AI Powered News Search App ,

TheSmartBridge

SMARTBRIDGE

REMOTE SUMMER INTERNSHIP

PROJECT REPORT

ON

AI POWERED NEWS SEARCH APP

BY MOHAMMED ABDUL RAHMAN

TABLE OF CONTENTS

PROJECT REPORT

1. INTRODUCTION

1.1 Overview

1.1.1 Project Summary

The web is home to massive amounts of data, with more being created every day. 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.

In this project, we built a NEWS mining web application starting with the basics, using Node-RED / Python Web App and the IBM Watson Discovery Service.

To do this we start by building a Server-Side Application using Node-RED then we use the pre-built Watson Discovery News collection and access the Watson Discovery Service through the Discovery API. Optionally, we can choose to use a Slack interface to query the data, push news alerts out to web notification and deploy the app on IBM Cloud.

1.1.2 Purpose

The purpose of the project is to resolve the issues generated due to false news being transmitted and to improve news fetching methodologies. Today, businesses want to know what buyers say about their brand and how they feel about their products. However, with all of the “noise” filling our email, social and other communication channels, listening to customers has become a difficult task.

In this project, sentiment analysis helps to learn how a bot uses artificial intelligence based approach can provide a medium which showcases how true the news being circulated actually is.

This project comprises of a UI using node red and a bot it is the same as a regular app: it can access the same range of APIs and do all of the magical things that a Slack App can do. A bot is a type of Slack App designed to interact with users via conversation. But when you build a bot for your Slack App, you're giving that app a face, a name, and a personality and encouraging users to talk to it.

The bot can send DMs, it can be mentioned by users, it can post messages or upload files, and it can be invited to channels - or kicked out. Since bot is capable of doing everything that a Slack App can do, we're going to limit our focus to a common use-case for bots.

2. LITERATURE SURVEY

2.1 Existing problem

a. *Too much information* - It goes without saying that the distributed content production of individuals is an amazing and powerful phenomenon, but relying on search engines to index, sort and rank externally-produced content that isn't standardized for format doesn't work for the newsreader. There's still too much information and the organization is distorted by advertising dollars.

b. *News is geographically sensitive information* - If you live in Dallas, TX do you want to have to sort through local content from Boston, MA or New York, NY or vice versa to find local news? No, you want to see time-sensitive information about what's happening in your community and NOW! While Geo RSS is out there, it's still nowhere near being adopted as a widely-used content standard.

c. .*News is time-sensitive information* - There should be constant organizing externally-produced content and figure out that new content was produced and, then organize all of the content that comes in many different structural formats and it has to passively (algorithmically) distinguish between news information (time-sensitive) and historical information that isn't time-sensitive.

d. *Distilling quality news* - It is difficult as link-backs aren't necessarily reflective of quality or demand. One has to filter the glut of information hitting the internet 24/7 while distilling the time-sensitive and geographically-sensitive content a newsreader wants.
e. Costly and Slow: using popular search engines like google, bing or duckduckgo, we get only the search results listed by the SEO, but we dont get any other metadata like sentiment, score, the url from where it is submitted. And if we want to implement that, we have to take explicit permission and apply for api or even have to pay huge amount of money to just get data. Even after getting the data, we have to create and deploy a model to perform sentiment analysis and then its integration in any web framework like Flask, Django or Node.js makes the website loading very slow and unresponsive.

2.2 Proposed solution

IBM has upgraded its Watson AI platform to boost the ability to understand business lingo that could be used to analyze the English language text and conversations.

The tool is designed to help develop arguments that could ultimately be used to improve machine understanding and help analyze complex human speech. The integration of Project Debater technology with Watson includes a sentiment analysis and “summarization” tools along

with the ability to cluster data by topic.

The natural language processing capability represents the first commercialization of IBM Research's Project Debater, billed as an AI system capable of debating humans.

The sentiment analyzer being released this week can detect sentiment shifters because a negative modifier can reverse a speaker's meaning using a phase like "hardly helpful." The analysis tool would also enable the Watson API to identify common idioms like "calling it a day."

The briefing tool extracts text data from different sources to provide a summary of speech and text on a specific topic. The clustering feature allows users to organize data and generate topics.

IBM Node-RED: Node-RED is a flow-based development tool for visual programming developed originally by IBM for wiring together hardware devices, APIs and online services as part of the Internet of Things. Node-RED provides a web browser-based flow editor, which can be used to create JavaScript functions.

IBM's Watson Discovery enterprise search platform, the data organizer is aimed at subject experts in specific industries like

a. Health care
b. Insurance
c. Technology

3. THEORITICAL ANALYSIS

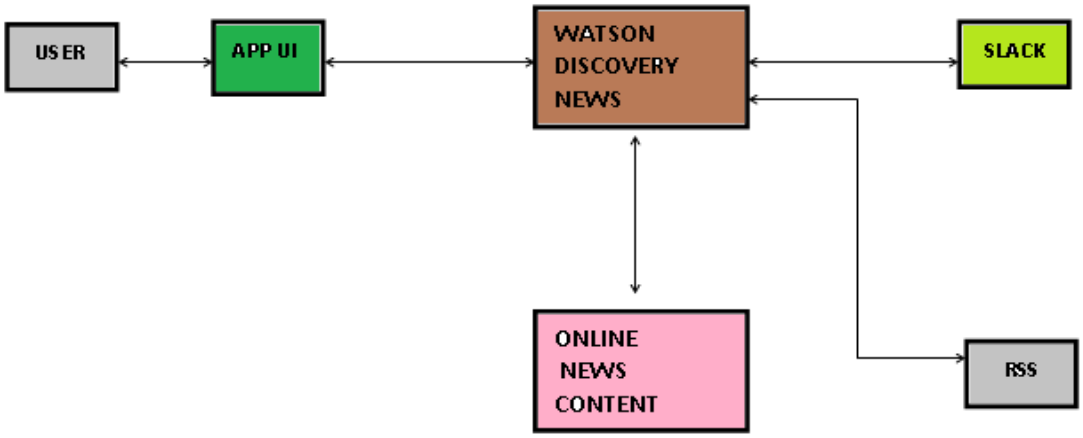
3.1 Block diagram

The project revolves around:

1. Building a Server Side Application using Node-RED
2. Using the pre-built Watson Discovery News collection
3. Accessing the Watson Discovery Service through the Discovery API

Additionally, we include:

1 Deploy the app on IBM Cloud.
2. Use a Slack interface to query the data
3. Push news alerts out to web notification



3.2 Hardware / Software designing

Elements involved:

1. IBM Cloud
2. IBM Watson Discovery
3. Node Red
4. Node.js
5. GitHub
6. RSS
7. Slack
8. Html
9. Zoho Writer

SOFTWARE :

10. Code is written in Node.js, with the server-side using the Express framework and the client using ReactJS.
11. The pre-built Watson Discovery News collection was used.
12. Access the Watson Discovery Service through the Discovery API.
13. Use a Slack interface to query the data

4. EXPERIMENTAL INVESTIGATIONS

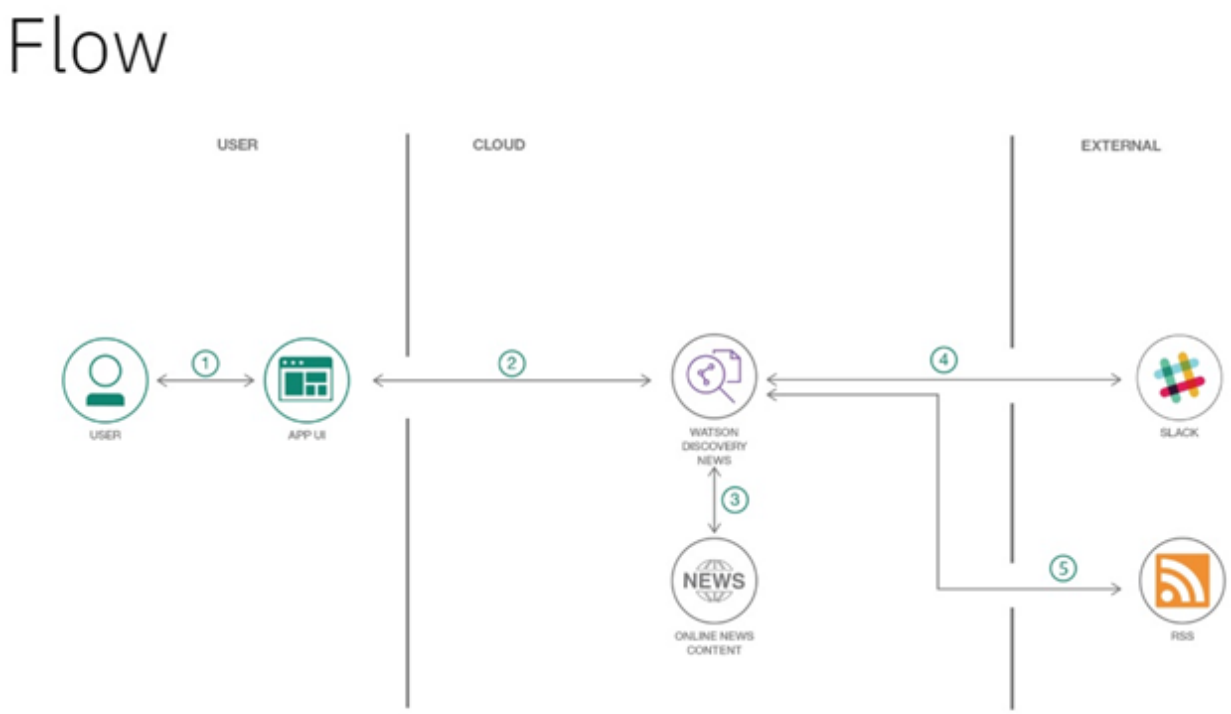
This project used is an experiment to the use of Watson Discovery to query news and display it to the user via UI and a slack bot. It was taken into consideration after examining differences in credibility assigned to news read in paper form and the same news read on a web site.

Manipulating the meta_data to get score, url, hostname, titles and so. Playing with msg.payload (debug) to get various outputs .

The bot can send DMs, it can be mentioned by users, it can post messages or upload files, and it can be invited to channels or kicked out. Since bot is capable of doing everything that a Slack App can do, we're going to limit our focus to a common use-case for bots.

This project includes a bot it is the same as a regular app it can access the same range of APIs and do all of the magical things that a Slack App can do. A bot is a type of Slack App designed to interact with users via conversation. But when you build a bot for your Slack App, you're giving that app a face, a name, and a personality and encouraging users to talk to it.

5. FLOWCHART

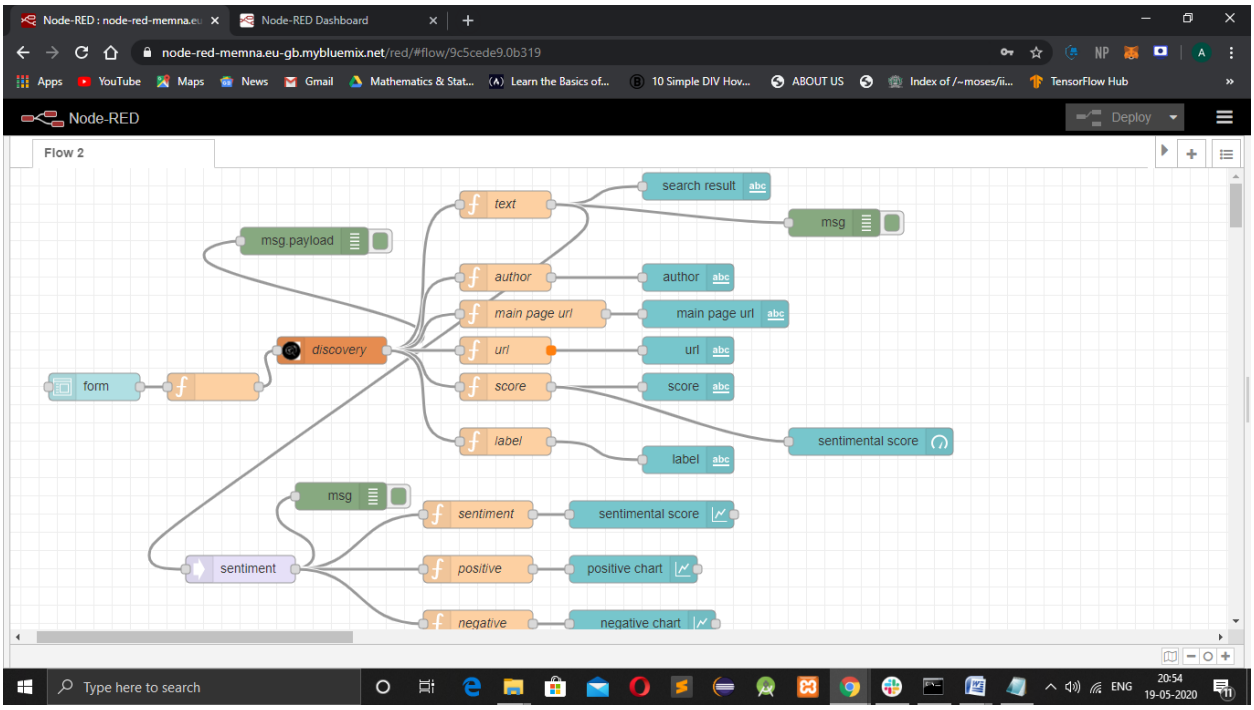

Flow
USER
CLOUD
EXTERNAL
1
2
3
4
5
USER
APP UI
WATSON DISCOVERY NEWS
NEWS
ONLINE NEWS CONTENT
SLACK
RSS

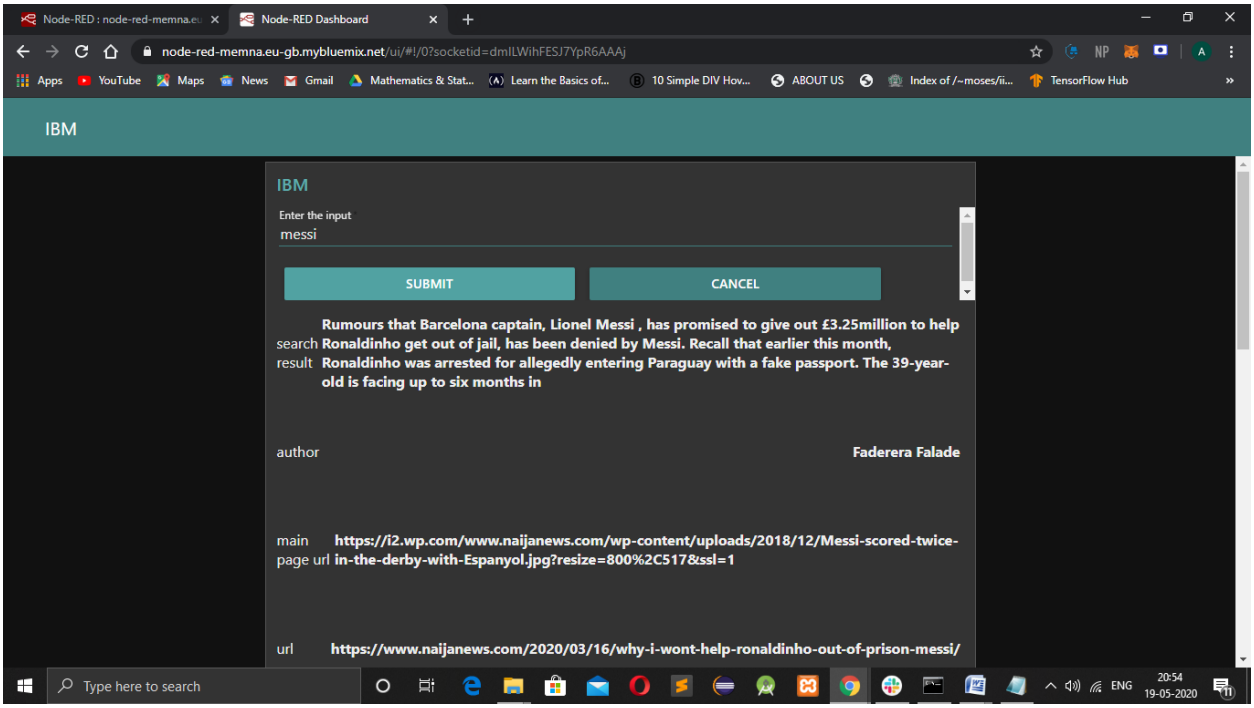
6. RESULT

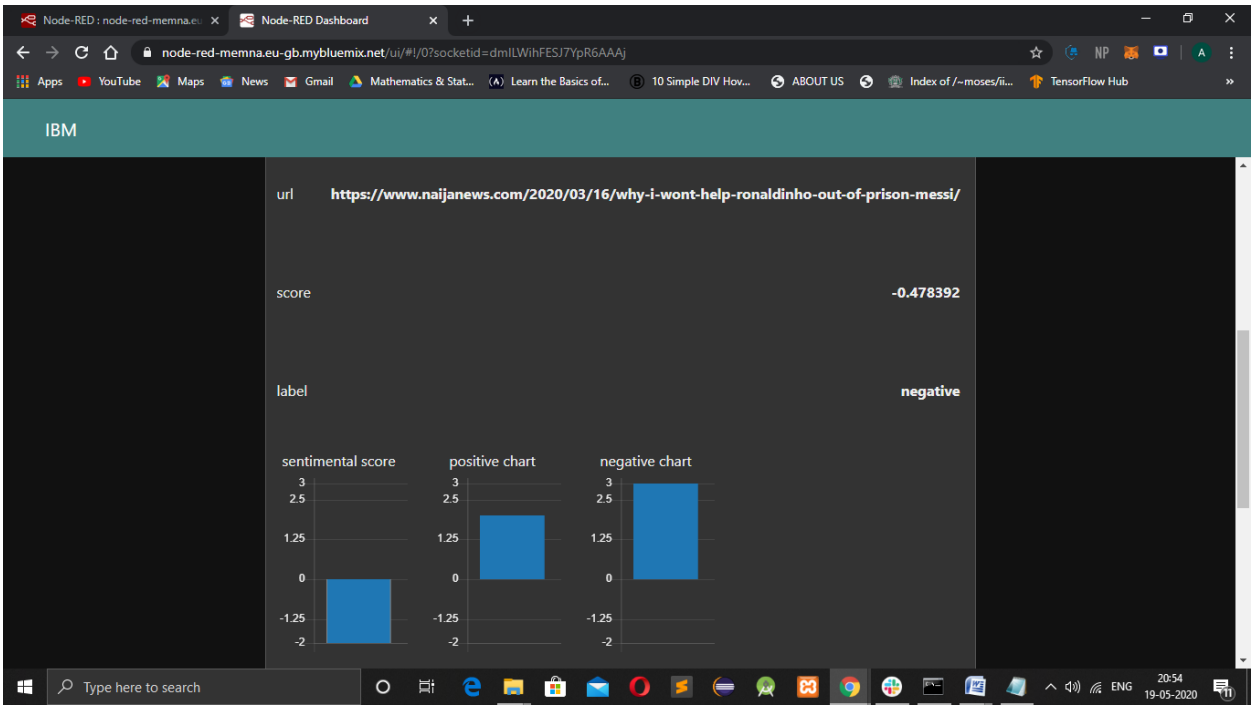
The final application has

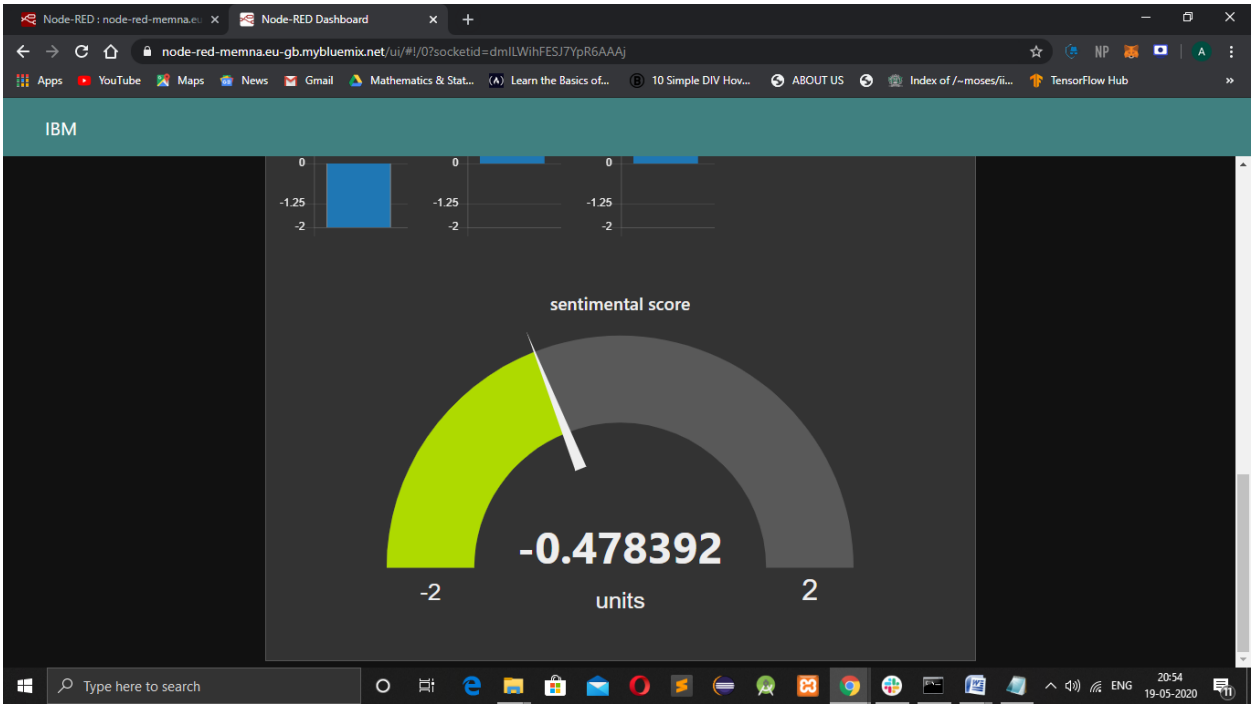
a. A user-interface is used to request the extraction of required news content.
b. The app generates the latest news regarding request generated by user using the Watson Discovery News.
c. It generates the URL it used to extract the news from.
d. It displays the author who had submitted the news report.
e. It provides with sentimental analysis that separate positive or negative opinion from good or bad news.
f. Slack Integration

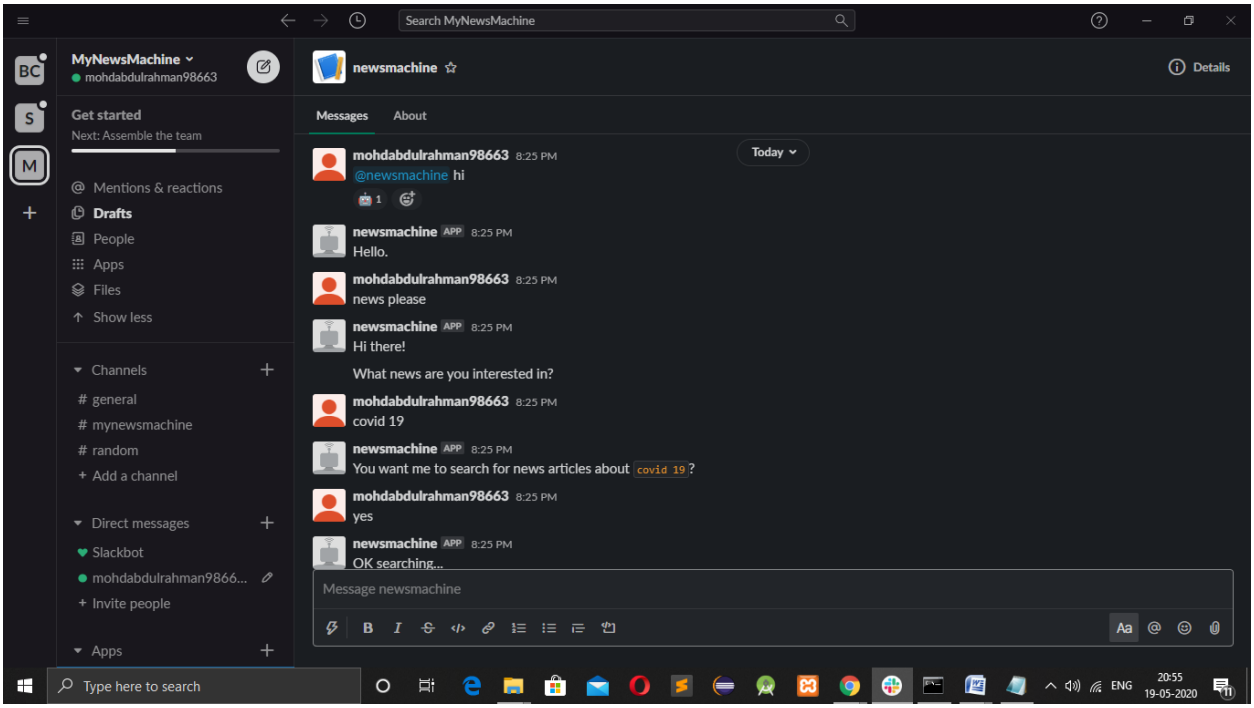
URL Link of Project:https://node-red-memna.eu-gb.mybluemix.net/ui/

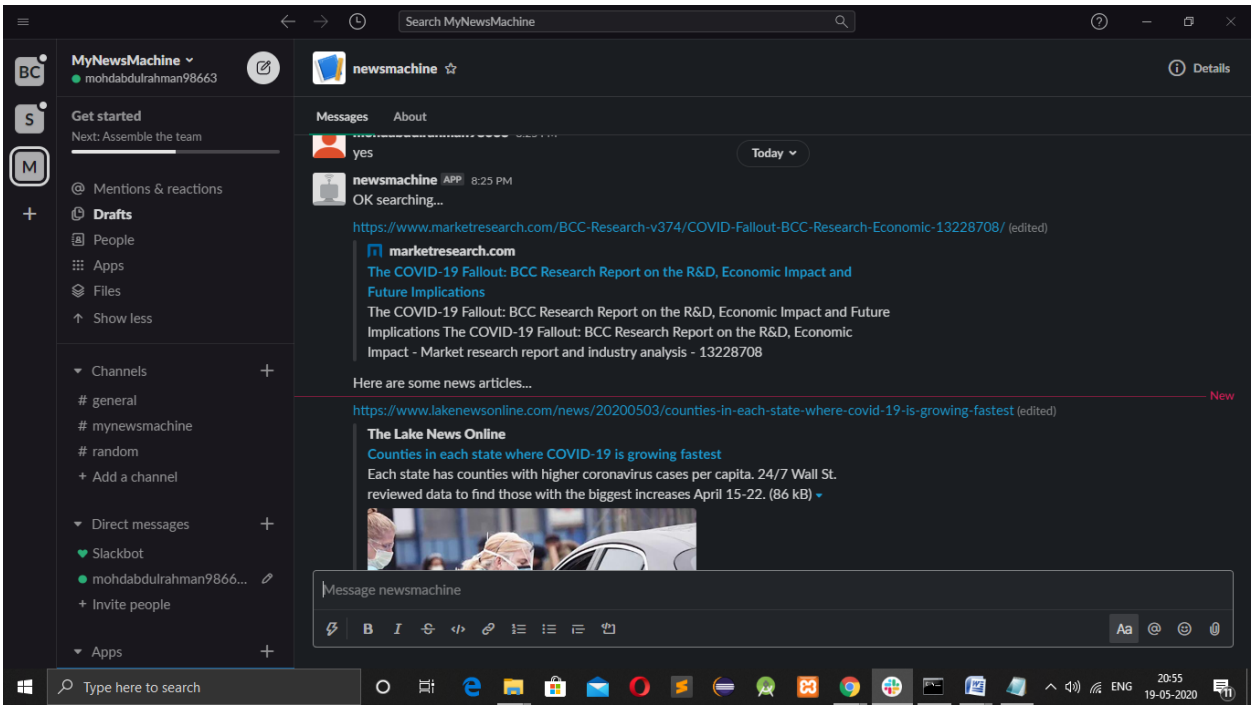


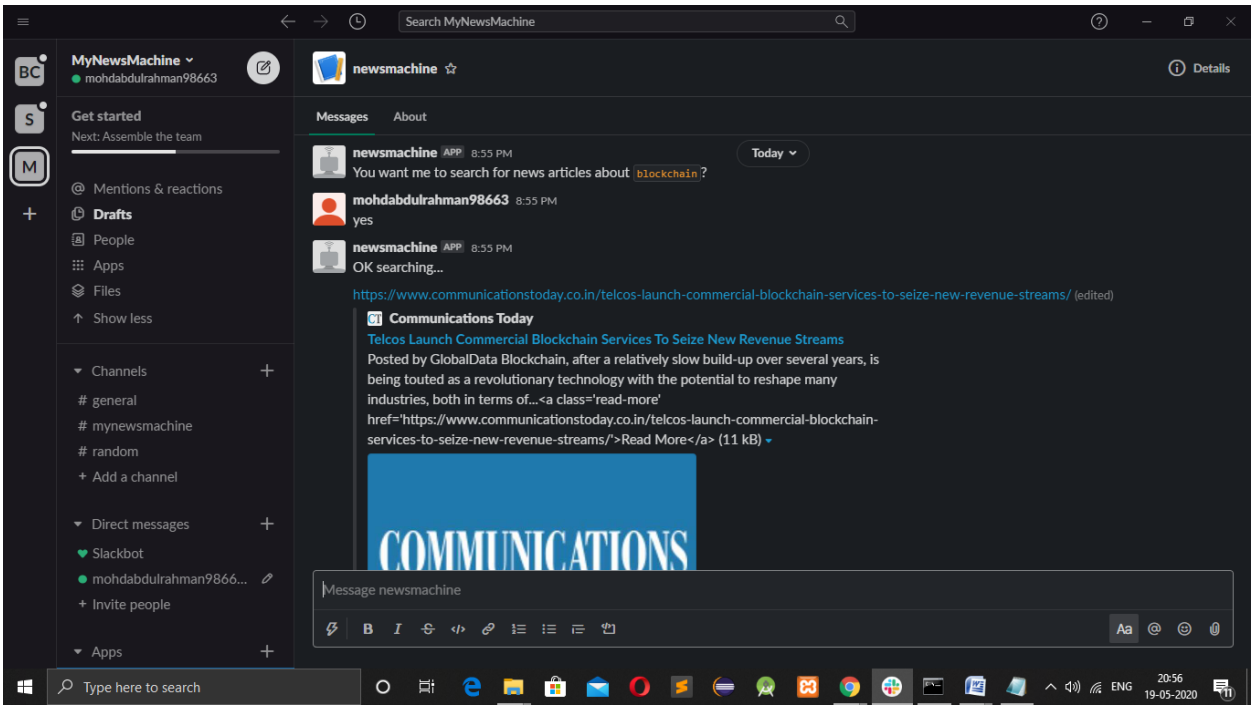

Node-RED : node-red-memna.eu
Node-RED Dashboard
node-red-memna.eu-gb.mybluemix.net/ui/#!/0?socketid=dmILWihFESJ7YpR6AAAj
IBM
IBM
Enter the input
messi
SUBMIT
CANCEL
search result
Rumours that Barcelona captain, Lionel Messi , has promised to give out £3.25million to help Ronaldinho get out of jail, has been denied by Messi. Recall that earlier this month, Ronaldinho was arrested for allegedly entering Paraguay with a fake passport. The 39-year-old is facing up to six months in
author
Faderera Falade
main page url
https://i2.wp.com/www.naijanews.com/wp-content/uploads/2018/12/Messi-scored-twice-in-the-derby-with-Espanyol.jpg?resize=800%2C517&ssl=1
url
https://www.naijanews.com/2020/03/16/why-i-wont-help-ronaldinho-out-of-prison-messi/
Type here to search
ENG
20:54
19-05-2020


Node-RED : node-red-memna.eu
Node-RED Dashboard
node-red-memna.eu-gb.mybluemix.net/ui/#!/0?socketid=dmILWihFESJ7YpR6AAAj
IBM
url
https://www.naijanews.com/2020/03/16/why-i-wont-help-ronaldinho-out-of-prison-messi/
score
-0.478392
label
negative
sentimental score
positive chart
negative chart
3
2.5
1.25
0
-1.25
-2
Type here to search
ENG
20:54
19-05-2020


Node-RED : node-red-memna.eu
Node-RED Dashboard
node-red-memna.eu-gb.mybluemix.net/ui/#!/0?socketid=dmILWihFESJ7YpR6AAAj
IBM
sentimental score
-0.478392
units
-2
2
Type here to search
ENG
20:54
19-05-2020


Search MyNewsMachine
MyNewsMachine
mohdabdulrahman98663
Get started
Next: Assemble the team
Mentions & reactions
Drafts
People
Apps
Files
Show less
Channels
general
mynewsmachine
random
Add a channel
Direct messages
Slackbot
Invite people
Apps
newsmachine
Details
Messages
About
Today
mohdabdulrahman98663 8:25 PM
@newsmachine hi
newsmachine APP 8:25 PM
Hello.
mohdabdulrahman98663 8:25 PM
news please
newsmachine APP 8:25 PM
Hi there!
What news are you interested in?
mohdabdulrahman98663 8:25 PM
covid 19
newsmachine APP 8:25 PM
You want me to search for news articles about covid 19?
mohdabdulrahman98663 8:25 PM
yes
newsmachine APP 8:25 PM
OK searching...
Message newsmachine
Type here to search
ENG
20:55
19-05-2020


Search MyNewsMachine
MyNewsMachine
mohdabdulrahman98663
newsmachine
Details
Get started
Next: Assemble the team
Messages
About
Mentions & reactions
Drafts
People
Apps
Files
Show less
Channels
general
mynewsmachine
random
+ Add a channel
Direct messages
Slackbot
mohdabdulrahman9866...
+ Invite people
Apps
yes
Today
newsmachine APP 8:25 PM
OK searching...
https://www.marketresearch.com/BCC-Research-v374/COVID-Fallout-BCC-Research-Economic-13228708/ (edited)
marketresearch.com
The COVID-19 Fallout: BCC Research Report on the R&D, Economic Impact and Future Implications
The COVID-19 Fallout: BCC Research Report on the R&D, Economic Impact and Future Implications The COVID-19 Fallout: BCC Research Report on the R&D, Economic Impact - Market research report and industry analysis - 13228708
Here are some news articles...
New
https://www.lakenewsonline.com/news/20200503/counties-in-each-state-where-covid-19-is-growing-fastest (edited)
The Lake News Online
Counties in each state where COVID-19 is growing fastest
Each state has counties with higher coronavirus cases per capita. 24/7 Wall St. reviewed data to find those with the biggest increases April 15-22. (86 kB)
Message newsmachine
Type here to search
ENG
20:55
19-05-2020


Search MyNewsMachine
MyNewsMachine
mohdabdulrahman98663
newsmachine
Details
Get started
Next: Assemble the team
Messages
About
Mentions & reactions
Drafts
People
Apps
Files
Show less
Channels
general
mynewsmachine
random
+ Add a channel
Direct messages
Slackbot
mohdabdulrahman9866...
+ Invite people
Apps
newsmachine APP 8:55 PM
You want me to search for news articles about blockchain?
Today
mohdabdulrahman98663 8:55 PM
yes
newsmachine APP 8:55 PM
OK searching...
https://www.communicationstoday.co.in/telcos-launch-commercial-blockchain-services-to-seize-new-revenue-streams/ (edited)
Communications Today
Telcos Launch Commercial Blockchain Services To Seize New Revenue Streams
Posted by GlobalData Blockchain, after a relatively slow build-up over several years, is being touted as a revolutionary technology with the potential to reshape many industries, both in terms of...Read More (11 kB)
COMMUNICATIONS
Message newsmachine
Type here to search
ENG
20:56
19-05-2020

7. ADVANTAGES & DISADVANTAGES

7.1 ADVANTAGES

The app uses standard search UI components such as filter lists, tag clouds, and sentiment graphs, but also more complex Discovery options such as the passages and highlight features.
With these two features, the app identifies the most relevant snippets in your data based on your query and is more likely to return the data that you're searching for.
News coming with sentiment analysis done and only singular commands are needed for slack to get the requested news.
The main benefit of using the Watson Discovery Service is its powerful analytics engine that provides cognitive enrichments and insights into our data. The app in this code pattern provides examples of how to showcase these enrichments through the use of filters, lists, and graphs.

The project makes use of slack that has multiple APIs to create bots and to create a connection to Slack. Any request to the bot user, to the bot directly, or from a Slack channel in which the bot is a member, is sent to the server that starts the connection.
Slack uses the bot token as a means to authenticate and all of these works due to the use of the Real Time Messaging API.

The key enrichments are:

a. *Entities:* People, companies, organizations, cities, and more
b. *Categories:* Classification of the data into a hierarchy of categories up to 5 levels deep

c. *Concepts:* Identified general concepts that aren't necessarily referenced in the data
d. *Keywords:* Important topics typically used to index or search the data
e. *Sentiment:* The overall positive or negative sentiment of each document

7.2 DISADVANTAGES

Application Programming Interface (API)

a. One of the key problems that most of the developer's encounter is the Application Programming Interface (API) keeps on changing at frequent intervals and does not remain stable.
b. At times, a new API appears having a number of backwards-incompatible changes. As a result, the developers are forced to make changes in the accessible code bases to match the compatibility with the latest version of the Node.js API.
c. Inability to Understand – Due to fixed programs, bot can be stuck if an unsaved query is presented in front of them. This can lead to customer dissatisfaction and result in loss.

Sentiment Analysis:

Sentiment analysis tools can identify and analyze many pieces of text automatically and quickly.

But computer programs have problems recognizing things like sarcasm

and irony, negations, jokes, and exaggerations.

The sorts of things a person would have little trouble identifying. And failing to recognize these can skew the results.

a. We would find it easy to recognize as sarcasm the statement "I'm really loving the enormous pool at my hotel!", if this statement is accompanied by a photo of a tiny swimming pool; whereas an automated sentiment analysis tool probably would not, and would most likely classify it as an example of positive sentiment.
b. With short sentences and pieces of text, for example like those you find on Twitter especially, and sometimes on Facebook, there might not be enough context for a reliable sentiment analysis.
c. Languages – Many services only operate in English, ignoring the other 6999 languages (yes – that’s how many there are) spoken in the world.

8. APPLICATIONS

A Chatbot is an artificial intelligence software that can simulate a conversation with a user in natural language through messaging applications,websites,app or through the telephone.

The following examples highlight interesting applications of Chatbots in businesses across various industries –

a. Order Food: Various fast food giants like KFC and Pizza Hut have invested in Chatbots that enable customers to place their orders through conversations. Taco Bell went a step further to improve the conversational experience by giving their Chatbot named Taco Bell some personality. It cracks jokes, uses emojis, answers trivia questions, and will even add a cup of water to an order if the customer mentions being hungover.
b. Content delivery:Media Publishers have realized that chatbots are a powerful way to engage with their audiences and monitor engagement to gain valuable insights on reader interests. Chat with the CNN and Wall Street Journal Chatbots on Facebook Messenger and receive the latest news directly in Messenger, without having to visit their websites.
c. A full fledged software or website with real-time applications
d. Getting news articles fast and reliable with score and other metadata informations
e. Can be used as an everyday-app to browse and get news real-time, anytime in the day
f. Sentiment analysis on the data

The following examples highlight interesting applications of Sentiment Analysis in businesses across various industries –

a. Sentiment analysis has many applications and benefits to your business and organization. It can be used to give your business valuable insights into how people feel about your product brand or service.

b. When applied to social media channels, it can be used to identify spikes in sentiment, thereby allowing you to identify potential product advocates or social media influencers.

c. It can be used to identify when potential negative threads are emerging online regarding your business, thereby allowing you to be proactive in dealing with it more quickly.

9. CONCLUSION

This application uses individual out-of-the-box UI components to extract and visualize the enriched data provided by the Discovery analytics engine. The web application includes Sentiment analysis,slack integration and embedded using Node-Red.

10. FUTURE SCOPE

Improving result relevance with the tooling:

a. When we provide a Discovery instance with training data, the service uses machine-learning Watson techniques to find signals in your content and questions. As you add more training data, the service instance becomes more accurate and sophisticated in the ordering of results it returns.

Integrating with other services:

The real magic of a bot comes when it is connected with external services, providing a seamless conversational interface for them from within Slack.

To use Continuous Relevancy Training:

a. Once trained, Continuous Relevancy Training is used to influence the results of a natural language query when using an environment-level query.

b. Continuous Relevancy Training can be used at query time by running a multi-collection natural language query across all collections in your environment.

Usage monitoring:

->Monitor and track usage of Discovery instance and use this data to improve applications. The Events API can be used to create log entries that are associated with specific natural language queries and actions. For example, one can record which documents in a results set were "clicked" by a user, and when that click occurred.

Making more complex bots:

a. News search using audio and converting that using speech-to-text. .
b. Using language translator to view news in regional languages.
c. Adding language translator with audio to enable search and results in regional language

Tracking conversations:

a. Effective Tracking of a conversation, the participants involved, and the progress through the flow. For example, when the user first mentions the bot, a database entry is created that identifies that user and the open workflow with them.

b. As the user progresses through the flow, the database records this, and the user is unable to repeat earlier steps in the conversation. The workflow is completed, the database entry is also marked as complete, and the bot waits for another mention before starting anew.

Variations in phrasing:

a. The bot can get more complex by broadening its understanding of natural language queries to capture a wider range of potential trigger phrases. Alternatively, we can be more prescriptive about the exact phrasing to use, and provide user education to train correct usage.

11. BIBLIOGRAPHY

a. https://www.youtube.com/watch?v=LOCkV-mENq8&feature=youtu.be

b. https://github.com/

c. https://www.youtube.com/watch?v=w3jLJU7DT5E&feature=youtu.be

d. https://www.youtube.com/watch?v=7YUTc4Cigc8&feature=youtu.be

e. https://www.ibm.com/cloud/get-started

f. http://nodeme.blogspot.com/2017/02/getting-started-with-node-red-and-iot.html

g. https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/

h. https://www.w3schools.com/howto/howto_make_a_website.asp

i. https://www.youtube.com/watch?v=W3iPbFTAAds&feature=youtu.be

j. https://www.youtube.com/watch?v=W3iPbFTAAds&feature=youtu.be

k. https://discovery-news-demo.ng.bluemix.net/

APPENDIX

Source code can be found at:
https://eu-gb.git.cloud.ibm.com/160417733042/NodeREDMEMNA.git

Link for the video:

----------xxx----------

○