

Lab Center – Hands-On Lab

Session 6958

IBM Watson Natural Language Processing Services
and Discovery in action



DISCLAIMER

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results like those stated here.

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.

IBM products are manufactured from new parts or new and used parts. In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates. Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

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

OpenShift is a trademark of Red Hat, Inc.

UNIX is a registered trademark of The Open Group in the United States and other countries.

© 2020 International Business Machines Corporation. No part of this document may be reproduced or transmitted in any form without written permission from IBM.

U.S. Government Users Restricted Rights — use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.

We Value Your Feedback

Don't forget to submit your Think 2020 session and speaker feedback! Your feedback is very important to us – we use it to continually improve the conference.

Access the Think 2020 agenda tool to quickly submit surveys from your smartphone or laptop.

Table of Contents

1 Getting Started	6
1.1 Overview	6
1.2 Working with IBM Cloud.....	6
1.2.1 IBM Cloud	6
1.2.2 AI services.....	6
1.2.3 Watson Discovery	6
1.3 Prerequisites	7
1.3.1 Obtaining a IBM Cloud account.....	8
1.3.2 Downloading OS-specific coding-friendly editing tool	8
2 Creating Your Watson Discovery Instance	9
3 Configuring Your Watson Discovery Service Instance.....	13
3.1 Using Smart Document Understanding to identify, manage, and enrich fields within your index ..	15
3.1.1 Identify Fields.....	15
3.1.2 Manage Fields.....	18
3.1.3 Enrich Fields	19
4 Obtaining data insights/Querying the collection.....	22
4.1 Search for Answers from Relevant Documents.....	23
5 Relevancy Training.....	25
5.1 Overview and Requirements.....	25
5.1.1 Train Watson to Improve Results	25

Lab Structure

When completing this lab, we included explanations for reference to support the general understanding of the subject matter and to get a better sense of what you are actually doing.

To complete the exercises, you only need to perform numbered steps.

This lab exercise is designed for every student to work independently using a Mac or PC. Each student requires their own IBM Cloud account to create an IBM Watson Discovery instance in Cloud.

Exercise Environment

The lab exercise will be performed on the IBM Watson Discovery provided as part of the AI catalog on IBM Public Cloud.

Course Focus

The lab consists of 4 main sections.

1. Getting started with Watson Discovery and creating your own instance and environment.
2. How to ingest content within IBM Watson Discovery for Agent Assist use cases.
3. How to enrich your data within IBM Watson Discovery.
4. How to extract relevant insight from your data set by querying the collection.

Preparation

To prepare for the lab exercises, locate the Lab folder provided to you by your instructor or open the github repository on your machine.

The lab folder contains the following files:

File name	Description
WDS Lab Exercise for Banking	Lab manual
Banking_LabSample PDF	Real Chase FAQ document available as a PDF
Banking Process PPT	Mock PPT to describe internal mock processes
LabSampleDoc2	Additional PDF document (for future use)

1 Getting Started

While this Lab can be performed using Linux, the current version of this guide assumes the use of a Windows or iOS computer to complete all tasks.

To perform these lab exercises, follow these instructions to set up your computer.

1.1 Overview

This lab guides you through the configuration steps required to use Watson Discovery Service to satisfy an Agent Assist scenario. Watson Discovery Service will be configured to help knowledge workers find relevant insight from complex documents. The lab is meant to guide Tanya, the ABC Bank Solution Admin, on how to configure WDS to satisfy the use case. The end-user is Henry, an ABC Bank Agent, responsible for quickly providing answers to customers.

The lab walks through the development activities required to configure IBM Watson Discovery Service to satisfy these use cases. What this lab does not cover are the broader best practices and considerations that you should follow when creating an enterprise-grade solution.

1.2 Working with IBM Cloud

1.2.1 IBM Cloud

IBM Cloud is an implementation of IBM's Open Cloud Architecture, leveraging Cloud Foundry to enable developers to rapidly build, deploy, and manage their cloud applications, while tapping a growing ecosystem of available services and runtime frameworks. The purpose of this guide is not to introduce you to IBM Cloud, which you should already be familiar with, at least on a high level.

1.2.2 AI services

IBM Watson is a platform of AI services, designed to help developers build solutions and extract insight from Big Data. AI computing systems learn and interact naturally with humans to augment their ability to make better decisions from data.

As such, Watson offers a variety of AI services that cover various aspects of natural interaction including text (Watson Discovery, Natural Language Classifier, Watson Assistant), images (Visual Recognition), and speech (Speech To Text and Text To Speech).

Furthermore, IBM offers services to understand a user's personality (Personality Insights) and emotional/social tone (Tone Analyzer) in a scalable manner.

1.2.3 Watson Discovery

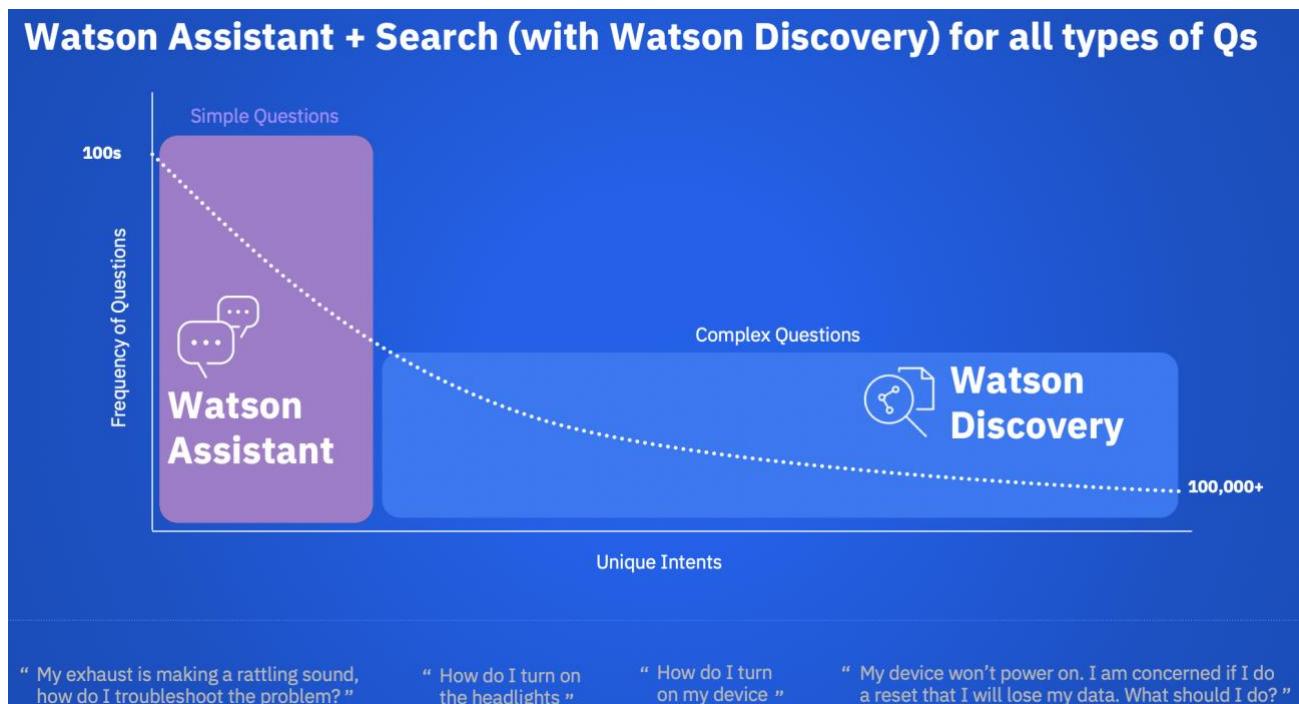
This guide takes an instructional approach to working with the IBM Watson Discovery, a cognitive search and content analytics engine that allows applications to identify patterns, trends and actionable insights that drive better decision-making.

Watson Discovery Service comes with an easy-to-use graphical interface to connect, ingest, configure, and query content indexed into collections. Creating and querying your first collection using the IBM Watson™ Discovery (WD) entails the following steps:

- Creating your Watson Discovery Service environment;
- Configure the Watson Discovery Service pipeline to convert, enrich, and normalize your content (involves using a sample data file);
- Ingest your content in a Collection as a searchable text index; and
- Creating queries to extract relevant insight from your Collection.

IBM Watson Discovery is designed to securely unify structured and unstructured data with pre-enriched content, and uses a simplified query language to eliminate the need for manual filtering of results. The input data can be a CSV, PPT, HTML, DOC, PDF, JPG, or JSON document. WDS returns the output in the form of JSON which can be easily queried to gain insights into the data.

IBM Watson Discovery service can also be used to augment Watson Assistant and/or address use cases where “long tail” less frequently asked questions require answers obtained from documentation. This use case will not be covered in this Lab.



1.3 Prerequisites

This section provides instructions to help you get started quickly with the IBM Watson Services.

1.3.1 Obtaining a IBM Cloud account

IBM Cloud is a cloud PaaS (Platform as a Service), which allows you to host your application on-line and bind it to a variety of SaaS service offerings from IBM including Watson services.

Learn more at <https://www.ibm.com/cloud/> and if you are new to IBM Cloud, you can create a trial account at <https://cloud.ibm.com/registration>.

1. Direct your browser to the Cloud home page: <https://cloud.ibm.com/> to access your dashboard
2. If you do not yet have a Cloud account, click Sign Up on the top right
3. Enter requested information and click Create Account

If you use your personal e-mail address, you have 30 days to evaluate Cloud. Some services, such as Watson Discovery Service, are free for limited use during the trial period.

1.3.2 Downloading OS-specific coding-friendly editing tool

4. If you are using a PC, you can use Notepad ++ (<https://notepad-plus-plus.org/download>)
5. For Mac, use Sublime Text (<http://www.sublimetext.com/3>)→also works for Windows

2 Creating Your Watson Discovery Instance

1. Navigate to IBM Cloud - : <https://cloud.ibm.com/> and Log In

The screenshot shows the IBM Cloud homepage with a dark blue header. The main heading is "IBM Cloud" with the tagline "Start building immediately using 190+ unique services." Below the tagline is a button labeled "Create an IBM Cloud account". A section titled "Get a \$200 credit when you upgrade" provides information about upgrading to a Pay-As-You-Go account. At the bottom left, there's a "Follow us on" section with social media links (Facebook, Twitter, YouTube, LinkedIn). The bottom right corner includes copyright information ("© Copyright IBM Corp. 2014, 2019. All rights reserved.") and a "Cookie Preferences" link.

Welcome to
IBM Cloud

Start building immediately using
190+ unique services.

Create an IBM Cloud account

Get a \$200 credit when you upgrade
After you upgrade to a Pay-As-You-Go account, you can use the credit to try new services or scale your projects. The credit is valid for one month and can be used with any of our IBM Cloud offerings.

Learn more:
[Pricing](#) [Catalog](#) [Docs](#) [Status](#)

Follow us on
[f](#) [Twitter](#) [YouTube](#) [LinkedIn](#)

© Copyright IBM Corp. 2014, 2019. All rights reserved.
Cookie Preferences

2. Use your organization credentials obtained in Step 0.3.1 to Log in to IBM Cloud.
3. Once logged in, navigate to the Cloud Catalog by clicking on Catalog in the top right-hand corner.
4. Search for Discovery service in the IBM AI catalog:

The screenshot shows the IBM Cloud Catalog interface. The top navigation bar includes "IBM Cloud", search, "Catalog", "Docs", "Support", "Manage", and user account information. The main area is titled "Catalog" with a "Search the catalog..." input field and a "Filter" button. On the left, a sidebar lists "All Categories" with "AI" selected. The main content area displays the "AI" category with several service cards. The "Discovery" service card is highlighted with a blue border, indicating it has been selected.

IBM Cloud

Catalog Docs Support Manage 1781269 - IBM

Dashboard Create resource

IBM Cloud Search resources and offerings... Catalog Docs Support Manage 1781269 - IBM

Catalog Search the catalog... Filter

All Categories AI

- Watson Assistant** Lite • IBM • IAM-enabled
- Compare and Comply** Lite • IBM • IAM-enabled
- Discovery** Lite • IBM • IAM-enabled
- Knowledge Catalog** Lite • IBM • IAM-enabled
- Knowledge Studio** Lite • IBM • IAM-enabled
- Language Translator** Lite • IBM • IAM-enabled

5. Click on the Discovery tile
6. Within this window, you can provide a unique Service name (e.g. Discovery_TeamName). Click on the blue box under Service name to provide a unique name to your instance. Do not make any changes to the Credential name
7. Create a trial service instance by selecting the Lite Plan option (default setting) and by clicking Create.

The screenshot shows the IBM Cloud service catalog interface. At the top, there's a navigation bar with 'IBM Cloud', a search bar, and various account management links. Below the navigation is a service card for 'Discovery'. The card includes a brief description of the service: 'Add a cognitive search and content analytics engine to applications to identify patterns, trends and actionable insights that drive better decision-making. Securely unify structured and unstructured data with pre-enriched content, and use a simplified query language to eliminate the need for manual filtering of results.' It also lists basic metadata: AUTHOR (IBM), PUBLISHED (03/13/2019), and TYPE (Service). The main body of the card is titled 'Features' and lists two bullet points: 'Rapid results' and 'Domain intelligence'. Each feature has a detailed description below it. At the bottom of the card, there are 'Need Help?' and 'Contact IBM Cloud Support' links, along with 'Add to estimate' and 'Create' buttons.

8. IBM Cloud will now instantiate your personal Watson Discovery Service instance. It will automatically redirect you to your Resources Dashboard while the system is being provisioned. Once it has been provisioned you can click on it.

The screenshot shows the 'Resource list' dashboard. At the top, there's a navigation bar with 'IBM Cloud', a search bar, and various account management links. The main area displays a table of services. The columns are: Name, Group, Location, Offering, Status, and Tags. There are filters for each column. A section titled 'Services (19)' is expanded, showing two entries: 'Discovery-n1' (Group: Default, Location: Dallas, Offering: Discovery, Status: Provision in progress) and 'Knowledge Studio-sv' (Group: Customer Care, Location: Dallas, Offering: Knowledge Studio, Status: Provisioned). At the top right of the table area, there's a 'Create resource' button. Below the table, there are links for 'Collapse all | Expand all'.

If you close the tab and would want to re-access it, you can navigate to the Resources List dashboard and by selecting the created discovery service from the list of services created in your account. The list of the services you have created can be found at <https://cloud.ibm.com/resources>.

9. When you click on Discovery instance from your Resource List you will be directed to the service dashboard.

The screenshot shows the IBM Cloud Resource List interface. On the left, there's a sidebar with 'Manage' selected. The main area displays a service instance named 'Discovery-n1'. Below the service name, it shows 'Resource Group: Default' and 'Location: Dallas'. There are buttons for 'Add Tags' and a three-dot menu. A central panel has a welcome message: 'Get started by launching the tool.' It includes a 'Launch tool' button, a 'Getting started tutorial' link, and an 'API reference' link. To the right, there's a 'Plan: Lite' section with an 'Upgrade' link. Below this, a 'Credentials' section shows an 'API Key' field containing a redacted value and a 'URL' field containing 'https://gateway.watsonplatform.net/discovery/api'. There are download and show credentials links next to these fields.

10. Launch the discovery tool by clicking “Launch Tool” button from the discovery service instance
 11. Launching the tool takes you to the dashboard with welcome message overlay. Click on the cross button on the right top to close the overlay window.

The screenshot shows the Watson Discovery Service dashboard. On the left, there's a sidebar titled 'Manage data' with sections for 'Create a new data collection' and 'Watson Discovery News'. A welcome overlay is displayed in the center. The overlay has a dark background with white text. It says 'Welcome to Watson Discovery Service!' and 'This guide will help you to set up, enrich, and query your data with Discovery.'. Below this is a blue 'Let's get started' button. At the bottom of the overlay, there's a note: 'If you don't need help, turn off these useful tips.' In the top right corner of the overlay, there's a large white 'X' button used to close the overlay.

12. In the dashboard, you can see the collections present in your Watson Discovery Service environment. An option is provided to create a new collection.

IBM Watson Discovery

Cookie Preferences Instance: Discovery-n1

Manage data Collections of your private data and pre-enriched data to configure and query against. [Learn more.](#)

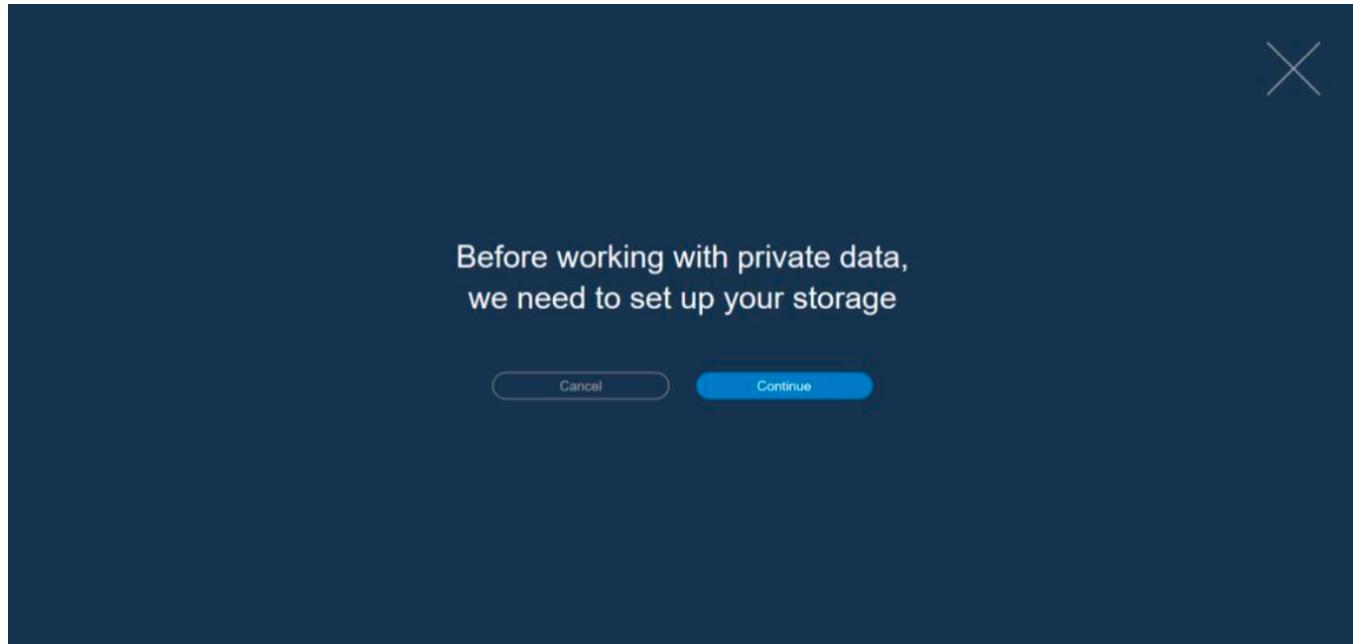
Create a new data collection [Upload your own data](#) [Connect a data source](#)

PRE-ENRICHED DATA
Watson Discovery News
News sources: English

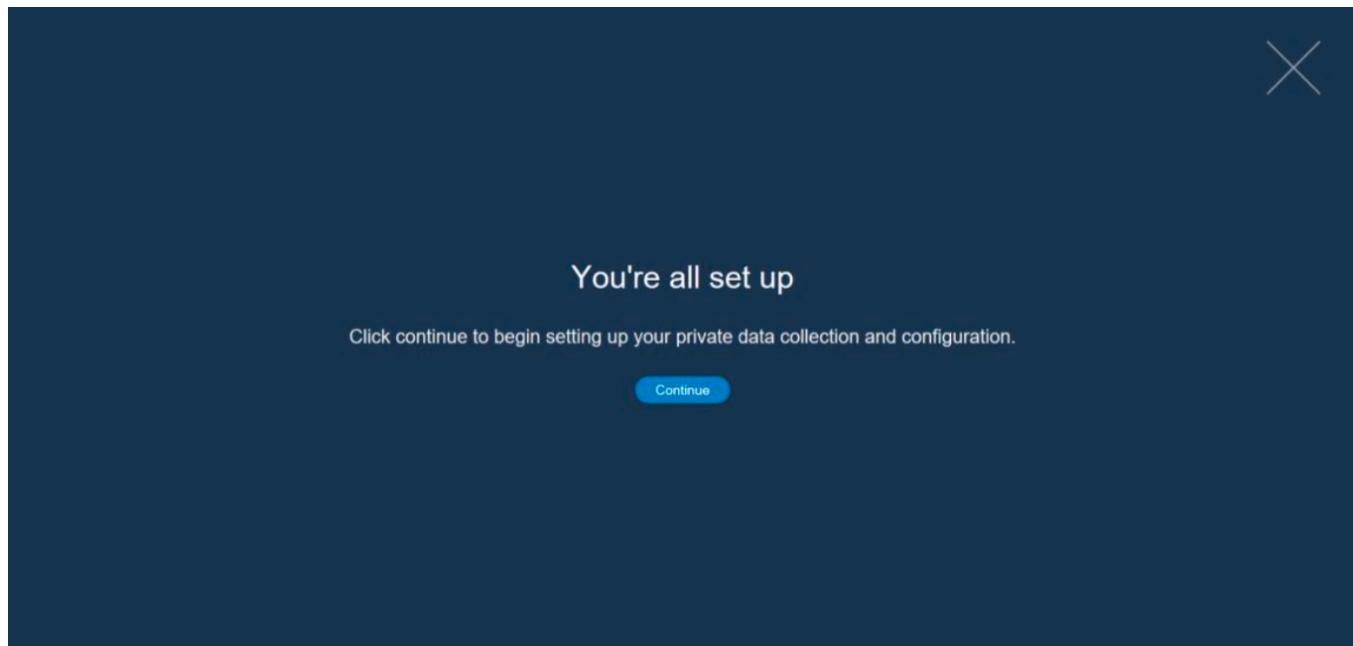
A new Watson Discovery Instance will come with a default environment, configuration and a pre-enriched collection, Watson Discovery News. To upload data, you will have to create a new collection.

3 Configuring Your Watson Discovery Service Instance

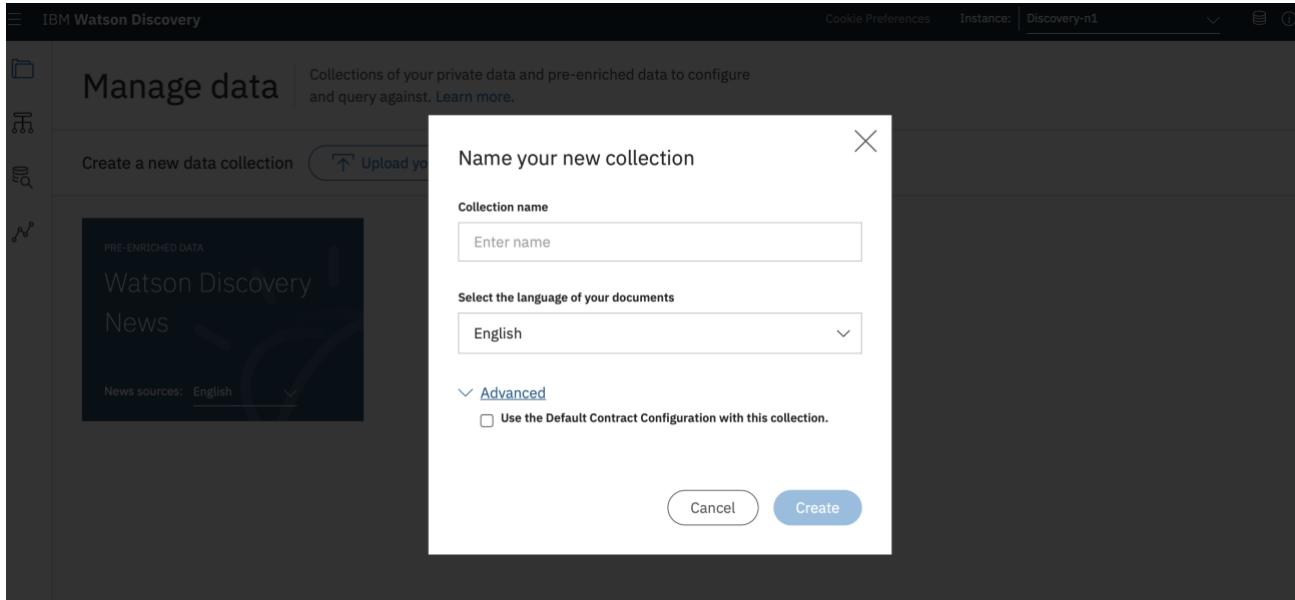
1. Create “Agent Assist” collection by clicking on “Upload Your Own Data” button.
2. When you create your first collection, the tooling will indicate that the storage needs to be setup before working with private data. Click “Continue” button. This process can take 1-2 minutes



3. You get a message saying, “You’re all set up” once the storage is provisioned and the collection is created. Click “Continue” button



4. Specify the name for your Collection. In this Lab, you will create a collection that you will want to name “Agent Assist”.
5. Notice that English language is selected by Default. Under “Advanced” you can select a Contract specific NLP model. This will not be used in this Lab. Click the Create button.



6. The collection landing page opens up automatically. The collection landing page displays information like configuration id, environment id, collection id and the current configuration that was selected for your collection.

Upload data to get started

Drag and drop your documents here, or [select documents](#)

PDF, HTML, JSON, Word, Excel, PowerPoint, PNG, TIFF, JPG
50MB max per document

Before ingesting data, it is important that you create a configuration tailored to your content format and structure.

7. To create a new configuration, click on “select documents” button, locate the Chase PDF document and the PPT shared for this Lab, and click “Choose” to ingest the two.
8. Once ingestion begins, you can click on “Configure Data” in the top right hand corner. You are now ready to build your Configuration pipeline.

3.1 Using Smart Document Understanding to identify, manage, and enrich fields within your index

In the “Configure Data” windows, you have access to Smart Document Understanding (SDU) capabilities which allows Watson Discovery to visually learn the structure of documents. It allows you to train on examples of the document structure, and then apply the model to more data, to take something like a pdf and extract structure and meaning. This becomes extremely valuable when trying to create well-formatted, metadata-rich answers.

Familiarize yourself with this window:

- On the right are “Labels” these are the tags that you can assign to different parts of the document. (Note this is training on the structure, so it ignores the words on the page).
- In the center you see a view of the original PDF and a view with each section block highlighted, this is where you can make annotations by selecting a label and then a block
- On the left are all the pages of the document.
- For each page, SDU will show a preview of what the model thinks the content should contain, as you add annotations, you will see these predictions update for future pages.

The screenshot shows the IBM Watson Discovery interface with the following details:

- Header:** IBM Watson Discovery, Cookie Preferences, Instance: Discovery-n1, Apply changes to collection.
- Left Sidebar:** Agent Assist / Configure data, Identify fields (selected), Manage fields, Enrich fields.
- Central View:** Shows page 1/17 of 'Banking_LabSamplePDF.pdf'. The page contains sections like 'CHASE GUIDE TO BENEFITS', 'CHASE SAPPHIRE Visa Signature', and 'For questions, call 1-888-320-9961.' with page numbers 2, 3, 4, and 12.
- Right Panel:** Field labels panel with a list of categories: answer, author, footer, header, question, subtitle, table_of_contents, text, title, image, and table (BETA). Each category has a corresponding colored circle and a 'Create new' button.
- Bottom:** Submit page button and a note: Viewing: Applied predictions of collection model.

3.1.1 Identify Fields

Make sure the Banking_LabSample PDF is selected in the drop-down menu.

- Notice that Watson Discovery is automatically converting the PDF content into text form. Click on the “page” icon to overlay the text extraction view with the original PDF.

Let's disregard the title page as it is somewhat unique. Starting on page two, you can see the structure of the document. It has a Title section in blue and then a number of FAQs as subtitles. So we can use the labels to mark the document as such. There is also a consistent footer on every page, we want to identify that footer so we can potentially remove it.

- Use the labels on the right to highlight 1) questions using the question label; 2) footers using the footer label; 3) section title using the title label; and 4) bolded section using subtitle field. Consistency is key.

3. Click on Submit page to submit the training. The tool now moves to Page 3.
4. On Page 3, the Smart Document Understanding model is already trying to apply its learnings although no results are shown. The model needs more training data. Let's go ahead and provide more training input. Highlight question and footer fields as shown below.

The screenshot shows the IBM Watson Discovery interface. The top navigation bar includes 'Cookie Preferences', 'Instance: Discovery-n1', and a user icon. Below the navigation is a toolbar with icons for file operations, search, and other functions. The main workspace is titled 'Banking_LabSamplePDF.pdf' and shows page 3 of 17. The document content is displayed with several text blocks highlighted in green, indicating they have been identified as 'question' or 'footer' fields. To the right of the document is a sidebar titled 'Field labels' with a list of categories: answer, author, footer (which is highlighted in green), header, question, subtitle, table_of_contents, text, title, image, and table (Beta). A button at the bottom right of the sidebar says 'Learn more about how to use.'

5. Click on Submit page. The tool will now move to Page 4. Notice that the Smart Document Understanding model has already applied its predictions based on what it has learnt so far. We are now getting good results!

The screenshot shows the IBM Watson Discovery interface. At the top, there's a navigation bar with 'IBM Watson Discovery', 'Cookie Preferences', 'Instance: Discovery-n1', and other icons. Below the navigation is a sidebar with icons for 'Agent Assist / Configure data', 'Identify fields' (which is selected), 'Manage fields', and 'Enrich fields'. The main area displays a PDF titled 'Banking_LabSamplePDF.pdf' with page numbers 1/17. On the left, a vertical list shows pages 3, 4, 5, and 6. The right side shows the content of page 1 with several text segments highlighted in yellow, green, and blue, corresponding to different field labels. A legend on the right lists field labels: answer, author, footer (highlighted in green), header, question, subtitle, table_of_contents, text, title, image, and table (BETA). A button 'Apply changes to collection' is at the top right, and a 'Submit page' button is at the bottom right.

6. Go ahead and click on Submit Page to provide one more training input. When you move to Page 5 you will notice we are still missing the Footer. Highlight the footer using footer label and Submit Page 5.

3.1.2 Manage Fields

After training the visual Smart Document Understanding model, you will need to define how Watson Discovery should manage the fields highlighted within your documents. Each label will appear as a separate field for the doc in Discovery.

The screenshot shows the 'Manage fields' section of the IBM Watson Discovery interface. It has a navigation bar and sidebar similar to the previous screen. The main area is divided into two sections: 'Identify fields to index' on the left and 'Improve query results by splitting your documents' on the right. In the 'Identify fields to index' section, there's a note that all fields are indexed by default and a list of toggle switches for 'answer', 'author', 'footer', 'header', 'image', 'question', 'subtitle', and 'table', all of which are currently set to 'On'. In the 'Improve query results' section, there's a note about splitting documents into segments and a button labeled '+ Split document'.

1. Click on the Manage Fields button.
2. Turn off the Footer field to avoid it being indexed (in this way we can keep the index “clean”). Then select “question” as the field we will use to split documents.

IBM Watson Discovery

Agent Assist / Configure data

Identify fields **Manage fields** Enrich fields

Cookie Preferences Instance: Discovery-n1

Identify fields to index

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

answer	<input checked="" type="checkbox"/> On
author	<input checked="" type="checkbox"/> On
footer	<input type="checkbox"/> Off
header	<input checked="" type="checkbox"/> On
image	<input checked="" type="checkbox"/> On
question	<input checked="" type="checkbox"/> On
subtitle	<input checked="" type="checkbox"/> On

Improve query results by splitting your documents

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

Split document on each occurrence of

question

3.1.3 Enrich Fields

Watson Discovery automatically processes the ingested data through a Natural Language Processing pipeline. As the data goes through the pipeline it is tagged with relevant metadata. The metadata added to the content is what we refer to as “enrichments”. Enrichments can include relevant concepts, entities, and keywords extracted as well as the sentiment and emotion detected. The enrichments play a key role in content analytics use cases such as Voice of the Customer.

IBM Watson Discovery

Agent Assist / Configure data

Identify fields **Manage fields** **Enrich fields**

Cookie Preferences Instance: Discovery-n1

Enrich your data with additional Watson insights

Set up rules for which fields you want to apply enrichments to. [Learn more.](#)

Add a field to enrich

Fields to be enriched	Enrichments
text	categories <input type="button" value="X"/> concepts <input type="button" value="X"/> entities <input type="button" value="X"/> sentiment <input type="button" value="X"/>

[+ Add enrichments](#)

1. Click on “Add enrichments”. You will get a pop-up window allowing you to select the enrichments you want to apply.

Add Enrichments

text: categories X concepts X entities X sentiment X

- Keyword Extraction**
Determines important keywords in this field, ranks them, and optionally detects the sentiment.
[Learn more](#) [Add](#)
- Sentiment Analysis**
Identifies the overall positive or negative sentiment within this field.
[Learn more](#) [Added!](#)
- Concept Tagging**
Identifies general concepts that aren't necessarily directly referenced in this field.
[Learn more](#) [Added!](#)
- Category Classification**
Classifies this field into a hierarchy of categories that's five levels deep.
[Learn more](#) [Added!](#)
- Semantic Role Extraction**
Parses sentences into subject, action, and object form and returns additional semantic information.
[Learn more](#) [Add](#)
- Emotion Analysis**
Analyzes the emotions (anger, disgust, fear, joy, and sadness) in this field.
[Learn more](#) [Add](#)

2. Remove “categories” by pressing X on the bubble.
3. Add Keyword Extraction. For the time being these are the relevant enrichments for the data we are ingesting. Click on Done to apply these enrichments.

IBM Watson Discovery

Cookie Preferences Instance: Discovery-n1

Agent Assist / Configure data

Identify fields Manage fields Enrich fields

Enrich your data with additional Watson insights

Set up rules for which fields you want to apply enrichments to. [Learn more](#).

Add a field to enrich Enter a field

Fields to be enriched	Enrichments
text	concepts X entities X sentiment X keywords X

[+ Add enrichments](#)

4. Now that you've set up the configuration pipeline for the PDF file, let's make sure that it will get applied properly to the PPT as well. To do so, click on Identify fields.
5. Select the PPT file from the drop down menu

IBM Watson Discovery

Agent Assist / Configure data

Identify fields Manage fields Enrich fields

Online Credit Card Application Gu... 1/2 ▾

Online Credit Card Application Guideline...

Banking_LabSamplePDF.pdf

Customers Apply for Credit Card

Viewing: Your training

Submit page

Field labels

Identify document elements using the labels below.

+ Create new Upgrade

- answer
- author
- footer
- header
- question
- subtitle
- table_of_contents
- text
- title
- image
- table BETA

Learn more about how to use.

- Notice that the “question” labels are being applied incorrectly to the PPT. That is ok – let’s use the “text” label to train the model further. Click on “Text” and highlight Slide 1 content.
- Click on Submit page to advance to Slide 2. Click on “Text” and highlight Slide 2 content.
- Click on Submit page to advance to Slide 3 – the structure should now be correct.
- Click “Apply changes to collection” and Choose the PDF and PPT files to re-ingest the documents and apply the new configuration pipeline we just created.

You will be directed back to the Collection landing page and you will see the ingestion process kick in and the number of documents increase. Each question/text pair is now its own document!

IBM Watson Discovery

Agent Assist Configure data

Overview Errors and warnings (0)

50 documents

0 documents failed View details

Created on 3/19/2019 2:31:22 pm EDT

Last updated 3/19/2019 2:31:22 pm EDT

Upload documents

Identified 3 fields from your data

- question
- subtitle
- text

Need to identify more fields? [Add fields](#)

Added 4 enrichments to your data

Entity Extraction

Benefit Administrator (20) | theft (14) | United States (9) | Chase (5) | Tour Operator (5)

Sentiment Analysis

44%	14%	42%
positive	neutral	negative

Concept Tagging

Insurance (11) | United States (10) | Automobile (8) | Travel (6) | Vehicle (6)

Keyword Extraction

Benefit Administrator (16) | claim (16) | benefit (15) | damage (12) | days (12)

Now you're ready to query!

Entities of type **JobTitle** which have positive sentiment Run

Documents that contain **Insurance**, but not **United States** Run

Documents about **Benefit Administrator** as a **JobTitle** with a very negative sentiment Run

[Build your own query →](#)

5 enrichments available. [Add enrichments](#)

4 Obtaining data insights/Querying the collection

After the discovery instance is configured and data is ingested, the collection will be queried to gain insights into the data. The data schema will also be provided to gain an overall view of the collection and documents.

1. To build a query, you can either click “Build queries→” in the bottom right corner or click on the build queries icon on the left hand side bar. Option 2 is circled in red below. If you click the icon, you will have to choose the collection you want to query which is “Agent Assist”.

The screenshot shows the IBM Watson Discovery interface. At the top, there's a dark header bar with the text "IBM Watson Discovery". Below it is a sidebar with several icons: a folder icon, a magnifying glass icon (circled in red), a network icon, and a line graph icon. The main area has tabs for "Overview" (which is active and underlined) and "Errors and warnings (0)". A large central section displays the number "49" and the word "documents". To the right of this, there's a warning icon and the text "0 d" followed by a link "View details". Below this, a message says "Identified **4 fields** from your data" with a list of four items: "question" (green square), "subtitle" (pink square), "text" (yellow square), and "title" (purple square). At the bottom, a message says "Need to identify more fields? [Add fields](#)".

2. In the “Build queries” page, you can choose to “Search for documents,” “Include analysis of your results,” or “Filter which documents you query.”

4.1 Search for Answers from Relevant Documents

In this section, we will walk through queries that an SME, Solution Admin, or developer (Tanya) would configure within the application in order to gain insight from the data. The application end user for this first use case is Henry, an ABC Bank Agent. He is responsible for assisting customers with their ABC bank

products and offering. In this day and age, client trust means everything. And Clients rely on Henry for timely and accurate advice. He is constantly tasked with digging up information hidden in company policies, documents, FAQs in order to provide the best recommendations to clients. In a study of how agents spend their time, we found that agents spend up to 35% of their days searching for information from a knowledge base or CRM. Additionally, customer service agents use dozens of disconnected applications in the course of resolving a single customer issue, often duplicating data from application to application or performing repetitive, manual tasks —actions that take up 15% of an agent's day. There's a lot of pressure to provide quality answers to his Clients in little time. He begins his day by accessing the Watson Agent Assist app just in time as customer inquiries start coming in.

The Discovery application provides Tanya an option to build their own customer query and execute them. The query results will be shown in the right window of “Build your own query” page.

1. Click on “Search for documents”. Ensure “Use natural language” is underlined.

2. Type “does auto rental coverage apply in other countries” and click Run Query. By default Watson Discovery will return relevant passages AND documents that pertain to the query. The next step is to fine tune the queries to get even better results.
3. Click “More Options” to explore more advanced querying techniques.
4. Under Passages, click No to turn off Passages.

5. Under Documents, click Specify and select the “text” field. Also update the Number of documents to return and set it to 3.

The screenshot shows the IBM Watson Discovery interface with the following settings:

- Passages** section: Includes "Include relevant passages" with "No" selected.
- Documents** section:
 - Fields to return**: "Specify" is selected, and "text" is entered in the input field.
 - Number of documents to return**: The value is set to 3.
 - Number of query results to skip at the beginning**: The value is set to 0.
- Buttons**: "Run query" and "Close".

We are building a query that leverages the smart structure we built using Smart Document Understanding. We want the query to return only the “text” field (not the questions), and we want to showcase the top 3 results back to the user.

6. Click Run Query and then click JSON to see the results.

You will notice that the relevant document is now being shown in second position. The answer returned contains the full text (not just a snippet) – which is exactly what an end user needs to answer that question.

5 Relevancy Training

During this lab exercise, we have facilitated the configuration steps by providing you with sample content data and example queries. In this section, we will go over how you can improve your answer results by training the model. Relevancy training (RT) is optional; if the results of your queries meet your needs, no further training is necessary. In cases with really hard questions to answer where there are a lot of possible documents linked to queries, relevancy training is most useful. Relevancy training can be done programmatically or in the tooling. In this section, we will show you how to train the model in the tooling.

5.1 Overview and Requirements

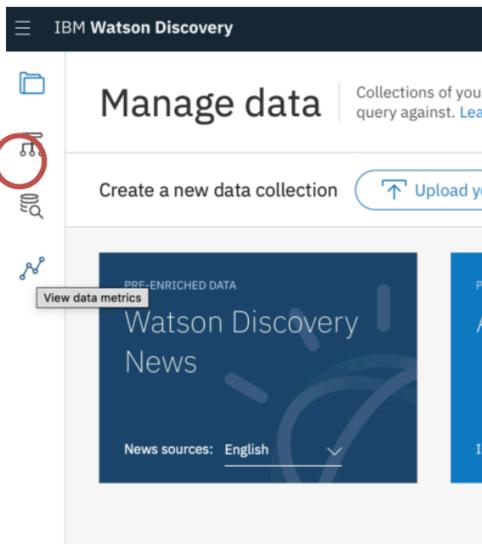
The relevance of natural language query results can be improved in Watson Discovery service with training. In order to train Watson, you'll need to provide example queries that are representative of queries your user enters and provide ratings to say which results for each query are relevant and not relevant. The requirements for Discovery to begin applying your ratings include training a minimum of 49 queries or possibly more (Watson will give you feedback if it needs more queries in order to train) and applying both Relevant and Not relevant ratings. Only rating Relevant documents will not provide the data needed.

In this section, we will only train two example queries to show the training process flow. In real life instances, a minimum of 49 queries must be trained for Discovery to begin applying rating.

5.1.1 Train Watson to Improve Results

We will train the results for the query “does auto rental coverage apply in other countries” to help the model rank the right answer to be more relevant. Remember to rate them with both Relevant and Not relevant.

1. In order to train your model, start by reviewing the logs. To do so, click the “View Data Metrics” icon on the left-hand side bar.



2. Choose “View all and perform relevancy training” to take the latest queries and perform Relevancy Training.

The screenshot shows the IBM Watson Discovery interface. On the left, there's a sidebar with icons for file management, search, and other functions. The main area is titled "Performance" with a "Learn more" link. Below it, there's a section titled "Fix queries with no results by adding more data" which includes a "QUERIES WITH NO RESULTS" box containing the message "All queries have returned results." To the right, there's a "MOST RECENT QUERIES" section with three entries: "how long will it take to submit the application?", "how long will it take to submit the application?", and "how long will it take to submit the application?". At the bottom right of this section is a blue button labeled "View all and perform relevancy training".

3. Choose “Agent Assist” as the collection that you want to train. Then click “Train this collection”

A modal dialog box is displayed over the background. The title of the dialog is "Select a collection to train". Inside, there is a dropdown menu showing "Agent Assist" as the selected option. At the bottom of the dialog are two buttons: "Cancel" on the left and a larger blue "Train this collection" button on the right. The background of the dialog is white, while the rest of the screen is dark.

4. Now click “Add recent queries from Watson Discovery to Agent Assist”

The screenshot shows the IBM Watson Discovery interface. At the top, there's a dark header bar with the text "IBM Watson Discovery". Below it, on the left, is a sidebar with four icons: a folder, a network, a magnifying glass, and a gear. The main content area has a breadcrumb navigation "Agent Assist / Train Watson". A message says "Watson will learn which are the best results for your queries after you've rated enough." Below this are three buttons: "Add more queries", "Rate more results", and "Add more variety to your ratings". Underneath this, there's a section titled "Queries (0)". It includes a sub-instruction "Train Watson by adding natural language queries and rating the results. [Learn more.](#)". There are two blue links: "+ Add recent queries from Watson Discovery to Agent Assist" and "+ Add a natural language query".

5. Select “does auto rental coverage apply in other countries ” and click on “Add to training List”. Then exit the small window.

A modal window titled "Select recent queries from Watson Discovery for Agent Assist to train" is shown. It contains a list of checkboxes. One checkbox is checked: "does auto rental coverage apply in other countries". At the bottom of the modal, there are navigation arrows (< 1 >), a "Cancel" button, and a blue "Add to training list" button.

<input type="checkbox"/> how long will it take to submit the application?
<input type="checkbox"/> how long will it take to submit the application
<input type="checkbox"/> how long will it take to submit the application?
<input checked="" type="checkbox"/> does auto rental coverage apply in other countries
<input type="checkbox"/> does auto rental coverage apply in other countries?
<input type="checkbox"/> does auto rental coverage apply in other countries?
<input type="checkbox"/> does auto rental coverage apply in other countries?

6. For that question, click on “Rate Results”

IBM Watson Discovery

Agent Assist / Train Watson

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

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

Queries (1)

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

Add recent queries from Watson Discovery to Agent Assist

Add a natural language query

does auto rental coverage apply in other countries Rate results Not rated yet

7. Rate “Not Relevant” or “Relevant” based on the answers returned.

IBM Watson Discovery

Agent Assist / Train Watson

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

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

does auto rental coverage apply in other countries

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

Banking_LabSamplePDF.pdf View document

"... The Auto Rental CDW benefit provides reimbursement for damage due to collision or theft up to the actual cash value of most rental vehicles. Within your country of residence, Auto Rental CDW ..."

"... It does not duplicate insurance provided by or purchased through the auto rental company. Auto Rental CDW will not pay for theft or damage reimbursable by your own insurer, employer, employer's insurance, or any other valid and

Show more Relevant Not relevant

Banking_LabSamplePDF.pdf View document

"... your Auto Rental CDW will apply. This benefit is in effect while the rental vehicle remains in your control or in the control of another Authorized Person. Coverage ends when the rental company reassumes control of the vehicle. ..."

Relevant Not relevant

Banking_LabSamplePDF.pdf View document

"... • Reasonable and customary towing charges related to a covered loss to take the vehicle to the nearest qualified repair facility Auto Rental CDW is secondary coverage and provides reimbursement up to the actual cash value of the vehicle as it was originally manufactured. ..."

Relevant Not relevant

As mentioned previously, one example is not sufficient to update the Relevancy model. You will need a minimum of 50 questions to answer mapping. This exercise, if conducted by SMEs, usually takes 30-60 minutes and can significantly improve the Watson Discovery Results.