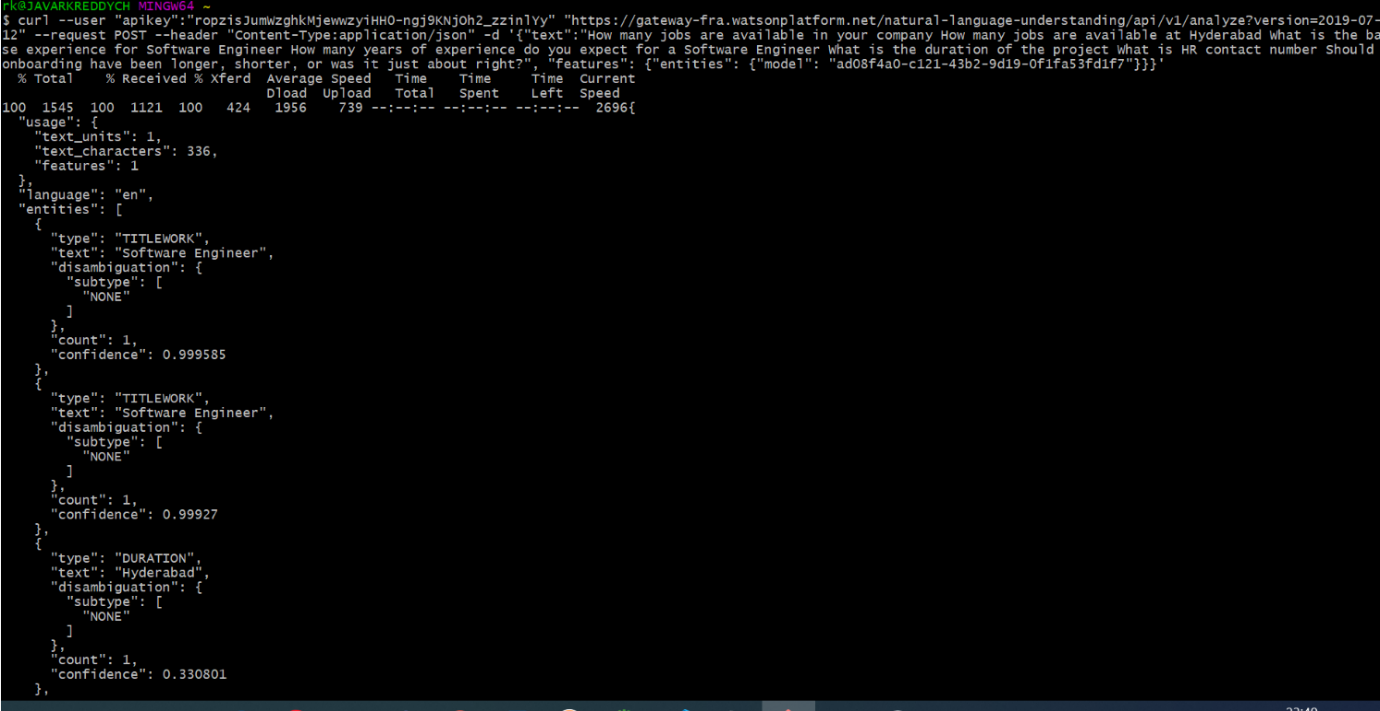
**Entities:**

The following command finds the entities that are present in the sentence that is passed by using the text parameter:

// Curl command

curl --user "apikey":"ropzisJumWzghkMjewwzyiHH0-ngj9KNjOh2\_zzinlYy" "https://gateway-fra.watsonplatform.net/natural-language-understanding/api/v1/analyze?version=2019-07-12" --request POST --header "Content-Type:application/json" -d '{"text":"How many jobs are available in your company How many jobs are available at Hyderabad What is the base experience for Software Engineer How many years of experience do you expect for a Software Engineer What is the duration of the project What is HR contact number Should onboarding have been longer, shorter, or was it just about right?", "features": {"entities": {"model": "ad08f4a0-c121-43b2-9d19-0f1fa53fd1f7"}}}'

**Result:** Using Gitbash, run the above curl command,

****