PROJECT REPORT

ON

"AI POWERED NEWS SEARCH APP"

By

Girish Allam

Contents:

1. INTRODUCTION

The Internet is home to huge amounts of data, and more is being created everyday.Organizations can harness this constant stream of information to gain understanding, plan strategies, and find opportunities. Enriched news data can help your application make dynamic connections across current events faster.A standard search for a site can return too many results for someone to want to go through. However, it's possible to quickly build out a search interface that is not only efficient in returning queries, but also provides the latest news regarding the query that is being searched.

1.1 OVERVIEW

The basic idea of the app is to create a simple UI that interacts with the user and provides the latest news on the topic that is queried by the user.The interface makes use of the Watson Discovery Service to return the news feeds along with the details of the author and the article it is being taken from.

We also perform the sentimental analysis of the news feeds and integrate the app to slack so that we can directly interact with the slack bot to get the required news information.

1.2 PURPOSE

In order to get any information regarding a topic, we search the web to get our answers.Many organizations all around the world need useful data to plan their strategies and find opportunities.So, enriched news data plays a vital role in the efficient functioning of the organizations to gain a good understanding of the current affairs and build dynamic connections for providing better services.

This creates a need for an enriched news mining web application to provide the user with the latest news regarding any topic.For this purpose, we make use of the pre-built Watson Discovery news collection that discovers trends and patterns and provides latest information about product and brand perception, events and more.

2. LITERATURE SURVEY

The main idea behind creating this project is to minimize the problems that are prevalent in many news mining applications like retrieving irrelevant data and providing information that is not up to date.The web crawlers downloads pages that do not update their database from time to time, as a result it provides outdated information to the users.The Watson Discovery Service is a solution to all these problems as it is updated continously with new articles from various sources.It also uses techniques like news alerting and event detection to provide the trending news in any field that is being searched.Thus, we can track recent historical trends and stay abreast of latest information about various products, events and many more.

2.1 EXISTING PROBLEM

The main problems with the existing web mining applicatons are:

1. Many web mining applications do not provide results that are relevant to the subject being searched.Most of them use personalized searches that raise privacy concerns.
2. The next problem is that the users do not specify the exact search words that are required, thereby resulting in poor queries.
3. The applications that are presently used are poor in performance.They are quite inefficient in returning the results, thus decreasing the precision and recall of the data that is being retrieved.

These are few of the current problems that hinder the applications in providing better user experience and create better opportunities to make connections across current events in a faster way.

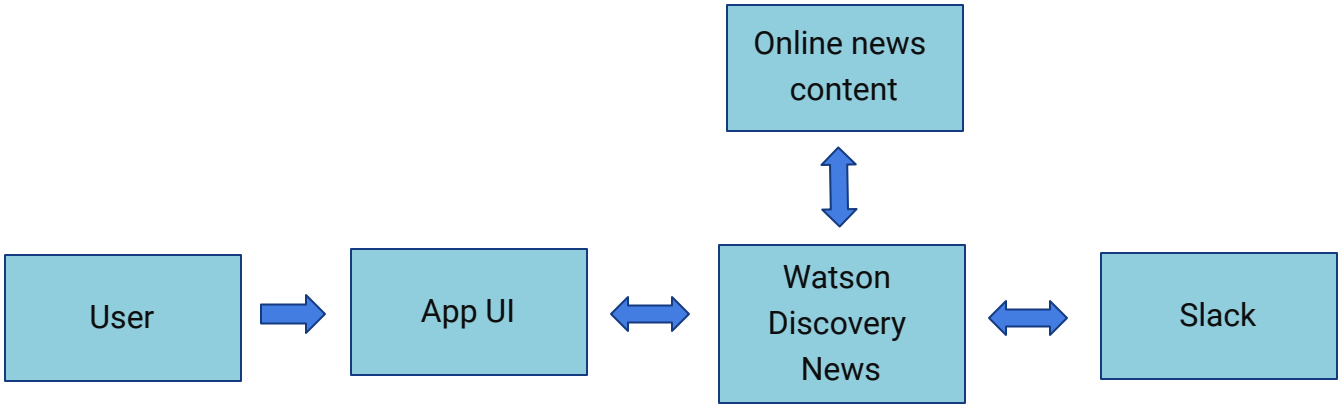
2.2 PROPOSED SOLUTION

In the proposed solution, we are building a web mining application using Node-Red and Watson Discovery Service.The user enters the query, then the Watson Discovery News collection searches for the trending news across millions of articles and stories regarding that particular topic and returns the latest news.We build the server side

application using Node Red by making use of the configuration nodes and embedding them in the flows.Then, we use the pre-built Watson Discovery News collection and access the Watson Discovery Service through the Discovery API to retrieve the result and also perform sentimental analysis on the data to gain an overview of the wider public opinion behind certain topics.We finally integrate the Watson discovery service with slack, and query the data using slack bot.

3. THEORETICAL ANALYSIS

3.1 BLOCK DIAGRAM



3.2 HARDWARE / SOFTWARE DESIGNING

There are few hardware and software requirements that are necessary for the exection of the project.

Hardware Requirements:

1. Processor: Minimum 1GHz.Recommended 2GHz or more.
2. Ethernet Connection (LAN) or Wireless adapter (WiFi)
3. Memory (RAM): Minimum 3GB.Recommended 4GB or more.

Software Requirements:

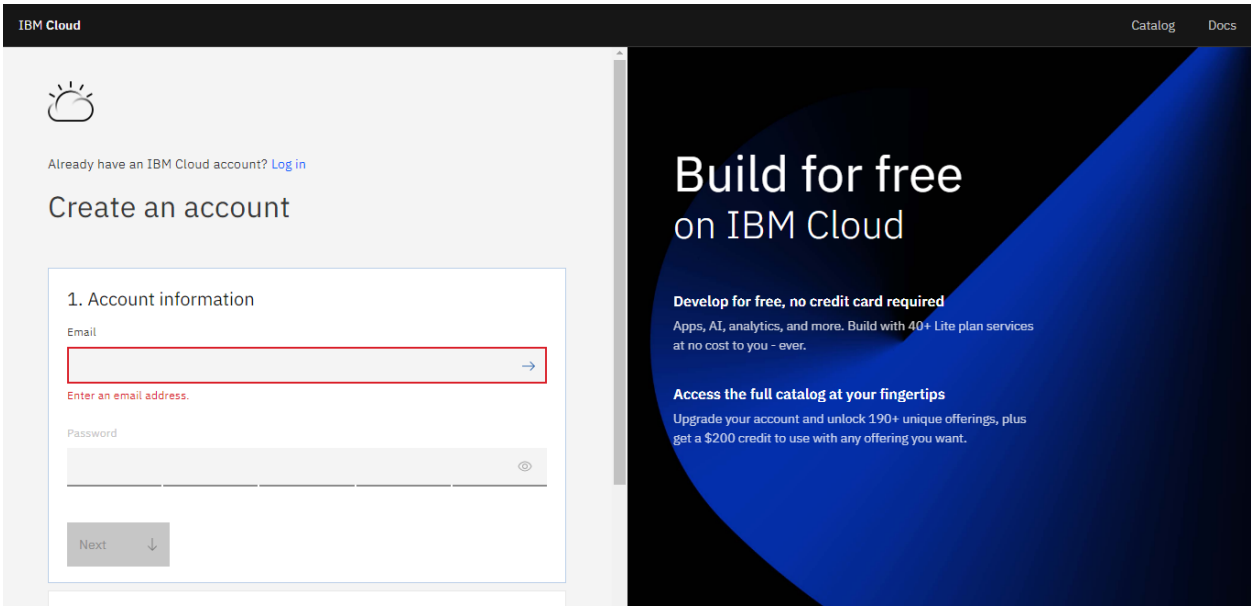
1. Recommended OS: Windows, MAC or Linux
2. Git control system
3. Node.js runtime environment
4. Tools like Slack and Node-Red

4. EXPERIMENTAL INVESTIGATIONS

To create our own News Mining Web Application, we initially create an IBM Cloud account and provide the cloudant details to create a cloud foundry app.We then create a Watson Discovery Service and then choose the default Watson Discovery News Collection to be used with our application.We can even create a custom collection by providing our own dataset.And, then we launch the discovery service to make use of the collection.Now, to create the UI, we use a flow based development tool called Node-Red that is used to create various flows by including a wide range of nodes present in thee palette.In this tool, we create flow using the discovery node to query the data and provide the output.We also perform sentimental analysis on the retrieved information to find out the latest trends.Later, we integrate our application with Slack wherein, we can make conversation with the slack bot and also get information from it.The slack bot listens to our messages and also responds to it by providing suitable replies.Thus, not only making it interactive but also interesting.

5. FLOWCHART

Step-1: Initially, create an IBM Cloud Account by filling out the details and complete the registration by providing your personal details and signing in.

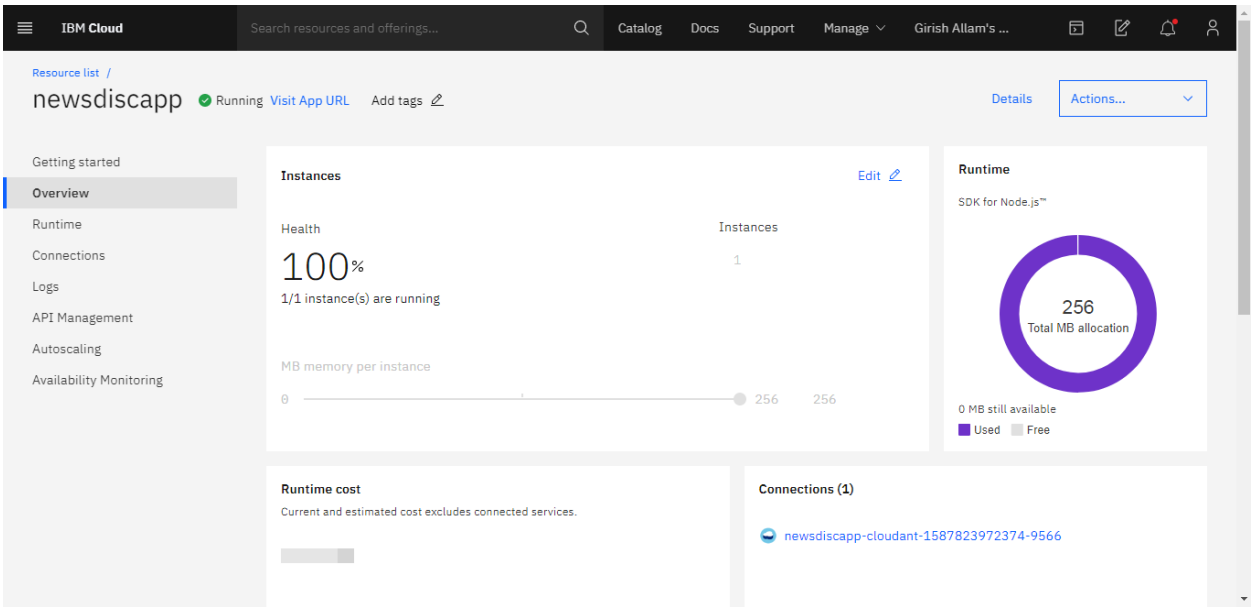


Link: https://cloud.ibm.com/registration

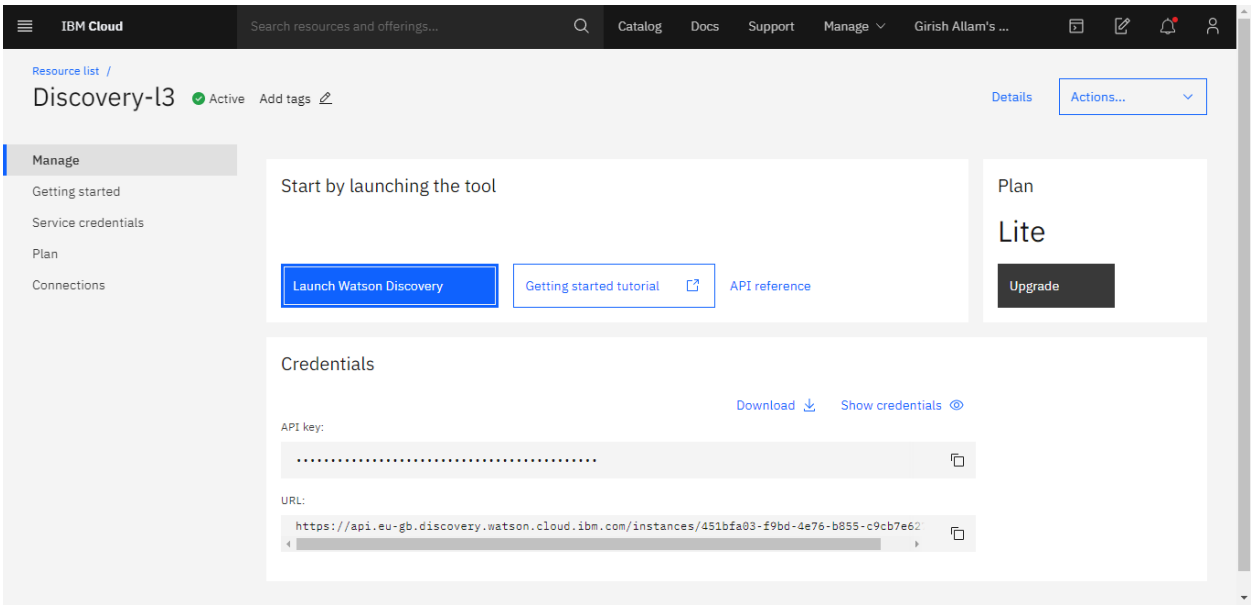
Step-2: Later, create an IBM cloud app by giving the details of the app and also the server details like location etc. to get started with your custom application.

After the app is created, you can find it under the resource list and access it from there.

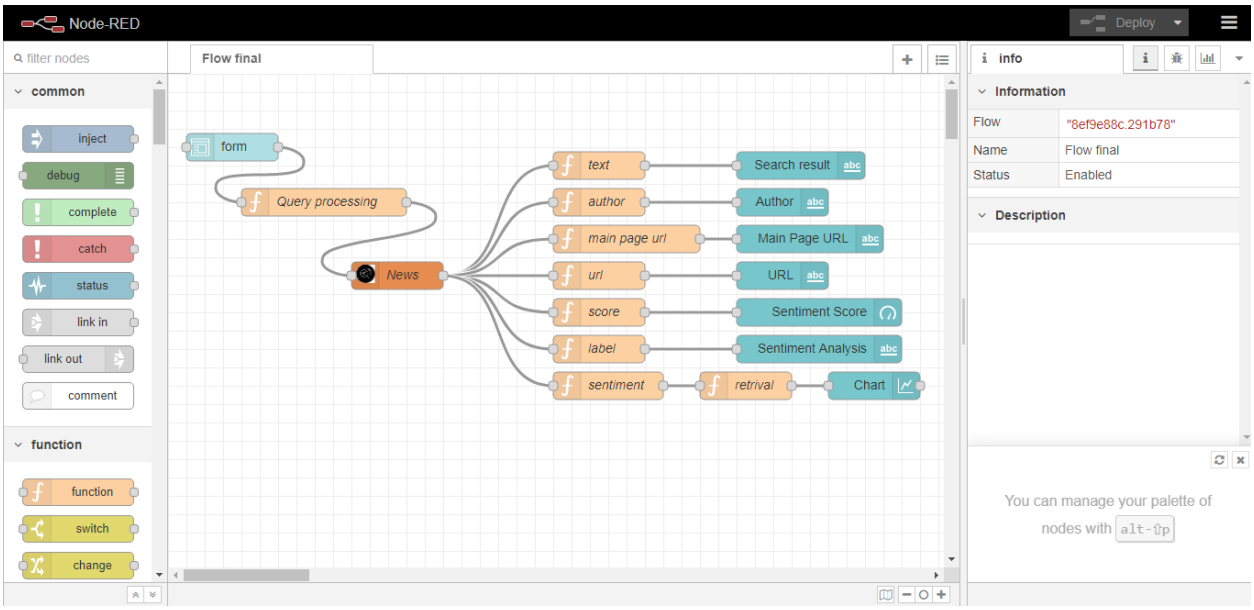
The app will look like below



Step-3: Once, the app is created, go to the AI sub-section of the services section under catalog, and select "Discovery" service.Then, choose the pre-built Watson Discovery News Collection and then launch the service.



Step-4: After launching the Discovery service, click the URL of the app to create the flow using Node-Red tool.Go to the Node-Red flow editor and use the nodes to create the flow.



Step-5: Finally, we have to integrate the application with slack by following the below steps.

1. Download Git for windows by using the link below.

   https://git-scm.com/download/win

And clone the watson-discovery-news Github repository by executing the command.

git clone https://github.com/ibm/watson-discovery-news

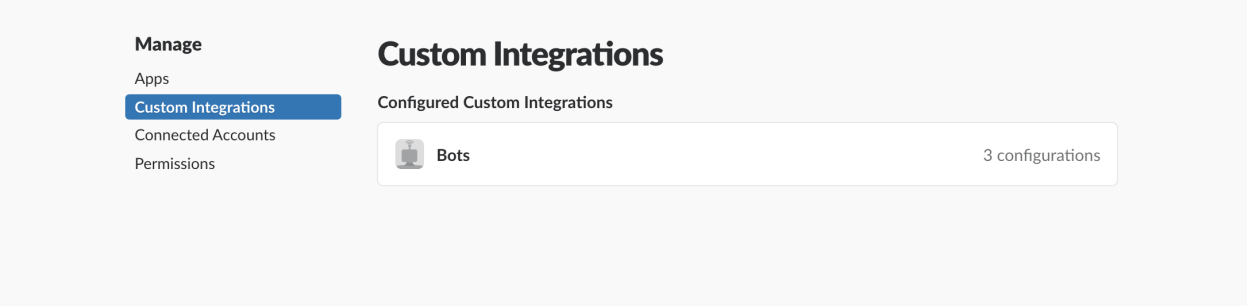
2. From the home directory of your cloned local repo, create a .env file by copying it from the sample version.

copy env.sample .env #Command for Windows user

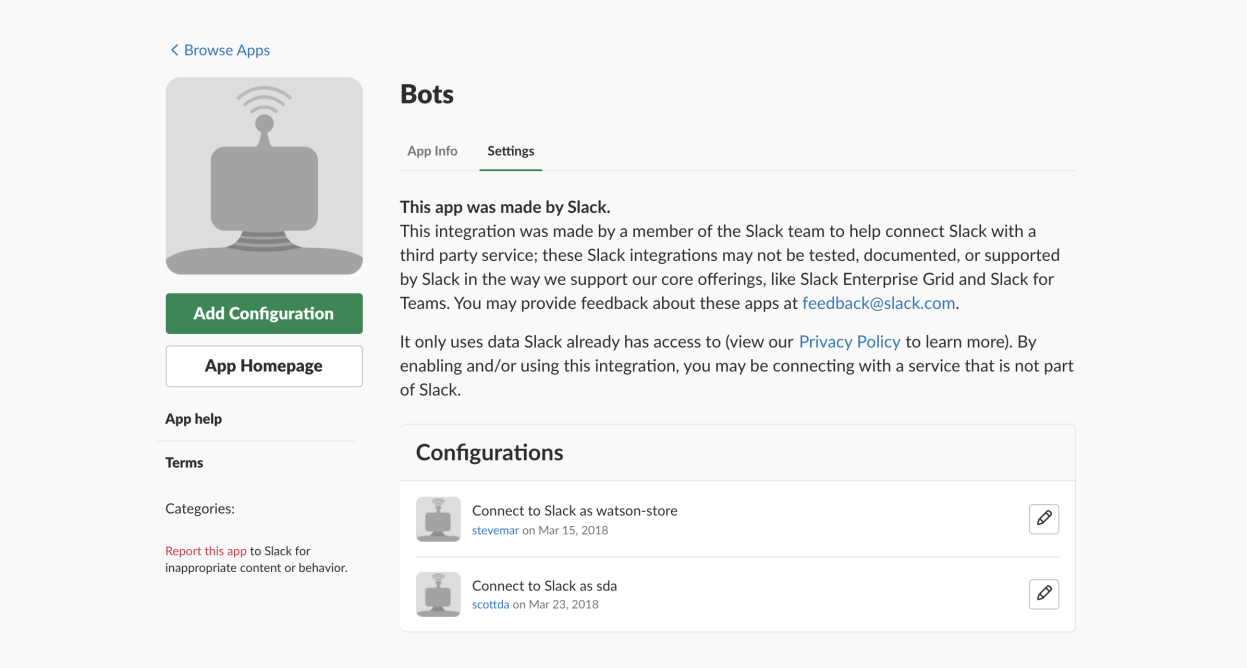
3. Copy all the service credentials listed on the homepage of yor Discovery service and paste it in the .env file.

4. To integrate a new Slack Bot into your existing Slack team, navigate to https://<my.slack.com>/apps/manage/custom-integrations, where <my.slack.com> is the Slack workspace you want to customize.

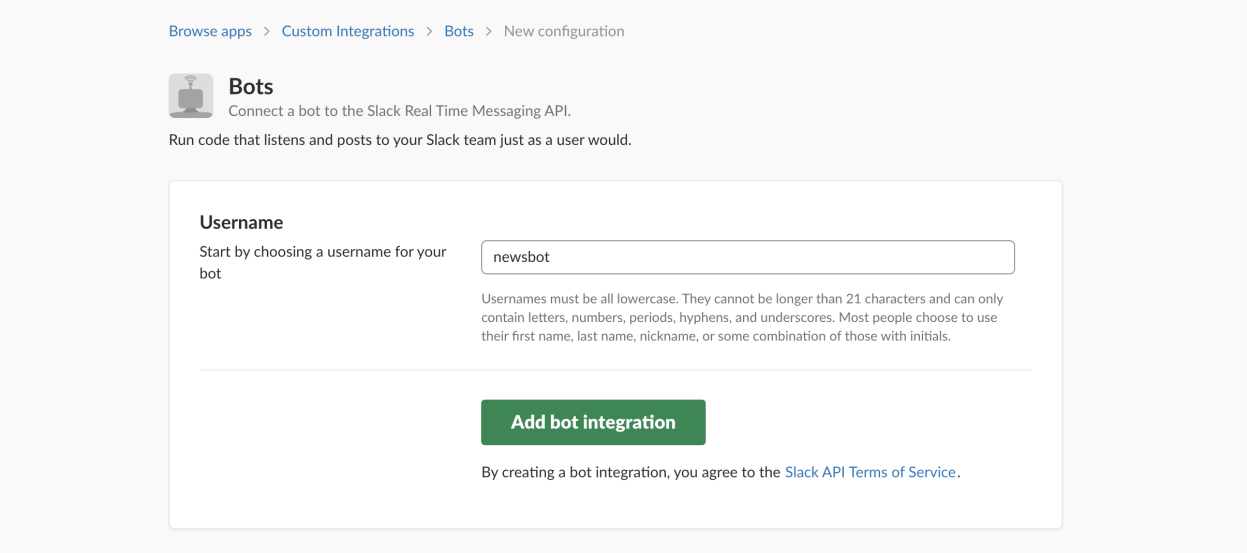
- From the Custom Integrations page, select the Bots option.



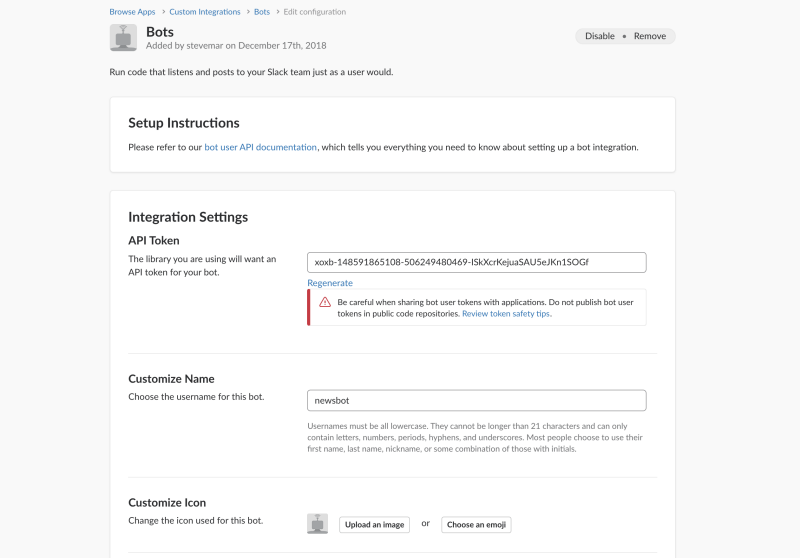
- To add a new bot, select the Add Configuration button.



- Enter a username for the bot and click Add bot integration.



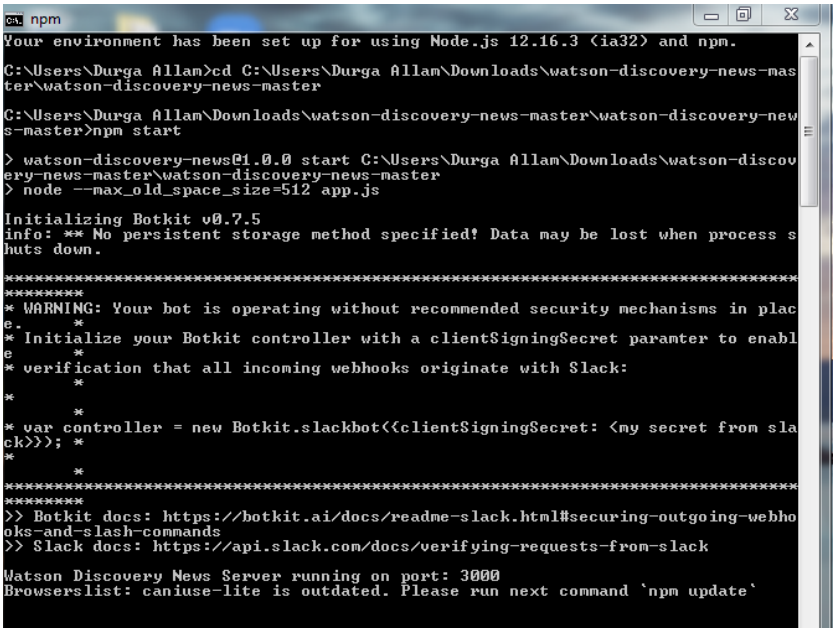
- Once created, save the API Token that is generated.



5. Edit the .env file and enter the Slack Bot API Token saved in the previous step.
6. Finally, run the application by executing the following commands.

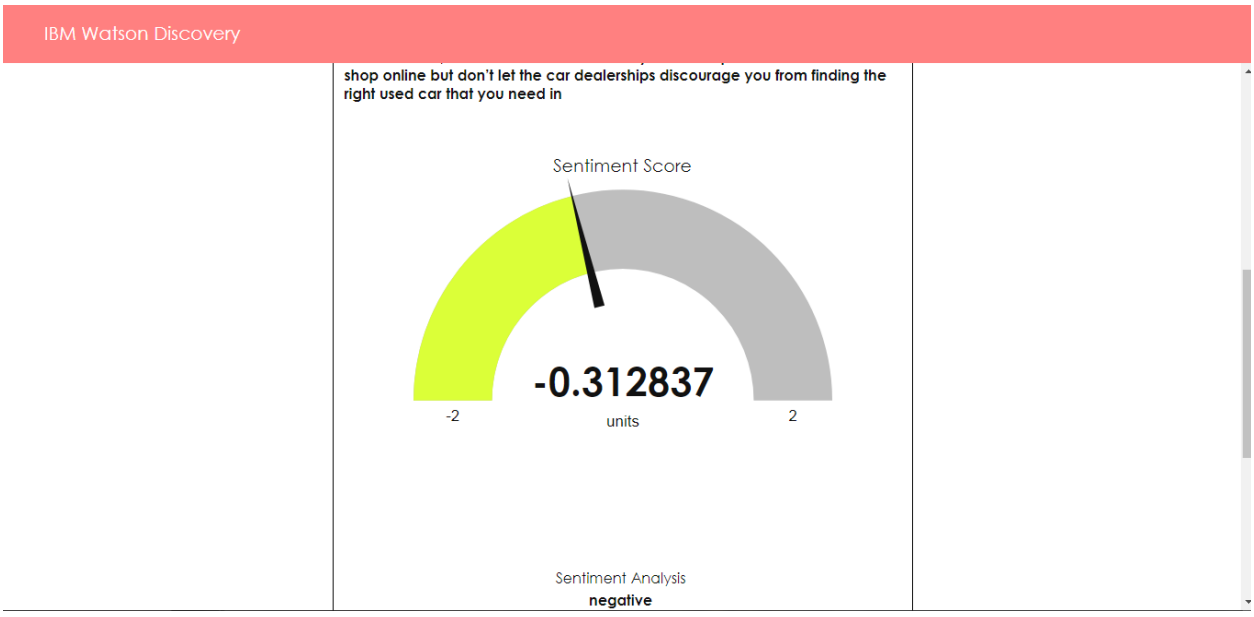
   npm install

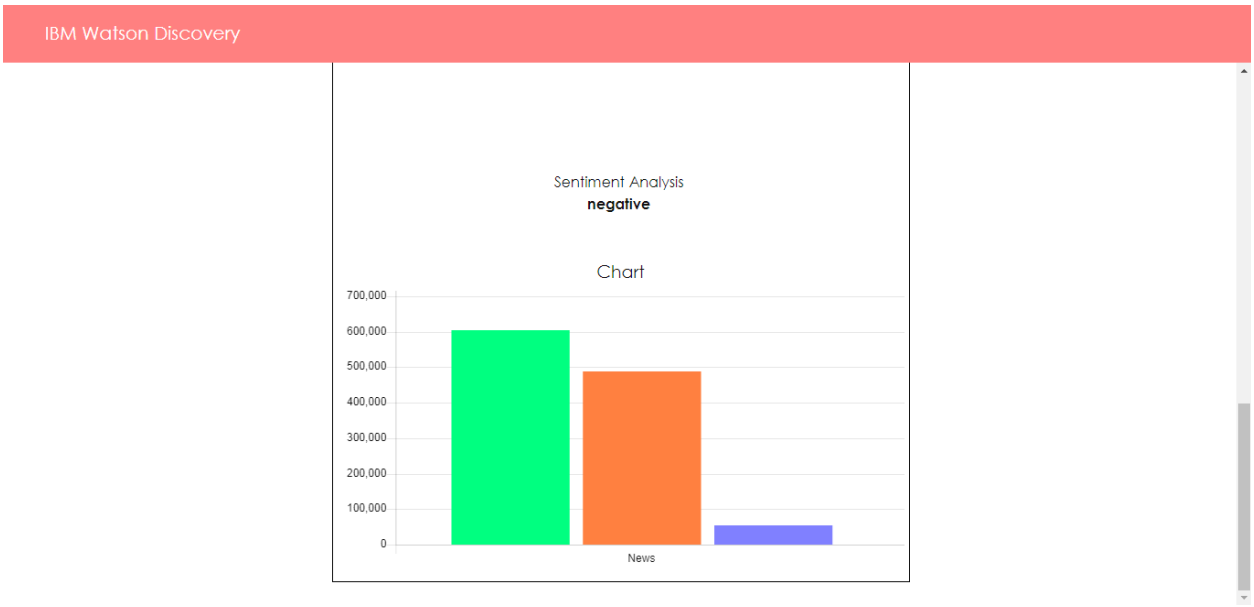
   npm start



6. RESULT

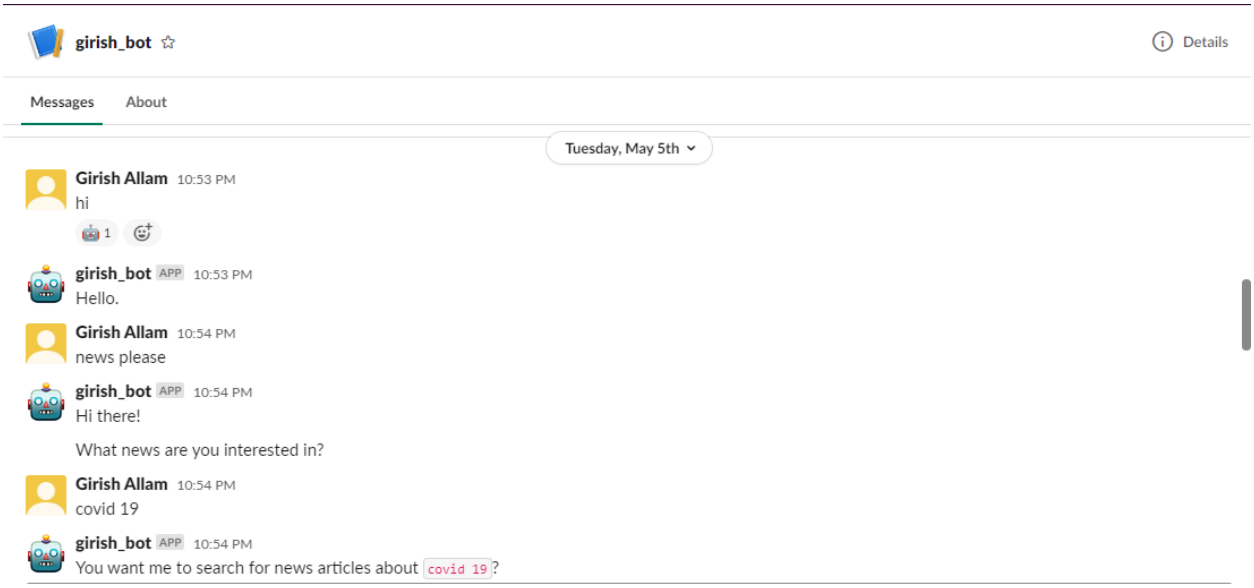
Images of the UI output:

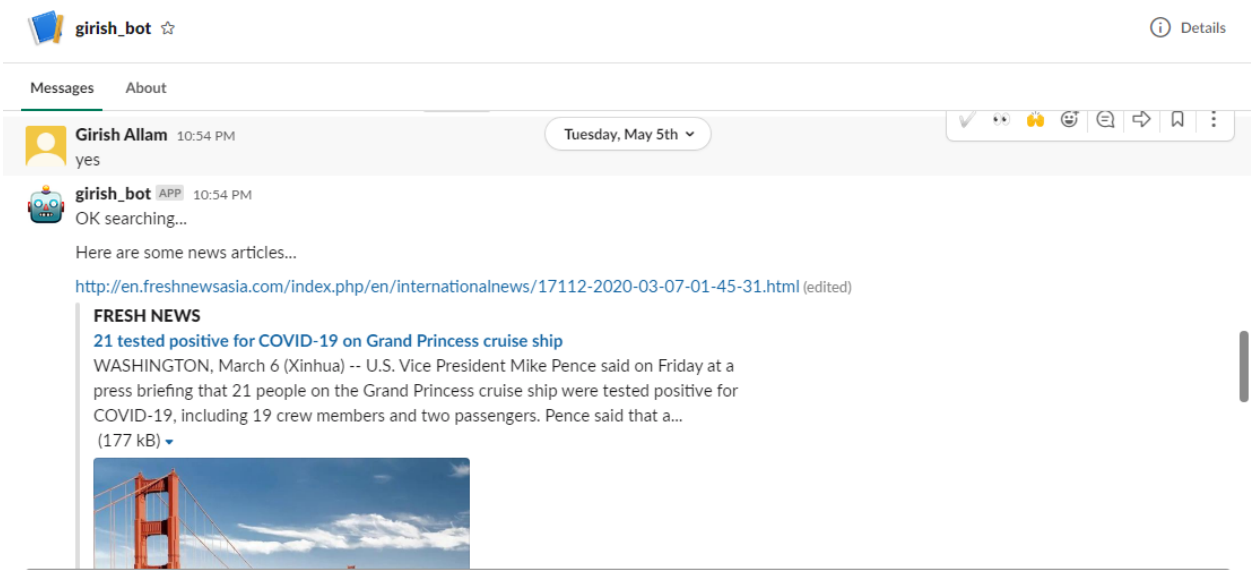


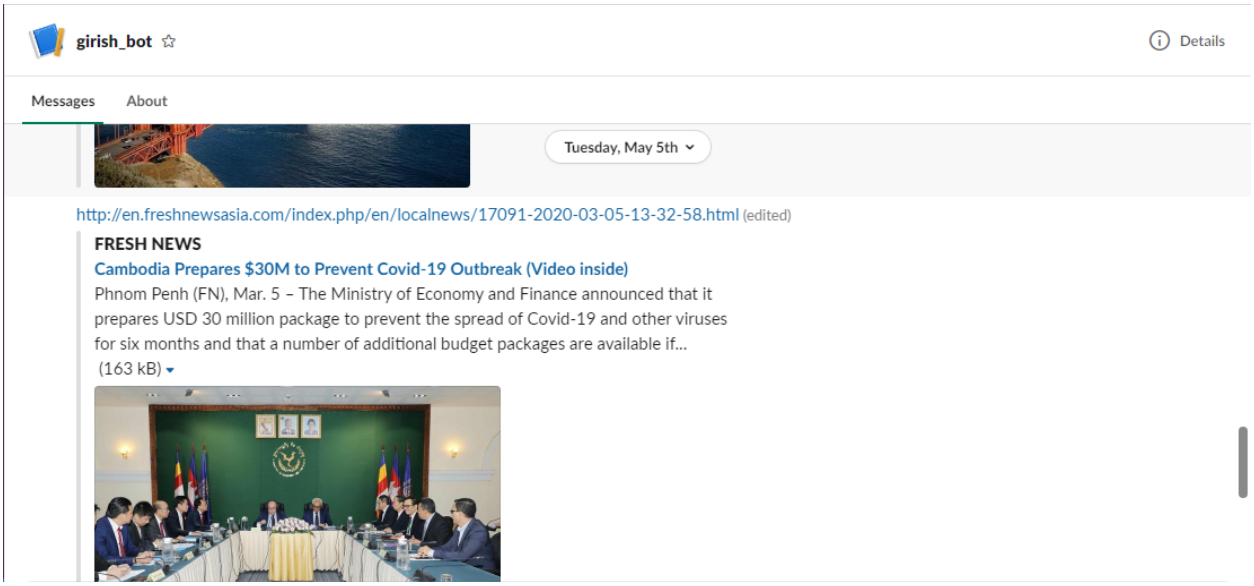


URL Address of the UI: https://newsdiscapp.apic.eu-gb.mybluemix.net/ui

Images of the Slack bot output:







7. ADVANTAGES AND DISADVANTAGES

The various adavantages of this news mining application are:

- Flexible access - The users can access from any part of the world.
- Always up-to-date - Since, Watson Discovery Service keeps updating the data from time to time, the user always gets the latest information.
- With the availability of cloud, the storage space can be increased anytime.

- Highly accurate data is retrieved everytime the query is processed.
- Integrating to slack is also highly efficient as anny request to the bot is sent to the server that starts the connection.

The disadvantages of the application are:

- Whilst 4G & Wi-Fi internet access is available in many locations, if you happen to lose connection you will not be able to access the app.
- The application, at present, can only work in one language.
- The application does not process structured data directly, thus resulting in slower performance.

8. APPLICATIONS

The main application of this project is to provide enriched news data about any particular topic.It not only helps to gain an understanding about it, but also proves helpful to make dynamic connections across the events in a faster way.The user can query for the most relevant new articles about a specific topic or subject. Because the news collection is pre-enriched with natural language processing, you can query not just on keywords or categories but also on concepts, sentiment, and relations to get richer search responses.We can also identify popular topics over the past 24 hours. Topics can be general or specific to an industry or category.

The Watson Discovery service also identifies important meta-information – like authors, publication dates, and relevant keywords in addition to sentimental analysis of the data.By discovering trends and patterns in sentiment with aggregate analysis, you'll see new perspectives on how news unfolds across the globe.We can even create news alerts by taking advantage of the support for entities, keywords, categories.

9. CONCLUSION

This project gives you some basic working knowledge of the Watson Discovery Service and shows you how to use Discovery along with JavaScript and Node.js to build your own news mining web application.With open source toolkit Botkit, it’s easier than ever to create your own chatbot to respond to the user by giving them all the latest news.In this project, which demonstrates how to serve news articles with the Watson Discovery API, we use Botkit to create a chatbot that acts as a middleman to facilitate queries to the Watson Discovery Service.This project satisfies it's objective of creating a news mining application that provides the user with alll the latest trends and patterns.

10. FUTURE SCOPE

This application can further be improved by including multiple language querying.That is, the user can ask queries in multiple languages and the application can retrieve the data by interpreting the query.We can also include customer feedback option to make changes to the UI and make it much better.The application can even be integrated with Watson Assistant so that chatbots can access the Discovery service and reply to the user's questions.Even, the layout of the UI can be optimized to make it more simple and clean.Thus, making the UI more attractive.We can also develop a more Mobile optimized design so that the users can access the application even on their mobiles.The load time of the graphics can also be decreased by minifying the page and making the user experience far more pleasurable.

11. BIBLIOGRAPHY

References:

1. https://developer.ibm.com/technologies/artificial-intelligence/patterns/create-a-cognitive-news-search-app/
2. https://www.ibm.com/watson/services/discovery-news/

3. https://developer.ibm.com/tutorials/integrating-slack-with-watson-discovery-news/
4. https://developer.ibm.com/articles/introduction-watson-discovery/

Appendix A:

Source Code

[{"id":"8ef9e88c.291b78","type":"tab","label":"Flow final","disabled":false,"info":""},{"id":"98564723.dc2918","type":"ui_form","z":"8ef9e88c.291b78","name":"","label":"","group":"31ade637.1ae3fa","order":1,"width":6,"height":2,"options":[{"label":"Search for News","value":"Input","type":"text","required":true,"rows":null}],"formValue":{"Input":""},"payload":"","submit":"submit","cancel":"cancel","topic":"","x":70,"y":80,"wires":[["2c81a3ac.23eecc"]]},{"id":"2c81a3ac.23eecc","type":"function","z":"8ef9e88c.291b78","name":"Query processing","func":"msg={\ndiscoveryparams:\n{\n\"environment_id\":\"system\",\n\"query\":msg.payload.Input\n }\n}\nreturn msg;","outputs":1,"noerr":0,"x":170,"y":140,"wires":[["2b70b443.a9dadc"]]},{"id":"269bb725.7177d8","type":"function","z":"8ef9e88c.291b78","name":"text","func":"msg.payload=msg.search_results.results[0].text\nreturn msg;","outputs":1,"noerr":0,"x":470,"y":100,"wires":[["a5c09950.b178f8"]]},{"id":"2b70b443.a9dadc","type":"watson-discovery-v1","z":"8ef9e88c.291b78","name":"News","environmentname":"","environmentId":"system","collectionId":"news-en","configurationname":"","configurationId":"","language_code":"en","collection_name":"","count":"1","passages":true,"nlp_query":true,"query":"","filter":"","aggregation":"term(enriched_text.sentiment.document.label,count:10)","return":"","description":"","size":"LT","discovery-method":"query","service-endpoint":"https://api.eu-gb.discovery.watson.cloud.ibm.com/instances/451bfa03-f9bd-4e76-b855-c9cb7e627d91","x":250,"y":220,"wires":[["269bb725.7177d8","4cc33878.37c118","b733cd8e.06b17","d2c1d8b9.10a3d8","3bd55baf.c46104","5b66fb16.9888d4","b1d35d5a.811f9"]]},{"id":"a5c09950.b178f8","type":"ui_text","z":"8ef9e88c.291b78","group":"31ade637.1ae3f

a","order":8,"width":6,"height":1,"name":"","label":"Search result","format":"{{msg.payload}}","layout":"col-center","x":690,"y":100,"wires":[]},{"id":"d2c1d8b9.10a3d8","type":"function","z":"8ef9e88c.291b78","name":"url","func":"msg.payload=msg.search_results.results[0].url\nreturn msg;","outputs":1,"noerr":0,"x":470,"y":220,"wires":[["9d7c0f27.1eb1b"]]},{"id":"4cc33878.37c118","type":"function","z":"8ef9e88c.291b78","name":"author","func":"msg.payload=msg.search_results.results[0].author\nreturn msg;","outputs":1,"noerr":0,"x":470,"y":140,"wires":[["6d52518c.454d2"]]},{"id":"5b66fb16.9888d4","type":"function","z":"8ef9e88c.291b78","name":"main page url","func":"msg.payload=msg.search_results.results[0].main_image_url\nreturn msg;","outputs":1,"noerr":0,"x":500,"y":180,"wires":[["3c169d99.355fd2"]]},{"id":"b733cd8e.06b17","type":"function","z":"8ef9e88c.291b78","name":"score","func":"msg.payload=msg.search_results.results[0].enriched_text.sentiment.document.score\nreturn msg;","outputs":1,"noerr":0,"x":470,"y":260,"wires":[["43792c25.381804"]]},{"id":"3bd55baf.c46104","type":"function","z":"8ef9e88c.291b78","name":"label","func":"msg.payload=msg.search_results.results[0].enriched_text.sentiment.document.label\nreturn msg;","outputs":1,"noerr":0,"x":470,"y":300,"wires":[["35f74572.ed330a"]]},{"id":"35f74572.ed330a","type":"ui_text","z":"8ef9e88c.291b78","group":"31ade637.1ae3fa","order":12,"width":6,"height":2,"name":"","label":"Sentiment Analysis","format":"{{msg.payload}}","layout":"col-center","x":710,"y":300,"wires":[]},{"id":"9d7c0f27.1eb1b","type":"ui_text","z":"8ef9e88c.291b78","group":"31ade637.1ae3fa","order":6,"width":6,"height":1,"name":"","label":"URL","format":"{{msg.payload}}","layout":"col-center","x":670,"y":220,"wires":[]},{"id":"6d52518c.454d2","type":"ui_text","z":"8ef9e88c.291b78","group":"31ade637.1ae3fa","order":2,"width":6,"height":1,"name":"","label":"Author","format":"{{msg.payload}}","layout":"col-center","x":670,"y":140,"wires":[]},{"id":"3c169d99.355fd2","type":"ui_text","z":"8ef9e88c.291b78","group":"31ade637.1ae3fa","order":4,"width":6,"height":1,"name":"","label":"Main Page URL","format":"{{msg.payload}}","layout":"col-center","x":700,"y":180,"wires":[]},{"id":"43792c25.381804","type":"ui_gauge","z":"8ef9e88c.291b78","name":"","group":"31ade637.1ae3fa","

order":10,"width":6,"height":6,"gtype":"gage","title":"Sentiment Score","label":"units","format":"{{value}}","min":"-2","max":"2","colors":["#1cff1c","#ffff3e","#ff8040"],"seg1":"","seg2":"","x":710,"y":260,"wires":[]},{"id":"32e1ef9.7ff3a1","type":"function","z":"8ef9e88c.291b78","name":"retrival","func":"msg.payload=\n[\n   {\n \"series\":[\"News\"],\n   \"data\":[msg.payload[0],msg.payload[1],msg.payload[2]],\n \"labels\":[\"Positive\",\"Negative\",\"Neutral\"]\n   }\n]\nreturn msg;","outputs":1,"noerr":0,"x":630,"y":340,"wires":[["1fe434c0.46685b"]]},{"id":"1fe434c0.46685b","type":"ui_chart","z":"8ef9e88c.291b78","name":"","group":"31ade637.1ae3fa","order":13,"width":6,"height":6,"label":"Chart","chartType":"bar","legend":"false","xformat":"HH:mm:ss","interpolate":"linear","nodata":"","dot":false,"ymin":"","ymax":"","removeOlder":1,"removeOlderPoints":"","removeOlderUnit":"3600","cutout":0,"useOneColor":false,"useUTC":false,"colors":["#00ff80","#ff8040","#8080ff","#2ca02c","#98df8a","#d62728","#ff9896","#9467bd","#c5b0d5"],"useOldStyle":false,"outputs":1,"x":770,"y":340,"wires":[[]]},{"id":"b1d35d5a.811f9","type":"function","z":"8ef9e88c.291b78","name":"sentiment","func":"var val=[msg.search_results.aggregations[0].results[0].matching_results];\nval.push(msg.search_results.aggregations[0].results[1].matching_results);\nval.push(msg.search_results.aggregations[0].results[2].matching_results);\nmsg.payload=val;\nreturn msg;","outputs":1,"noerr":0,"x":480,"y":340,"wires":[["32e1ef9.7ff3a1"]]},{"id":"31ade637.1ae3fa","type":"ui_group","z":"","name":"Watson Discovery News Search","tab":"bd70c946.4f1318","order":1,"disp":true,"width":"6","collapse":false},{"id":"bd70c946.4f1318","type":"ui_tab","z":"","name":"IBM Watson Discovery","icon":"dashboard","disabled":false,"hidden":false}]