

Hands-on Workshop

IBM Developer Advocacy
i.carrada@ibm.com

Módulo: Discovery.

Agenda

1. Prework.....p. 2
2. Realizar el scrapping de los datosp. 3
3. Guardar el scrapper en un archivo de texto o .csv.....p. 4
4. Siguientes pasos.....p. 5

0. Prework:

Pre-requisitos:

1. Tener una cuenta en IBM Cloud
2. Tener instalado el CLI de IBM Cloud
3. Tener instalado GIT
4. Tener instalado NODE.JS
5. descargar o realizar un “fork” y un “clone” el siguiente repositorio:
<https://github.com/ibmdeveloperxm/watson-discovery>

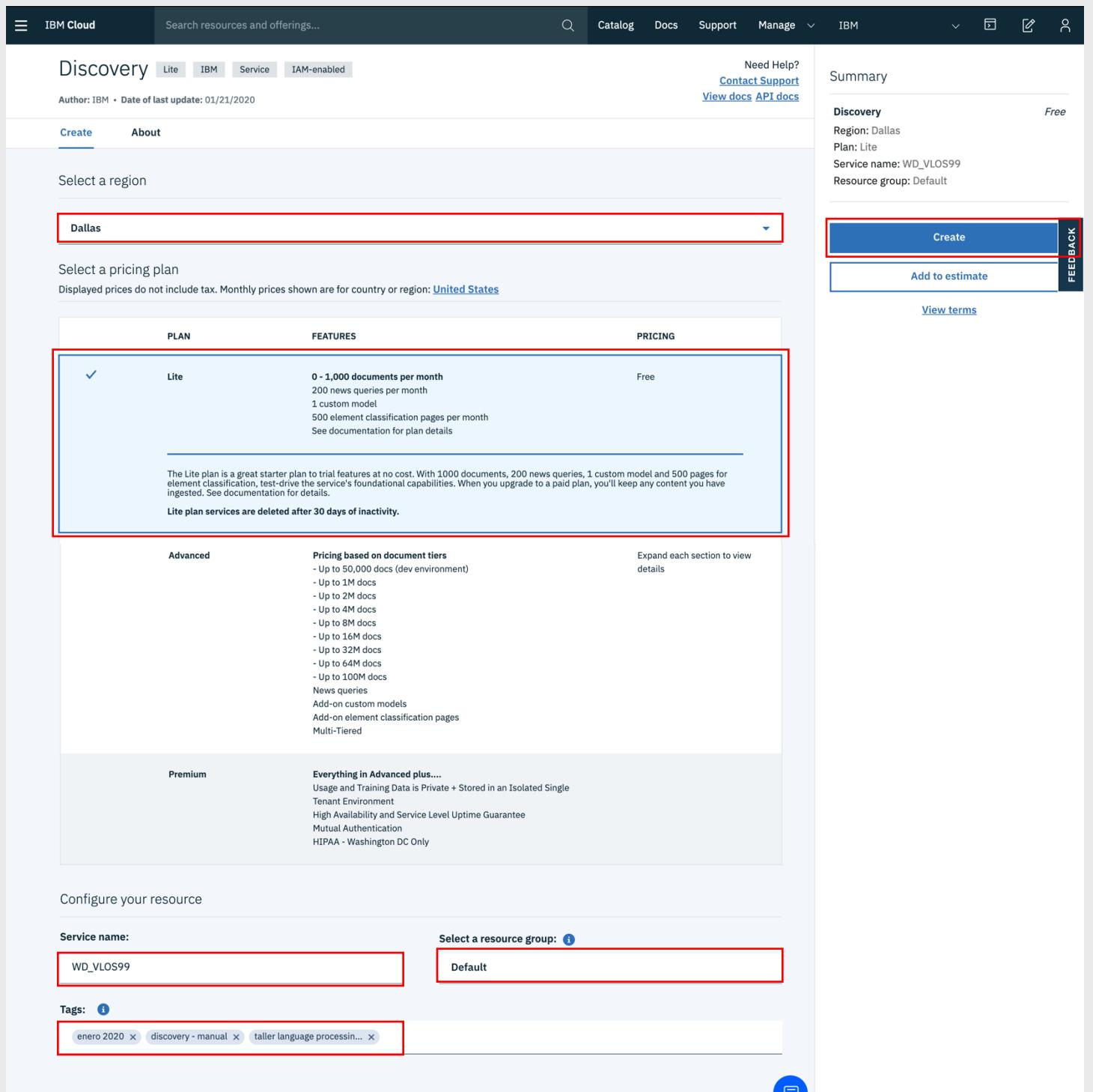
1. Crear una instancia en el servicio de Discovery:

1. En la parte superior derecha, damos clic en catalog o catálogo.

2. En el menú del lado izquierdo, selecciona la opción de AI y posteriormente, el servicio de Discovery.

Service	Description	Category
Watson Assistant	Watson Assistant lets you build conversational interfaces into any application, device, or channel.	AI
Watson Studio	Embed AI and machine learning into your business. Create custom models using your own data.	AI
Compare and Comply	Process governing documents to convert, identify, classify, and compare important elements.	AI
Discovery	Add a cognitive search and content analytics engine to applications.	AI
Knowledge Catalog	Discover, catalog, and securely share enterprise data.	AI
Knowledge Studio	Teach Watson the language of your domain.	AI

3. Le damos nombre al servicio, seleccionamos el plan de pago que querramos y damos clic en Create o Crear.



Discovery Lite IBM Service IAM-enabled

Author: IBM • Date of last update: 01/21/2020

[Need Help?](#) [Contact Support](#) [View docs](#) [API docs](#)

[Create](#) [About](#)

Select a region

Dallas

Select a pricing plan
Displayed prices do not include tax. Monthly prices shown are for country or region: [United States](#)

PLAN	FEATURES	PRICING
<input checked="" type="checkbox"/> Lite	0 - 1,000 documents per month 200 news queries per month 1 custom model 500 element classification pages per month See documentation for plan details	Free
The Lite plan is a great starter plan to trial features at no cost. With 1000 documents, 200 news queries, 1 custom model and 500 pages for element classification, test-drive the service's foundational capabilities. When you upgrade to a paid plan, you'll keep any content you have ingested. See documentation for details. Lite plan services are deleted after 30 days of inactivity.		
Advanced	Pricing based on document tiers - Up to 50,000 docs (dev environment) - Up to 1M docs - Up to 2M docs - Up to 4M docs - Up to 8M docs - Up to 16M docs - Up to 32M docs - Up to 64M docs - Up to 100M docs News queries Add-on custom models Add-on element classification pages Multi-Tiered	Expand each section to view details
Premium	Everything in Advanced plus... Usage and Training Data is Private + Stored in an Isolated Single Tenant Environment High Availability and Service Level Uptime Guarantee Mutual Authentication HIPAA - Washington DC Only	

Configure your resource

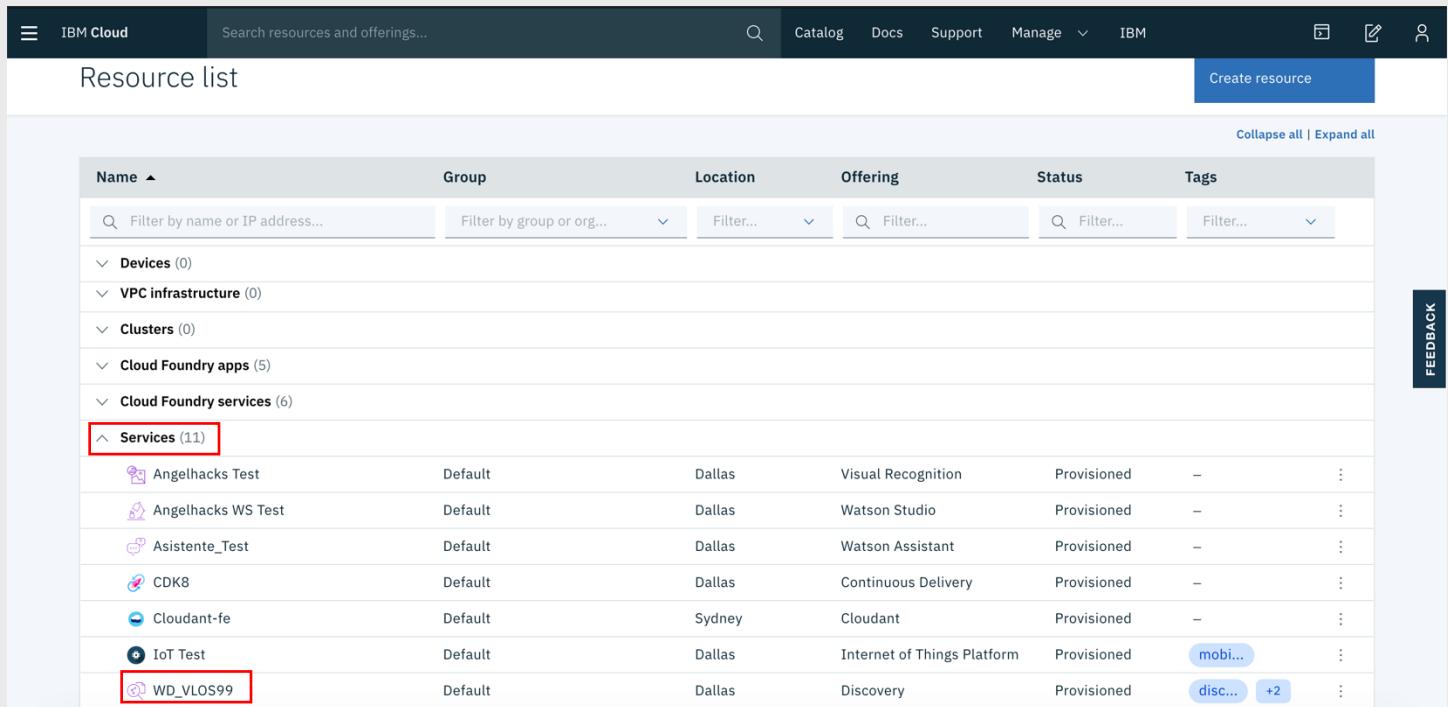
Service name:

Select a resource group:

Tags: enero 2020 discovery - manual taller language processin...

[Create](#)

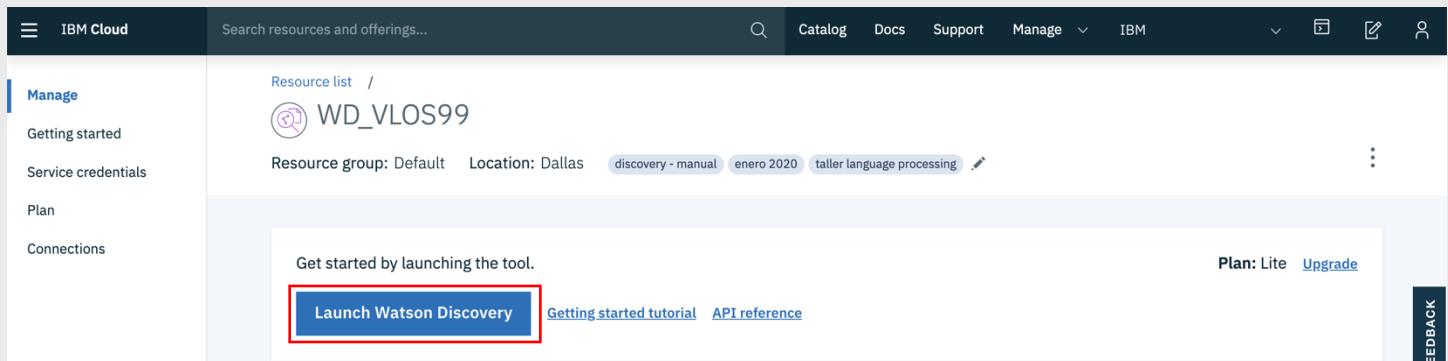
4. Una vez que nos encontramos en nuestro Dashboard o Panel de control, seleccionamos el servicio que acabamos de crear desde la pestaña de Servicios o Services.



The screenshot shows the IBM Cloud Resource list interface. At the top, there are search bars and navigation links for Catalog, Docs, Support, Manage, and IBM. A blue button labeled "Create resource" is visible. Below the header, a table lists resources under various categories: Devices (0), VPC infrastructure (0), Clusters (0), Cloud Foundry apps (5), Cloud Foundry services (6), and Services (11). The "Services" category is highlighted with a red box. Under "Services", there are eleven entries, each with a small icon, name, group, location, offering, status, and more details. The entry "WD_VLOS99" is also highlighted with a red box.

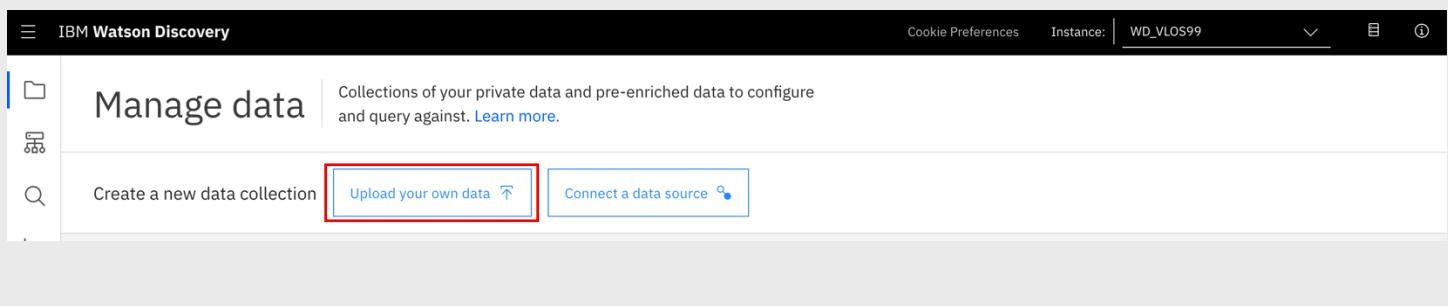
Name	Group	Location	Offering	Status	Tags
Devices (0)					
VPC infrastructure (0)					
Clusters (0)					
Cloud Foundry apps (5)					
Cloud Foundry services (6)					
Services (11)					
Angelhacks Test	Default	Dallas	Visual Recognition	Provisioned	-
Angelhacks WS Test	Default	Dallas	Watson Studio	Provisioned	-
Asistente_Test	Default	Dallas	Watson Assistant	Provisioned	-
CDK8	Default	Dallas	Continuous Delivery	Provisioned	-
Cloudant-fe	Default	Sydney	Cloudant	Provisioned	-
IoT Test	Default	Dallas	Internet of Things Platform	Provisioned	mobi...
WD_VLOS99	Default	Dallas	Discovery	Provisioned	disc... +2

5. Una vez dentro de la página del servicio, damos clic en Launch Watson Discovery o Iniciar Watson Discovery.



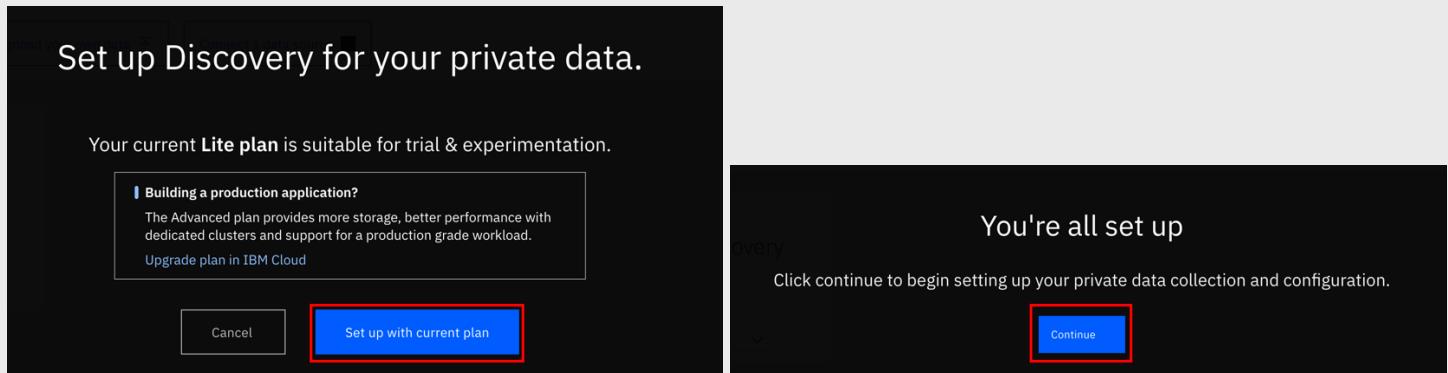
The screenshot shows the service page for "WD_VLOS99". The left sidebar has a "Manage" section with options: Getting started, Service credentials, Plan, and Connections. The main area shows the service name "WD_VLOS99", its resource group (Default), location (Dallas), and some tags. Below this, a message says "Get started by launching the tool." and features a large blue button labeled "Launch Watson Discovery" which is highlighted with a red box. There are also links for "Getting started tutorial" and "API reference".

6. Ahora crearemos una nueva colección dando clic en “Cargar tus propios datos” o “Upload your own data”.

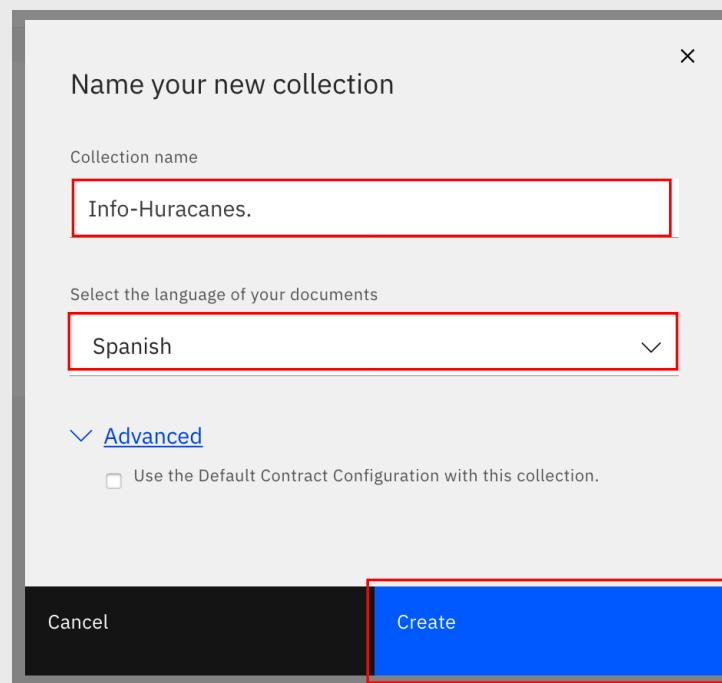


The screenshot shows the "Manage data" page of IBM Watson Discovery. On the left, there's a sidebar with icons for file, database, and search. The main area has a heading "Manage data" and a sub-instruction: "Collections of your private data and pre-enriched data to configure and query against. [Learn more.](#)". Below this, there are two buttons: "Create a new data collection" (highlighted with a red box) and "Upload your own data" (highlighted with a red box). Another button, "Connect a data source", is also visible.

7. Damos clic en “Configurar con el plan actual” o “Set up with current plan” y posteriormente en Continue o Continuar.



8. Le damos un nuevo nombre a nuestro servicio, seleccionamos el idioma español y damos clic en crear.



Name your new collection

Collection name

Info-Huracanes.

Select the language of your documents

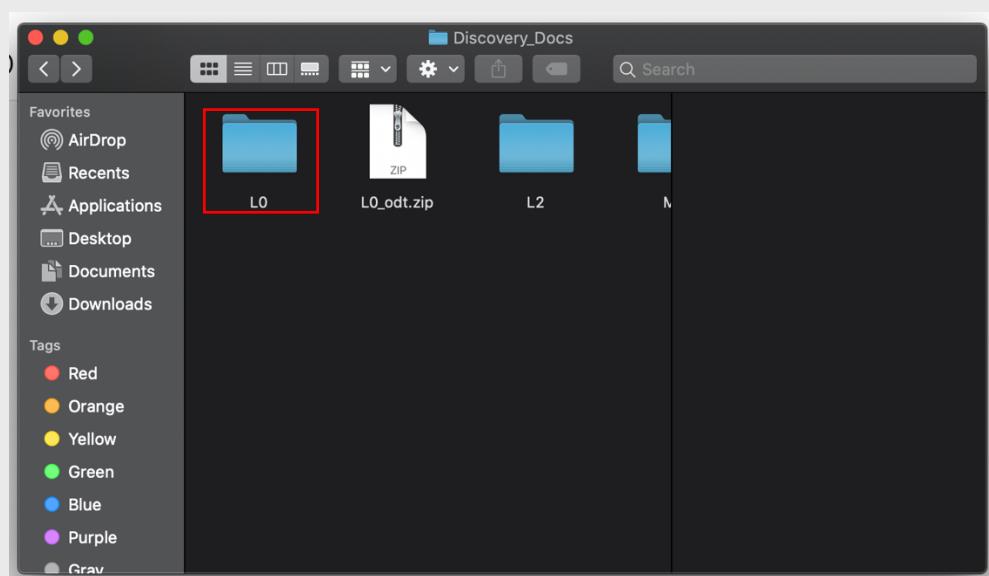
Spanish

Use the Default Contract Configuration with this collection.

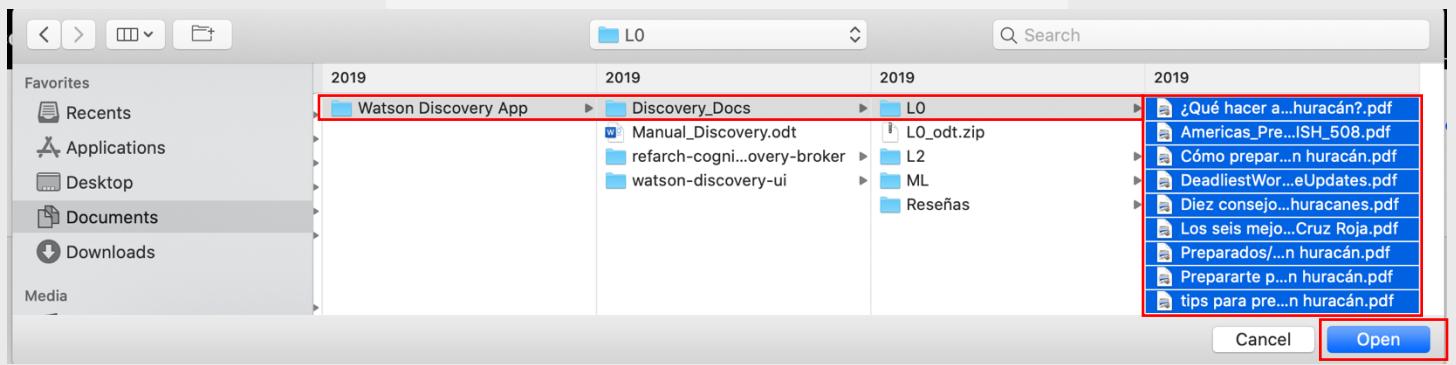
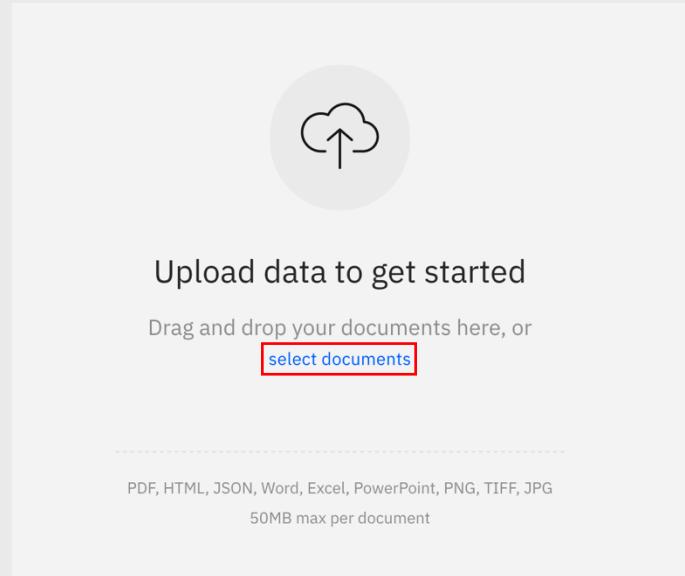
Cancel Create

2. Preparar datos y documentos.

1. Ingresa a la carpeta L0 del repositorio clonado. (Discovery_Docs/L0)



- Regresa al servicio de Discovery y arrastra y suelta los documentos de la carpeta antes mencionada a la parte que dice "Upload data to get started" o da clic en "Select documents" o "Seleccionar documentos".



- Una vez hecho esto, espera a que los datos sean procesados.

Reseñas ↗

Configure data ⚙️

API ↴

Overview Errors and warnings (1) Search settings

9

documents

● 0 documents failed
[View details](#)Created on
Last updated
1/27/2020 11:07:57 am EST
1/27/2020 11:07:57 am EST[Upload documents](#) ↴

Identified 1 field from your data

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

Added 4 enrichments to your data

Entity Extraction

huracán (7) | Los (4) |
Estados Unidos (3) | los (3) |
24 horas (2)

Sentiment Analysis

56% positive
0% neutral
44% negativeNow you're ready to **query!**Most common entity types
and their top entities[Run](#)Top entities with their
average, min, max
sentiment score[Run](#)Top people related to
/science/weather
/meteorological
/disaster/hurricane[Run](#)5 enrichments available. [Add enrichments](#)[Build your own query →](#)

3. Realizar una consulta a la base de datos

1. Realizar una consulta simple.

- a) Da clic en “Build your own query” o “Crea tu propia consulta”.

IBM Watson Discovery

Overview Errors and warnings (1) Search settings

9 documents

0 documents failed View details

Created on 1/27/2020 11:07:57 am EST Last updated 1/27/2020 11:07:57 am EST

Upload documents

Identified 1 field from your data

text

Need to identify more fields? Add fields

Added 4 enrichments to your data

- Entity Extraction
 - huracán (7) | Los (4)
 - Estados Unidos (3) | los (3)
 - 24 horas (2)
- Sentiment Analysis

56%	0%	44%
positive	neutral	negative
- Concept Tagging
 - Ciclón tropical (4)
 - Estados Unidos (3) | Viento (3)
 - Agencia Federal para el Ma... (2)
- Category Classification
 - science → weather → meteorological disasters... hurricane

Now you're ready to query!

Most common entity types and their top entities

Top entities with their average, min, max sentiment score

Top people related to /science/weather /meteorological disaster/hurricane

Build your own query →

- b) Da clic en “More Options” o “Más opciones” para desplegar los demás campos.

Reseñas / Build queries

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

- c) En Passages o Pasajes, seleccionamos la opción no en el apartado “Include relevant passages” o “Incluir pasajes relevantes”.
- d) Da clic en “Run query” o “Realizar consulta”

▽ [More options](#)

Passages

Include relevant passages

Yes No

Documents

Fields to return

All Specify

Number of documents to return

10

Default: 10

Number of query results to skip at the beginning

0

This option is for pagination

[Run query](#)

[Close](#)

De esta forma, lograrás ver en el documento JSON que se genera de la consulta, los diferentes campos, como entidades (entities), conceptos (concepts) y sentimientos (sentiment) de la colección de datos que se consultó.

4. Crear una consulta.

Este paso puedes realizarlo en un editor de textos.

Primero necesitamos escribir: **enriched text**, seguido de un punto.

Después escribimos el nombre del objeto que deseemos buscar.(entities, concepts, sentiment, keywords). En este caso, buscaremos en **entities:** Seguido de dos puntos.

Ahora escribimos el nombre del valor que se busca, entre paréntesis (text, label, type). En este caso, utilizaremos **text:** seguido de dos puntos.

Finalmente, después de los dos puntos, escribimos el valor que deseamos buscar. Para esta ocasión, utilizaremos **tormenta** y cerramos paréntesis.

La consulta entonces debería quedar así:

enriched text.entities:(text:tormenta)

5. Realizar una consulta.

Ahora vamos a probar la consulta que acabamos de crear.

1. Primero debemos crear una nueva consulta con “Build your own query” o “Construye tu propia consulta”. Desplegamos el apartado “Search for documents” o “Buscar documentos”. Damos clic en la pestaña “Use the Discovery Query language” o “Usar el lenguaje de consultas de Discovery”. Podemos crear la consulta directamente como se muestra en la imagen o hacerlo de forma manual con el lenguaje de consultas de Discovery.

Reseñas [🔗](#)

[Configure data](#) [API](#) [Delete](#)

Overview Errors and warnings (1) Search settings

9 documents

0 documents failed [View details](#)

Created on 1/27/2020 11:07:57 am EST
Last updated 1/27/2020 11:07:57 am EST

[Upload documents](#)

Identified 1 field from your data: **text**
Need to identify more fields? [Add fields](#)

Added 4 enrichments to your data:

- Entity Extraction:** huracán (7) | Los (4) | Estados Unidos (3) | los (3) | 24 horas (2)
- Sentiment Analysis:** 56% positive, 0% neutral, 44% negative
- Concept Tagging:** Ciclón tropical (4) | Estados Unidos (3) | Viento (3) | Agencia Federal para el Ma... (2)
- Category Classification:** science → weather → meteorological disasters... hurricane

Now you're ready to **query!**

Most common entity types and their top entities [Run](#)

Top entities with their average, min, max sentiment score [Run](#)

Top people related to /science/weather /meteorological disaster/hurricane [Run](#)

[Build your own query →](#)

Search for documents [Edit in query language](#)

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

Satisfy [all](#) of the following rules

Field	Operator	Value
enriched_text.entities.text	is	tormenta

+ Add rule
+ Add group of rules

```
enriched_text.entities.text::"tormenta"
```

- Para realizar la consulta manual (como la armamos en el [paso 4](#)), damos clic en “Edit in Query Language” o “Editar en lenguaje de consultas”.
- En la caja de texto que dice “Enter query here” o “Ingresa consulta aquí”, escribimos o pegamos la consulta que creamos en el paso anterior y damos clic en “Run query” o “Generar consulta”.

Search for documents

Build in visual mode

Use natural language Use the Discovery Query Language

enriched_text.entities.text::"tormenta"



4. Al realizar la consulta, nos dará como retorno la serie de documentos que cumplan con los parámetros de la consulta.

The screenshot shows two views of the Watson Discovery interface. On the left, the 'Summary' view displays a list of 9 matching documents with titles like 'Cómo prepararse ante un huracán.pdf', 'Diez consejos para la temporada de huracanes.pdf', and 'Prepararte para un huracán.pdf'. On the right, the 'JSON' view shows the raw API response. The JSON object contains a 'matching_results' array with 9 items, each representing a document. One item is expanded to show its full text, which discusses hurricane preparedness, including tips like 'Tome medidas básicas ahora para garantizar su seguridad en caso de que una tormenta azote su región', 'Abasteza su casa y su auto con suministros', and 'Evite las áreas inundadas'.

```

{
  "matching_results": [
    {
      "id": "3cb2e20e19df49ee2d5a168a89a75331",
      "result_metadata": {...},
      "extracted_metadata": {...},
      "text": "Cómo prepararse para un huracán o una tormenta tropical. Usted no puede detener una tormenta tropical o un huracán, pero puede tomar medidas ahora para protegerse y proteger a su familia. Si usted vive en un área en riesgo de ser afectada, los Centros para el Control y la Prevención de Enfermedades (CDC) le aconsejan que se empiece a preparar para la temporada de huracanes. La temporada de huracanes en el Atlántico se extiende desde el 1 de junio hasta el 30 de noviembre de cada año. Por favor siga estos consejos importantes que los CDC ofrecen para prepararse para un huracán: • Preparación para un huracán: Tome medidas básicas ahora para garantizar su seguridad en caso de que una tormenta azote su región. • Suministros de emergencia que necesitará: Abasteza su casa y su auto con suministros. • Elabore un plan: Cree un plan familiar para casos de desastre. • Evite las áreas inundadas: Tome medidas de precaución antes, durante y después de una inundación. • Prepárese para evacuar: Nunca ignore una orden de ...",
      "enriched_text": {
        "sentiment": {...},
        "entities": [...],
        "concepts": [...],
        "categories": [...]
      }
    },
    ...
  ],
  "id": "6ccc8143547409a5c4ee13539f932a13",
  "result_metadata": {...},
  "extracted_metadata": {...},
  "text": "10 Consejos para la Temporada de Huracanes Conversa con tu familia y distribuyan las responsabilidades de preparación para un huracán. Determina si las ventanas, puertas y techos son seguros para resistir la fuerza del viento y del agua. Haz los"
}

```

5. Desplegar una app que consuma el servicio de Watson Discovery.

1. Ingresamos a la página principal de nuestro servicio de discovery y damos clic en “Upload your own data” o “Cargar tus propios datos”, para crear una nueva colección.

IBM Watson Discovery

Manage data

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

Create a new data collection

[Upload your own data](#)

[Connect a data source](#)

PRE-ENRICHED DATA

Watson Discovery News

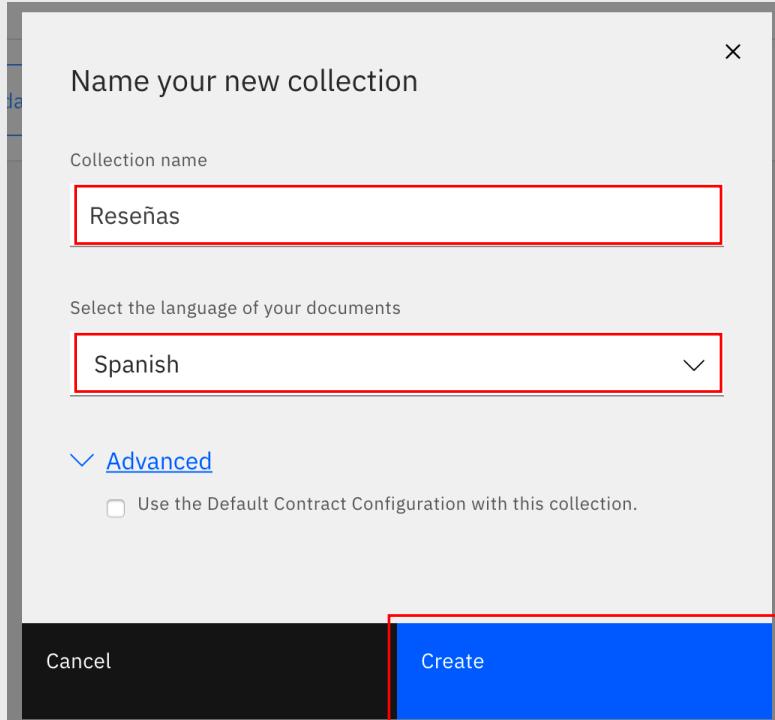
News sources: English

PRIVATE DATA

Reseñas

Indexed on 1/27/2020

- Le damos el nombre de “Reseñas”, seleccionamos el idioma “Español” y damos clic en “Create” o “Crear”.



- Vamos a configurar la colección. En la parte superior derecha, damos clic en “Configure Data” o “Configurar documentos”.

IBM Watson Discovery

Reseñas

[Configure data](#)

- Vamos a la pestaña de “Enrich Fields” o “Enriquecer campos” y debajo de “Enrichments” damos clic en “Add Enrichments”.

Reseñas / Configure data

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

Enrich your data with additional Watson insights

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

Add a field to enrich ▾

Fields to be enriched	Enrichments
text	categories × concepts × entities × sentiment ×

+ Add enrichments

5. Agregamos el campo “Keyword extraction” o “Extracción de palabras clave” dando clic en “Add” o “Agregar”.

Add Enrichments

x

text: categories × concepts × entities × sentiment ×

Keyword Extraction

Determines important keywords in this field, ranks them, and optionally detects the sentiment.

[Learn more](#) **Add**

Sentiment Analysis

Identifies the overall positive or negative sentiment within this field.

[Learn more](#) **Added!**

6. Cerramos el apartado para agregar enriquecimientos y damos clic en “Apply changes to the collection” o “Aplicar cambios a la colección”.

IBM Watson Discovery

Cookie Preferences Instance: WD_VLOS99

Manage data Reseñas / Configure data

View data schema Identify fields Manage fields **Enrich fields**

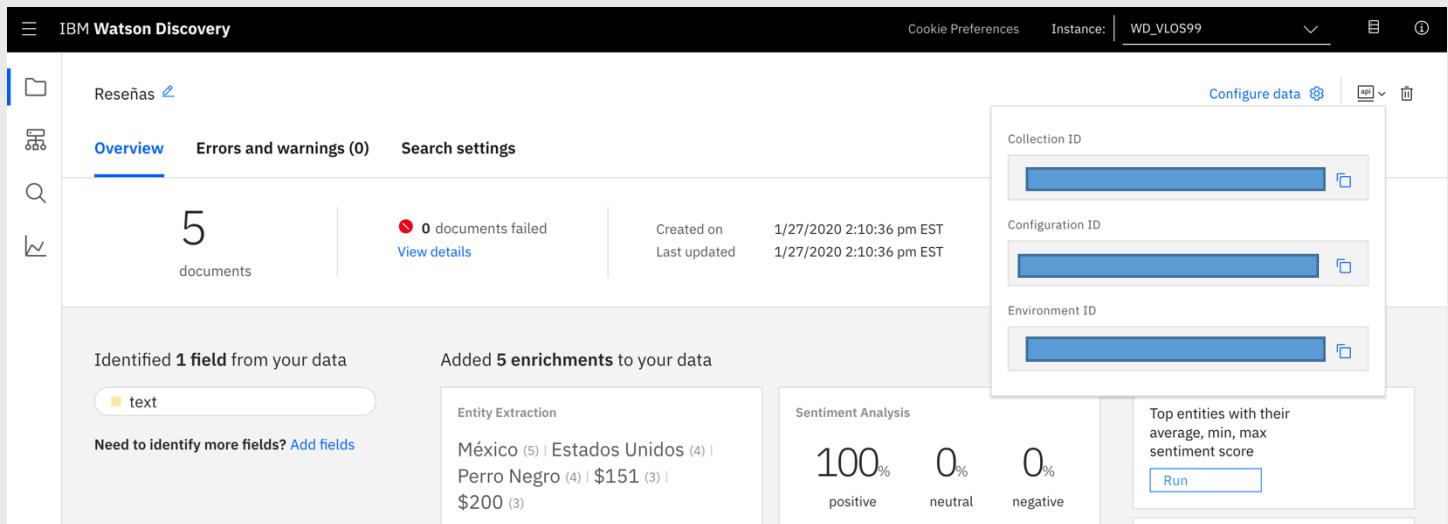
Apply changes to collection

7. Ahora cargamos los documentos necesarios a nuestra colección.
De la carpeta Seleccionamos todos los archivos que se encuentran en la carpeta de nuestro

repositorio(watson-discovery-master/Discovery_docs/Reseñas), de la misma forma que hicimos en el [paso 2](#).

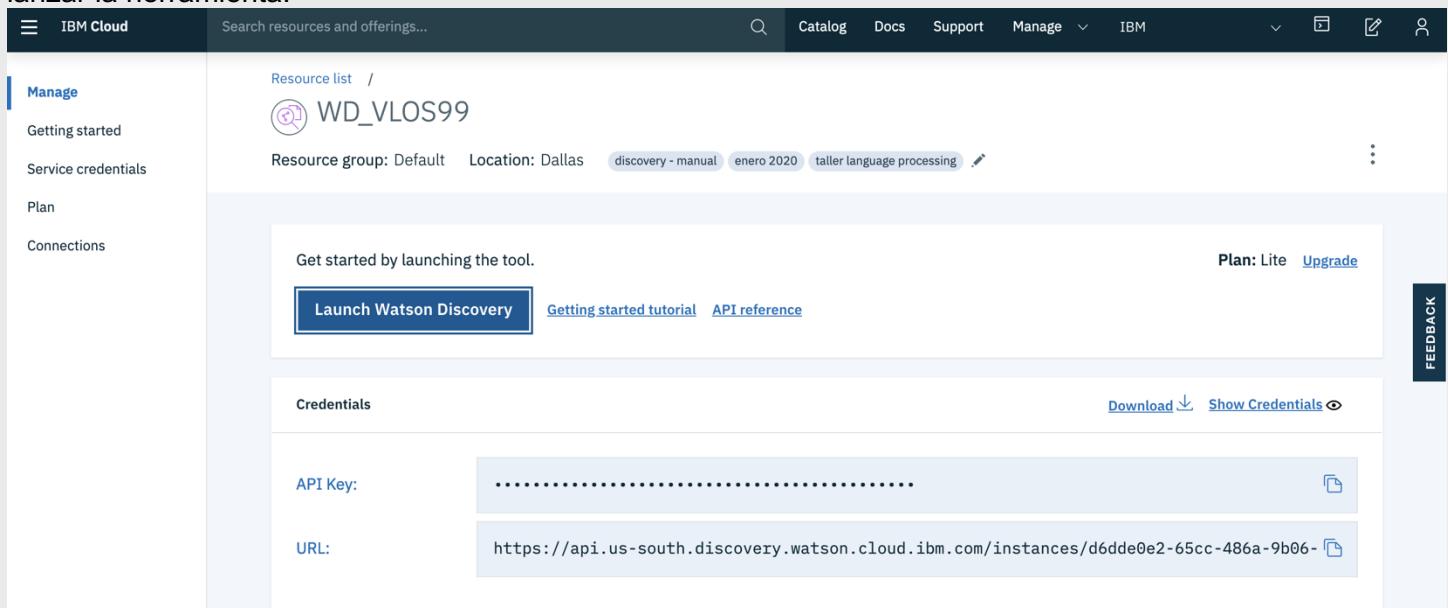
Una vez que hemos cargado los documentos, procederemos a modificar el archivo .env que se encuentra en la carpeta de nuestro repositorio, agregando las credenciales correspondientes a nuestra colección.

- En el apartado superior derecho de la página principal de nuestra colección, encontraremos un ícono que dice “API”, al poner el mouse sobre él, nos muestra un mensaje que dice “View API details” o “Ver detalles de API”. Le damos clic para copiar cada una de las credenciales.



The screenshot shows the IBM Watson Discovery interface. At the top, it displays "Collection ID: WD_VLOS99". Below this, there are sections for "Entity Extraction" (showing Mexico (5) | Estados Unidos (4), Perro Negro (4) | \$151 (3), \$200 (3)), "Sentiment Analysis" (100% positive, 0% neutral, 0% negative), and "Top entities with their average, min, max sentiment score" (with a "Run" button). On the left, a sidebar shows "Reseñas" with 5 documents, and tabs for "Overview", "Errors and warnings (0)", and "Search settings".

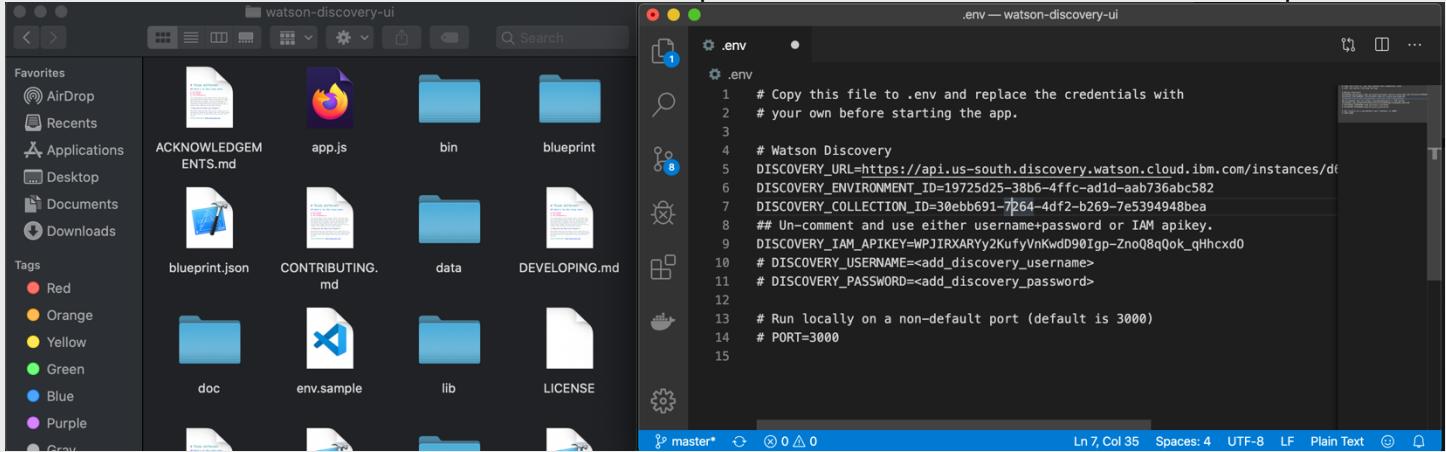
*Para obtener el apikey de Discovery y la URL, accedemos a la página del servicio de discovery antes de lanzar la herramienta.



The screenshot shows the IBM Cloud Resource list. It displays a resource named "WD_VLOS99" under the "Manage" category. The resource group is Default, and the location is Dallas. It includes links for "discovery - manual", "enero 2020", and "taller language processing". A "Launch Watson Discovery" button is present. Below this, there are sections for "Credentials" (with download and show credentials links) and "API Key:" (with a redacted value). The "URL:" is listed as <https://api.us-south.discovery.watson.cloud.ibm.com/instances/d6dde0e2-65cc-486a-9b06->.

- Accedemos a la carpeta de nuestro repositorio clonado “watson-discovery-ui” y accedemos al archivo “env.sample” y compiamos todo el contenido de texto.

10. Abrimos un nuevo documento de texto pegamos lo que copiamos del archivo anterior y modificamos las credenciales, agregando las que copiamos de nuestra colección de datos.
11. Una vez que modificamos las credenciales pertinentes, guardamos el archivo como ".env" sin ningún carácter adicional, solo ".env" en la misma carpeta donde se encuentra el archivo "env.sample"



12. Para el siguiente paso es necesario tener instalado NODE.JS o NPM del CLI de IBM Cloud, si no lo has hecho, al principio de este manual, se encuentran los enlaces necesarios.
13. Ahora vamos a desplegar la aplicación utilizando el comando "npm install" seguido de "npm start", claro que esto debe hacerse desde la carpeta de nuestro repositorio.

```
viridianas-mbp:Watson Discovery App viridianasalvatierra$ cd watson-discovery-ui
[viridianas-mbp:watson-discovery-ui viridianasalvatierra$ npm install
npm WARN babel-loader@8.0.5 requires a peer of webpack@>=2 but none is installed. You must install peer dependencies yourself.

audited 972415 packages in 6.305s
found 18692 vulnerabilities (2 low, 5 moderate, 18685 high)
  run `npm audit fix` to fix them, or `npm audit` for details

New minor version of npm available! 6.12.0 > 6.13.6
Changelog: https://github.com/npm/cli/releases/tag/v6.13.6
Run npm install -g npm to update!

[viridianas-mbp:watson-discovery-ui viridianasalvatierra$ npm start

> watson-discovery-ui@1.0.0 start /Users/viridianasalvatierra/Documents/Workshops/Watson Discovery App/watson-discovery-ui
> node --max_old_space_size=256 app.js

Found Discovery environment using DISCOVERY_ENVIRONMENT_ID.
{
  environment_id: '19725d25-38b6-4ffc-ad1d-aab736abc582',
  name: 'byod',
  description: '',
  created: '2020-01-23T18:49:01.772Z',
  updated: '2020-01-23T18:49:01.772Z',
```

```

}
createDiscoveryEnvironment - 19725d25-38b6-4ffc-ad1d-aab736abc582
findDiscoveryCollection
Found Discovery collection using DISCOVERY_COLLECTION_ID.
{
  collection_id: '30ebb691-7264-4df2-b269-7e5394948bea',
  name: 'Reseñas',
  configuration_id: 'd37baaa81-6bc4-4543-ac24-3c7cc43db03f',
  language: 'es',
  status: 'active',
  description: null,
  created: '2020-01-27T19:10:36.739Z',
  updated: '2020-01-27T19:10:36.739Z'
}
Created configuration with id: a6e72943-3d56-48d5-b082-54315c434245
Discovery is ready!
environment_id: 19725d25-38b6-4ffc-ad1d-aab736abc582
collection_id: 30ebb691-7264-4df2-b269-7e5394948bea
Begin loading 300 json files into discovery. Please be patient as this can take several minutes.
Get collection to check its status.
isDev: true
Watson Discovery UI Server running on port: 3000
Browserslist: caniuse-lite is outdated. Please run next command `yarn upgrade caniuse-lite browserslist`
Checking status of Discovery collection: { available: 5, processing: 0, failed: 0, pending: 0 }
Collection is already loaded with docs.
Discovery collection loading has completed!

```

14. Una vez hecho esto, en el navegador, introducimos “localhost:3000” en la barra de direcciones y damos “enter” para desplegar la aplicación en el navegador.

Reseñas de la pizzería "Las Pizzas del Perro Negro"

Enter search string...

 Búsqueda de lenguaje natural
 Búsqueda de pasajes
 Limitar a 100 resultados

Filtro
Por Lista

- Entidades**
- México (5)
- Estados Unidos (4)
- Perro Negro (4)
- \$151 (3)
- \$200 (3)
- 107 \$ (3)
- Mexico City (3)
- Pizza del (3)
- Trip De Marcia H. 29 (3)
- \$275 (2)

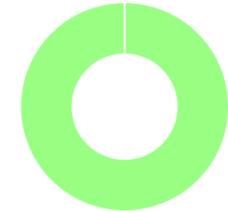
- Categorías**
- Conceptos
- Palabras clave
- Tipos de Entidades

Resultados

5 RESEÑAS	5 POSITIVAS	0 NEUTRALES	0 NEGATIVAS
Highest Score ▾			
Sin Título Las cookies permiten ofrecer nuestros servicios. Al utilizar nuestro servicio aceptas el uso de nuestras cookies. Más información Acceder Buscar Cerca de Madrid Restaurantes y bares Hogar Servicios de más... Puntaje: 1.0000 Fecha: Sentimiento: 0.59			
Sin Título Acceder Buscar Cerca de Ciudad de México, CDMX Restaurantes Hogar Servicios para autos Más Escribir una reseña Para negocios Reseñas recomendadas de Pizza del Perro Negro Tu confianza es nuestra mayor más... Puntaje: 1.0000 Fecha: Sentimiento: 0.50			
Sin Título Las cookies permiten ofrecer nuestros servicios. Al utilizar nuestro servicio aceptas el uso de nuestras cookies. Más información Acceder Buscar Cerca de Madrid Restaurantes y bares Hogar Servicios de más... Puntaje: 1.0000 Fecha: Sentimiento: 0.35			

Cuadro de sentimientos
Puntajes de sentimientos por porcentajes.

Entities ▾ All Terms ▾



% Positivo % Neutral % Negativo

Podemos darnos cuenta de que algunos conceptos aún se manejan en el idioma inglés. Esto es porque son etiquetas dinámicas que se generan desde el servicio de Discovery.

Siguientes Pasos

Te invitamos a explorar otros talleres y manuales en el siguiente enlace de Github.

<https://github.com/ibmdevelopermx>