IBM Training

Student Exercises

IBM Watson Knowledge Studio Hands-On

IBM Watson Technical Enablement

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. The following are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide:

AIX®	BigInsights™	Cognos®
DataStage®	DB™	DB2®
developerWorks®	Domino®	FileNet®
Global Business Services®	GPFS™	Guardium®
IBM Business Partner®	IBM Watson™	IMS™
Informix®	InfoSphere®	Insight™
iSeries®	LanguageWare®	Lotus Notes®
Lotus®	Notes®	OmniFind®
Optim™	Power®	PowerHA®
Rational Team Concert™	Rational®	Redbooks®
SPSS®	System Storage®	System x®
System z®	Tivoli®	Watson™
WebSphere®	z/OS®	

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java™ and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

VMware and the VMware "boxes" logo and design, Virtual SMP and VMotion are registered trademarks or trademarks (the "Marks") of VMware, Inc. in the United States and/or other jurisdictions.

Netezza® is a trademark or registered trademark of IBM International Group B.V., an IBM Company.

SoftLayer® is a trademark or registered trademark of SoftLayer, Inc., an

IBM Company. Other product and service names might be trademarks of

IBM or other companies.

January 2017 edition

Exercises description

This lab guide is intended to provide you with hands-on experience in using Watson Knowledge Studio to build a machine learning annotator.

This course includes the following exercises:

- Exercise 01: Create a project. Create and edit a Type System
- Exercise 02: Upload training corpus
- Exercise 03: Create dictionary annotator
- Exercise 04: Dictionary pre-annotation
- Exercise 05: Human annotation
- Exercise 06: Adjudication
- Supplemental Exercise 07: Rules Annotator

The exercises should be completed in the same order as listed before moving on to the next exercise.

Pre-requisites:

a) Complete the Watson Knowledge Studio Foundations & Methodology course
b) Open an IBM Account. You can open one here:- https://www.ibm.com/account/us-en/
c) Create a Bluemix (IBM Cloud) account. You can sign up for for a free account: https://console.bluemix.net/registration/
d) Install Chrome or Firefox to work with the Watson Knowledge Studio interface
e) Download the files available for performing the labs.

Note: The illustrating screenshots provided in this lab guide could be slightly different from what you see in the WKS interface you are using because of updates to WKS since the lab guide was created.

Exercise 01: Create a Type System

Goals:

In this lab you will

- create a project (workspace)
- Import a type system
- Modify the type system by adding new entity and relation types

At the end of this lab, you should be able to use the WKS service to create a type system

Exercise Instructions

Lesson 1.0 Create a Watson Knowledge studio instance

1. Login to the Bluemix.
2. You will be dropped to your *Dashboard* page that displays any apps or services you might have created.
3. Create a new service instance of Watson Knowledge studio “WKS Lab” from the Watson catalog and choosing the “Lite” plan.

Lesson 1.1 Create a Workspace

Pre-requisites:

- For this step, you should have logged into Bluemix and created an instance of Watson Knowledge studio named “*WKS Lab”.*

1. Click on your instance of Watson Knowledge studio (WKS) and launch WKS.
2. If this is the first time you launch your instance, it will provide you with a notification that says that no user is registered to the WKS service instance. Add yourself as the administrator.
3. You will then be dropped to the “Workspaces” page. Click “Create workspace” to create your new workspace.
4. In the “Create Workspace” pop up window, enter the name of the new workspace as *<LoginShortName>_WKS*. Add an optional “Description” if you choose to.
5. Choose “English” as the language of your documents because the documents for this lab are in English.

6. You can use a dictionary based Tokenizer if you would like. For the purposes of this lab, use the “Default Tokenizer”.
7. You can add more other project managers to the workspace, however, for this lab, do not select any other project managers.
8. Once the details have been filled, create your workspace.

Lesson 1.2 Import a Type System

1. Click your workspace (if you are not already there) to go the “Entity Types” page in order to create your Type System.
2. You can either manually create your Type System by defining the entity types and the relation types or import into your workspace a json file which has a pre-defined Type System specified. Importing an existing type system will also allow you to adapt the type system to the needs of your new project.
3. Open the TIRTypeSystem.json file from the folder to which you downloaded the files related to the lab and review the file in a text editor. The file displays a type system with three entity types and one relation type.
4. Click the option to upload a Type System.
5. Drag and drop the TIRTypeSystem.json file from your explorer window to the window that prompts you with the json file to upload.
6. You should see the entity and relation types from the existing type system imported.

Lesson 1.3 Edit a TypeSystem

1.3.1 Add new entity types

Edit the type system that has been imported to add new entity types and relation types. Add the following:

Entity types:

- Vehicle
- Incident
- Person
- Carpart

1. Click on “Add Entity Type” from the “Type System” page -> Entity Types tab.
2. Add the entity type “Vehicle” and hit the return key. You will see a new entry for vehicle. Similarly add the remaining entity types Incident, Person and Carpart.

3. Edit the entity types Model and Manufacturer and add a role “Vehicle” to both. E.g. to edit Model entity type, check the entity type and click on the “Edit Entity Type” button. In the “Select a role” drop down menu, pick “Vehicle” in order to add a role to Model entity type.

1.3.2 Add new relation types:

- locatedOn (CarPart, Vehicle)
- occupantOf (Person, Vehicle)

1. Click on the “Relation Types” link on the left.
2. Click on “Add Relation Type” button
3. Type the relation type name as “locatedOn” and choose the first entity type as CarPart and the second entity type as Vehicle.
4. Similarly, add a relation type “occupantOf” and choose the first entity type as Person and the second entity type Vehicle
5. Delete the existing relationship between manufacturer and model
6. Once the relation types have been added, the Relation Types tab should look like this:

You have completed Lab 01.

Exercise 02: Upload documents

Goals:

In this lab you will:

- Upload documents to your workspace
- Create annotation sets

At the end of this lab, you should be able to upload representative documents to that will be used to train your machine learning model.

Exercise Instructions

Lesson 2.0 Upload Documents

1. Login to Bluemix and click the instance of your WKS service,*"WKS Lab"* and launch WKS.
2. Click the workspace <*LoginShortName*>*_WKS.*
3. Click the *"Documents"* link under the *"Assets & Tools"* menu on the left.
4. Use the *"Upload Document Sets"* button to upload your documents.
5. Click *"Show Details about the file formats*" link to see the supported file formats.
6. Drag and drop the *Corpus.csv* file with traffic incident report descriptions. The file should be available in the directory to which you downloaded the lab related documents. You will see a total of 5 documents that have been uploaded.

Lesson 2.1 Create 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. Open the documents page for your workspace
2. Click the button to create annotation sets
3. Choose the overlap as 60% and create 2 sets:-

   DocSet1 assigned to yourself
   DocSet2 assigned to yourself

You can add two sets by clicking on the '+' icon

4. Click Generate. You will see DocSet1 and DocSet2 each containing 4 documents

You have now completed Lab 02

Exercise 03: Create Dictionaries and dictionary pre-annotator

Goals:

In this lab you will:

- Create a dictionary
- Create a dictionary annotator to pre-annotate your documents

At the end of this lab, you should be able to create dictionaries and a dictionary annotator

Exercise Instructions

Lesson 3.0 Create a dictionary of manufacturers

1. Launch WKS for your instance and click your workspace ie. *<LoginShortName>_WKS*.
2. Click the *"Pre-annotators"* link under the *"Assets & Tools"* menu item on the left.
3. Go to the *"Dictionaries*" tab
4. Click *"Manage Dictionaries"* to create a new dictionary.
5. Create a new dictionary called Mfr_dict.
6. Choose the entity type as *"Manufacturer"*. Click *"Upload"* to upload a dictionary. Drag and drop *manufacturer.csv* file from your list of files that you downloaded for the lab.
7. Similarly create a dictionary for the entity type by uploading the models.csv file.
   Note: Other than entity types, you can also map classes to dictionaries.
8. You will now see the dictionary and all the entries. Further you could make additional entries if necessary by clicking on the "Add Entry" button.

Lesson 3.1 Create a dictionary pre-annotator

1. Go back to the *"Pre-annotators"* page by clicking the left arrow next to "Manage Dictionaries".

2. Click *"Apply This Pre-annotator"* button to apply the dictionaries to the annotation sets.

3. Check *"DocSet1"* and *"DocSet2"* and click *"Run"* to apply the pre-annotations. You will be notified when the dictionary pre-annotations have been applied to your documents.

You have now completed Lab 03

Exercise 04: Create a human annotation task

Goals:

In this lab you will:

- Create a human annotation task

At the end of this lab, you should be able to create a human annotation task and include the annotation sets in this task.

Exercise Instructions

Lesson 4.0 Create a human annotation task

1. For your workspace, click the *"Documents"* link under the *"Assets & Tools"* menu on the left.
2. Go to the *"Tasks"* tab.
3. Click on the *"Add Task"* button to add a new task.
4. Enter the title *DictTask1*. Leave the deadline as-is. Click the *"Create"* button
5. Check the document sets *"DocSet1"* and "*DocSet2*" and click on *"Create Task"* button on the top right corner
6. Click on the DictTask1 task.
7. Click on "Annotate" button for DocSet2. Click on any of the documents. You should see the Manufacturer and Model already annotated.

You have now completed Lab 04

Exercise 05: Human Annotation

Goals:

In this lab you will:

- Annotate mentions, co-references and relations for the documents in the document set assigned to you

Use the Annotation_Guidelines document provided as a reference to annotate the documents. At the end of this lab, you should know how to annotate documents for training a machine learning model.

Exercise Instructions

Login to the Bluemix and launch WKS for your service instance. Click on the workspace you have created. You will annotate documents that are assigned to you in this project.

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 you may want to purposely annotate the overlapping documents slightly differently just to be able to go through the process of adjudicating.

Lesson 5.0 Annotate Mentions

1. Click on the WKS working workspace <*LoginShortName*>*_WKS*.
2. Click the *"Documents"* link under the *"Assets & Tools"* menu on the left.
3. Go to the *"Tasks"* tab on the top.
4. Click on the "*DictTask1"* task to begin annotation. You will see the document sets assigned for you to annotate. In this case, since you are the only one working on the project, you will see two document sets both assigned to you. Click on the *"Annotate"* button for one of the document sets.
5. You will see the documents that have been assigned to you. Click on one of the documents to begin the annotation task. The default section that comes up is 'Mentions'. For annotating mentions, click the word or phrase that you would like to annotate and choose the entity type (which is color coded) on the right.
   Note: The dictionary pre-annotator you created in Lab 03 should have already labeled mentions of manufacturer and model.

6. As and when you annotate, you can save the annotations by clicking on the ‘Save’ button at the top.
7. To annotate a role for a mention, select the mention, go to the ‘Role’ tab on the right. You will see the list of roles that are applicable to the entity type that has been chosen for the mention. Select the role you would like.

Lesson 5.1 Annotate co-references

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

1. To annotate co-references, go to the co-reference page by clicking co-references on the left.
2. Choose the mentions that belong to the same entity and double click the last mention. You will see the number key on the right with co-referenced mentions having the same number.

Lesson 5.2 Annotate relations

1. To annotate relations, go to the relations page by clicking relations on the left.
2. Choose the mention which belongs to the first entity type and then the mention that belongs to the second entity type in a relation.
3. Select the relation on the right to annotate the relation.

Lesson 5.3 Save and submit your annotations

1. Once you have completed annotating mentions, co-references and relations on all the documents, mark each document as ‘Complete’ and save.
2. Close the document by clicking the ‘x’ button to the right of the status drop down box to close the current document and annotate the remaining documents.
3. Once you 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.
4. Carry out annotation tasks on the second document set also that has been assigned to you.

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 you may want to purposely annotate the overlapping documents slightly differently just to be able to go through the process of adjudicating.

You have completed Lab 05

Exercise 06: Adjudication

Goals:

In this lab you will:

- Adjudicate the overlapping documents among human annotators
- Promote annotated documents to ground truth

At the end of this lab, you should know how to interpret the Inter-Annotator agreement, adjudicate and create ground truth.

Lesson 6.1 Adjudicate and build SIRE model

Once the documents have been submitted, the project manager needs to promote the documents to ground truth if the annotations are acceptable. To do so, the first step is to review the IAA score to see if the F measure, precision and recall are acceptable.

6.1.1 Annotate documents for <LoginShortName>_WKS

1. if you have not already done as part of lab 5, go to the workspace < *LoginShortName*>_WKS where <LoginShortName> is your short IBM login. Annotate and submit the documents from both document sets that have been assigned to you.

6.1.2 Set the IAA threshold

2. Click the *"Settings"* menu item on the left.

3. Go to "IAA Settings" tab.

4. Set the threshold IAA value as 0.5 and save.

6.1.3 Adjudicate and promote to ground truth

1. Go back to the *"Tasks"* tab from *"Assets & Tools" -> Documents.*
2. Click *"DictTask1"*.
3. Click *"Calculate Inter-Annotator Agreement"* button
4. 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 lab, assume that the IAA value is acceptable.

5. Click the *"Back to DictTask1 Task"* link.
6. Since the IAA is acceptable, you are now ready to create ground truth that will be used in training the machine learning annotator. Check *DocSet1* and *DocSet2* and click *"Approve".*
7. Once approve is clicked, the documents that do not overlap are automatically promoted to ground truth. The documents that do overlap need to be adjudicated.
8. To adjudicate conflicts, click on *"Check Overlapping Documents for Conflicts"* button.
9. To start resolving conflicts, click *"Resolve Conflicts"*. The adjudication tool shows how many mention, relation, and co-reference chain conflicts exist. Click "Accept" or "Reject" for each individual annotation
10. Continue 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.

Lesson 6.2 Create a machine learning annotator

1. The steps to follow to build a ML annotator is documented in the user guide. It can be accessed from:

   https://console.bluemix.net/docs/services/knowledge-studio/tutorials-create-ml-model.html#wks_tutml_intro

2. A set of 5 documents is not enough of a training corpus to build the machine learning annotator. If you would like practice building a machine learning model, you will need more documents, at least about 10 to 20. Even with those, the model may not perform well since the training corpus is too small. You can obtain more NHTSA traffic incident reports from: http://www-nass.nhtsa.dot.gov/nass/sci/SearchForm.aspx

You have completed Lab 06

Exercise 07: Rule Annotator

Goals:

In this lab you will:

- Create a rule based annotator to pre-annotate documents

At the end of this lab, you should know how to create a rule based annotator using the Rules editor within WKS and use it to pre-annotate documents.

Lesson 7.1 Use rules editor to create new rule

Watson Knowledge Studio version includes a new rules editor that can be used to build a rule based annotator. This annotator can be used to pre-annotate documents that are used for training the machine learning model. In this lab, we will use the manufacturer and models dictionary to create a rule to identify the mentions of model year of vehicles found in your documents. The rule-based annotator can also be independently deployed to Watson Explorer or Watson Discovery service. However, deploying a rule based annotator is only available as an experimental feature.

NOTE: Applying a rule based annotator for pre-annotations on documents that are already annotated will erase the existing annotations (human annotated and any other pre-annotations). In an actual project, you can apply only one pre-annotator. Any pre-annotator that is applied after annotating documents will erase existing annotations.

7.1.1 Create a rule using rules editor

1. Launch Watson Knowledge studio from Bluemix and click the < *LoginShortName>_WKS workspace.*
2. Go to the **Rules** link under the *"Document Annotation"* menu on the left.
3. On the right, click the '+' sign to create a class called Manufacturer.
4. Similarly create two other classes Model and ModelYear.
5. Click the '+' next to "**Documents**" in order to add a sample text that can be used to create the rule. Enter the **Title** as "SampleText" and enter the following passage in the **Text** field:

"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. The Hyundai was equipped with seat back mounted side impact air bags for the front seat positions and curtain air bags for the four outboard positions. In addition to the side impact air bags, the Hyundai was equipped with a Certified Advanced 208-Compliant frontal air bag system. The Hyundai was involved in an intersection crash with a 2005 Ford Explorer that resulted in left side damage to the Sonata. A 2004 Ford Focus subsequently struck the Explorer as the vehicle came to rest. As a result of the crash, the Hyundai's left seat back mounted side impact air bag and the left side curtain air bag deployed. The driver of the Hyundai sustained minor injuries during crash and was transported to a hospital where she was treated and released."

---

6. Click "*Add*".
7. Go to the *Dictionaries* link on the left.
8. You will see two dictionaries – Mfr_dict and Model_dict
9. Click on the Mfr_dict and map the dictionary to the Manufacturer class
10. Similarly map the Model_dict dictionary to the Model class.
11. Now if you expand the document and make sure Manufacturer and Model classes on the right checked, those mentions in the document will be annotated.
12. Move to the *"Rules"* link under the *"Document Annotation"*
13. Highlight the text *2007 Hyundai Sonata* in the first sentence and click the *'+'* sign next to *Rules* to create a new rule.
14. Click on the **tick mark** on 2007 and make sure that the repeat setting is set to "**Required (Exactly 1)"**
15. Click the "**Text**" icon on 2007 and make sure you select the "Character Type: Numeric" condition. This ensures that the token that should be annotated is actually a number. Similarly make sure that the length of the token is exactly 4. Make sure that "Text: 2007" is also not highlighted.
16. Click on the **tick mark** icon above manufacturer. Make sure **"Required (Exactly 1)"** is selected.
17. For model, keep the **Repeats** setting as "**Occuring 0 or 1 time**" and the **Text** setting to "**Rule Match: model"**
18. Click the rectangular box above 2007 and choose the class as **ModelYear** from the **Class** drop down.
19. Change the name of the rule to **Rule_ModelYear** and click **Save**.
20. Notice now that the remaining ModelYears automatically get annotated.

7.1.2 Create a new document set for using the rules pre-annotator

1. Make a copy of the corpus.csv file and call it corpus1.csv.
2. From the *"Documents"* page upload the corpus1.csv file
3. Create a new annotation set from the corpus1.csv base set. Name the new annotation set as *RuleSet* and allocate *100%* of the documents to yourself.

7.1.3 Create and run the rule pre-annotator

1. Click the *"Versions"* link below the *"Model Management"* menu on the left.
2. Go to the "Rule-based model type mapping" tab and map the entity types **Manufacturer, Model and Model_Year** to the classes **Manufacturer, Model and Model_Year** respectively.
3. Go to the *"Rule-based"* tab.
4. Click the *"Run this model"* button. Choose the document set on which the rules pre-annotator needs to be run as *"RuleSet".*
5. Once the pre-annotator has completed running on the RuleSet documents, you will see a confirmation on the top right corner.

7.1.4 Create a new human annotation task

1. Create a new task called RuleTask and include the RuleSet annotation set in the task.
2. Click on the RuleTask and click the "Annotate" button for the RuleTask.
3. You should now see mentions of entity types "Manufacturer", "Model" and "Model_Year" annotated.

You have completed Lab 07