

# Learn Session Training



## Watson Discovery Service

### Lab Exercise

Topic: Expert Services Learn – Watson Discovery Service

Version: **1.6**

Updated: February 2, 2018

For questions about the Watson workshop and materials, contact:  
*Giulio Soliani E-mail: [gsolian@us.ibm.com](mailto:gsolian@us.ibm.com)*

## Lab Overview

This lab guides you through the configuration steps required to use Watson Discovery Service to satisfy a Voice of the Customer user scenario. Watson Discovery Service will be configured to help end-users visualize and analyze product feedback obtained via customer product reviews.

## Change Log

Revision History: Changes to this document are summarized in the following table in reverse chronological order (latest version first).

Revision	Date	Author	Short Description of Changes
1.6	02/02/18	Raffi K.	Edited Section 5.4 to align with WDS enrichment updates
1.5	12/06/17	Raffi K.	Added Section 7: Relevancy Training
1.4	11/22/17	Raffi K.	Updated WDS photos and query collection steps
1.3	05/20/17	Giulio Soliani	Published Lab v1.3 with updated use case description
1.2	05/03/17	Chethan R	Published Lab v1.2: incorporated review comments
1.1	04/24/17	Chethan R	Published Lab v1.1
1.0	04/24/17	Chethan R.	Published Lab v1.0
0.3	04/21/17	Giulio S.	Reviewed Data Crawler Section
0.2	04/20/17	Giulio S.	Added Section 1
0.1	04/17/17	Chethan R.	Created Section 2-6

## Contents

0	Getting Started .....	6
0.1	Overview.....	6
0.2	Working with Watson Developer Cloud.....	6
0.2.1	IBM Cloud .....	6
0.2.2	Watson Developer Cloud (WDC) .....	6
0.2.3	Watson Discovery Service (WDS) .....	7
0.3	Prerequisites.....	8
0.3.1	Obtaining a IBM Cloud account.....	8
0.3.2	Installing the node.js runtime .....	8
0.3.3	Installing cf (Cloud Foundry) CLI.....	9
0.3.4	Verifying Prerequisites .....	9
0.3.5	Downloading OS-specific coding-friendly editing tool .....	10
0.3.6	Installing the Watson Discovery Service Data Crawler .....	10
1	Creating Your Watson Discovery Instance .....	11
2	Configuring Your Watson Discovery Service Instance.....	15
2.1	Selecting attributes for conversion, enrichment, and normalization in the configuration .....	18
2.1.1	Configuring Enrichment Attributes .....	19
2.1.2	Configuring Conversion Attributes (Optional for this lab) .....	24
2.1.3	Configuring Normalization Attributes (Optional for this lab).....	31
3	Data Ingestion using data crawler .....	37
3.1	Data Crawler Overview.....	37
3.2	Data Crawler Configuration.....	37
3.3	Uploading Data to WDS collection .....	39
4	Data Ingestion using Discovery Tool and API calls .....	40
4.1	Adding documents to the collection .....	40
4.1.1	Uploading documents using the Tooling interface .....	40
4.1.2	Process to Upload documents using API calls (Optional).....	42
5	Obtaining data insights/Querying the collection .....	44
5.1	Search for Documents .....	48
5.2	Include Analysis of Your Results .....	50
5.3	Filtering Documents You Query .....	51
5.4	Top keywords .....	53
5.5	General Sentiments Analysis .....	59
5.6	Top Entities.....	61
6	Useful Considerations: Data Collection .....	64
6.1	Collecting Content for Watson Discovery Service .....	64
6.1.1	Data formats supported .....	64
6.1.2	Database creation and access .....	65
6.1.3	Database table schema for our Amazon Product Review use case .....	66
7	Relevancy Training.....	68
7.1	Overview and Requirements .....	68
7.1.1	Train Watson to Improve Results.....	68

# Course overview

---

## Credits

This project is as a collaboration between many experts within and outside of Watson in the hopes of broadening skills within IBM, our clients, partners and the developer community.

If you have suggestions for how to improve this guide, please send a note to [gsolian@us.ibm.com](mailto:gsolian@us.ibm.com).

## Objectives

During these lab exercises, you will configure IBM Watson Discovery Service to fit a sample discovery use case scenario. You will learn how to use the new IBM Watson Discovery service tooling to walk through the main configuration steps required to setup, build, and refine a Watson Discovery Service instance.

The tasks and tools that are covered within the Lab allow you to perform and complete all necessary steps to efficiently and effectively ingest Amazon Product Reviews data, configure your Watson Discovery Service Collection to convert, enrich, and normalize the data, and Query the collection to discover insights.

## 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 Discovery Service instance in Cloud.

## Exercise Environment

The lab exercise will be performed on the IBM Watson Discovery Service provided as part of the Watson Developer Cloud.

## Course Focus

The lab consists of 5 main sections.

1. Getting started with Watson Discovery Service and creating your own instance and environment.
2. How to configure you through ingesting content within IBM Watson Discovery Service.
3. Ingesting data using the Watson Discovery Service Data Crawler.
4. Presenting an alternative to data ingestion: using the Watson Discovery Service to ingest your data.
5. Extracting 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. Make sure to unzip the Lab Folder to a directory on your local hard drive.

The lab folder contains the following files:

File name	Description
WDS Lab Exercise v1.0	Lab manual
Amazon_Product_Reviews Folder	Folder containing 999 sample Amazon Product Reviews

## 0 Getting Started

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

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

### 0.1 Overview

This lab guides you through the configuration steps required to use Watson Discovery Service to satisfy a Voice of the Customer user scenario. Watson Discovery Service will be configured to help end-users visualize and analyze product feedback obtained via customer product reviews. The data set is a collection of reviews for Amazon Kitchen and Dining products. The application end-user is Henry, an Amazon Business Analyst, responsible for reading and analyzing consumer reviews to:

- Identify all relevant feedback on product functionality;
- Address product pain points and upsell product success point; and
- Make informed suggestions and provide actionable insight to the product development team.

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

## 0.2 Working with Watson Developer Cloud

### 0.2.1 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. You can view a short introductory video here: <http://www.ibm.com/developerworks/cloud/library/cl-bluemix-dbarnes-ny/index.html>

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.

### 0.2.2 Watson Developer Cloud (WDC)

IBM Watson Developer Cloud (**WDC**) is a platform of cognitive services, designed to help developers build solutions to help users extract insight from Big Data. Cognitive computing systems learn and interact naturally with humans to augment their ability to make better decisions from data.

As such, the Watson Developer Cloud services offer a variety of services that cover various aspects of natural interaction including text (**Natural Language Understanding**, **Natural Language Classifier**, **Conversation**), images (**Visual Recognition**), and speech (**Speech To Text** and **Text To Speech**).

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

### 0.2.3 Watson Discovery Service (WDS)

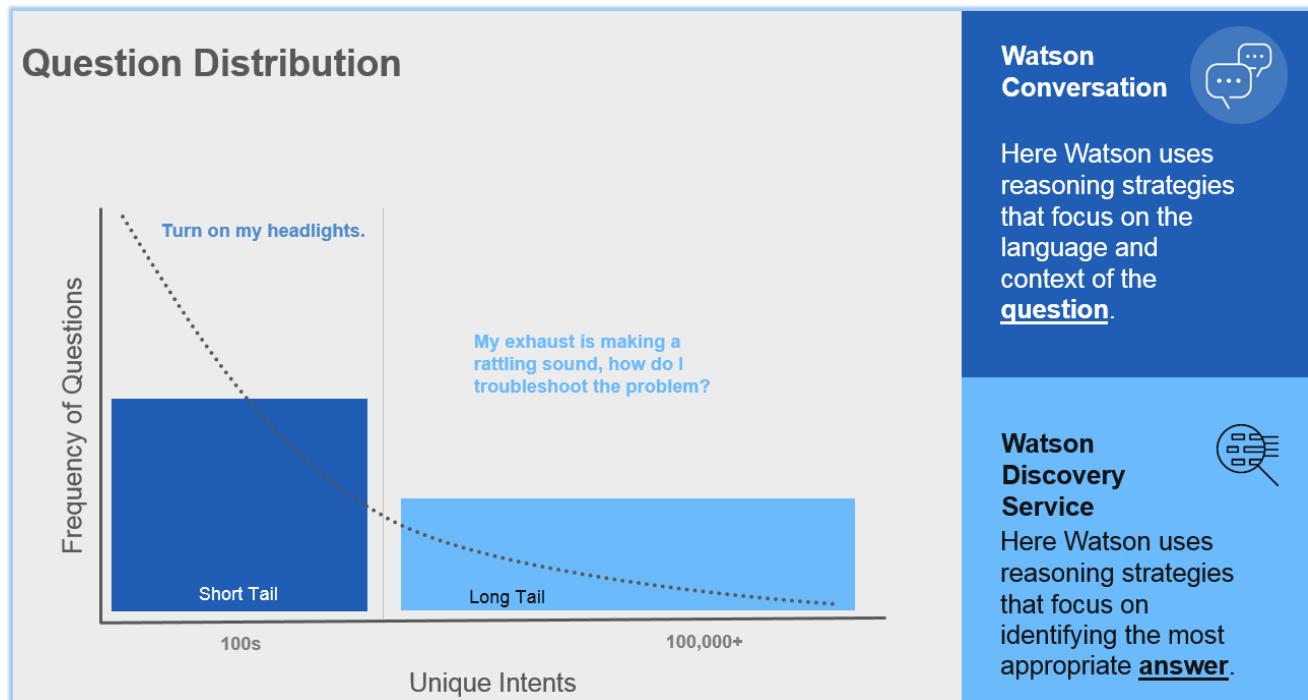
This guide takes an instructional approach to working with the IBM Watson™ Discovery Service, a cognitive search and content analytics engine that allows applications to identify patterns, trends and actionable insights that drive better decision-making.

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

- Creating your Watson Discovery Service environment;
- Configure the WDS 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 Service 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 HTML, DOC, PDF, 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 Conversation Service 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.



## 0.3 Prerequisites

This section provides instructions to help you get started quickly with the IBM Watson™ Developer Cloud services using Node.js as your programming runtime environment. To make it easy to get up and running with a functional application that uses the REST Application Programming Interface (API) for any Watson service, IBM provides a Node.js package with wrappers that simplify application development. The package includes simple command-line example applications to let you experiment with any of the available services.

1. To get set up, complete the following steps:

### 0.3.1 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 analytic services.

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

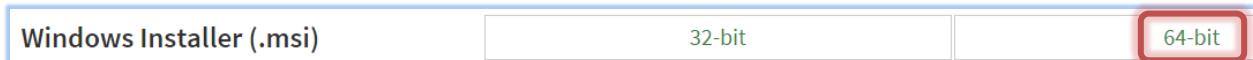
2. Direct your browser to the Cloud home page: <https://console.ng.bluemix.net/home/> to access your dashboard
3. If you do not yet have a Cloud account, click **Sign Up** on the top right
4. Enter requested information and click **Create Account**

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.

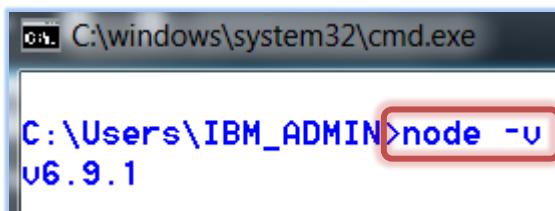
### 0.3.2 Installing the node.js runtime

The default installation includes both the runtime and package manager. Make sure to include the installed binaries on your PATH environment variable after installation (typically, the default installation locations that the installer selects does the inclusion).

5. Direct your browser to the nodejs.org web site: <https://nodejs.org/en/>
6. Click **DOWNLOADS** in the menu bar
7. Select and install the installer (not the binary), appropriate for your operating system (i.e. Windows Installer.msi (64 bit))



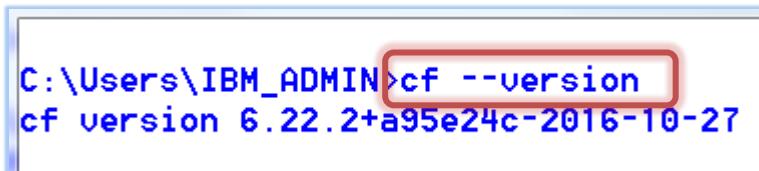
8. Complete the installation
9. Verify that node.js has been installed by opening a command line and issuing `node -v`



```
C:\Windows\system32\cmd.exe
C:\Users\IBM_ADMIN>node -v
v6.9.1
```

### 0.3.3 Installing cf (Cloud Foundry) CLI

10. Direct your browser to a GitHub repository: <https://github.com/cloudfoundry/cli/releases>
11. Download and install the most recent installer appropriate for your operating system
12. You may need to open Preferences > Security and Privacy > General tab (in Mac)
13. Unlock and change the Allow applications downloaded from Anywhere
14. Verify that installation was successful by issuing `cf --version` from a command line



```
C:\Windows\system32\cmd.exe
C:\Users\IBM_ADMIN>cf --version
cf version 6.22.2+a95e24c-2016-10-27
```

### 0.3.4 Verifying Prerequisites

Before you start the following section, please run the following commands to verify you have the required packages installed on your machine.

1. Open a Terminal window
2. Check that git is installed by typing `git --version`

If this returns an error, then git is not installed. Please download and install git at <https://git-scm.com/download/win>

3. Check that node.js is installed by typing `node -v`

If this returns an error, then node is not installed. Please download and install Node from [nodejs.org](http://nodejs.org)

4. npm should be installed with nodejs, verify it is installed on your machine by typing `npm -v`
5. Check cf (cloud foundry command line interface) is installed by typing `cf -v`

If this returns an error, then cf is not installed.

### 0.3.5 Downloading OS-specific coding-friendly editing tool

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

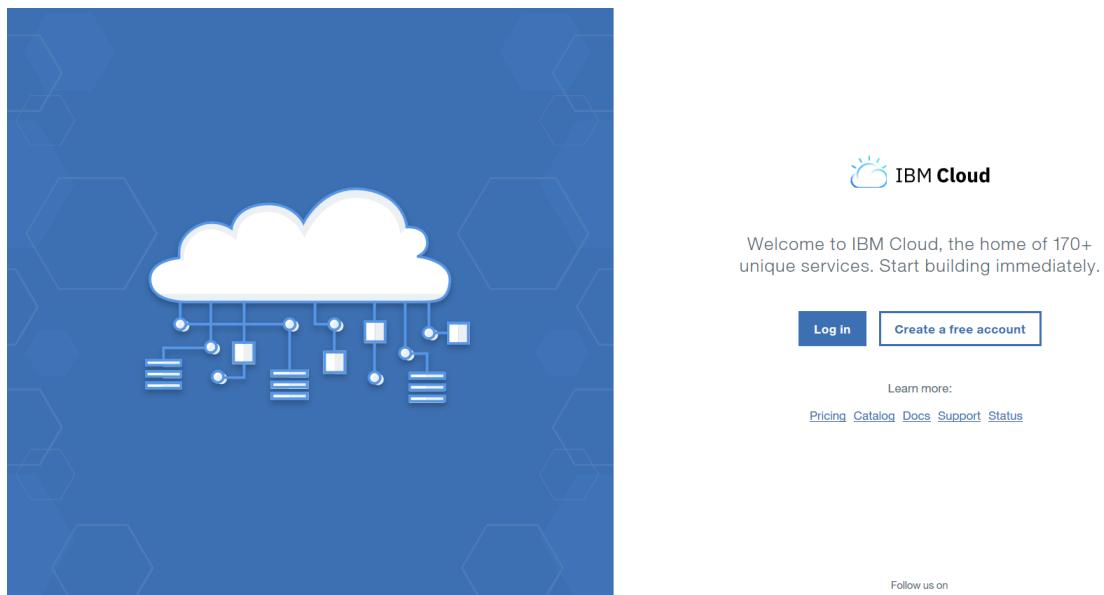
### 0.3.6 Installing the Watson Discovery Service Data Crawler

Watson Discovery Service provides a Data Crawler tool that will help users take their documents from the repositories where they reside and push them to the cloud, to be used by the Discovery Service. The Data Crawler is recommended when users are managing an upload of a significant number of files from a remote system, or trying to extract content from a supported repository (such as a DB2 database).

8. Installation instructions for the Data Crawler tool can be found [here](#).
9. Please review the Data Crawler supported platforms available [here](#).
10. Supported repositories can be found [here](#).

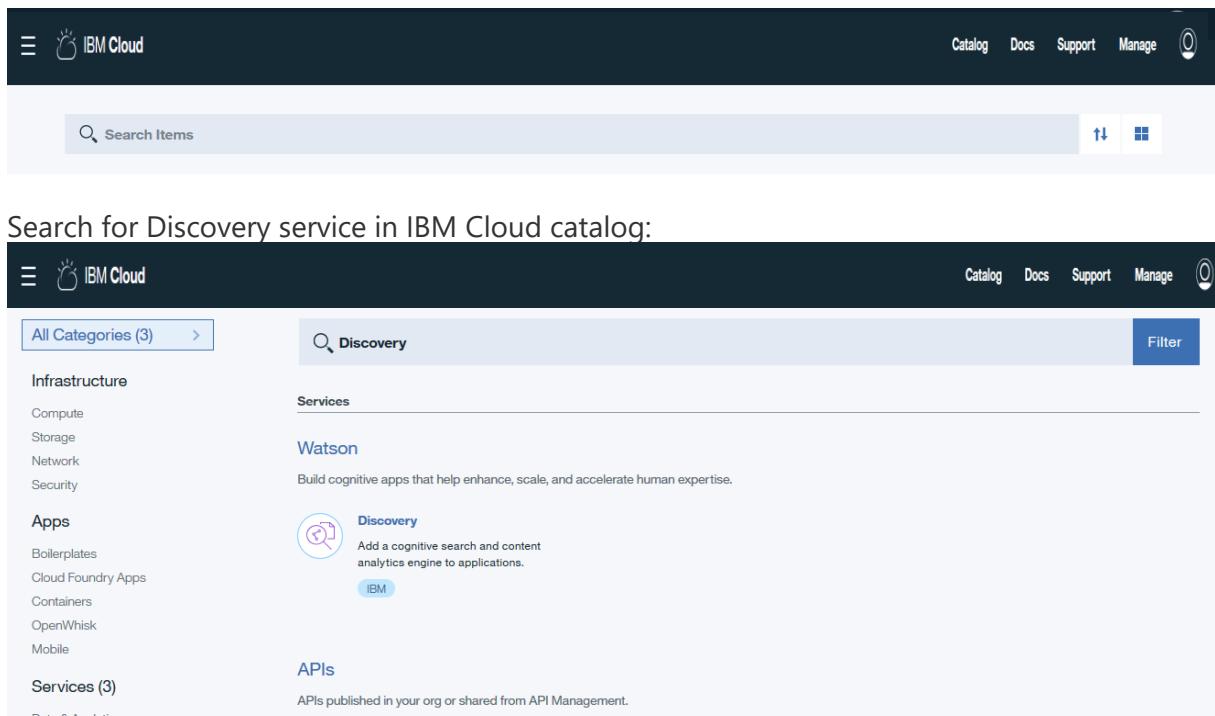
## 1 Creating Your Watson Discovery Instance

1. Navigate to IBM Cloud - <https://console.ng.bluemix.net/> and click on Log In



The screenshot shows the IBM Cloud homepage. It features a large central graphic of a white cloud with blue lines connecting it to various small icons below. The background is a dark blue with a hexagonal pattern. At the top right is the "IBM Cloud" logo. Below the logo is a welcome message: "Welcome to IBM Cloud, the home of 170+ unique services. Start building immediately." Underneath this are two buttons: "Log in" and "Create a free account". Below these buttons is a "Learn more:" section with links to "Pricing", "Catalog", "Docs", "Support", and "Status". At the bottom right, there is a "Follow us on" link with icons for social media platforms.

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 left-hand corner. You can alternatively follow this link - <https://console.ng.bluemix.net/catalog/>



The screenshot shows the IBM Cloud Catalog search results for "Discovery". The search bar at the top contains the word "Discovery". On the left, there is a sidebar with categories: "All Categories (3)" (selected), "Infrastructure" (Compute, Storage, Network, Security), "Apps" (Boilerplates, Cloud Foundry Apps, Containers, OpenWhisk, Mobile), and "Services (3)" (Data & Analytics). The main search results area shows a "Discovery" service card. The card has a circular icon with a magnifying glass and the word "Discovery". Below the icon, it says "Add a cognitive search and content analytics engine to applications." and "IBM". To the right of the card, there is a "Filter" button. At the bottom of the page, there is a "Catalog", "Docs", "Support", and "Manage" navigation bar.

5. Click on the Discovery tile

## Watson Discovery Service

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 Free Plan option (default setting) and by clicking Create.

**Service name:** Discovery-0v

**Credential name:** Credentials-1

**Pricing Plans**

PLAN	FEATURES	PRICING
✓ Free	30-day trial 1GB RAM, 2GB storage Unlimited enrichments 1000 news queries Custom domain model	Free
Standard	Environment Size 1: 2GB RAM, 48GB storage, 4,000 enrichments Environment Size 2: 8GB RAM, 192GB storage, 16,000 enrichments	\$960.00 USD/Env/Month \$3,480.00 USD/Env/Month

Monthly prices shown are for country or region: United States

Need Help? Contact Bluemix Sales      Estimate Monthly Cost Cost Calculator      Create

8. IBM Cloud will now instantiate your personal Watson Discovery Service instance. It will automatically guide you to the service dashboard.

**Manage**    **Service credentials**    **Connections**

**Discovery**

Add a cognitive search and content analytics engine to applications to identify patterns, trends and actionable insights that drive better decision-making

Launch tool ↗

**Developer resources:**

- Documentation
- Demo

### Discovery tooling

Visually configure your service. Upload, convert, normalize and enrich sample content and then use that

Launch tool ↗

If you close the tab and would want to re-access it, you can navigate to the service dashboard directly 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://console.ng.bluemix.net/dashboard/apps>

9. In the service dashboard click on the "Service Credentials" tab and go to service credentials section. Click on "View Credentials" dropdown to view Username, password and URL
10. Select and Copy the service credentials to a local file and save it for future reference

Service credentials		
	KEY NAME	DATE CREATED
<input type="checkbox"/>	Credentials-1	Dec 20, 2016 - 03:24:30

11. Once you've recorded Username and Password. Click on the Manage tab to return to the service dashboard.
12. Launch the discovery tool by clicking "Launch Tool" button from the discovery service instance

Discovery

Add a cognitive search and content analytics engine to applications to identify patterns, trends and actionable insights that drive better decision-making

Developer resources:

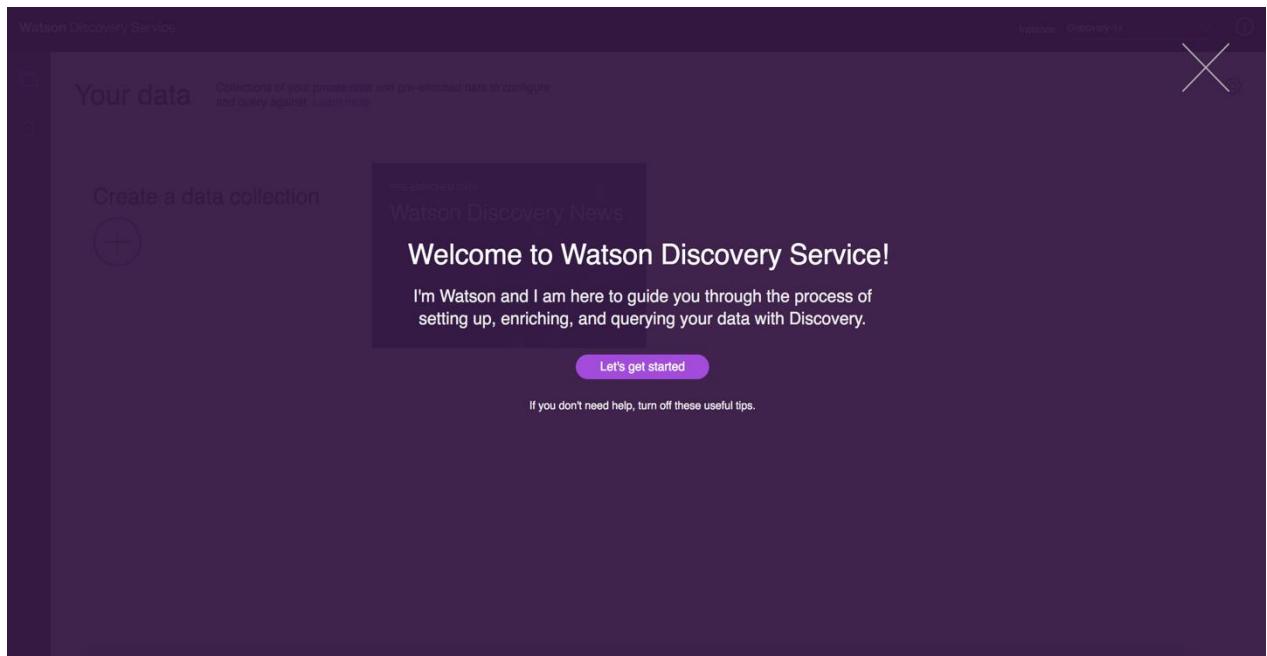
- Documentation
- Demo

Discovery tooling

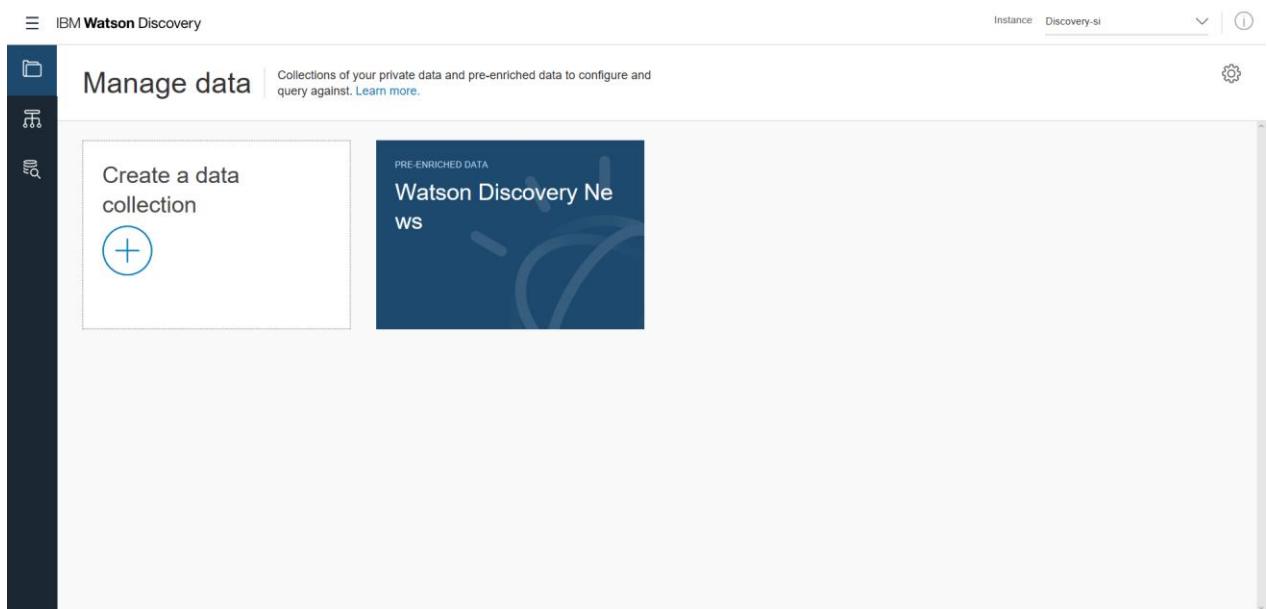
Intended Use

13. 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.

## Watson Discovery Service



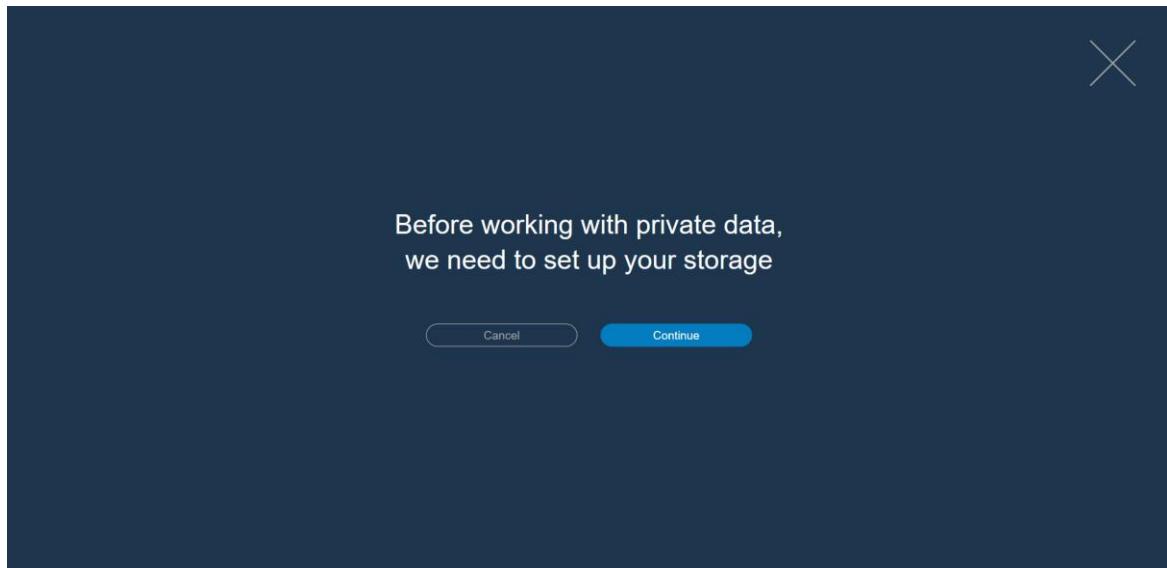
14. In the dashboard, you can see the collections present in your Watson Discovery Service environment. An option is provided to create a new collection.



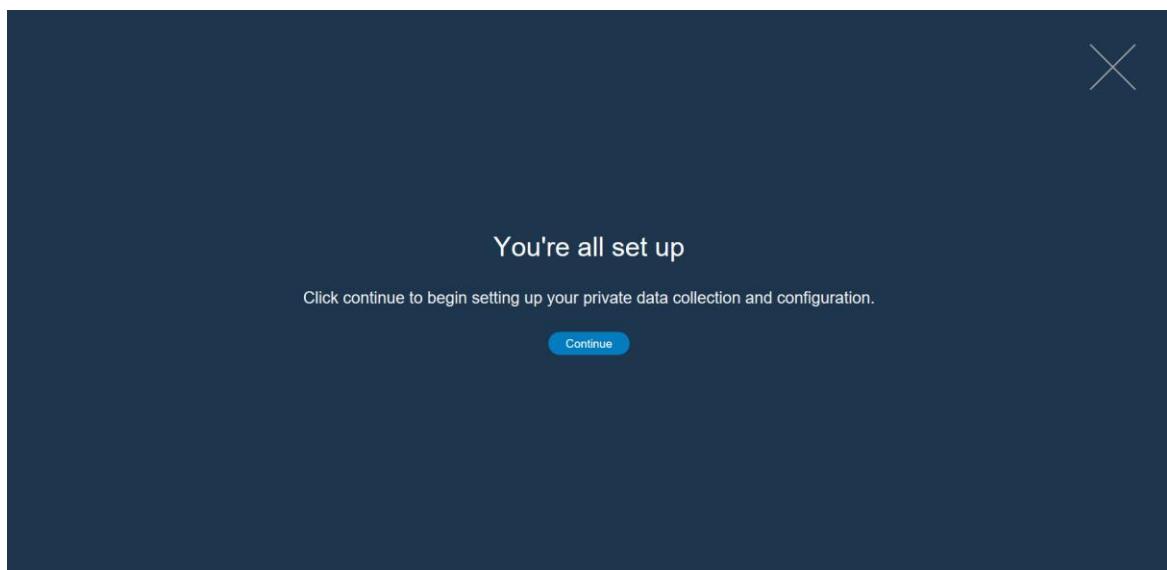
A new Watson Discovery Instance will come with a default environment, configuration and a pre-enriched collection, Watson News. To upload Amazon review data, you will have to create a new collection.

## 2 Configuring Your Watson Discovery Service Instance

1. Create "Amazon Product Reviews" collection by clicking on Create a data collection 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

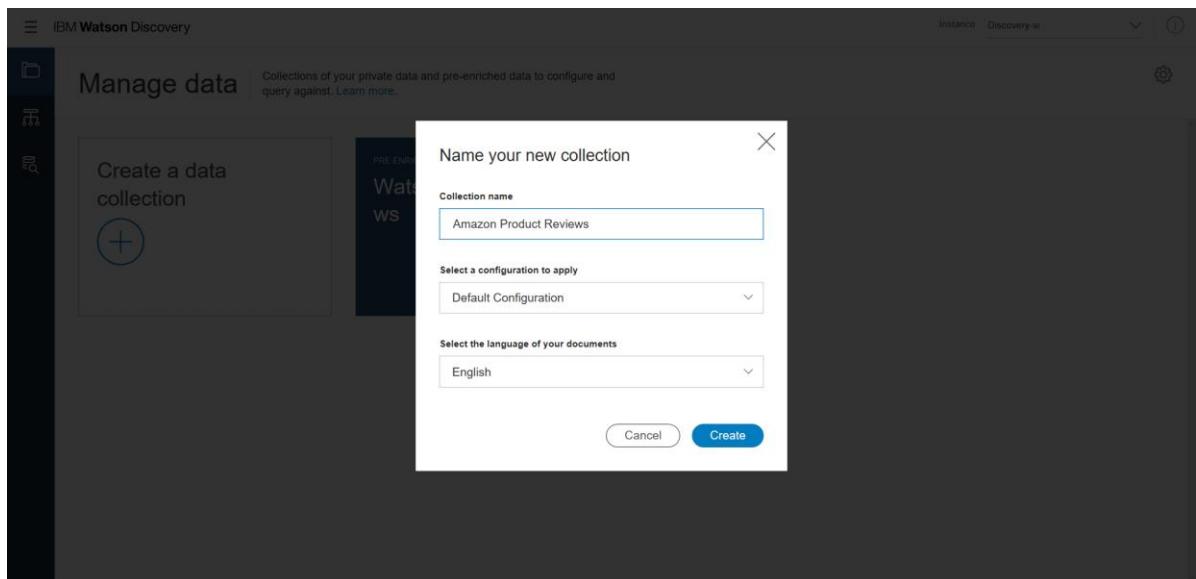


3. You get a message saying, "You're all set up" once the storage is provisioned and the collection is created. Click "Continue" button

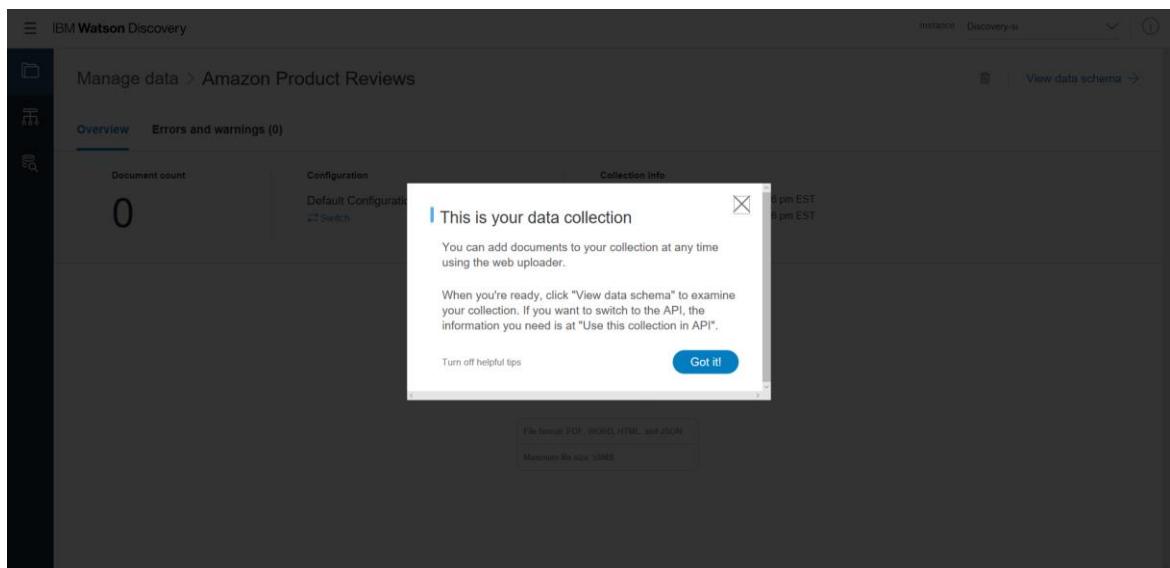


4. Again, click on Create a new collection and specify the name for your Collection. In this Lab, you will create a collection that you will want to name "Amazon Product Reviews".
5. Apply the Default Configuration to the collection (which is already selected for you) and click the Create button.

## Watson Discovery Service



6. If you are creating a collection for the first time, it automatically creates a hosting Environment which may take up to 2 minutes to provision the storage for environment
7. 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.



# Watson Discovery Service

The screenshot shows the 'Amazon Product Reviews' collection in the Watson Discovery Service. The 'Overview' tab is selected. Key details include:

- Document count:** 0
- Configuration:** Default Configuration (with a 'Switch' button)
- Collection info:**
  - Created on: 11/20/2017 12:12:06 pm EST
  - Last updated: 11/20/2017 12:12:06 pm EST
  - [Use this collection in API](#)
- A central area for document upload with a cloud icon and instructions: "Drag and drop your documents here or browse from computer".

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

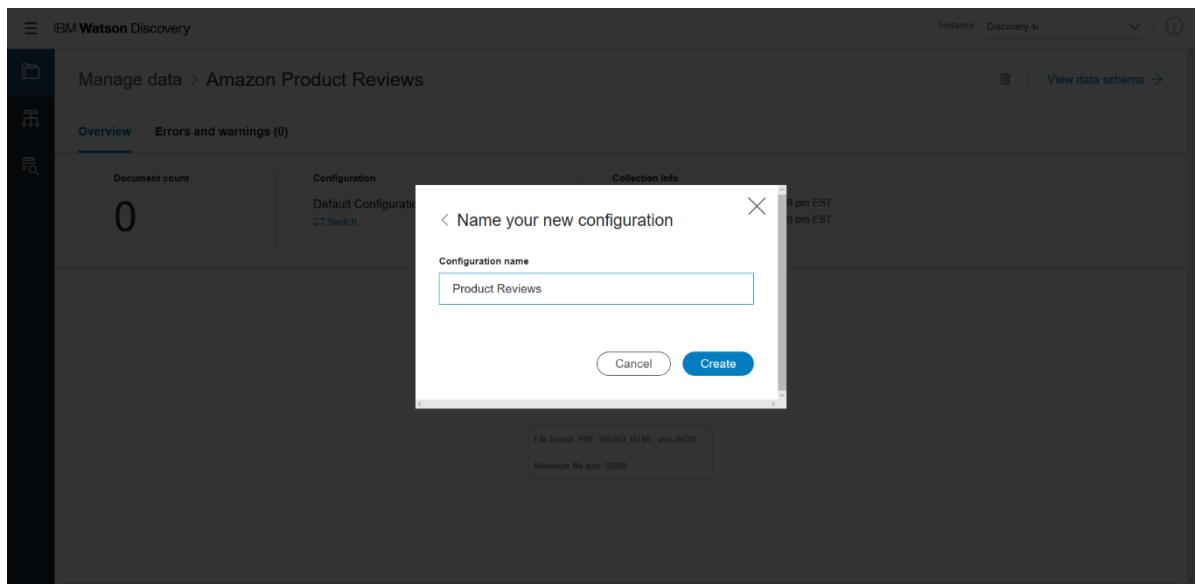
8. To create a new configuration, click on "Switch" button next to Default Configuration
9. Select Create a new configuration.

The screenshot shows the 'Switch configuration' dialog box. It contains the following elements:

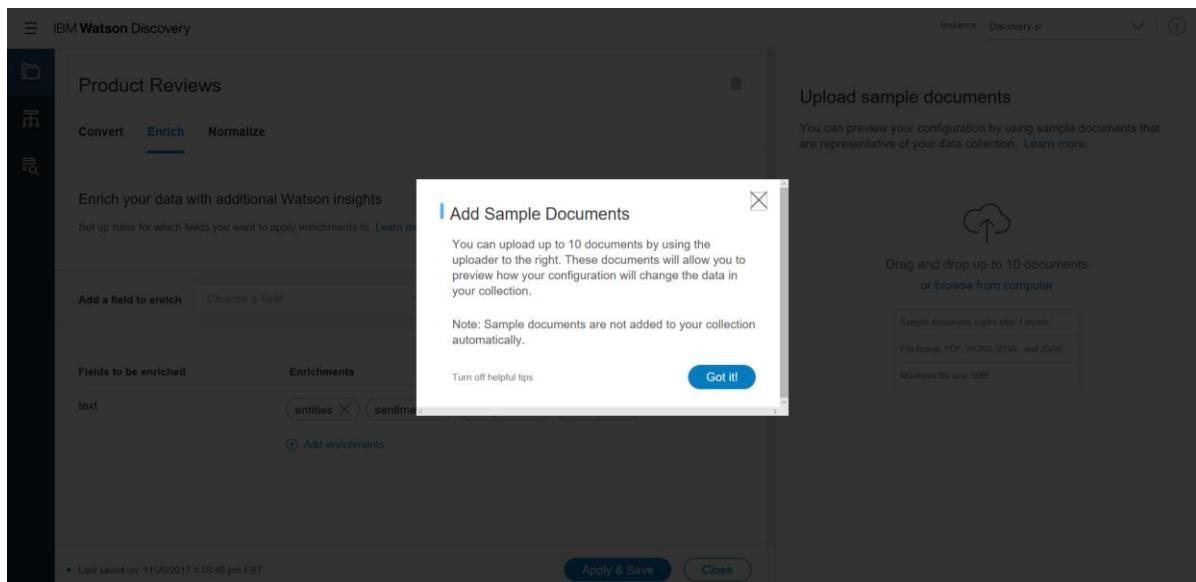
- Switch configuration to:** Default Configuration (dropdown menu)
- Create a new configuration:** Link
- Buttons:** Cancel, Switch

10. Name the configuration: "Product Reviews"

## Watson Discovery Service



Note: If there is already a preexisting configuration which you had created before, you can select it from the dropdown when you are switching the configuration



11. You are now ready to create a Configuration that is specific to Amazon Product Reviews. Before moving to the next Section, you will need to confirm that the Amazon data set is present within your Lab folder directory. Make sure you locate the Amazon Data files on your machine.

## 2.1 Selecting attributes for conversion, enrichment, and normalization in the configuration

User should identify the data field from the Amazon Product Review data which needs to be converted, enriched, and normalized by the Watson Discovery Service. NLU and NLP is embedded in WDS through the enrichments so you can receive seamless data insights with only a few clicks. The identified field is

then selected for enrichment during creating/modifying configuration. In general, the data will be analyzed to extract features out of it against which the analysis will be made. But, the Amazon review data is available in a structured format(JSON) and data is arranged in the form of attribute:value pair. Users will need to identify which field to be attributed. For this use case, we are enriching review\_text from the review documents. As mentioned earlier, it is essential to set the configuration and the attributes to be enriched before uploading all your data files.

### 2.1.1 Configuring Enrichment Attributes

1. Remove the existing field "text" from the fields to be enriched section since the review does not have any attribute called "text". You can remove the field by clicking the "minus" button next to the field which appears when you scroll over on the field – the delete button will be to the far right.

The screenshot shows the Watson Discovery Service interface for configuring enrichment. The left sidebar has icons for folder, content, and search. The main panel title is 'Product Reviews'. Below it are tabs: 'Convert', 'Enrich' (which is underlined), and 'Normalize'. A sub-section titled 'Enrich your data with additional Watson insights' includes a link to 'Learn more'. Below this is a section to 'Add a field to enrich' with a dropdown menu 'Choose a field'. The main configuration area shows a table with 'Fields to be enriched' (text) and 'Enrichments' (keywords, entities, sentiment, emotion, categories, relations, concepts, semantic\_roles). A 'Delete text field' button is at the top of this list. At the bottom of the configuration area is a note 'Last saved on: 11/20/2017 2:38:33 pm EST' and buttons 'Apply & Save' and 'Close'. To the right, there's a sidebar for 'Upload sample documents' with instructions, a cloud icon for file upload, and details about sample document expiration and file formats.

2. Download the amazon review files and store it in a new folder and name it as "contents". You can find the review files at <https://ibm.ent.box.com/folder/23913798838>
3. Identify a sample Amazon Product Review from the Content folder.
4. Upload Review1.json from the Content Folder by browsing to the Content Folder using the "browse from computer" button.
5. Select Review1.json and click Upload to complete the ingestion of the sample document.

# Watson Discovery Service

The screenshot shows the 'Enrich' tab selected in the top navigation bar. The main area displays 'Product Reviews' and an 'Enrich' configuration panel. The 'Fields to be enriched' section is currently empty, showing 'No fields added yet'. To the right, a separate panel titled 'Upload sample documents' provides instructions for previewing configurations using sample documents, including a file upload icon and details about document formats and expiration.

- Once it has been uploaded, the Review1.json is shown under Upload sample documents section which can be clicked to open in preview mode

The screenshot shows the 'Enrich' tab selected in the top navigation bar. The main area displays 'Product Reviews' and an 'Enrich' configuration panel. The 'Fields to be enriched' section is currently empty, showing 'No fields added yet'. To the right, a separate panel titled 'Upload sample documents' provides instructions for previewing configurations using sample documents, including a file upload icon and details about document formats and expiration. A file named 'review1.json' is listed in the document list.

- Click on Review1.json to open up a preview of the document. The Amazon Product Review data fields are displayed within the window.

# Watson Discovery Service

The screenshot shows the 'Enrich' tab selected in the top navigation bar. In the main area, there's a section titled 'Enrich your data with additional Watson insights' with a link to 'Learn more'. Below this is a dropdown menu labeled 'Add a field to enrich' with the placeholder 'Choose a field'. A preview pane on the right displays a JSON document:

```

{
  "rating": "5",
  "uuid": "2",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "A1JQTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206"
}

```

At the bottom, a status message says 'Last saved on: 11/20/2017 2:38:33 pm EST' and two buttons: 'Apply & Save' and 'Close'.

8. The automatically identified fields from the sample review document are available to be selected for enrichment from the dropdown menu next to Add a field to enrich.
9. Click on the drop-down menu and select “review\_text” for enrichment

The screenshot shows the 'Enrich' tab selected. In the 'Fields to be enriched' section, the 'review\_text' field is highlighted in blue in the dropdown menu. The preview pane on the right shows the same JSON document as before.

10. After selecting the field to be enriched, you need to select the type of enrichments you want applied to your data. Click on “Add Enrichments” next to the field to be enriched

# Watson Discovery Service

Fields to be enriched	Enrichments
review_text	No enrichments applied yet
	<a href="#">+ Add enrichments</a>

11. Select all enrichment types by clicking Add on all tiles. The “added” message will appear once the enrichment is selected. Once you’ve selected all enrichments, click Done.

The screenshot shows the 'Add Enrichments' dialog box. At the top, it lists the fields to be enriched: 'review\_text'. Below this, there are six enrichment tiles, each with a 'Learn more' link and an 'Added!' button. The tiles are: Keyword Extraction, Sentiment Analysis, Concept Tagging, Category Classification, Semantic Role Extraction, and Emotion Analysis. To the right of the dialog, a preview window shows a JSON document with enriched data, including sentiment analysis results like 'positive' and 'neutral'.

Note: If you have a WKS model, you can deploy it in enrichments by adding the model ID

The screenshot shows the 'Add Enrichments' dialog box. At the top, it lists the fields to be enriched: 'review\_text'. Below this, there are five enrichment tiles: Semantic Role Extraction, Emotion Analysis, Entity Extraction, Relation Extraction, and Custom Model ID. The 'Custom Model ID' tile has a tooltip: 'Custom Models are created in Watson Knowledge Studio and imported into Watson Discovery.' To the right, a preview window shows a JSON document with enriched data, including entity extraction results like 'entity': 'zoku quick pop' and 'model\_id': '0615391206'.

12. The selected type of enrichments is now shown in the Enrich page. Now click “Apply and Save” button to save the configuration. This configuration will reflect against any documents added to the

## Watson Discovery Service

collection from now on.

```

{
  "rating": "5",
  "uuid": "72",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "A1JVQTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206"
}
  
```

```

{
  "rating": "5",
  "uuid": "72",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "A1JVQTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206",
  "enriched_review_text": {
    "sentiment": [...],
    "semantic_roles": [...],
    "relations": [...],
    "keywords": [...],
    "entities": [...],
    "emotion": [...],
    "concepts": [...],
    "categories": [...]
  }
}
  
```

13. Click on the Close button to return to the Your Data dashboard.

14. Click on the Amazon Product Review blue collection tile to open the collection dashboard. From the collection info, click on 'Use this collection in API' to receive your Collection ID, Configuration ID, and your Environment ID. Make sure you record in a separate text file the Environment ID, Configuration ID, and Collection ID. This will be used later for data crawling.

The screenshot shows the Watson Discovery Service interface for the 'Amazon Product Reviews' collection. At the top, there are tabs for 'Instance' and 'Discovery-si'. Below the tabs, there's a button to 'Upload documents' and a link to 'View data schema'. The main area displays '0' errors and warnings. Under 'Configuration', it shows 'Product Reviews' with options to 'Edit' or 'Switch'. The 'Collection info' section provides details like 'Created on' (11/20/2017 12:12:06 pm EST) and 'Last updated' (11/20/2017 1:58:41 pm EST). A link to 'Use this collection in API' is also present. A tooltip for 'Collection Id' highlights the value '10b9d8b5-7ffe-453b-978d-0e40b9dad1c5'. The 'General sentiments' section describes identifying overall positive or negative sentiment. A tooltip for 'Environment Id' shows the value '7199833b-c6d3-46be-8d4f-2e279cc27b0f'. Other environment details like 'Id' and 'Name' are also visible.

15. Now that you have defined the Enrichment configuration, you are ready to ingest data within the collection. For more complex use cases, you will also have to set Conversion and Normalization configuration attributes. For this lab, we are working with clean Amazon Product Review data so you are not required to set Conversion and Normalization attributes. However, we do cover those tasks as Optional in Section 2.1.2 and Section 2.1.3. Feel free to explore these optional Tasks.
16. Alternatively, if you are a developer, please refer to section 3 to learn how to use data crawler to ingest files.
17. If you are a general user, please refer to section 4 to learn how to use tooling interface to upload files.
18. Users can either use crawler tool or tooling interface but not both.

### 2.1.2 Configuring Conversion Attributes (Optional for this lab)

You can perform the following conversion steps after completing Section 2.1.1. These steps are optional and meant to illustrate the Conversion capability that is provided with Watson Discovery Service. Keep in mind that users should configure the conversion attributes prior to selecting the enrichment attributes. The configuration pipeline can be visualized as going left to right: from Convert to Enrich and then to Normalize.

The screenshot shows the 'Convert' tab for the 'Product Reviews' collection. On the left, there's a sidebar with icons for folder, file, and search. The main area has tabs for 'Convert' (which is selected), 'Enrich', and 'Normalize'. Below the tabs are buttons for PDF, WORD, HTML, and JSON. A section titled 'Clean up your JSON' allows users to set rules for conversion, with a link to 'Learn more'. Another section below it lets users 'Move, merge, copy or remove fields' and includes an 'Add field' button. A toggle switch for 'Remove empty fields' is set to 'Off'. At the bottom, a message says 'Last saved on: 11/20/2017 2:38:33 pm EST' and there are 'Apply & Save' and 'Close' buttons. To the right, a preview window titled 'Preview with: review1.json' shows a JSON object with various fields including rating, uid, review\_date, helpful, reviewer\_id, not\_helpful, review\_text, summary, product\_name, and product\_id.

```

{
  "rating": "5",
  "uid": "2",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "A1JVGQTAGH0L7F",
  "not_helpful": "0",
  "review_text": "bought this zoku quick pop for my daughter who loves zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206"
}
  
```

The ingested content (PDF, Word, HTML, JSON) data can be converted to JSON to suit your requirements by removing, copying, moving attributes.

1. Click on the Edit button next to "Product Reviews" in the collection dashboard.
- ### Amazon Product Reviews

#### 1d warnings (0)

This screenshot shows the collection dashboard for 'Amazon Product Reviews'. On the left, under 'Errors and warnings', it says '0 documents failed' and has a 'View details' link. On the right, under 'Configuration', it shows the collection name 'Product Reviews' and two buttons: 'Edit' and 'Switch'.

2. In this lab scenario, we are already ingesting JSON data and there are no attributes that need to be modified within the Amazon Product Review data. We are therefore going to walk through some sample steps to illustrate how the functionality works. Start by selecting 'Convert' tab from "Product Reviews" configuration page

# Watson Discovery Service

```

{
  "rating": "5",
  "uuid": "2",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "A1VQTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206"
}

```

3. Click on “Add field” link under “Move, merge, copy or remove fields” section. You will get a dropdown menu with the following options: remove, copy, move or merge the attributes.
4. Select “remove” from the dropdown and select “review\_date” field from the dropdown next to it (for this example we are assuming that review\_date field is not needed to satisfy the use cases)
5. Click “Apply & Save” button. You could see in the right window that the attribute “review\_date” is removed from review1.json

```

{
  "rating": "5",
  "uuid": "2",
  "helpful": "0",
  "reviewer_id": "A1VQTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206"
}

```

6. To see an example of how the merge functionality works, click the “Add field” link (underneath Move, merge, copy or remove fields section) and select “merge” from the dropdown menu.

## Watson Discovery Service

The screenshot shows the 'Convert' tab of the Watson Discovery interface. On the left, there's a sidebar with icons for file, document, and search. The main area has tabs for 'Convert', 'Enrich', and 'Normalize'. Below these are buttons for 'PDF', 'WORD', 'HTML', and 'JSON'. A section titled 'Clean up your JSON' contains instructions to set up rules for conversion. Under 'Move, merge, copy or remove fields', there are dropdown menus for 'remove' (set to 'review\_date') and 'merge' (set to 'reviewer\_id' and 'to' 'uuid'). A 'Remove empty fields' toggle switch is turned off. At the bottom, a status bar says 'Last saved on: 11/20/2017 4:05:51 pm EST' and buttons for 'Apply & Save' and 'Close'.

Preview with review1.json

```
{
  "rating": "5",
  "uuid": "2",
  "helpful": "0",
  "reviewer_id": "A1JVQTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206"
}
```

- Now you will have the option to select two fields: each representing the fields that you want to merge with one another. Select “reviewer\_id” from the first dropdown menu and “uuid” from the second dropdown menu.

This screenshot is similar to the previous one but shows the result of the merge operation. The 'merge' dropdown now has 'reviewer\_id' selected and 'to' 'uuid'. The JSON preview window shows the 'reviewer\_id' field removed and its value ('A1JVQTAGHYOL7F') copied into the 'uuid' field.

Preview with review1.json

```
{
  "rating": "5",
  "uuid": "2",
  "helpful": "0",
  "reviewer_id": "A1JVQTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206"
}
```

- Click “Apply & Save” button. In the right window, you will see that the “reviewer\_id” field is removed and the data field is merged into “uuid” field.
- You can also copy one data field to another field. To do so click the “Add field” button and select “copy” from the dropdown menu.

# Watson Discovery Service

IBM Watson Discovery

Product Reviews

**Convert** Enrich Normalize

PDF WORD HTML JSON

Clean up your JSON

Set up rules for how you want your documents to be converted. [Learn more](#).

Move, merge, copy or remove fields

remove review\_date

merge reviewer\_id to uuid

copy rating to helpful

Add field

Remove empty fields

Last saved on: 11/20/2017 4:10:06 pm EST

Apply & Save Close

10. Select “rating” field from the first dropdown menu and the “helpful” field from the second dropdown menu. Click “Apply & Save” to see that the “rating” field is copied into the “helpful” field

IBM Watson Discovery

Product Reviews

**Convert** Enrich Normalize

PDF WORD HTML JSON

Clean up your JSON

Set up rules for how you want your documents to be converted. [Learn more](#).

Move, merge, copy or remove fields

remove review\_date

merge reviewer\_id to uuid

copy rating to helpful

Add field

Remove empty fields

Off

Last saved on: 11/20/2017 4:39:04 pm EST

Apply & Save Close

11. And finally, you can try moving one data field to another field. Click the “Add field” button and select “move” from the dropdown menu.

## Watson Discovery Service

The screenshot shows the 'Convert' tab in the IBM Watson Discovery interface. On the left, there's a sidebar with icons for PDF, WORD, HTML, and JSON. The main area has tabs for 'Move, merge, copy or remove fields' and 'Remove empty fields'. Under 'Move, merge, copy or remove fields', there are four rules:

- remove review\_date
- merge reviewer\_id to uuid
- copy rating to helpful
- move product\_id to product\_name

On the right, a preview window titled 'Preview with review1.json' shows the JSON document. The 'product\_id' field has been replaced by the 'product\_name' field.

```

{
  "rating": "5",
  "uuid": [
    "2",
    "A1JVQTAGHYOL7F"
  ],
  "helpful": "5",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206"
}
  
```

12. Select the “product\_id” field from the first dropdown menu and the “product\_name” field from the second dropdown menu. Click “Apply & Save” to see that the contents of “product\_id” field replaces the content within the “product\_name” field and the “product\_id” field content is removed from review1.json.

This screenshot shows the same 'Convert' tab interface after the changes made in step 12. The 'move' rule now maps 'product\_id' to 'product\_name'.

On the right, the preview window shows the updated JSON object. The 'product\_id' field has been replaced by the 'product\_name' field.

```

{
  "rating": "5",
  "uuid": [
    "2",
    "A1JVQTAGHYOL7F"
  ],
  "helpful": "5",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book"
}
  
```

13. Any changes made in “Convert” tab attributes are reflected on all the data files ingested.

Now that you have learned how to configure conversion attributes, you can remove all the changes. Remember, the Amazon Product Review data does not need any conversion attributes for this Lab.

14. You should revert all the changes you made in Convert section. To delete the Convert configuration, click on “minus” button on far right of each field added for conversion. The minus sign is displayed when you scroll over the field with your mouse.

## Watson Discovery Service



15. After removing all the conversion attributes, click the "Apply & Save" button to return to the original review1.json. Make sure your preview file looks like what is shown in the following snapshot.

The screenshot shows the Watson Discovery Service interface with the 'Convert' tab selected. On the right, there is a JSON preview window titled 'Preview with review1.json' containing the following JSON data:

```
{
  "rating": "5",
  "uuid": "2",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "A1JVQTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206"
}
```

At the bottom, there are 'Apply & Save' and 'Close' buttons.

16. Additional item for consideration. By switching the "Remove empty fields" switch to "on", any empty fields from the JSON data will be removed during conversion of the ingested documents.

The screenshot shows the Watson Discovery Service interface with the 'Convert' tab selected. The 'Remove empty fields' switch is turned 'On'. On the right, there is a JSON preview window titled 'Preview with review1.json' containing the same JSON data as the previous screenshot.

At the bottom, there is a note 'Last saved on: 11/20/2017 4:44:05 pm EST' and 'Apply & Save' and 'Close' buttons.

More details about converting attributes can be found by clicking "Learn more" link under "Clean up your JSON" header

### 2.1.3 Configuring Normalization Attributes (Optional for this lab)

Attributes normalization follows a similar process to that of conversion. However, conversion attributes are only applied to the original data field in the JSON file. Normalization is performed for both the original and enriched data field. It is useful when you want to remove certain enrichment attributes, copy them to another field etc..

1. Start by selecting ‘Normalize’ tab from “Product Reviews” configuration page

The screenshot shows the 'Product Reviews' configuration page in the IBM Watson Discovery interface. The 'Normalize' tab is currently selected. A modal window titled 'Normalize your data' is displayed, containing text about enriching data and a 'Got it!' button. The main panel shows a preview of a JSON document with fields like rating, review\_text, and product\_name.

```

{
  "rating": "5",
  "uuid": "2",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "ALV0TAGHOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206"
}

```

2. Click on “Add field” link under “Move, merge, copy or remove fields” section. You will get a dropdown menu with the following options: remove, copy, move or merge the attributes.

The screenshot shows the 'Product Reviews' configuration page in the IBM Watson Discovery interface. The 'Normalize' tab is selected. A modal window titled 'Normalize your data' is displayed, containing text and a 'Got it!' button. The main panel shows a preview of a JSON document with fields like rating, review\_text, and product\_name.

```

{
  "rating": "5",
  "uuid": "2",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "ALV0TAGHOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206"
}

```

3. Select “remove” from the dropdown and select “enriched\_review\_text.emotion.document.joy” field from the dropdown next to it (for this example we are assuming that the joy field is not necessary for the emotional analysis)

# Watson Discovery Service

Last chance! Clean up your data before it gets indexed  
Set up rules to clean up your data. [Learn more.](#)

**Move, merge, copy or remove fields**

remove

Add field

Remove empty fields

Last saved on: 11/27/2017 11:27:27 am EST

[Apply & Save](#) [Close](#)

Preview with review1.json

```

+ {
  "rating": "5",
  "uuid": "2",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "A1JVGTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206",
  "enriched_review_text": {
    "sentiment": {...},
    "semantic_roles": [...],
    "relations": [...],
    "keywords": [...],
    "entities": [...],
    "emotion": {
      "document": {
        "emotion": {
          "sadness": 0.103632,
          "joy": 0.846737,
          "fear": 0.012378,
          "disgust": 0.016771,
          "anger": 0.014654
        }
      },
      "concepts": [...],
      "categories": [...]
    }
  }
}

```

- Click "Apply & save" button. In the right window, you can see that the "joy" field under enriched\_review\_text.emotion.document.emotion is removed

Last chance! Clean up your data before it gets indexed  
Set up rules to clean up your data. [Learn more.](#)

**Move, merge, copy or remove fields**

remove

Add field

Remove empty fields

Last saved on: 11/27/2017 11:53:39 am EST

[Apply & Save](#) [Close](#)

Preview with review1.json

```

+ {
  "rating": "5",
  "uuid": "2",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "A1JVGTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206",
  "enriched_review_text": {
    "sentiment": {...},
    "semantic_roles": [...],
    "relations": [...],
    "keywords": [...],
    "entities": [...],
    "emotion": {
      "document": {
        "emotion": {
          "sadness": 0.103632,
          "fear": 0.012378,
          "disgust": 0.016771,
          "anger": 0.014654
        }
      },
      "concepts": [...],
      "categories": [...]
    }
  }
}

```

- To see an example of how the merge functionality works, click the "Add field" link (underneath Move, merge, copy or remove fields section) and select "merge" from the dropdown menu.

# Watson Discovery Service

IBM Watson Discovery      Instance: Discovery-si      [Edit](#)

## Product Reviews

**Convert**   **Enrich**   **Normalize**

Last chance! Clean up your data before it gets indexed  
Set up rules to clean up your data. [Learn more.](#)

Move, merge, copy or remove fields

remove       merge  to

+ Add field

Remove empty fields  Off

Last saved on: 11/27/2017 11:53:39 am EST      [Apply & Save](#)      [Close](#)

Preview with review1.json

```
{
  "rating": "5",
  "uuid": "2",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "A1JVTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206",
  "enriched_review_text": {
    "sentiment": {...},
    "semantic_roles": [...],
    "relations": [...],
    "keywords": [...],
    "entities": [],
    "emotion": {
      "document": {
        "emotion": {
          "sadness": 0.103632,
          "fear": 0.012378,
          "disgust": 0.016771,
          "anger": 0.014654
        }
      },
      "concepts": [...],
      "categories": [...]
    }
  }
}
```

- Now you will have the option to select two fields: each representing the fields that you want to merge with one another. Select “enriched\_review\_text.language” from the first dropdown menu and “enriched\_review\_text.status” from the second dropdown menu.

IBM Watson Discovery      Instance: Discovery-si      [Edit](#)

## Product Reviews

**Convert**   **Enrich**   **Normalize**

Last chance! Clean up your data before it gets indexed  
Set up rules to clean up your data. [Learn more.](#)

Move, merge, copy or remove fields

remove       merge  to

+ Add field

Remove empty fields  Off

Last saved on: 11/27/2017 11:53:39 am EST      [Apply & Save](#)      [Close](#)

Preview with review1.json

```
{
  "rating": "5",
  "uuid": "2",
  "review_date": "2014-06-17",
  "helpful": "0",
  "reviewer_id": "A1JVTAGHYOL7F",
  "not_helpful": "0",
  "review_text": "I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream.",
  "summary": "zoku",
  "product_name": "Zoku Quick Pops Recipe Book",
  "product_id": "0615391206",
  "enriched_review_text": {
    "sentiment": {...},
    "semantic_roles": [...],
    "relations": [...],
    "keywords": [...],
    "entities": [],
    "emotion": {...},
    "concepts": [
      {
        "text": "Maker",
        "relevance": 0.01136,
        "dbpedia_resource": "http://dbpedia.org/resource/Maker"
      }
    ],
    "categories": [...]
  }
}
```

- Click “Apply & Save” button. In the right window, you will see that the “enriched\_review\_text.concepts” field is removed and the data field is merged into “enriched\_review\_text.keywords” field.

# Watson Discovery Service

IBM Watson Discovery

Instance Discovery-si

The screenshot shows the 'Normalize' tab selected in the top navigation bar. On the left, there's a sidebar with icons for folder navigation and search. The main area is titled 'Product Reviews' and contains sections for 'Move, merge, copy or remove fields' and 'Remove empty fields'. In the 'Move, merge, copy or remove fields' section, there are dropdown menus for 'remove', 'merge', and 'copy'. The 'copy' menu has 'enriched\_review\_text.keywords' selected as the source field and 'enriched\_review\_text.categories' selected as the target field. To the right, a preview window shows a JSON document with several objects, each containing a 'text' field and a 'sentiment' field with a score and label. One object also includes a 'keywords' field with a URL to a DBpedia resource.

8. You can also copy one data field to another field. To do so click the "Add field" button and select "copy" from the dropdown menu. Select "enriched\_review\_text.keywords" field from the first dropdown menu and the "enriched\_review\_text.categories" field from the second dropdown menu

IBM Watson Discovery

Instance Discovery-si

This screenshot is similar to the previous one but shows the 'copy' operation completed. The 'copy' dropdown menu now has 'enriched\_review\_text.keywords' selected as the source field and 'enriched\_review\_text.categories' selected as the target field. The preview window shows the JSON document with the 'enriched\_review\_text.keywords' field copied into the 'enriched\_review\_text.categories' field.

9. Click "Apply & Save" to see that the "enriched\_review\_text.keywords" field is copied into the "enriched\_review\_text.categories" field

# Watson Discovery Service

The screenshot shows the 'Normalize' tab selected in the top navigation bar. On the left, there's a sidebar with icons for file management and search. The main area has sections for 'Move, merge, copy or remove fields' and 'Remove empty fields'. A preview window on the right shows the JSON structure of 'review1.json' with multiple documents, each containing fields like 'text', 'sentiment', 'score', 'label', and 'relevance'. Below the preview are 'Apply & Save' and 'Close' buttons.

10. You can try moving one data field to another field. Click the "Add field" button and select "move" from the dropdown menu. Select the "enriched\_review\_text.concepts.firebaseio" field from the first dropdown menu and the "enriched\_review\_text.concepts.opencyc" field from the second dropdown menu.

This screenshot shows the same interface after a move operation. The 'Move, merge, copy or remove fields' section now includes a 'move' option. In the preview window, a single document is shown with its fields rearranged. The 'enriched\_review\_text.semantic\_roles' field has been moved to replace the 'enriched\_review\_text.relations' field. The 'enriched\_review\_text.semantic\_roles' field is now present in the JSON structure where the relations field used to be.

11. Click "Apply & Save" to see that the contents of "enriched\_review\_text.semantic\_roles" field replaces the content within the "enriched\_review\_text.relations" field and the "enriched\_review\_text.semantic\_roles" field content is removed from review1.json

The screenshot shows the IBM Watson Discovery interface. On the left, there's a sidebar with icons for folder, file, and search. The main area has a header "Last chance! Clean up your data before it gets indexed" and a sub-header "Set up rules to clean up your data. [Learn more.](#)". Below this is a section titled "Move, merge, copy or remove fields" with four rows of operations:

- remove
- merge  to
- copy  to
- move  to

Below these is a button "+ Add field". Under "Remove empty fields", there's a toggle switch set to "Off". At the bottom are buttons "Apply & Save" and "Close". A timestamp "Last saved on: 11/27/2017 2:19:39 pm EST" is also present. On the right, a preview window titled "Preview with review1.json" shows JSON code representing two sentences and their analysis.

12. You can remove the normalization attributes using the “button” on to the far sight of the fields
13. Remove the normalization attributes since it is not required for Amazon use case lab
14. You can click on Close to go back to Your data tooling homepage.
15. Once the configuration setup is completed, users are now ready to upload the Amazon Product Review data files.
16. If you are a developer, please refer to section 3 to learn how to use data crawler to upload files.
17. If you are a general user, please refer to section 4 to learn how to use tooling interface to upload files.
18. Users can either use crawler tool or tooling interface but not both.

## 3 Data Ingestion using data crawler

Data crawler can either upload data files located in the local file system to the discovery collection or crawl data from the data base and upload them as JSON to the collection. For this lab, we will walk through the steps of crawling data from your local file system.

### 3.1 Data Crawler Overview

The data crawler is useful when you have a directory or database with a large number of documents to ingest. Data crawler can be executed from command line and be configured with ease. Data crawler defaults to run 14 concurrent threads to speed up upload time. For this lab, we assume users have followed the prerequisites documents and have the data crawler already installed. If that is not the case, check the IBM Cloud documentation for instructions to install the data crawler. The documentation can be found at

<https://www.ibm.com/watson/developercloud/doc/discovery/data-crawler-install.html>

### 3.2 Data Crawler Configuration

1. Locate the 'crawler-x.x.x/share/examples/config' folder in your machine
2. Copy the 'crawler-x.x.x/share/examples/config' folder into the same directory as the Amazon Product Review content folder.
3. Once copied to the new location, open the copied folder.
4. Open the config/discovery/discovery\_service.conf file. Modify the discovery\_service.conf file by copying and pasting your Discovery Service credentials (password and username) to the #Your Discovery Service credentials section of the file.
5. As a reminder, your Discovery Service credentials can be found in the Service credentials tab in the same window that you used to open the WDS Tooling.

```
# Your Discovery Service credentials
credentials {
    username = "1d48f5a5-1b42-431a-bdc2-3e7fe1665070",
    password = "fDpVhbAcQtDy"
},
```

6. You will also need the Environment ID, Collection ID and Configuration ID for the "Amazon Product Reviews" collection you saved in Section 2. If you missed that step, the IDs can always be retrieved from the collection dashboard.
7. Copy and paste the Environment ID, Collection ID, and Configuration ID within their respective fields identified with "#" (e.g. #The Collection in which documents should be indexed. Collection id = ...)

```

# The environment to use.
environment_id = "c90e85d6-691a-4b9c-b1c0-aba9a75c3dd6",

# The collection in which documents should be indexed.
collection_id = "9ec6362e-ae84-4d51-96c1-27a62697c2b2",

# The conversion configuration to use.
configuration_id = "91e7b8a8-3ed2-4cbe-9ca3-3b76e6bd5675",

# Or you can provide a file with your config
#configuration = "file://PATH_TO_CONFIG"
# Check if the document was successfully processed at the endpoint.
# This will reduce the perceived performance of the crawler but will
# produce reliable notification of a successful document upload and conversion.
# When enabling this, it is wise to increase concurrent_upload_connection_limit
# to your WDS ingestion conversion concurrency limit in order to fully utilize
# the resources available to you.
check_for_completion = false

# Your Discovery Service credentials
credentials {
    username = "1d48f5a5-1b42-431a-bdc2-3e7fe1665070",
    password = "fDpVhbAcQtDy"
},

```

8. Save and exit out of discovery\_service.conf file.
9. Open the crawler.conf file located in the top directory of the configuration folder. Under the output\_adapter section, make sure that class is set to **"com.ibm.watson.crawler.discoveryserviceoutputadapter.DiscoveryServiceOutputAdapter"** and the config is set to "discovery\_service"

```

output_adapter {
    # Select the Output Adapter to use by specifying its class. See crawler.conf(5) for more information or read below.
    class = "com.ibm.watson.crawler.discoveryserviceoutputadapter.DiscoveryServiceOutputAdapter",

    # Select the configuration key to be passed to the Output Adapter. The string here must correspond to a key within this configuration object.
    config = "discovery_service",

```

10. Save and close the crawler.conf file.
11. Navigate to the seeds directory folder within your configuration folder.
12. Open config/seeds/filesystem-seed.conf
13. Modify the value attribute directly under: name="url" attribute to the file path of the Amazon Product Review folder that you want to crawl. This url will be specific to the location in which you stored the content. Example: value="sdk-fs:///Users/chethan/Downloads/WDS/Crawler/data/" Note that the third slash is needed as well as a slash at the end of the directory url (e.g. data/).

```

# The third / is necessary! Do not prepend file:// to the listed paths.
{
    name ="url",
    value="sdk-fs:///Users/chethan/Downloads/WDS/Crawler/data/"
},
{
    name = "bops"

```

14. Save and exit the file.
15. Open config/crawler.conf and make sure the crawl\_config\_file is set to "connectors/filesystem.conf" and the crawl\_seed\_file is "seeds/filesystem-seed.conf"

```
# The configuration for the Connector Framework.  
connector_framework = {  
    # The file containing the connector configuration for the crawl.  
    crawl_config_file = "connectors/filesystem.conf",  
  
    # The file containing the seed URL for the crawl.  
    crawl_seed_file = "seeds/filesystem-seed.conf",  
  
    # The file containing the encryption key used for decrypting passwords.  
    # Generate a new key with the 'vcrypt' command included with Crawler.  
    id_vcrypt_file = "id_vcrypt",
```

### 3.3 Uploading Data to WDS collection

1. Open terminal/command line and navigate to top level folder that contains the Data Crawler config folder
2. Run the command **crawler testit** to verify that crawler is crawling the URL as expected

Ignore the error message "*Could not find path configuration in the Discovery config. There was an error trying to find required path in the Discovery config*". The testing is successful if the command executes without any error other than the mentioned one. If this error message does not show – please get the attention of the instructor.

3. Run the command **crawler crawl** to push your files to discovery service

You will get the same error message that you got while testing crawler. If there is no other error message and the command executes successfully, then you can see that the documents are uploaded to the collection. You can revert back to the Watson Discovery Service tooling interface.

4. You can see the number of documents uploaded by accessing the collection dashboard from the tooling interface. This step will be completed in 2-5 minutes.
5. The crawling will come to a halt. You can now exit from the terminal.

## 4 Data Ingestion using Discovery Tool and API calls

Data files can also be added to the collection using either the discovery tool or the API calls. Before uploading any documents, it is suggested that you create configuration and add enrichment attributes into the configuration (this should have already been completed in Section 2). Any data files uploaded before configuration updates will not reflect the changes.

### 4.1 Adding documents to the collection

The system detects the type of the document being uploaded and will only accept the following types:

- i. text/html (.html)
- ii. application/pdf (.pdf)
- iii. application/JSON (.JSON)
- iv. application/msword (.doc)
- v. application/vnd.openxmlformats-officedocument.wordprocessingml.document (.docx)
- vi. application/xhtml+xml (.xhtml)

#### 4.1.1 Uploading documents using the Tooling interface

The tooling should not be used to upload large amounts of data as it is single threaded, but it is a good way to get started when you have smaller sets of data.

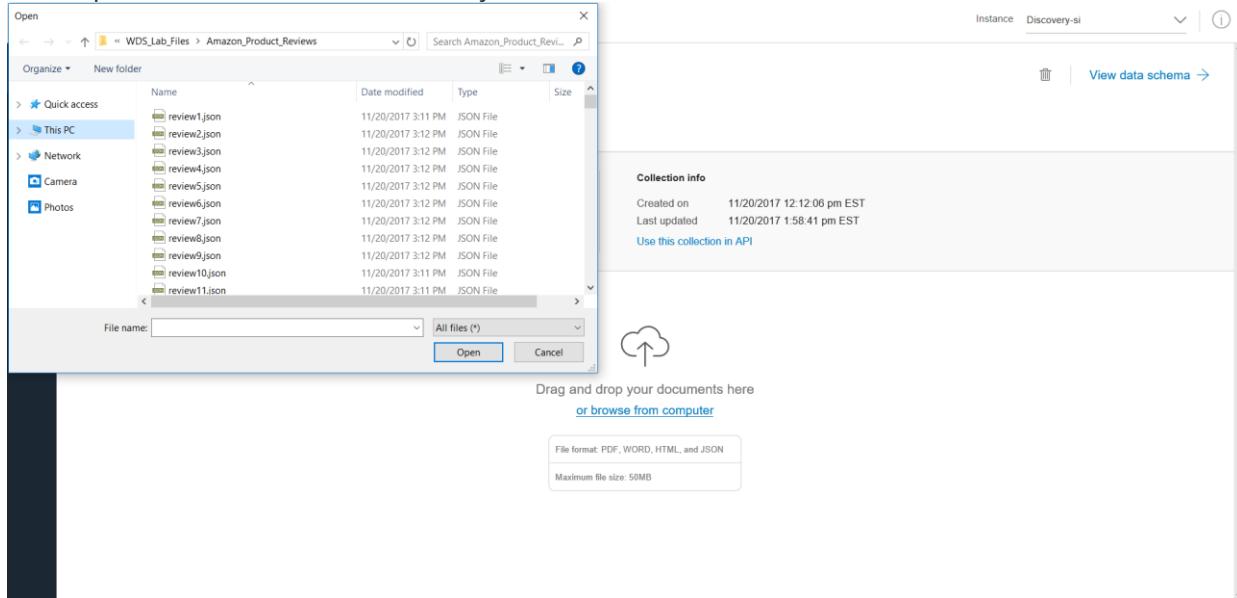
1. Open the Amazon Product Reviews collection from Your Data tooling homepage by clicking on the Amazon Product Reviews light blue tile.

The screenshot shows the 'Manage data' section of the Watson Discovery Tooling interface. On the left, there's a sidebar with icons for 'Manage data', 'Data sets', and 'Enrichments'. The main area has a dark header bar with the IBM Watson logo and navigation links. Below the header, the 'Manage data' section is visible, featuring a 'Create a data collection' button and a list of collections. One collection, 'Amazon Product Reviews', is highlighted with a light blue background. It displays the collection name, a small icon, and the text 'Indexed on 11/20/2017'. The overall interface is clean and modern, using a dark theme with light-colored cards for data items.

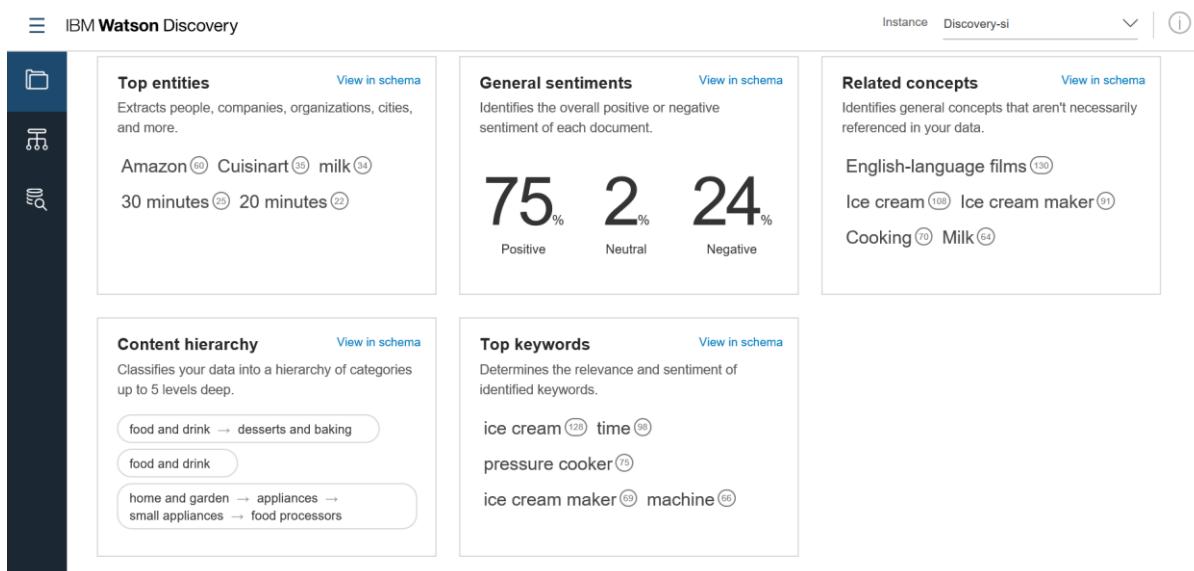
## Watson Discovery Service

You can ingest content by either dragging and dropping your document in the "Add data to this collection" box or by clicking "browse from computer".

- Click on 'browse from computer', navigate to the Amazon Reviews Content folder (this was part of the Zip folder which you have previously downloaded in Section 2), select all Amazon Reviews, and click Open. For Windows Machines, you can Choose CTRL-A, for MAC users it is Command-A.



- Once you click Open (or drag and drop depending on the method you used), the tooling will begin the ingestion process. It will take 5-10 minutes to upload the Amazon Review content to your collection.
- Once the selected documents are uploaded successfully, you will receive a popup saying, "Documents Uploaded". You can also refresh the page (by pressing Ctrl + R for Windows or Command + R for Macs) to see the number of documents available, processing and the failed uploads.



#### 4.1.2 Process to Upload documents using API calls (Optional)

Please note: this is not a step to be performed during the lab. It is additional information that users could find helpful. Developers can ingest documents using direct API calls. Documents can be uploaded into the collection using Node JS, python, java or any other programming language. The client should know the file handling process and creating HTTP post request process before trying out the API calls method

1. To upload data using the Node JS application, use the following code

```
var watson = require('watson-developer-cloud');
var fs = require('fs');

var discovery = new DiscoveryV1({
  username: '{username}',
  password: '{password}',
  version_date: '2016-12-01'
});

var file = fs.readFileSync('{/path/to/file}');

discovery.addDocument('{environment_id}', '{collection_id}', file),
  function(error, data) {
    console.log(JSON.stringify(data, null, 2));
});
```

2. If you are using java or C++ make sure you do not set the document type to "multipart/form-data" as you will encounter the error saying the format is not supported. Instead use the document type which is exactly matching with that of the document format you are uploading. For example, for an html document use "text\_html". A java example to upload document is shown below

```
package com.ibm.watson.developer_cloud.discovery.v1;

import java.io.FileInputStream;
import java.io.InputStream;

import com.ibm.watson.developer_cloud.discovery.v1.model.document.CreateDocumentRequest;
import com.ibm.watson.developer_cloud.discovery.v1.model.document.CreateDocumentResponse;
import com.ibm.watson.developer_cloud.http.HttpMediaType;

public class WDSDocumentUploader {

    public static void main(String[] args) {

        try {

            Discovery discovery = new Discovery("2016-12-01");

            discovery.setEndPoint("https://gateway.watsonplatform.net/discovery/api");
            discovery.setUsernameAndPassword("Username", "Password");

            InputStream documentStream = new FileInputStream("sample.html");

            CreateDocumentRequest.Builder builder = new CreateDocumentRequest.Builder(
                "environment id", "collection id");
            builder.inputStream(documentStream, HttpMediaType.TEXT_HTML);
            CreateDocumentResponse createResponse = discovery.createDocument(builder.build()).execute();

        } catch(Exception ex) {

        }
    }
}
```

## 5 Obtaining data insights/Querying the collection

After the discovery instance is configured and amazon review data is uploaded and enriched, 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. NLP is running behind the scenes to provide clear data insights.

1. In the "Overview" section, you can view the Amazon review analysis results. Data insights obtained for Amazon review data is shown in the below figure.

Document count: 999

Errors and warnings: 0 documents failed

General sentiments: 75% Positive, 2% Neutral, 24% Negative

Related concepts: English-language films, Ice cream, Ice cream maker, Cooking, Milk

Top entities: Amazon, Cuisinart, milk, 30 minutes, 20 minutes

Top keywords: ice cream, time, pressure cooker, ice cream maker, machine

Content hierarchy: food and drink → desserts and baking; food and drink; home and garden → appliances → small appliances → food processors

2. Click "View data schema →" in the top right corner of the "Overview" section or in the top right corners of any of the data insight boxes for exploration purposes of the collection and documents. You also can click the data schema icon on the lefthand side bar, but if you do that, you must reselect the collection you are working in which is "Amazon Product Reviews." The links are circled in red below.

View data schema →

View in schema

View in schema

View in schema

View in schema

## Watson Discovery Service

3. You can view the data schema in the "Collection view" or the "Document view." "Collection view" will display all fields in your collection and "Document view" will display a sample of documents from your data set with all fields included in the left side. Both representations are included below.

**View data schema** | Collection: Amazon Product Reviews

**Collection view** Document view

**Fields**

- review\_text
- enriched\_review\_text (Watson Enrichment)
- semantic\_roles
- emotion
- sentiment
- relations
- entities
- concepts
- keywords

**review\_text**

**Top values**

Value	Count
A good tool for home use. Shows several types of simple, easy to perform techniques. Would recommend for beginners and home bakers.	1 mention
A piece of trash for ice crushing. Motor sounds like it's about to explode. The cup is just frustration if you're trying to make smoothies. Replaced it with the Oster Beehive classic. Way better and actually quieter because of the sturdy base.	1 mention
ALTHOUGH ATTRACTIVE TO LOOK AT, I THOUGHT THIS WOULD BE LARGER THAN IT IS. I HAVE A \$10.00 PRODUCT THAT WORKS BETTER THAN THIS ONE. I EXPECTED MORE FROM THIS BRAND.	1 mention
After just over a week my unit began to make a clicking noise like a moving part was hitting something. The back on all these also gets quite hot.	1 mention
After making the purchase and using a few times, decided the Cooker was too much of a challenge. Purchased one that was electric that does not need me to watch the timer.	1 mention

**View data schema** | Collection: Amazon Product Reviews

**Collection view** Document view

**Documents**

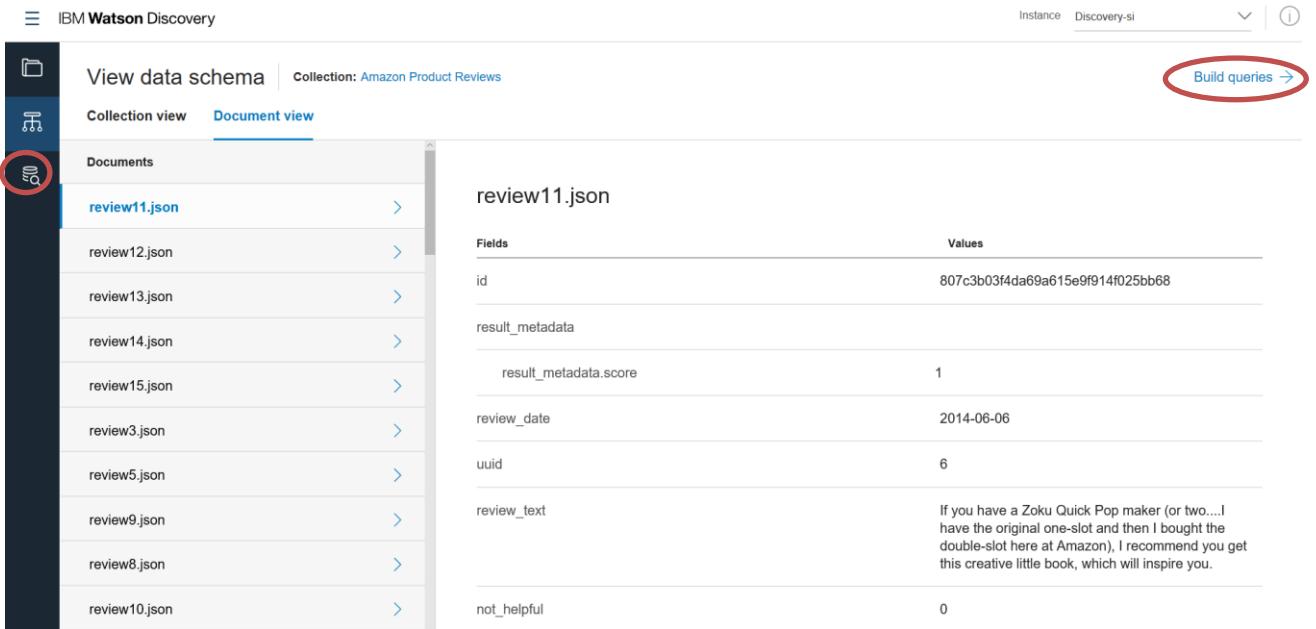
- review11.json
- review12.json
- review13.json
- review14.json
- review15.json
- review3.json
- review5.json
- review9.json
- review8.json
- review10.json

**review11.json**

Fields	Values
id	807c3b03f4da69a615e9f914f025bb68
result_metadata	
result_metadata.score	1
review_date	2014-06-06
uuid	6
review_text	If you have a Zoku Quick Pop maker (or two....I have the original one-slot and then I bought the double-slot here at Amazon), I recommend you get this creative little book, which will inspire you.
not_helpful	0

4. To build a query, you can either click "Build queries →" in the top right corner or click on the build queries icon on the lefthand side bar. Both options are circled in red below. If you click the icon, you will have to choose the collection you want to query which is "Amazon Product Reviews."

# Watson Discovery Service



**View data schema** | Collection: Amazon Product Reviews

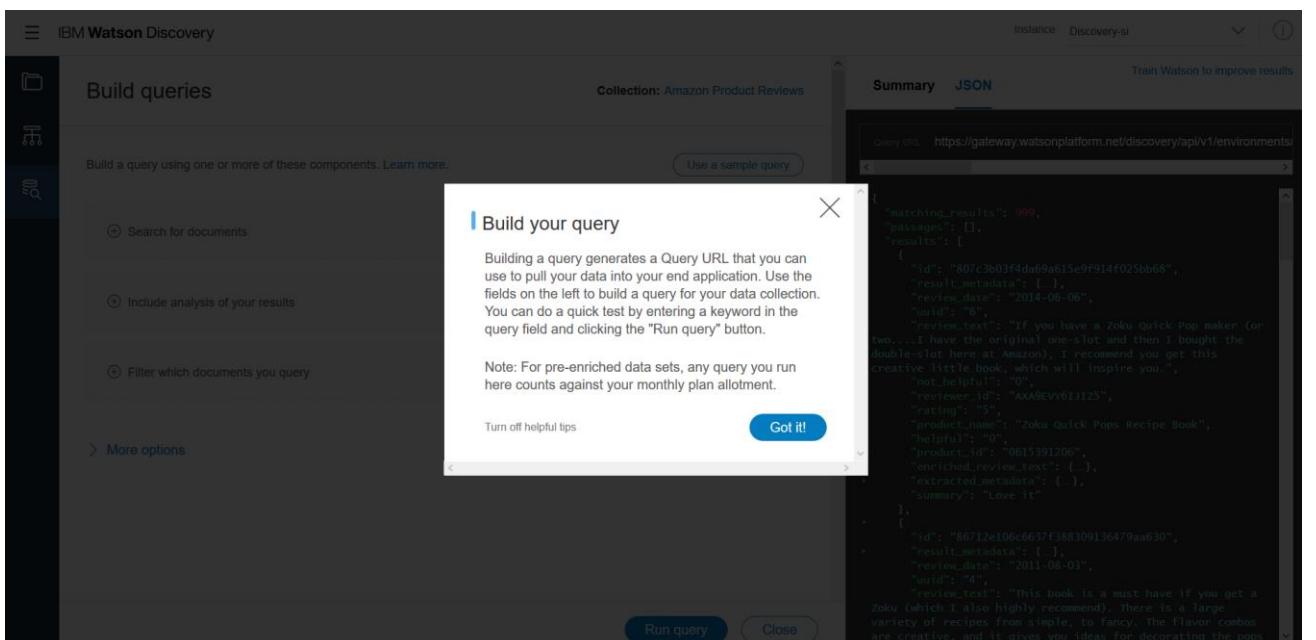
Collection view Document view

Documents

- review11.json >
- review12.json >
- review13.json >
- review14.json >
- review15.json >
- review3.json >
- review5.json >
- review9.json >
- review8.json >
- review10.json >

review11.json

Fields	Values
id	807c3b03f4da69a615e9f914f025bb68
result_metadata	
result_metadata.score	1
review_date	2014-06-06
uuid	6
review_text	If you have a Zoku Quick Pop maker (or two....I have the original one-slot and then I bought the double-slot here at Amazon), I recommend you get this creative little book, which will inspire you.
not_helpful	0

Build queries

Collection: Amazon Product Reviews

Summary JSON

Train Watson to improve results

Query URL: https://gateway.watsonplatform.net/discovery/api/v1/environments/

Build a query using one or more of these components. Learn more.

Use a sample query

Search for documents

Include analysis of your results

Filter which documents you query

More options

Build your query

Building a query generates a Query URL that you can use to pull your data into your end application. Use the fields on the left to build a query for your data collection. You can do a quick test by entering a keyword in the query field and clicking the "Run query" button.

Note: For pre-enriched data sets, any query you run here counts against your monthly plan allotment.

Turn off helpful tips Got it!

Run query Close

```
[{"searching_results": 999, "passages": [], "results": [ {"id": "807c3b03f4da69a615e9f914f025bb68", "result_metadata": {}, "review_date": "2014-06-06", "uuid": "6", "review_text": "If you have a Zoku Quick Pop maker (or two....I have the original one-slot and then I bought the double-slot here at Amazon), I recommend you get this creative little book, which will inspire you.", "not_helpful": "0", "reviewer_id": "AA5EVY6J11Z5", "rating": "5", "product_name": "Zoku Quick Pops Recipe Book", "helpful": "0", "product_id": "70615391206", "enriched_review_text": {}, "extracted_metadata": {}, "summary": "Love it!"}, {"id": "80712e106c6637f388309136479aa630", "result_metadata": {}, "review_date": "2011-08-01", "uuid": "4", "review_text": "This book is a must have if you get a Zoku (which I also highly recommend). There is a large variety of recipes from simple, to fancy. The flavor combos are creative, and it gives you ideas for decorating the pops"}, {"id": "80712e106c6637f388309136479aa630", "result_metadata": {}, "review_date": "2011-08-01", "uuid": "4", "review_text": "This book is a must have if you get a Zoku (which I also highly recommend). There is a large variety of recipes from simple, to fancy. The flavor combos are creative, and it gives you ideas for decorating the pops"}]
```

5. In the "Build queries" page, you can choose to "Search for documents," "Include analysis of your results," or "Filter which documents you query." Click "More options" to turn on passage retrieval.

# Watson Discovery Service

The screenshot shows the 'Build queries' panel of the Watson Discovery Service. On the left, there's a sidebar with icons for folder, list, and search. The main area has a title 'Build queries' and a subtitle 'Build a query using one or more of these components. [Learn more](#)'. Below this are three input fields: 'Search for documents', 'Include analysis of your results', and 'Filter which documents you query'. A link 'More options' is located below the first field. At the bottom right are 'Run query' and 'Close' buttons. The top right shows 'Collection: Amazon Product Reviews', 'Summary' (which is underlined), and 'JSON'. A note 'Your results will appear here when you run a query.' is also present.

The screenshot shows the 'Passages' configuration panel. It includes sections for 'More options' (with a dropdown arrow), 'Passages', 'Include relevant passages' (radio buttons for 'Yes' and 'No'), 'Fields to return' (radio buttons for 'All' and 'Specify'), 'Number of passages to return' (input field set to '5', with 'Default: 5' below it), 'Max character count for each passage' (input field set to '400', with 'Default: 400' below it), and 'Documents' (with a 'Fields to return' section). At the bottom right are 'Run query' and 'Close' buttons.

You can click "Use a sample query" to check out a few pre-built queries.

Below sections helps in understanding specific analysis performed on the dataset

## 5.1 Search for Documents

In this section, we will walk through custom queries that a developer would configure within their application in order to gain insight from the data. The application end user is Henry, a Business analyst hired by Amazon. Henry is responsible for consolidating product feedback and presenting it back to the product development team for continuous enhancements. Going through all the reviews by himself would be a tedious task. He therefore takes advantage of the WDS based application which can sort through huge number of reviews and provide valuable insight. As a developer, we need to help Henry generate actionable feedback by surfacing insight from the customer data. Below are some examples of the custom queries that a developer would create to help Henry achieve his goal.

The Discovery application provides users 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. In this example, passage retrieval is turned on.

1. Click on "Build queries" button on the lefthand bar and select the "Amazon Product Reviews" collection. In "Build queries" page, text fields are provided for filling in the keyword, count, sentiment types and other parameters against which the queries will be executed

2. Click "More options" to make sure passage retrieval is included.
3. Click "Search for documents" to build queries.
4. The below figure shows an example of a custom query, "ice cream maker" executed against the Amazon review data

# Watson Discovery Service

The screenshot shows the IBM Watson Discovery Service interface. On the left, there's a sidebar with icons for file management and search. The main area is titled "Build queries" and shows a search bar with the query "ice cream maker". Below the search bar are sections for "Include analysis of your results" and "Filter which documents you query". At the bottom are "Run query" and "Close" buttons. To the right, under "Summary", it shows the "Query URL" as <https://gateway.watsonplatform.net/discovery/api/v1/environments>. The "Passages" section lists several snippets related to ice cream makers. Under "Results", it says "Showing 10 of 263 matching documents" and provides a breakdown of sentiment, keywords, and emotions for one document named "review367.json".

You can view the Summary or the corresponding JSON of the results on the righthand side.

In this query, the enriched reviews are first filtered against "ice cream maker." The query searches for "ice cream maker" in all the fields of the review unless you specify a particular field to look for. By default, the query returns 10 matches, but as a user you can change that number by editing the field "Number of results to return".

The custom queries are powerful and can get deeper insights which can be exploited by users to understand where does their business stand and what could be done to reach a better ground.

5. Click the trash icon to clear the query.
6. On the Build your own query page "Search for documents" section, click the "Use the Discovery Query Language" button and add to the field, operator, and value.

This screenshot shows the same interface as above, but the "Search for documents" section has been modified. The "Use natural language" button is now grayed out, and the "Use the Discovery Query Language" button is highlighted. A dropdown menu is open over the "Field" input field, showing options like "enriched\_review\_text.categories.label", "operator", and "value". Other dropdowns show "is" as the operator and "food and drink/desserts and baking" as the value. The rest of the interface remains the same, showing the "Passages" and "Results" sections on the right.

This shows another way you can build queries.

- To clear some fields from dropdown, click on the cross mark next to each field if you have multiple selected. To clear the query click on the trash icon in the upper righthand side of the white space.

## 5.2 Include Analysis of Your Results

Henry may want to see a further analysis of the distribution of sentiment in reviews. He also may want to dive deeper into the keywords to see if there are negative trends with certain products. Section 5.4 covers keyword analysis using queries. To do this, he can use aggregations in the build queries page. Below is a quick briefing on how to use the analysis component.

- Click "Include analysis of your results"

The screenshot shows the 'Build queries' interface for the 'Amazon Product Reviews' collection. On the left, there's a sidebar with icons for 'Build queries', 'Search for documents', and 'Include analysis of your results'. The main area has three components: 'Search for documents', 'Include analysis of your results' (which is currently selected), and 'Filter which documents you query'. A 'More options' link is visible below the first two. On the right, there's a summary section with a message: 'Your results will appear here when you run a query.' Below the summary are 'Summary' and 'JSON' tabs, and a 'Train Watson to improve results' button. At the bottom are 'Run query' and 'Close' buttons.

- Type or select "enriched\_review\_text.sentiment.document.label" in the Field section and keep the count at 10.

The screenshot shows the 'Build queries' interface with the 'Include analysis of your results' component selected. In the 'Output' section, 'Top values' is chosen for 'Field' and 'enriched\_review\_text.sentiment.document.label' is selected. The 'Count' is set to 10. Below this, there are buttons for '+ Add condition', '+ Add child aggregation', '+ Add top-level aggregation', and a query term 'term(enriched\_review\_text.sentiment.document.label,count:10)'. On the right, the 'Summary' tab is active, showing a 'Query URL' and an 'Aggregations' section with a table:

	positive	negative	neutral
Count	746	238	15

The 'Results' section shows 10 matching documents, with the first one expanded to show sentiment ('positive'), emotions ('disgust: 0.02, joy: 0.73, anger: 0.02, fear: 0.02, sadness: 0.07'), and other review IDs.

Henry sees the distribution of positive, negative, and neutral reviews and can view the sentiment and emotions of 10 results.

- The analysis can be removed by clicking the trash icon.

## 5.3 Filtering Documents You Query

Henry might want to start his analysis by trying to understand how much positive and negative feedback is provided for a specific product. For example, he would want to know if the Cuisinart products are not doing so well so that he can take action and mitigate any risks. As a developer, we need to start by understanding what part of the data provides this insight. The enriched review JSON returned by Watson Discovery Service has lot of fields and we are only interested in querying the particular field that provides an answer to Henry's question.

The query results usually contain all the fields present in the original JSON files along with the fields generated under the enriched attribute. We can specify in the query to return just the fields we want in the result output.

- The filtering can be applied using the "Filter my data insights" section on the left side of the My data insights page.

The screenshot shows the IBM Watson Discovery interface. On the left, there's a sidebar with icons for folder, file, and search. The main area is titled 'Build queries' with a sub-section 'Collection: Amazon Product Reviews'. Below this, there are three components: 'Search for documents', 'Include analysis of your results', and 'Filter which documents you query'. A 'More options' link is also visible. On the right, there's a results panel with tabs for 'Summary' and 'JSON'. The 'JSON' tab is selected, showing a query URL and a large block of JSON data. The JSON data includes fields like 'matching\_results', 'passages', and 'results', with detailed information about individual reviews including IDs, dates, and enriched review text.

```

{
  "matching_results": 999,
  "passages": [],
  "results": [
    {
      "id": "807c3b03f4da69a615e9f914f025bb68",
      "result_metadata": {...},
      "review_date": "2014-06-06",
      "uuid": "6",
      "review_text": "If you have a Zoku Quick Pop maker (or two... I have the original one-slot and then I bought the double-slot here at Amazon), I recommend you get this creative little book, which will inspire you.",
      "not_helpful": "0",
      "reviewer_id": "AXA9EVY6IJIZ5",
      "rating": "5",
      "product_name": "Zoku Quick Pops Recipe Book",
      "helpful": "0",
      "product_id": "0615391206",
      "enriched_review_text": {...},
      "extracted_metadata": {...},
      "summary": "Love it"
    },
    {
      "id": "86712e106c6637f388309136479aa630",
      "result_metadata": {...},
      "review_date": "2011-08-03"
    }
  ]
}
  
```

- Select the type of enrichment you want to filter on the dropdown menu: keywords, concepts, sentiment, entity or taxonomy. Select enriched\_review\_text.sentiment.document.label.

# Watson Discovery Service

The screenshot shows the IBM Watson Discovery Service interface. On the left, there's a sidebar with icons for file management and search. The main area has a search bar at the top with placeholder text "Include analysis or your results". Below it is a section titled "Filter which documents you query" with a dropdown menu "Satisfy all" and a table for defining rules. The table has columns for "Field", "Operator", and "Value". One row is selected, showing "enriched\_review\_text.sentiment.document.label" as the field, "is" as the operator, and "positive" as the value. To the right of the filter section is a "Run query" button and a "Close" button. On the far right is a "Summary" tab and a "JSON" tab. The "JSON" tab is active, displaying a hierarchical tree of query results. One node is expanded, showing a document with fields like "id", "result\_metadata", "review\_date", "uuid", "review\_text", and "summary". The "review\_text" field contains a quote about Zoku Quick Pops.

3. From the “operator” dropdown select “is” and from the “value” dropdown select “positive.” Click apply filter to see the filtered insights.

This screenshot shows the same interface after applying the filter. The "Filter which documents you query" section now includes a new rule: "enriched\_review\_text.entities.sentiment.label::positive". The "Run query" button is visible at the bottom. To the right is a "Summary" tab and a "JSON" tab. The "JSON" tab is active, showing a list of 10 matching documents. Each document is represented by a collapsed node labeled with its ID (e.g., "review33.json"). Expanding one node reveals its contents, including "Sentiment" (positive), "Emotions" (disgust: 0.0, joy: 0.91, anger: 0.03, fear: 0.03, sadness: 0.01), and the "review\_text" field which contains a short review sentence.

4. The filtered insights show the number of matching documents along the top 10 documents emotions.
5. To filter against multiple types, click “add rule” on top of already filtered insights. In the below example, “enriched\_review\_text.keywords.text” “contains” “ice cream” is added so the results filter positive reviews with the mention of ice cream.

The screenshot shows the Watson Discovery Service interface. On the left, there's a sidebar with icons for folder, documents, and search. The main area has a header "IBM Watson Discovery" and a sub-header "Filter which documents you query". Below this, there's a section titled "Satisfy all of the following rules" with two filter conditions: "enriched\_review\_text.keywords.sentiment.label" is "positive" and "enriched\_review\_text.keywords.text" contains "ice cream". There are buttons for "+ Add rule" and "+ Add group of rules". At the bottom of this section is the query string: "enriched\_review\_text.keywords.sentiment.label::\"positive\",enriched\_review\_text.keywords.text:\"ice cream\"". To the right, there's a "Run query" button and a "Close" button. The right side of the screen shows the "Summary" tab selected, with a "Results" section displaying 10 of 177 matching documents. One document, "review1.json", is expanded, showing fields like Sentiment (positive), Keywords (ice cream), Emotions (disgust: 0.02, joy: 0.85, anger: 0.01, fear: 0.01, sadness: 0.1), and Review\_text ("I bought this zoku quick pop for my daughter with her zoku quick maker. She loves it and have fun to make her own ice cream"). Other documents listed are review192.json, review263.json, review264.json, and review262.json.

6. The insights now are filtered against all the reviews containing ice cream keyword and are classified under positive sentiment
7. The filtering can be removed by clicking the cross button next to the applied filter displayed on top of data insights.

These first steps allow developers to sift through the enriched JSON output from Watson Discovery Service. We are now ready to query the JSON fields to extract meaningful insight.

## 5.4 Top keywords

This segment displays the top keywords present in the reviews along with the count depicting the number of times they appear in all of the reviews. The Discovery service automatically identifies supported languages in your input content, and then identifies and ranks keywords in that content.

For the Amazon review data, the keywords include ice cream, time, pressure cooker, ice cream maker, and machine. This makes sense since the reviews are about kitchen appliances and ice cream machines.

1. When looking at the "Insights from your enriched data" on the Manage data page, click the "View in schema" button on the right top of the Top keywords box. Upon clicking the button, the user will be sent to the data schema collection view of the top keywords as discussed in the overview.

**Top keywords**[View in schema](#)

Determines the relevance and sentiment of identified keywords.

ice cream (128) time (98) pressure cooker (75)

ice cream maker (69) machine (66)

- The top keywords query is prepopulated with a few example values and the corresponding relevance and sentiment. Click "show more values" to view different relevant examples.

- After viewing the different values, click "Build queries" in the top right corner. The "Build queries" page is shown below.

# Watson Discovery Service

The screenshot shows the Watson Discovery Service interface. On the left is a sidebar with icons for file management, search, and other functions. The main area is titled "Build queries" and specifies "Collection: Amazon Product Reviews". Below this, there's a note: "Build a query using one or more of these components. [Learn more.](#)". A "Use a sample query" button is available. Three main options are listed in boxes: "+ Search for documents", "+ Include analysis of your results" (which is highlighted in blue), and "+ Filter which documents you query". At the bottom are "Run query" and "Close" buttons.

- Click "Include analysis of your results" to build a query associated with the top keywords.

This screenshot is identical to the one above, showing the "Build queries" interface for the "Amazon Product Reviews" collection. The "Include analysis of your results" option is again highlighted in blue, indicating it has been selected. The rest of the interface elements are the same: sidebar, main title, collection info, note, sample query button, and the three main query components.

- Type in "enriched\_review\_text.keywords.text" in the Field section and keep the count at 10 then click "Run query" at the bottom of the page. Below the top 10 keywords are aggregated.

# Watson Discovery Service

IBM Watson Discovery

Instance Discovery-si

The screenshot shows the Watson Discovery Service interface. On the left, there's a sidebar with icons for folder, documents, and search. The main area has a search bar with placeholder text "Search for documents". Below it, a section titled "Include analysis of your results" contains two aggregation configurations:

- Output:** Top values
  - Field:** enriched\_review\_text.keywords.text
  - Count:** 10
- Output:** Top values
  - Field:** enriched\_review\_text.sentiment.document.label
  - Count:** 10

The query editor at the bottom shows the constructed query: `term(enriched_review_text.keywords.text,count:10)`. To the right, the "Summary" tab is selected, showing the URL `https://gateway.watsonplatform.net/discovery/api/v1/environment`. The "Aggregations" section lists the top keywords with their counts:

- ice cream (128)
- time (98)
- pressure cooker (75)
- ice cream maker (69)
- machine (66)
- freezer (61)
- book (58)
- thing (56)
- bit (53)
- bowl (49)

The "Results" section indicates 10 of 999 matching documents.

6. You can change the query by increasing the count, adding a condition, adding child aggregation, or adding top-level aggregation. Click “+Add child aggregation” and type or select “enriched\_review\_text.sentiment.document.label” for the field. This way you can view the distribution of positive and negative reviews relating to the specific keyword. Click “Run Query

This screenshot shows the same interface after modifying the query. The "enriched\_review\_text.keywords.text" aggregation remains the same. The second aggregation now includes a child aggregation for the "enriched\_review\_text.sentiment.document.label" field:

- Output:** Top values
  - Field:** enriched\_review\_text.keywords.text
  - Count:** 10
- Output:** Top values
  - Field:** enriched\_review\_text.sentiment.document.label
  - Count:** 10
  - + Add child aggregation**

The query editor shows the updated query: `term(enriched_review_text.keywords.text,count:10).term(enriched_review_text.sentiment.document.label,count:10)`. The "Summary" tab shows the same URL and aggregation results as before, but the "Results" section is not visible in this specific screenshot.

The results show how there are 100 positive reviews about “ice cream” and 28 negative reviews. For “time,” there are 67 positive reviews and 30 negative reviews. It is worth drilling further into the reviews with “time” in them.

## Aggregations

term(enriched\_review\_text.keywords.text) **ice cream** (128)

- term(enriched\_review\_text.sentiment.document.label) **positive** (100)
- term(enriched\_review\_text.sentiment.document.label) **negative** (28)

term(enriched\_review\_text.keywords.text) **time** (98)

- term(enriched\_review\_text.sentiment.document.label) **positive** (67)
- term(enriched\_review\_text.sentiment.document.label) **negative** (30)
- term(enriched\_review\_text.sentiment.document.label) **neutral** (1)

term(enriched\_review\_text.keywords.text) **pressure cooker** (75)

- term(enriched\_review\_text.sentiment.document.label) **positive** (60)
- term(enriched\_review\_text.sentiment.document.label) **negative** (15)

- After reviewing the Aggregations results, click the trash can icon to move on to the "Search for documents" section.

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

- Search for documents:** A search bar at the top left.
- Include analysis of your results:** A section where users can define aggregations.
  - Output:** Top values (selected), Field: enriched\_review\_text.keywords.text, Count: 10. A trash can icon is circled in red.
  - Output:** Top values (selected), Field: enriched\_review\_text.sentiment.document.label, Count: 10. A trash can icon is also present here.
  - Add child aggregation:** A button to add nested aggregations.
  - Add top-level aggregation:** A button to add higher-level aggregations.
- Query URL:** https://gateway.watsonplatform.net/discovery/api/v1/environnem...
- Summary:** Shows the results of the aggregations:
  - term(enriched\_review\_text.keywords.text) **ice cream** (128)
    - term(enriched\_review\_text.sentiment.document.label) **positive** (100)
    - term(enriched\_review\_text.sentiment.document.label) **negative** (28)
  - term(enriched\_review\_text.keywords.text) **time** (98)
    - term(enriched\_review\_text.sentiment.document.label) **positive** (67)
    - term(enriched\_review\_text.sentiment.document.label) **negative** (30)
    - term(enriched\_review\_text.sentiment.document.label) **neutral** (1)
  - term(enriched\_review\_text.keywords.text) **pressure cooker** (75)
    - term(enriched\_review\_text.sentiment.document.label) **positive** (60)
    - term(enriched\_review\_text.sentiment.document.label) **negative** (15)
- JSON:** A tab showing the JSON representation of the query.

- Make sure "Passage Retrievals" is set to on by clicking "more options at the bottom of the page. After you confirm it is on, click "Search for documents."

Build a query using one or more of these components. [Learn more.](#)

[Use a sample query](#)

[⊕ Search for documents](#)

[⊕ Include analysis of your results](#)

[⊕ Filter which documents you query](#)

[› More options](#)

- Under the “Use natural language” tab, type in “time” and click “Run Query at the bottom of the page or press enter on your keyboard.

The screenshot shows the Watson Discovery Service interface. On the left, there's a sidebar with icons for file, database, and search. The main area has a title "Build queries" and a sub-section "Search for documents". Under "Search for documents", there are two tabs: "Use natural language" (which is selected) and "Use the Discovery Query Language". A text input field contains the word "time". Below the input field are three expandable sections: "Include analysis of your results", "Filter which documents you query", and "More options". To the right, the "Summary" tab is selected in a panel titled "Passages". The panel shows a "Query URL" and several snippets of text from reviews. One snippet reads: "...Cooking Times" may be a bit deceptive...." Another snippet discusses cooking times: "...the time. I had used altogether 3 times and every time it did not do the job and after tightening the screw for 3 times the screw worn out. I realize am azon will only refund partially according to the return policy I just read....". A third snippet says: "...The so-called "cooking times" seem short and very fast, but they don't take into account the time it takes to build up the pressure, which is often longer than the actual stated cooking times....". A fourth snippet is partially visible: "...One upon a time, many years ago... I had a very bad experience with a pressure cooker. I swore at that time to never let one in my house again. Well, time passed and my patience with cooking grew more thin....". A fifth snippet discusses cooking times: "...We used, frankly, a pretty cheap piece of meat but it came out tasting fantastic. It was quick but something to keep in mind is cooking time does not usually include the time to heat it up and the time to cool it down. Also you can't just walk away. Once the pressure seals the cooker most of the time you have to then reduce the heat....".

When looking into the passages taken out from reviews, you can see that from the passages that “Cooking Times may be a bit deceptive,” “seem short and very fast, but they don’t take into account the time it takes to build up the pressure, which is often longer than the actual stated cooking time...” and lastly, “cooking time does not usually include the time to heat it up and the time to cool it down.” These passages taken from reviews on the pressure cooker show that the pressure cooker does not do a great job offering an accurate estimate of the cooking time. This offers insights for Henry to consider adding a more accurate timer feature to the pressure cooker.

## Passages

"..."Cooking Times" may be a bit deceptive...."

"...the time. I had used altogether 3 times and every time it did not do the job and after tightening the screw for 3 times the screw worn out. I realize am azon will only refund partially according to the return policy I just read...."

"...The so-called "cooking times" seem short and very fast, but they don't take into account the time it takes to build up the pressure, which is often longer than the actual stated cooking times...."

"...One upon a time, many years ago.. I had a very bad experience with a pressure cooker. I swore at that time to never let one in my house again. Well, time passed and my patience with cooking grew more thin...."

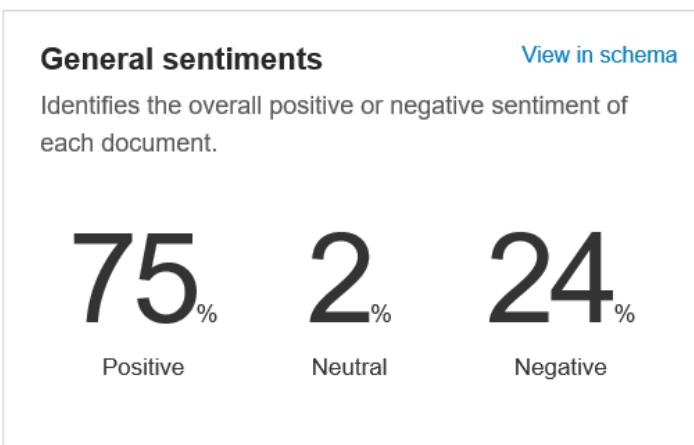
"...We used, frankly, a pretty cheap piece of meat but it came out tasting fantastic. It was quick but something to keep in mind is cooking time does not usually include the time to heat it up and the time to cool it down. Also you can't just walk away. Once the pressure seals the cooker most of the time you have to then reduce the heat...."

10. To return to the Query dashboard, click on the Collection name (Amazon Product Reviews) in the top right hand side of the white section.

## 5.5 General Sentiments Analysis

This segment displays the distribution of sentiment associated with the reviews broadly classifying them into positive, neutral and negative sentiment. Discovery service Identifies attitude, opinions, or feelings in the content that is being analyzed. A good rating score does not mean that the user is actually satisfied, since there is a possibility that the review carries negative sentiment with it. Sentiment analysis helps in identifying trends in the reviews and eliminating fake reviews

1. For the Amazon use case the classification says that 75% of the reviews fall in positive sentiment category, 2% and 24% fall in neutral and negative sentiment category respectively.
2. Click on the "View in schema" next to General Sentiments tile to navigate to view your data schema and then build your own query page.



- Example values are prepopulated. Click "show more values" to view different relevant examples. After viewing the different values, you can click "Build queries" in the top right corner to build out queries and gain further data analysis.

View data schema | Collection: Amazon Product Reviews

Collection view Document view

Fields

- review\_text >
- enriched\_review\_text (Watson Enrichment)
- semantic\_roles >
- emotion >
- sentiment >

enriched\_review\_text.sentiment (Watson Enrichment)

Fields	Example values	Show more values >
enriched_review_text.sentiment.document.score >0: positive; 0.0: neutral; <0: negative	0.910579	
enriched_review_text.sentiment.document.label	positive	

- Follow the steps you did in the last section to include an analysis of your results, but this time type in "enriched\_review\_text.sentiment.document.label." The query results show the number of reviews with positive, negative and neutral sentiments.

Build queries | Collection: Amazon Product Reviews

Build a query using one or more of these components. [Learn more](#). [Use a sample query](#)

Search for documents

Include analysis of your results [Edit in query language](#)

Output Field Count

Top values enriched\_review\_text.sentiment.document.label 10

+ Add condition

+ Add child aggregation

+ Add top-level aggregation

Run query Close

Summary JSON Train Watson to improve results

Query URL: <https://gateway.watsonplatform.net/discovery/api/v1/environ>

Aggregations

term(enriched\_review\_text.sentiment.document.label)

- positive (746)
- negative (238)
- neutral (15)

Results

Showing 10 of 999 matching documents

review11.json

Sentiment	positive
Emotions	disgust: 0.02, joy: 0.73, anger: 0.02, fear: 0.02, sadness: 0.07

review12.json

5. You can change the query by increasing the count or you can filter against categories, entities, concepts, emotion, keywords and more. You also can search for documents using the Discovery Query Language and selecting a specific sentiment value.

Build a query using one or more of these components. [Learn more](#)

[Use a sample query](#)

Search for documents

[Edit in query language](#) [Delete](#)

Use natural language [Use the Discovery Query Language](#)

Satisfy all of the following rules

Field	Operator	Value
enriched_review_text.sentiment.document.la...	is	positive negative neutral

+ Add rule  
+ Add group of rules

enriched\_review\_text.sentiment.document.label::

[Include analysis of your results](#)

[Filter which documents you query](#)

[Run query](#) [Close](#)

6. To return to the Query dashboard, click on the Collection name (Amazon Product Reviews) in the top righthand side of the white section.

## 5.6 Top Entities

This segment displays the top entities present in the review documents: Entities like person, place or organization. Entity extraction adds semantic knowledge to content to help understand the subject and context of the text that is being analyzed

1. For Amazon review data, the top entities identified in the Top entities tile were pressure cooker, Cuisinart, Amazon, Milk etc.,
2. Click on the "View in schema" link to navigate to view the data schema and then build your own query.

**Top entities** [View in schema](#)

Extracts people, companies, organizations, cities, and more.

Amazon <sup>(60)</sup> Cuisinart <sup>(35)</sup> milk <sup>(34)</sup>  
30 minutes <sup>(25)</sup> 20 minutes <sup>(22)</sup>

## Watson Discovery Service

3. The data schema sections show the example entities from the review documents and shows the relevance, count, and score. Click "Show more values" to view the other top entities and click "Build queries" in the top right corner after to build out your own queries.

View data schema | Collection: Amazon Product Reviews

**Collection view**   **Document view**

**Fields**

- review\_text
- enriched\_review\_text (Watson Enrichment)
- semantic\_roles
- emotion
- sentiment
- relations
- entities**
- concepts

**enriched\_review\_text.entities** (Watson Enrichment)

Fields	Example values
enriched_review_text.entities.text	Amazon
enriched_review_text.entities.type	Company
enriched_review_text.entities.relevance	0.33 Range: 0.0-1.0
enriched_review_text.entities.count	1
enriched_review_text.entities.sentiment.score	-- (>0: positive; 0: neutral; <0 negative)

Show more fields

4. You can run through steps similar to the previous lab sections for the entities enrichments.

Build a query using one or more of these components. [Learn more.](#)   [Use a sample query](#)

Search for documents

Include analysis of your results

Output	Field	Count
Top values	enriched_review_text.entities.text	10

term(enriched\_review\_text.entities.text,count:10)

Filter which documents you query

**Summary**   **JSON**   [Train Watson to improve results](#)

Query URL: <https://gateway.watsonplatform.net/discovery/api/v1/environments/Discovery-si>

**Aggregations**

term(enriched\_review\_text.entities.text)

- Amazon (60)
- Cuisinart (35)
- milk (34)
- 30 minutes (25)
- 20 minutes (22)
- Presto (22)
- 24 hours (18)
- two years (14)
- 10 minutes (13)
- 25 minutes (13)

**Results**

Showing 10 of 999 matching documents

5. To return to the Query dashboard, click on the Collection name (Amazon Product Reviews) in the top right hand side of the white section.
6. You can view the other entity sub fields by searching the various options under the field section. It's worth trying the different fields to gain a more in-depth analysis of your data.

## Watson Discovery Service

The screenshot shows the Watson Discovery Service interface. At the top, there are tabs for "Summary" (which is selected) and "JSON". Below the tabs, a message says "Your results will appear here when you run a query." On the left, there's a sidebar with icons for folder, search, and refresh. The main area has two tabs at the top: "Use natural language" (selected) and "Use the Discovery Query Language". A search bar below these tabs contains the placeholder text "e.g. IBM Watson in healthcare". The main content area is titled "Include analysis of your results". It shows a table with columns "Output", "Field", and "Count". The first row is selected and highlighted in blue, showing "enriched\_review\_text.categories.label" as the field and "10" as the count. Other rows include "enriched\_review\_text.concepts.dbpedia\_resource", "enriched\_review\_text.concepts.text", "enriched\_review\_text.entities.disambiguation.dbpedia\_resource", "enriched\_review\_text.entities.disambiguation.name", and "enriched\_review\_text.entities.disambiguation.subtype". There are buttons for "Edit in query language" and "Delete" next to each row. On the far left of the table, there are buttons for "Top values" (selected), "+ Add condition", "+ Add child aggregation", "+ Add top-level aggregation", and "term(enriched\_review\_...)" (partially visible). Below the table, there's a section for "Filter which document" and a dropdown for "Satisfy all" rules. At the bottom right are "Run query" and "Close" buttons.

More information to help with building your own query can be found at  
<https://www.ibm.com/watson/developercloud/doc/discovery/using.html> and  
<https://www.ibm.com/watson/developercloud/doc/discovery/query-reference.html>

## 6 Useful Considerations: Data Collection

### 6.1 Collecting Content for Watson Discovery Service

During this lab exercise, we have facilitated the configuration steps by providing you with sample content data. However, when working with your data, you will have to perform two additional activities: data collection and data preparation. Data collection activities are required to consolidate the data. Data preparation refers to the activities required to take the source data and prepare it in order to be supported for ingestion in Watson Discovery Service.

#### 6.1.1 Data formats supported

In this Lab, we provide you directly with a folder containing a set of clean Amazon Product Reviews. However, in real engagements, data will mostly likely be crawled from other sources such as internal databases. Imagine for that scenario that we were storing and maintaining Amazon Product Reviews within a database. Here are some considerations we would want to keep in mind:

1. Watson Discovery Service supports three ways of uploading data into your instance: through the tooling web interface, via API call, and by using the Data Crawler tool.
2. All ingestion methods support uploading Word, JSON, Pdf, or Html files;
3. Each Amazon Product Review should be converted into a single JSON document with the document structure clearly headlining the property name and its values. This step will allow us to enrich our data effectively and be able to extract the insight we need to meet the business scenario.
4. In this scenario, you might want to store the Amazon Product Reviews within a database. The database should contain the properties as columns and each review should be a row entry in the table.
5. An example database for the Amazon Product Review use case is shown below. The structure was created to be supported by the Data Crawler.

uuid	reviewer_id	product_id	helpful	not_helpful	review_text	rating	summary	product_name	review_date	time_inserted
2 A1JVQTAGHYOL7F	615391206	0	0	0	I bought this zoku quick pop for my daughter with her zoku	5	Zoku	Zoku Quick Pops Recipe	6/17/14	12/13/16
4 A2MHC7X43IMDZ	615391206	14	18	0	This book is a must have if you get a Zoku (which I also hig	5	Creative Combos	Zoku Quick Pops Recipe	8/3/11	12/13/16
5 AHA185T5C2DH3	615391206	0	0	0	This cookbook is great. I have really enjoyed reviewing all o	4	A must own if you own the Zoku	Zoku Quick Pops Recipe	8/6/14	12/13/16
6 AXA9EVY6LJIZ5	615391206	0	0	0	If you have a Zoku Quick Pop maker (or two....) have the or	5	Love it	Zoku Quick Pops Recipe	8/6/14	12/13/16
7 A1SW2D234X11MS	615391206	2	3	0	This book is so beautifully illustrated and easy to follow. It h	5	Beautifully illustrated recipe and ir	Zoku Quick Pops Recipe	7/1/12	12/13/16
9 A1V13XAWO98C1J	615391206	13	13	0	The Zoku accessories are pricey, but the recipes have the e	5	If you have the Zoku maker, buy t	Zoku Quick Pops Recipe	9/21/11	12/13/16
10 A3842PXNQ1QRM9	615391206	0	0	0	High quality book. Clear and helpful information, and then s	5	Satisfaction opinion	Zoku Quick Pops Recipe	5/31/14	12/13/16
11 A2BV75QJSEEQ1B	615391206	3	3	0	The recipes in this books are tedious to make. I found that e	4	I liked this book, but.....	Zoku Quick Pops Recipe	12/25/12	12/13/16
12 A3R9X003XW0LNR	689027818	0	0	0	This beautifully illustrated book featuring ten colorful,glittery	5	A sweet book for babies and todd	Good Night, Sweet Butter	7/26/09	12/13/16
14 A14BTJRH9VNLJJ	689027818	1	2	0	This is the story of a group of butterflies looking for a place	5	My little one absolutely loved this	Good Night, Sweet Butter	4/5/13	12/13/16
18 A3LHQ0LXICBLY2	912696591	1	1	0	I should not have bought it because it comes free from my c	5	I should not have bought it.	Wilton Decorating Cakes	6/6/13	12/13/16
19 A39RRTO760KGOY	912696591	0	0	0	had no idea about decorating cakes,by this booklet ,now i c	5	follow the instructions	Wilton Decorating Cakes	5/26/12	12/13/16
20 AZF13R8HB59E	912696591	0	1	1	excellent book, shipped on time, packaged nicely, helpful tip	5	book	Wilton Decorating Cakes	6/11/12	12/13/16
23 A2S3ZE0R19160A	912696591	0	0	0	helpful information on decorating basics. Good book for beg	4	This is a good book for anyone	Wilton Decorating Cakes	12/12/13	12/13/16
24 A33TB072UNBVR	912696591	1	1	1	Got this for my girlfriend who runs a baking business and sh	5	Handy to Have	Wilton Decorating Cakes	6/7/13	12/13/16
28 A2i5YZBMLU2R70	912696591	0	1	1	I am very happy I purchased this book as it contains so mu	4	good tips	Wilton Decorating Cakes	5/6/13	12/13/16
30 A1LAAZUR8YFRW8	912696591	0	0	0	I have always wanted to learn how to decorate a beautiful c	4	Great For Beginners!	Wilton Decorating Cakes	8/15/13	12/13/16

6. The Data Crawler tool can be used to push large numbers of records into the Watson Discovery Service collection;

7. The Data Crawler uses data stored in a database or the data files stored in file system and creates JSON documents out of each data entry and pushes it into the Watson Discovery Service instance;
8. The JSON documents created using the crawler tool will have a slightly different format from that of the ones uploading using tooling or API calls. The attributes present in the Amazon Product Review will be placed under a parent node defined as "metadata", in the JSON document
9. If you have crawled your data from a database, and you are trying to enrich your data fields, a prefix of 'metadata' will have to be added. The following figure shows the sample JSON document created by the Data Crawler after crawling data from database

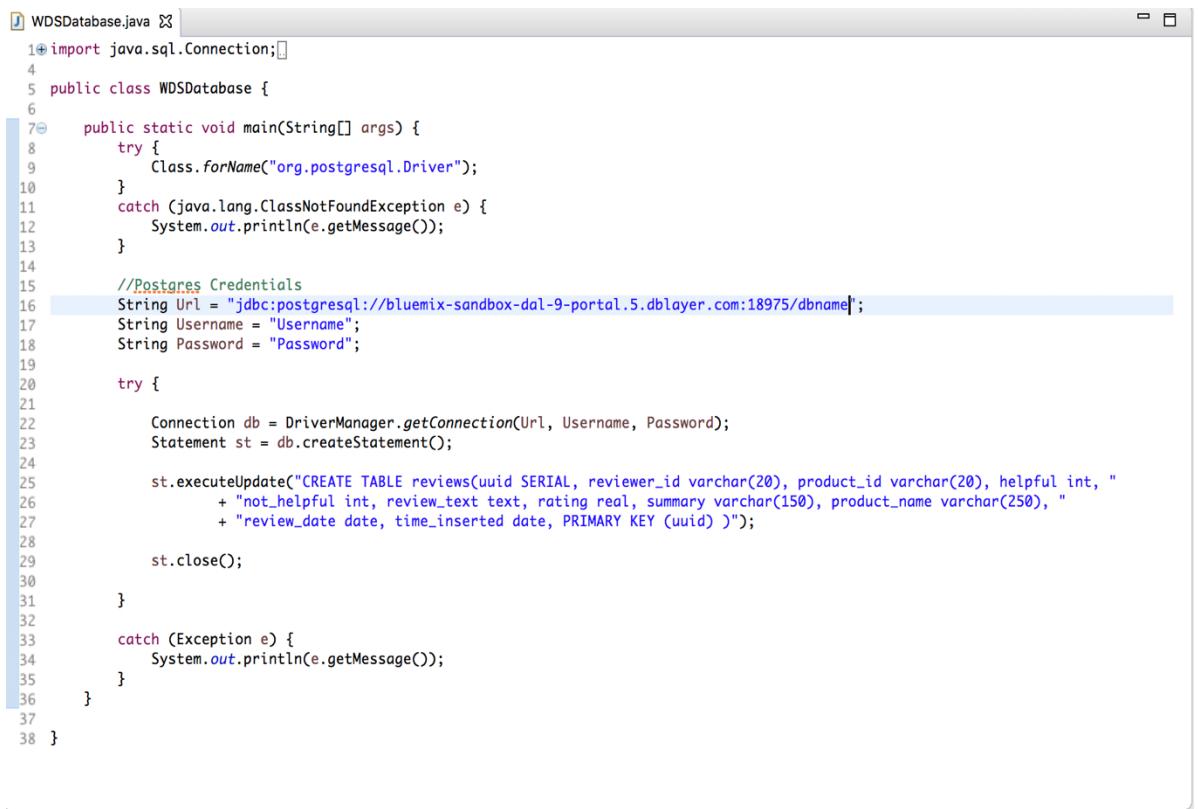
```
{  
    "id": "f29187a2eb6a7dcb3ea074725cb1167b49eb9785b947abb69f7046cc58b9ef5d",  
    "score": 1.2738007,  
    "metadata": {  
        "review_date": "2013-07-27",  
        "uuid": "347",  
        "time_inserted": "2016-12-13",  
        "review_text": "Worked great for the first year we had it, when we used it maybe 3-4 times. But after that, it seemed like nothing would ever really freeze up enough. I looked into the problem and decided that I might have damaged the cooling capability of the bowl by cleaning it with hot water (which is a no-no). So, I bought a replacement bowl, and once again, for the first year with a few uses it seemed to be OK. But now once again, it's just not getting cold enough - and this time I *know* it's not my fault since I have been very careful to only wash the bowl in cold water.",  
        "not_helpful": "0",  
        "id": "f29187a2eb6a7dcb3ea074725cb1167b49eb9785b947abb69f7046cc58b9ef5d",  
        "reviewer_id": "A2AXWQVA5V9GPJ",  
        "rating": "2",  
        "product_name": "Cuisinart ICE-20 Automatic 1-1/2-Quart Ice Cream Maker, White",  
        "helpful": "0",  
        "product_id": "B00000JGRT",  
        "summary": "Cooling capability wears out fast"  
    }  
}
```

10. The addition of the metadata field will have implications on how you configure and query the Amazon Product Reviews. On the configuration side, you will have to create a configuration file that enriches the metadata.review\_text field rather than just review\_text field.
11. A similar approach will have to be used for querying your data. For example, you will have to change your query to search metadata.review\_text rather than review\_text field.

### 6.1.2 Database creation and access

1. IBM Cloud database services can be used to store data for the crawler to crawl it and push into the WDS collection.
2. Postgres DB or DB2 is recommended if you are intending to use IBM Cloud Database services.

3. A sample Postgres DB can be created in IBM Cloud which can then host the Amazon Review Data. Below are the necessary credentials you will need to access the DB
  - i. Username: xxxxxxx
  - ii. Password: xxxxxxxxxxx
  - iii. URL: jdbc:postgresql://bluemix-sandbox-dal-9-portal.5.dblayer.com:18975/dbname
  - iv. Table Name: XXXXX
4. The database can be accessed/modified using API calls or using a simple java application with jdbc connection.
5. The application needs to reference a jdbc driver jar for postgres which can be found at <https://jdbc.postgresql.org/download.html>
6. A sample code for accessing and creating a table in the postgres service hosted in Cloud is shown below. In a similar way, an entry can be added to the row using simple SQL commands



```

1① import java.sql.Connection;
2
3 public class WDSDatabase {
4
5     public static void main(String[] args) {
6
7         try {
8             Class.forName("org.postgresql.Driver");
9         }
10        catch (java.lang.ClassNotFoundException e) {
11            System.out.println(e.getMessage());
12        }
13
14
15        //Postgres Credentials
16        String Url = "jdbc:postgresql://bluemix-sandbox-dal-9-portal.5.dblayer.com:18975/dbname";
17        String Username = "Username";
18        String Password = "Password";
19
20        try {
21
22            Connection db = DriverManager.getConnection(Url, Username, Password);
23            Statement st = db.createStatement();
24
25            st.executeUpdate("CREATE TABLE reviews(uuid SERIAL, reviewer_id varchar(20), product_id varchar(20), helpful int, "
26                            + "not_helpful int, review_text text, rating real, summary varchar(150), product_name varchar(250), "
27                            + "review_date date, time_inserted date, PRIMARY KEY (uuid )");
28
29            st.close();
30
31        }
32
33        catch (Exception e) {
34            System.out.println(e.getMessage());
35        }
36    }
37
38 }

```

### 6.1.3 Database table schema for our Amazon Product Review use case

1. The database table “reviews” should contain all the required properties of a review arranged as table columns
2. A typical schema of the table for our use case is shown below

```
CREATE TABLE reviews(
    uuid SERIAL,
    reviewer_id varchar(20),
    product_id varchar(20),
    helpful int,
    not_helpful int,
    review_text text,
    rating real,
    summary varchar(150),
    product_name varchar(250),
    review_date date,
    time_inserted date,
    PRIMARY KEY (uuid)
);
```

3. The column "uuid" which is of the serial datatype is used as an auto increment index for the table which is also a primary key
4. Indexing the table helps in fast retrieval of the data items

## 7 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. RT can be done programmatically or in the tooling. In this section, we will show you how to train the model in the tooling.

### 7.1 Overview and Requirements

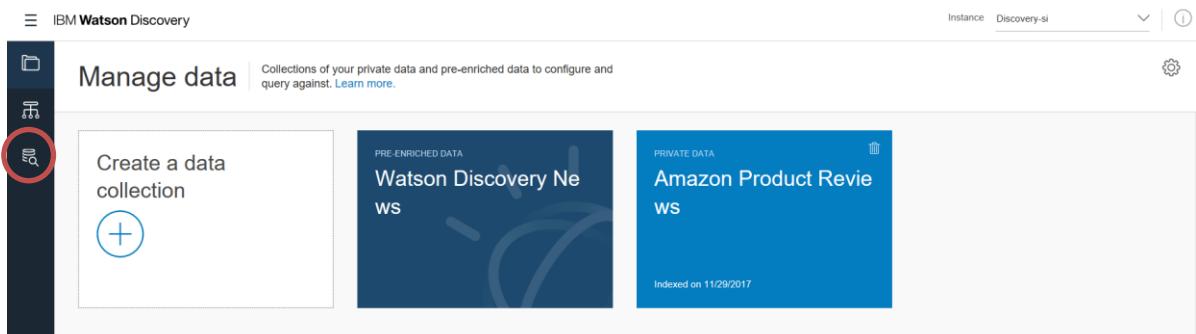
The relevance of natural language query results can be improved in Watson Discovery service with training. In order to train Watson, you'll need to provide example queries that are representative of queries your user enters and provide ratings to say which results for each query are relevant and not 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.

#### 7.1.1 Train Watson to Improve Results

We will train the results for the query "time" to help the model return results to "time" that relates to cooking times. We also will train the results to "problem" to see problems that reviewers have found. You can experiment and add more queries to train. Remember to rate them with both Relevant and Not relevant.

1. In order to train your model, you must head over to the "Build queries" page. To do so, click the "build queries" icon on the lefthand side bar.



2. Choose the "Amazon Product Reviews" collection to query from the dropdown menu and click "Get Started"

# Watson Discovery Service

IBM Watson Discovery

Build queries

Choose a collection to query

Amazon Product Reviews

Get started

- Click "Train Watson to improve results" in the top right corner to get to the relevancy training page

IBM Watson Discovery

Build queries

Collection: Amazon Product Reviews

Summary JSON Train Watson to improve results

Your results will appear here when you run a query.

Use a sample query

Search for documents

Include analysis of your results

Filter which documents you query

More options

Run query Close

<https://watson-discovery.bluemix.net/regions/us-south/services/5f411530-17c3-485b-b26d-7cd3d06ea9a7/environments/3ec76d35-94d7-4182->

- Click the "Add a natural language query" link that is in blue

IBM Watson Discovery

Build queries > Train Watson

Collection: Amazon Product Reviews

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

Queries (0)

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

+ Add a natural language query

## Watson Discovery Service

5. Click "Learn more" if you'd like view IBM Cloud documentation on improving result relevance with the tooling
6. Type in "time" into the query box and click the "Add" button

Build queries > Train Watson

Collection: Amazon Product Reviews

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

Queries (0)

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

7. Click "Rate results" on the bottom right hand side in blue

Build queries > Train Watson

Collection: Amazon Product Reviews

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

time

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

review707.json <a href="#">View document</a> Good solid design, time proven dependability.  <input checked="" type="radio"/> Relevant <input type="radio"/> Not relevant	review505.json <a href="#">View document</a> "... This is the first time I owned a pressure cooker. It's very easy to use and helps me cut down the time to cook beans, meat,... substantially..."  <input checked="" type="radio"/> Relevant <input type="radio"/> Not relevant	review623.json <a href="#">View document</a> "... Bought for my daughter and she has used it several times and is sooo glad to have it...she and her family have loved every thing she has cooked in it...she is very short on time and has found this to be a great time saver but still able to cook really super meals!!..."  <input checked="" type="radio"/> Relevant <input type="radio"/> Not relevant
--	--	---

8. Rate "Relevant" or "Not relevant" based on if the "time" mention is about cooking times

# Watson Discovery Service

IBM Watson Discovery      Instance: Discovery-si      Collection: Amazon Product Reviews

**Build queries > Train Watson**

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

**time**

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

<b>review707.json</b> <a href="#">View document</a> Good solid design, time proven dependability.  <input type="radio"/> Relevant <input type="radio"/> Not relevant	<b>review505.json</b> <a href="#">View document</a> "... This is the first time I owned a pressure cooker. It's very easy to use and helps me cut down the time to cook beans, meat... substantially..."  <input type="radio"/> Relevant <input type="radio"/> Not relevant	<b>review623.json</b> <a href="#">View document</a> "... Bought for my daughter and she has used it several times and is sooo glad to have it...she and her family have loved every thing she has cooked in it...she is very short on time and has found this to be a great time saver but still able to cook really super meals!..."  <input type="radio"/> Relevant <input type="radio"/> Not relevant
--	---	--

< 1 2 3 4 5 >

IBM Watson Discovery      Instance: Discovery-si      Collection: Amazon Product Reviews

**Build queries > Train Watson**

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

[Back to queries](#)

**time**

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

<b>review573.json</b> <a href="#">View document</a> "... "Cooking Times" may be a bit deceptive..." "... The so-called "cooking times" seem short and very fast, but they don't take into account the time it takes to build up the pressure, which is often longer than the actual stated cooking times..."  <input type="radio"/> Relevant <input type="radio"/> Not relevant	<b>review811.json</b> <a href="#">View document</a> but it's not meant for doing a whole house of hardwoods at one time.	<b>review498.json</b> <a href="#">View document</a> "... The handles came loose the first time I used it, so tightening them really good the second time seems to have done the trick..."  <input type="radio"/> Relevant <input type="radio"/> Not relevant
--	--	--

IBM Watson Discovery      Instance: Discovery-si      Collection: Amazon Product Reviews

**Build queries > Train Watson**

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

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

<b>review590.json</b> <a href="#">View document</a> "... One upon a time, many years ago.. I had a very bad experience with a pressure cooker. I swore at that time to never let one in my house again. Well, time passed and my patience with cooking grew more..."  <input type="radio"/> Relevant <input type="radio"/> Not relevant	<b>review567.json</b> <a href="#">View document</a> "... use all the time..."  <input type="radio"/> Relevant <input type="radio"/> Not relevant	<b>review747.json</b> <a href="#">View document</a> conventional cooking ... You not only gain the benefit of "speed" (about 1/3 the cooking time) of conventional cooking time) but little or nothing is lost as far as flavor and seasoning ... I think anyone who
---	--	--

< 2 3 4 5 6 >

## Watson Discovery Service

9. Click "back to queries" and then "add a natural language query"
10. Type in "problem" into the query box and click "Add" to the right of the box

The screenshot shows the 'Build queries > Train Watson' section of the IBM Watson Discovery interface. On the left is a sidebar with icons for folder, database, and search. The main area has a heading 'Build queries > Train Watson' and a note: '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'. Under the heading 'Queries (1)', there is a list with one item: 'problem'. To the right of the list are two buttons: 'Rate results' and '6 relevant 6 not relevant'. At the top right, it says 'Collection: Amazon Product Reviews'.

This screenshot shows the same 'Train Watson' section after a new query has been added. The 'Queries (2)' list now includes 'problem' and 'time'. The 'Rate results' button for 'problem' is highlighted, showing 'Not rated yet'. The other buttons remain the same: 'Rate results' and '6 relevant 6 not relevant'.

11. Click "Rate results" that corresponds to "problem" on the bottom right side

This screenshot shows the 'Train Watson' section after the 'problem' query has been rated. The 'Rate results' button for 'problem' is now greyed out and says 'Rated' with a checkmark. The other buttons remain the same: 'Rate results' and '6 relevant 6 not relevant'.

# Watson Discovery Service

Build queries > 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](#)

[Back to queries](#)

**problem**

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

review382.json	review199.json	review20.json
<a href="#">View document</a> I bought my ice cream maker years ago and have never had a <b>problem</b> . In fact, I bought one for my	<a href="#">View document</a> The box is really pretty, with its vintage style and its embossed scenes. The <b>problem</b>	<a href="#">View document</a> online. I had no <b>problem</b> with the recipes online working in the Zoku maker.
<input checked="" type="radio"/> Relevant <input type="radio"/> Not relevant	<input checked="" type="radio"/> Relevant <input type="radio"/> Not relevant	<input checked="" type="radio"/> Relevant <input type="radio"/> Not relevant

12. Rate “Relevant” or “Not relevant” based on if the “problem” mention is about an issue.

**problem**

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

review382.json	review199.json	review20.json
<a href="#">View document</a> I bought my ice cream maker years ago and have never had a <b>problem</b> . In fact, I bought one for my	<a href="#">View document</a> The box is really pretty, with its vintage style and its embossed scenes. The <b>problem</b>	<a href="#">View document</a> online. I had no <b>problem</b> with the recipes online working in the Zoku maker.
<input checked="" type="radio"/> Relevant <input type="radio"/> Not relevant	<input checked="" type="radio"/> Relevant <input type="radio"/> Not relevant	<input checked="" type="radio"/> Relevant <input type="radio"/> Not relevant

< 1 2 3 4 5 >

Review382.json mentions that the reviewer “never had a problem” and review20.json mentions that the review “had no problem” so both results should be rated Not relevant. Meanwhile review199 specifies a problem. To view the full review, click View document in blue under the review title.

**review199.json**

Relevant  Not relevant

The box is really pretty, with its vintage style and its embossed scenes. The problem is that the metal is extremely thin, so the box gets deformed incredibly easy, and the paint job scratches in seconds. So, it is wonderful to look at, but not at all resistant.

Click the X and continue on the next page to keep rating the results

# Watson Discovery Service

[Back to queries](#)

## problem

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

review512.json

[View document](#)

of the best, but I think they should address the handle **problem**.

Relevant

Not relevant

review184.json

[View document](#)

in permanent marker with no **problem** so the kids knew which cup was theirs. They are not BIG cups, but that is actually perfect for my students.

Relevant

Not relevant

review970.json

[View document](#)

to 3 minutes, after which i run cold water over the cooker to cool it down. The only **problem** i have

had is getting the lid to go on sometimes. I think it gives me **problems** if the black rubber gasket

. If you have a **problem** with the lid not shutting closed, then try rotating the black gasket

Relevant

Not relevant

## problem

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

review217.json

[View document](#)

"... Not for me. It's so small (of course) and it's really only good for nuts. It pulverizes other products. I returned it with no problem. ...."

Relevant

Not relevant

review378.json

[View document](#)

"... Great cooker. I have been using this one since about 2008. Very reliable, nice flat bottom that is very thick. I use it on a glass top stove without any problem. ...."

Relevant

Not relevant

review735.json

[View document](#)

It cooks very well, easy to use, zero **problem** and I use it regularly. Highly recommended. One

Relevant

Not relevant

< 1 2 3 4 5 >

[Back to queries](#)

## problem

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

review854.json

[View document](#)

a **problem** with spilling yet. This is the double set, so you get two of each color for a total of 12 cups.

Relevant

Not relevant

review489.json

[View document](#)

"... Have used this for several years now for the Christmas decor. Has worked perfectly without any problem. Would buy again without hesitation. ...."

Relevant

Not relevant

review888.json

[View document](#)

"... Cleaning takes a little more time than a regular maker, but they say it is well worth it for how fresh the coffee tastes. Just remember, .90% of people who post reviews are those who had problems. Most people without problems never think to bother Update - 12/9/08 - Apparently it finally died about 3 months ago, so it made it through 5 years of almost daily use. ...."

Relevant

Not relevant

< 2 3 4 5 6 >

## 13. Click "Back to queries"

The screenshot shows the 'Build queries > Train Watson' section of the IBM Watson Discovery interface. It features a sidebar with icons for file management and search. The main area displays three query cards:

- problem**: Watson will learn which are the best results for your queries after you've rated enough. Buttons: Add more queries, Rate more results, Add more variety to your ratings.
- time**: Train Watson by adding natural language queries and rating the results. Buttons: Rate results (4 relevant, 8 not relevant).
- time**: Train Watson by adding natural language queries and rating the results. Buttons: Rate results (6 relevant, 6 not relevant).

At the top right, there are buttons for Instance (Discovery-si), Collection (Amazon Product Reviews), and a refresh icon. A vertical scrollbar is visible on the right side of the main content area.