AI Powered News Search App

Project Report

Start Date: 27/04/2020
End Date: 17/05/2020

1. INTRODUCTION

Project ID: Sirat Baweja
Project Title: AI Powered News Search App
Maximum Duration: 30 Days
Project By: Sirat Baweja
Objective: To build a news mining web app using *Node-RED* to create the Web app. Using the *IBM Watson's Discovery Service* and *IBM Cloud* for the model's deployment and using *Slack* for the model's integration.

1.1 Overview

To start with this project, the basics of Node-RED and IBM Cloud's platform must be known. The end goal is to build your own news mining web application using Node-RED and the IBM Watson Discovery Service.

To do this, two different paths are taken:

PATH #1

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

- Deploying the app on IBM Cloud

PATH #2

- Using a Slack interface to query the data
- Pushing news alerts out to web notification
- Deploying the app on IBM Cloud

1.2 Purpose

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, which is why this app has been built.

Users can look up at the most popular article of the news topics they are interested in the click of a button and a few keystrokes using this application.

2. LITERATURE SURVEY

2.1 Existing Problem

Unfortunately, the web is under the control of huge cooperations that are hungry for more money and power. This means most of the content we see online is curated by these companies in a way that we consume more content, goods and services from these companies. This includes websites, news articles, youtube video recommendations and more. Even with unbiased web services, such as the web browser DuckDuckGo, we only get content that has better Search Engine Optimization (SEO) at the top of the results and sometimes, it just doesn't happen to be what we are looking for.

Analysing data is also a hard task. Just scraping data can be

such a huge task by itself, but then analysing it, creating a model out of it, figuring out how to deploy the model from the plethora of options ranging from Flask, Node.js, TFLite and more, and then finally distributing the hypothesis in a ready-to-use manner is not an easy task.

2.2 Proposed Solutions

To solve the above problem, a dashboard which collects information efficiently and easily is a requirement. Even bots on chat-apps such as Slack and Discord are great for accessing and extracting useful information from the web.

Creating a dashboard or an information scraping chatbot is not an easy task, but using less traditional methods can definitely reduce the workload of an individual.

This is where IBM's services come in to save the day.

IBM Cloud

The IBM cloud platform combines platform as a service (PaaS) with infrastructure as a service (IaaS) to provide an integrated experience. The platform scales and supports both small development teams and organizations, and large enterprise businesses. Globally deployed across data centers around the world, the solution you build on IBM Cloud spins up fast and performs reliably in a tested and supported environment you can trust.

The platform is built to support your needs whether it's working only in the public cloud or taking advantage of a multicloud deployment model. Whether you need to migrate apps to the cloud, modernize your existing apps by using cloud services, ensure data resiliency against regional failure, or leverage new paradigms and deployment topologies to innovate and build your cloud-native apps, the platform's open architecture is built to accommodate your use case.

IBM Watson (Discovery News)
Infusing IBM's Watson into your apps and workflows allows even those with no experience in the field of Artificial Intelligence to tap into organizational data and put AI to work across multiple departments – from finance, to customer care, to supply chain. With Watson, anyone can create better, more personalized experiences for customers and make smarter decisions based on deep insights from data. Using Watson's Discovery Service reduces the job of fetching data and creating complex models from scratch.

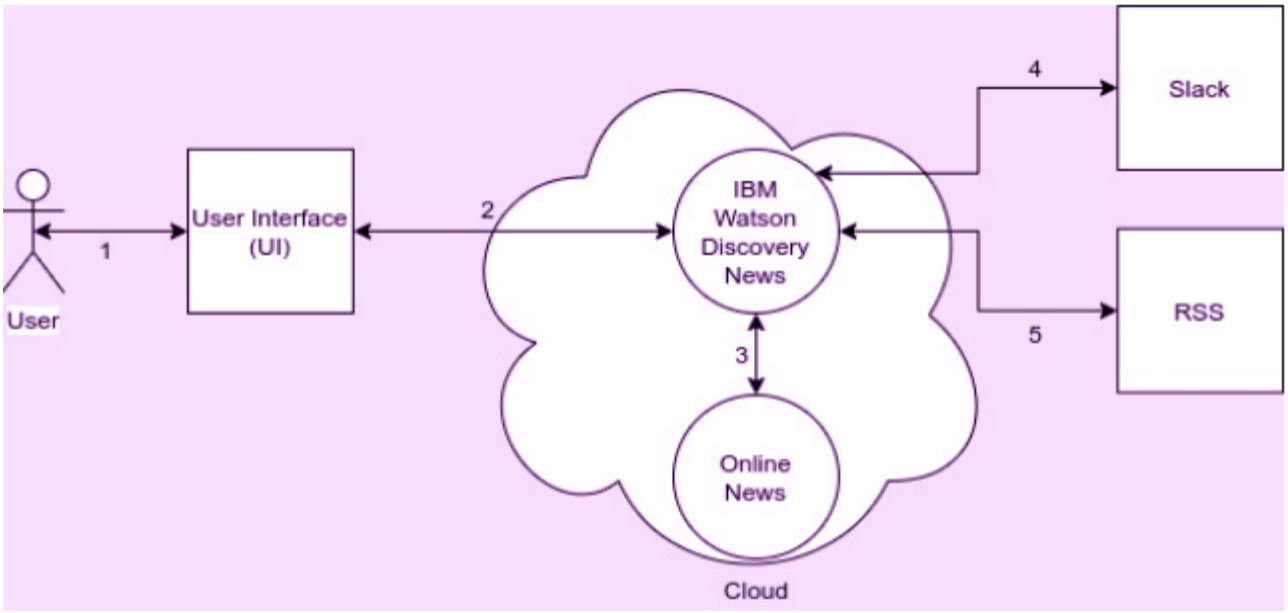
Node-RED
Node-RED is a programming tool for wiring APIs, hardware devices and online services using a browser-based editor which allows you to connects nodes with different functions together. Node-RED provides a browser-based flow editor that makes it easy to wire together flows using the wide range of nodes in the palette. Flows can be then deployed to the runtime in a single-click.

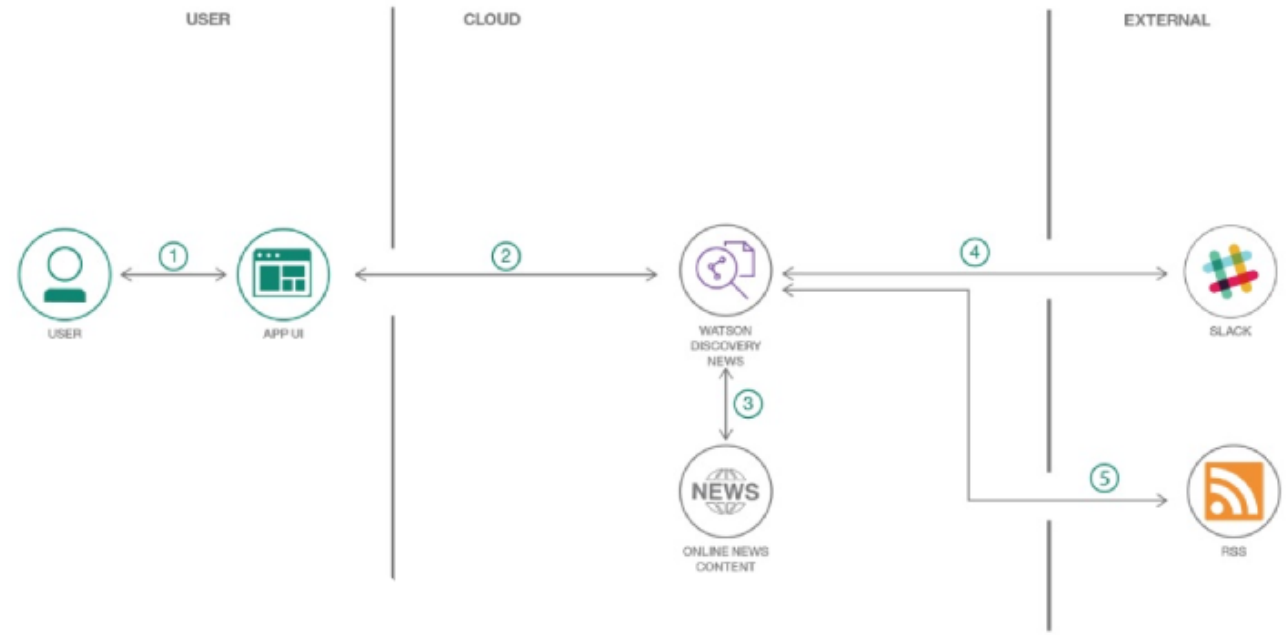
The flows created in Node-RED are stored using JSON which can be easily imported and exported for sharing with others.
An online flow library allows you to share your best flows with the world.

Slack (Service Integration)
Integrating this application's services with Slack just makes it more convinient, since you can just check the news while still working, without interrupting the flow of your thought.

3. THEORETICAL ANALYSIS

3.1 Block Diagram





3.2 Hardware/Software Design

Requirements and Specifications:

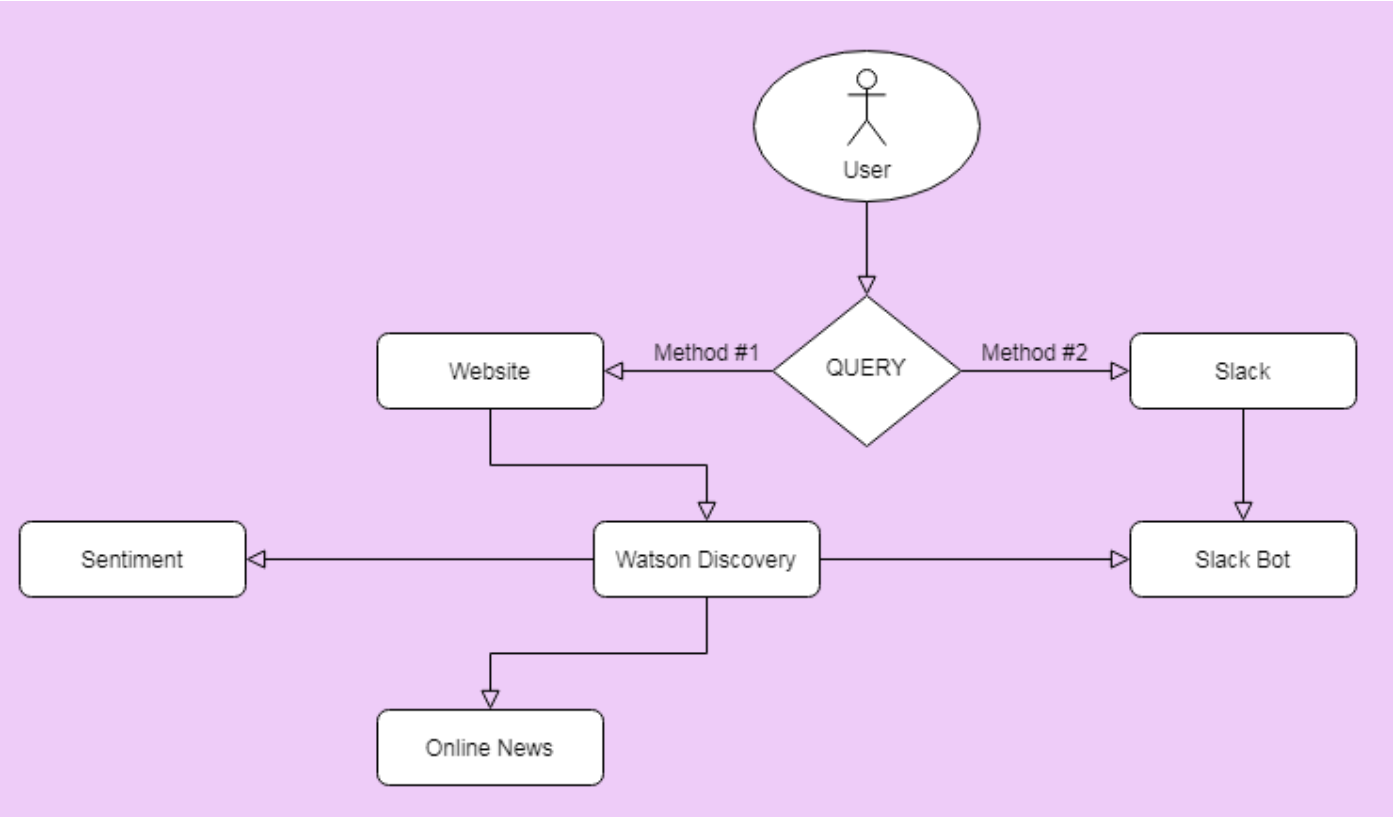
- An internet connection
- A PC (computer, laptop)
- Windows, MacOS or any recent Linux distribution
- Node 12.0.0^ (for Node-RED and local Slack Integration)
- npm 6.0.0^ (for Node-RED and local Slack Integration)
- IBM Cloud account (Lite, for all required services)
- Git (for version control)
- A Github account
- A Slack account

4. EXPERIMENTAL INVESTIGATIONS

While building this application, the following investigations were made:

- This application was built using Node-RED, which means that it's CORS (Cross-Origin Resource Sharing) functions and backend were made using Node.js and Express.js
- The frontend of this application was created with Angular.js, along with other frameworks like Chart.js and D3.js
- The IBM Watson Discovery News API is in use
- The IBM Cloud service (BlueMix) has been used to deploy the application
- Integration of Slack is done using the Node Package Manager

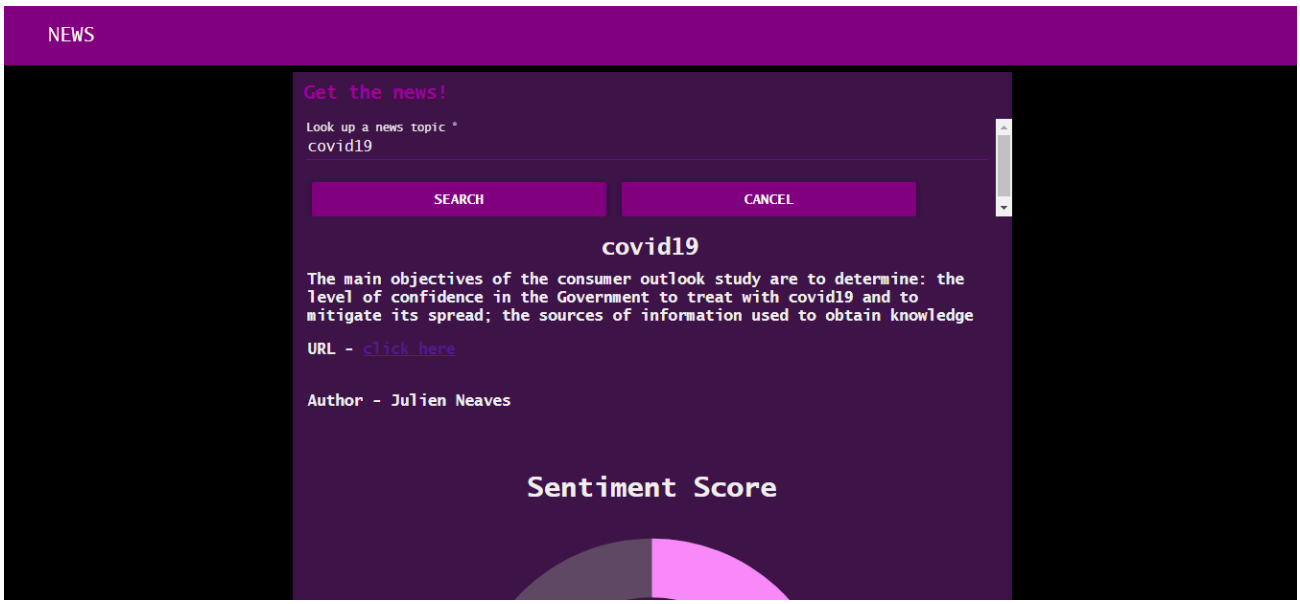
5. FLOWCHART

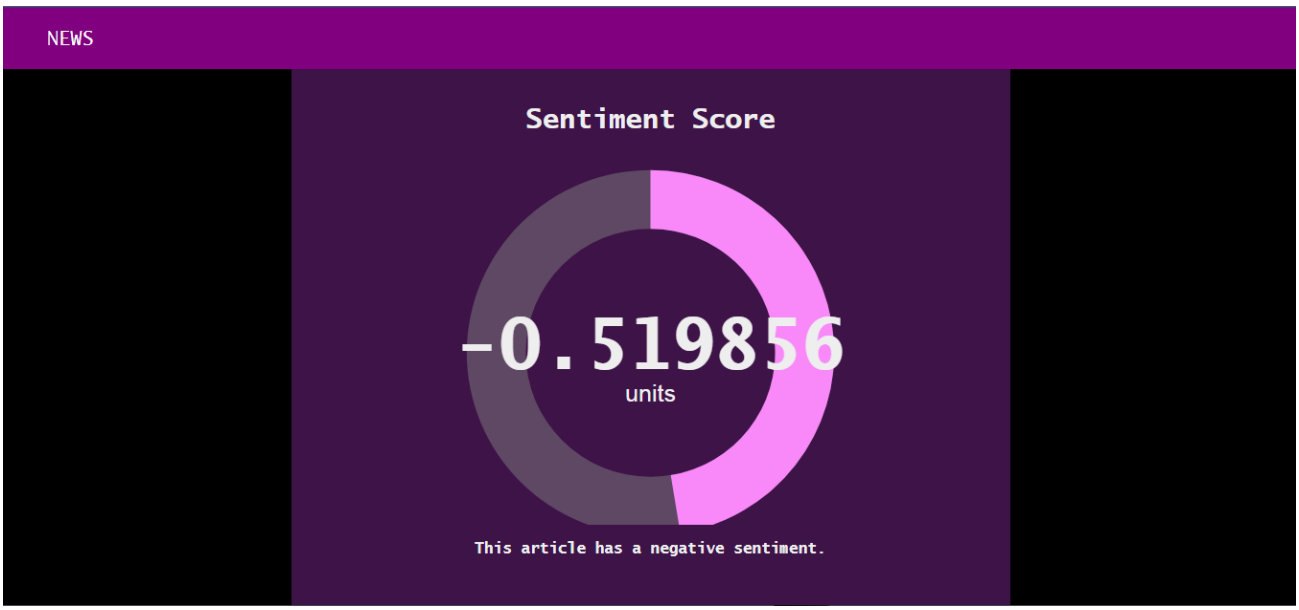


6. RESULT

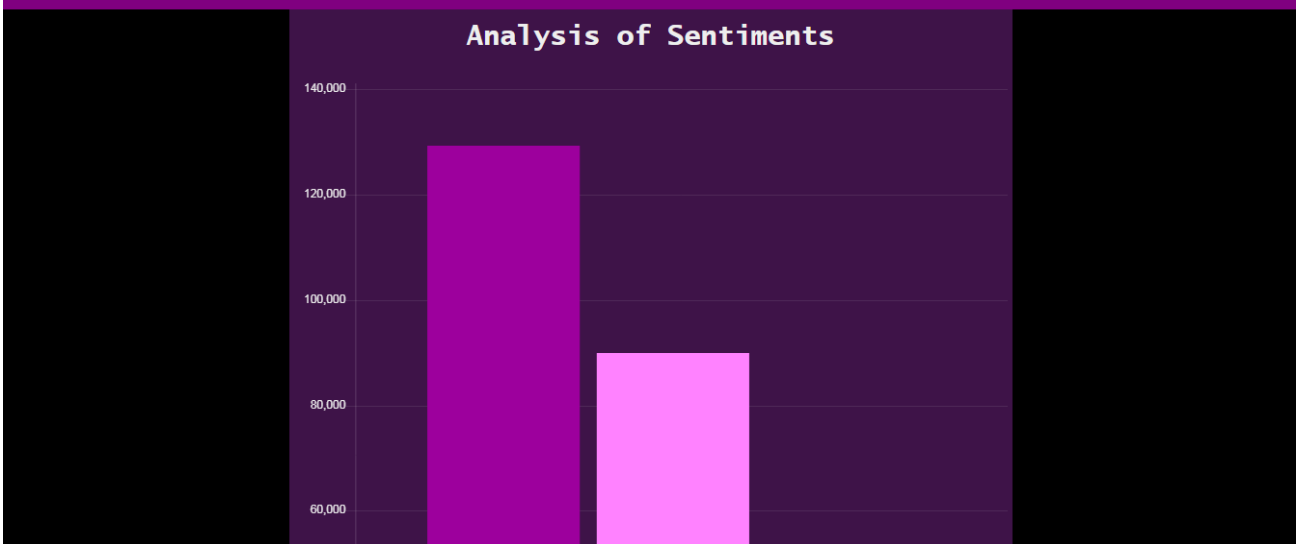
The top article relating to the keywords entered in the application, along with it's sentiment, is successfully found in a few keystrokes and a click of a button and integrated in a Slack bot.

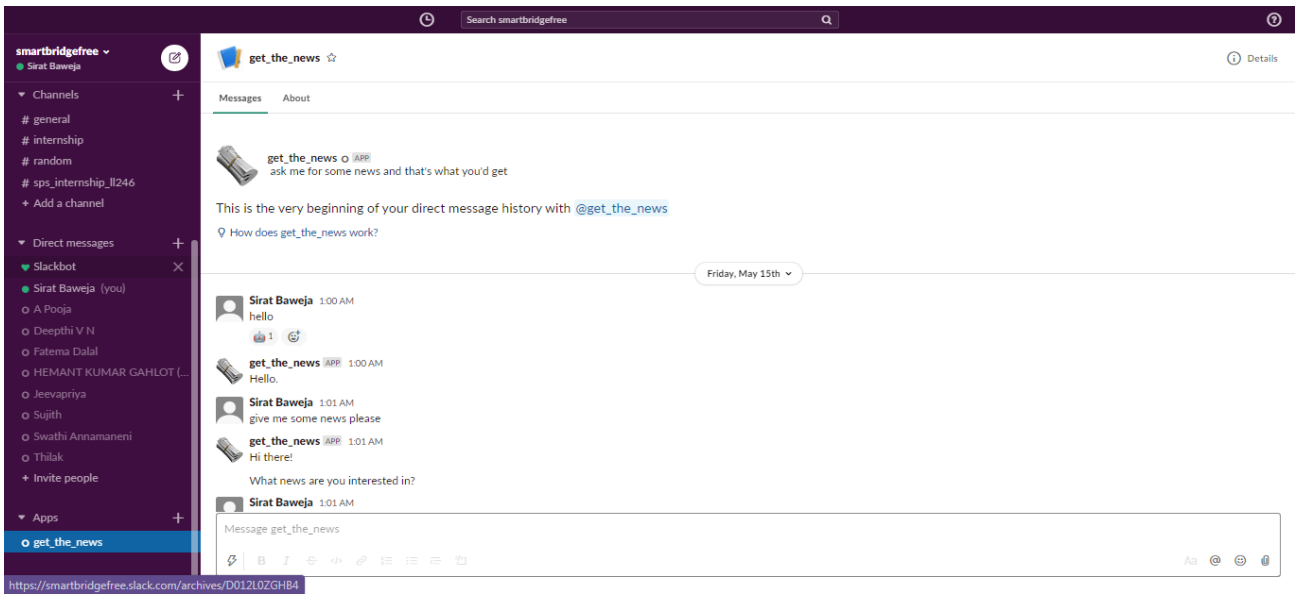
The app can be found at https://getthenews.eu-gb.mybluemix.net/ui

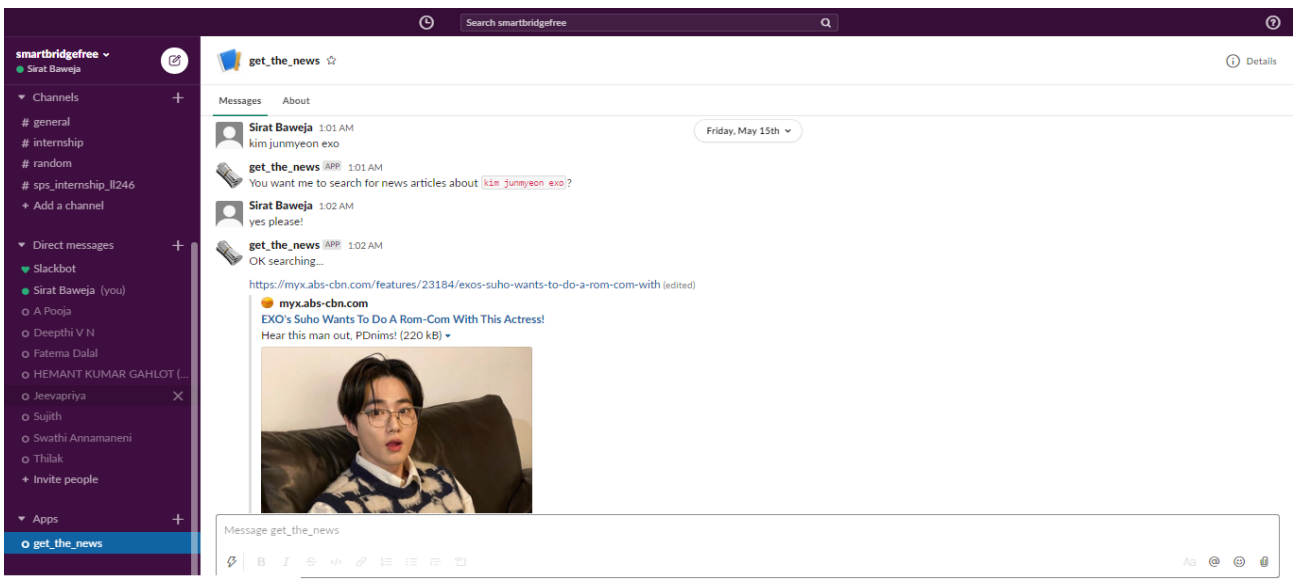




NEWS


Analysis of Sentiments
140,000
120,000
100,000
80,000
60,000


Search smartbridgefree
smartbridgefree
Sirat Baweja
Channels
general
internship
random
sps_internship_ll246
+ Add a channel
Direct messages
Slackbot
Sirat Baweja (you)
A Pooja
Deepthi V N
Fatema Dalal
HEMANT KUMAR GAHLOT (...
Jeevapriya
Sujith
Swathi Annamaneni
Thilak
+ Invite people
Apps
get_the_news
get_the_news
Details
Messages
About
get_the_news APP
ask me for some news and that's what you'd get
This is the very beginning of your direct message history with @get_the_news
How does get_the_news work?
Friday, May 15th
Sirat Baweja 1:00 AM
hello
get_the_news APP 1:00 AM
Hello.
Sirat Baweja 1:01 AM
give me some news please
get_the_news APP 1:01 AM
Hi there!
What news are you interested in?
Sirat Baweja 1:01 AM
Message get_the_news
https://smartbridgefree.slack.com/archives/D012L0ZGHB4


Search smartbridgefree
smartbridgefree
Sirat Baweja
Channels
general
internship
random
sps_internship_ll246
+ Add a channel
Direct messages
Slackbot
Sirat Baweja (you)
A Pooja
Deepthi V N
Fatema Dalal
HEMANT KUMAR GAHLOT (...
Jeevapriya
Sujith
Swathi Annamaneni
Thilak
+ Invite people
Apps
get_the_news
get_the_news
Details
Messages
About
Friday, May 15th
Sirat Baweja 1:01 AM
kim junmyeon exo
get_the_news APP 1:01 AM
You want me to search for news articles about kim junmyeon exo?
Sirat Baweja 1:02 AM
yes please!
get_the_news APP 1:02 AM
OK searching...
https://myx.abs-cbn.com/features/23184/exos-suho-wants-to-do-a-rom-com-with (edited)
myx.abs-cbn.com
EXO's Suho Wants To Do A Rom-Com With This Actress!
Hear this man out, PDnims! (220 kB)
Message get_the_news

7. ADVANTAGES AND DISADVANTAGES

Advantages

- The application is easy to navigate
- There is an improved interaction with customers
- The articles are not promoted forward by big cooperations, which means they are unbiased and accurate
- There is sentiment analysis provided for each article so that the user can analyse whether they want to read the article at all.

Disadvantages

- There are a limited number of requests all users can make, since this application has been created on IBM Cloud 'Lite' account.

8. APPLICATIONS

- This application can easily be scaled up to provide multiple top-trending articles.
- This application can be scaled up to be used in multiple different types of Slack workspaces.
- It can be modified to provide different types of services to keep customers engaged.

9. CONCLUSION

This application is created for the ease of looking up for specific interesting news topics quickly and efficiently, whether it be while you're just casually searching the internet, or if you're discussing important problems in your slack workspace. It runs on Artificial Intelligence algorithms created to optimize and find the best news articles, provided by IBM's Watson Discovery News service. Using IBM's Cloud's services such as the one above, its Cloudant and Node RED, this application was successfully developed.

10. FUTURE SCOPE

In the future, this application can be developed in a more customizable way, where the UI is responsive, there are text to speech (and vice verse) features provided, it can be developed into a cross-platform mobile application with notifications based on recommendation systems made using the user's data to find their news preferences.

11. BIBLIOGRAPHY

The following sources were used for the development of this application:

1. Project Kickoff Template - https://www.allbusinesstemplates.com/download/?filecode=2KBA4&lang=en&iuid=9f9faa69-9fab-40ee-8457-ea0e5df8c8de
2. How to Kickoff a Project - https://www.youtube.com/watch?v=qxjuo4Vnp6U
3. What is Github? - https://www.youtube.com/watch?v=w3jLJU7DT5E&feature=youtu.be
4. Slack for Beginners - https://www.youtube.com/watch?v=7YUTc4Cigc8&feature=youtu.be
5. Slack - https://slack.com/intl/en-in/
6. Zoho Document Writer - https://www.zoho.com/writer/help/working-with-text.html
7. IBM Cloud - https://cloud.ibm.com/login
8. IBM Student Initiative - https://my15.digitalexperience.ibm.com/b73a5759-c6a6-4033-ab6b-d9d4f9a6d65b/dxsites/151914d1-03d2-48fe-97d9-d21166848e65/home
9. Getting Started with IBM Cloud - https://www.ibm.com/cloud/get-started
10. Node-RED Starter Tutorial -

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

11. Node-RED - https://nodered.org/
12. Node-RED Labs - https://github.com/watson-developer-cloud/node-red-labs
13. APIs - https://www.youtube.com/watch?v=s7wmiS2mSXY&feature=youtu.be
14. Creating a Simple Webpage - https://www.w3schools.com/howto/howto_make_a_website.asp
15. IBM Watson (Products and Services) - https://www.ibm.com/watson/products-services
16. Watson at Work - https://www.youtube.com/watch?v=W3iPbFTAAds&feature=youtu.be
17. Introduction to Watson Discovery - https://developer.ibm.com/articles/introduction-watson-discovery/
18. Getting Started with Watson Discovery - https://cloud.ibm.com/docs/services/discovery?topic=discovery-getting-started
19. IBM’s Watson Discovery - https://www.y.com/watch?v=kwmqJRDbv98&feature=youtu.be
20. Querying News - https://cloud.ibm.com/docs/discovery?topic=discovery-query-concepts#querying-news
21. Watson Discovery Demo App - https://discovery-news-demo.ng.bluemix.net/

APPENDIX

Source Code

flow.json (the Node-RED flow)

```
[
    {
        "id": "afe6bac3.54c008",
        "type": "tab",
        "label": "main",
        "disabled": false,
        "info": ""
    },
    {
        "id": "ab45344c.5d1968",
        "type": "ui_form",
        "z": "afe6bac3.54c008",
        "name": "news ->",
        "label": "",
        "group": "5551d7cf.02f518",
        "order": 1,
        "width": "0",
        "height": "0",
        "options":
        [
            {
                "label": "Look up a news topic ",
                "value": "Input",
                "type": "text",
                "required": true,
                "rows": null
            }
        ],
        "formValue":
        {
            "Input": ""
        },
        "payload": "",
        "submit": "Search",
        "cancel": "Cancel",
        "topic": "",
        "x": 120,
        "y": 40,
        "wires":
        [
            [
                "190a1634.a676aa",
                "b06959eb.dc27c8"
            ]
        ]
    },
    {
```

```
        "id": "190a1634.a676aa",
        "type": "function",
        "z": "afe6bac3.54c008",
        "name": "query ->",
        "func": "msg={\ndiscoveryparams:\n
{\n\"environment_id\":\"system\",\n\"query\":msg.payload.Input\n  }\n}\nreturn msg;",
        "outputs": 1,
        "noerr": 0,
        "x": 180,
        "y": 100,
        "wires":
        [
            [
                "57e3d18d.d59dd"
            ]
        ]
    },
    {
        "id": "887f4ff5.2d7e",
        "type": "function",
        "z": "afe6bac3.54c008",
        "name": "article",
        "func": "msg.payload=msg.search_results.results[0].text\nreturn msg;",
        "outputs": 1,
        "noerr": 0,
        "x": 490,
        "y": 100,
        "wires":
        [
            [
                "68b81d6e.e12174"
            ]
        ]
    },
    {
        "id": "57e3d18d.d59dd",
        "type": "watson-discovery-v1",
        "z": "afe6bac3.54c008",
        "name": "discovery 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/a0048099-cf37-402e-b871-321
43f8cc94e",
        "x": 260,
        "y": 160,
        "wires":
        [
            [
                "887f4ff5.2d7e",
                "34fe010a.18dd5e",
                "f7ccfa88.086828",
                "d8e0a9cf.9e1488",
                "5a15eeb.b8a881",
                "1cc689aa.34edf6"
            ]
        ]
    },
    {
        "id": "68b81d6e.e12174",
        "type": "ui_text",
        "z": "afe6bac3.54c008",
        "group": "5551d7cf.02f518",
        "order": 3,
        "width": "0",
        "height": "0",
        "name": "-> article",
        "label": "",
        "format": "<br><br>{{msg.payload}}",
        "layout": "row-left",
        "x": 700,
        "y": 100,
        "wires": [],
        "info": "# The Article"
    },
    {
        "id": "d8e0a9cf.9e1488",
        "type": "function",
        "z": "afe6bac3.54c008",
        "name": "url",
        "func": "msg.payload=msg.search_results.results[0].url\nreturn msg;",
        "outputs": 1,
        "noerr": 0,
        "x": 490,
        "y": 160,
        "wires":
        [
            [
                "8efc2266.52018"
            ]
        ]
```

```
    },
    {
        "id": "34fe010a.18dd5e",
        "type": "function",
        "z": "afe6bac3.54c008",
        "name": "author",
        "func": "msg.payload=msg.search_results.results[0].author\nreturn msg;",
        "outputs": 1,
        "noerr": 0,
        "x": 490,
        "y": 220,
        "wires":
        [
            [
                "541d824d.d5bb1c"
            ]
        ]
    },
    {
        "id": "f7ccfa88.086828",
        "type": "function",
        "z": "afe6bac3.54c008",
        "name": "score",
        "func":
"msg.payload=msg.search_results.results[0].enriched_text.sentiment.document.score\nretu
rn msg;",
        "outputs": 1,
        "noerr": 0,
        "x": 490,
        "y": 280,
        "wires":
        [
            [
                "ab8f0932.c5b2e8"
            ]
        ]
    },
    {
        "id": "5a15eeb.b8a881",
        "type": "function",
        "z": "afe6bac3.54c008",
        "name": "label",
        "func":
"msg.payload=msg.search_results.results[0].enriched_text.sentiment.document.label\nretu
rn msg;",
        "outputs": 1,
        "noerr": 0,
        "x": 490,
        "y": 340,
        "wires":
        [
            [
                "220adda1.abf702"
```

```
            ]
        ]
    },
    {
        "id": "220adda1.abf702",
        "type": "ui_text",
        "z": "afe6bac3.54c008",
        "group": "5551d7cf.02f518",
        "order": 7,
        "width": "14",
        "height": "1",
        "name": "-> label",
        "label": "",
        "format": "This article has a {{msg.payload}} sentiment.",
        "layout": "row-center",
        "x": 700,
        "y": 340,
        "wires": [],
        "info": "# The Sentiment Label\n### This can either be positive, neutral or
negative"
    },
    {
        "id": "541d824d.d5bb1c",
        "type": "ui_text",
        "z": "afe6bac3.54c008",
        "group": "5551d7cf.02f518",
        "order": 5,
        "width": "14",
        "height": "1",
        "name": "-> author",
        "label": "",
        "format": "Author - {{msg.payload}}",
        "layout": "row-left",
        "x": 700,
        "y": 220,
        "wires": [],
        "info": "# The Author"
    },
    {
        "id": "ab8f0932.c5b2e8",
        "type": "ui_gauge",
        "z": "afe6bac3.54c008",
        "name": "-> score",
        "group": "5551d7cf.02f518",
        "order": 6,
        "width": "14",
        "height": "9",
        "gtype": "donut",
        "title": "<br><br><h1 style=\"font-weight:bold\">Sentiment Score</h1>",
        "label": "units",
        "format": "{{value}}",
        "min": "-10",
        "max": "10",
```

```
        "colors":
        [
            "#9d009d",
            "#ff91ff",
            "#d700d7"
        ],
        "seg1": "",
        "seg2": "",
        "x": 700,
        "y": 280,
        "wires": [],
        "info": "# The Sentiment Score \n### Ranging from -2 to +2"
    },
    {
        "id": "518c773a.00c868",
        "type": "function",
        "z": "afe6bac3.54c008",
        "name": "->",
        "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": 650,
        "y": 400,
        "wires":
        [
            [
                "99e240c8.28a57"
            ]
        ]
    },
    {
        "id": "99e240c8.28a57",
        "type": "ui_chart",
        "z": "afe6bac3.54c008",
        "name": "analysis",
        "group": "5551d7cf.02f518",
        "order": 8,
        "width": "14",
        "height": "16",
        "label": "<h1 style=\"font-weight:bold\">Analysis of Sentiments</h1>",
        "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":
        [
            "#9d009d",
            "#ff82ff",
            "#d900d9",
            "#2ca02c",
            "#98df8a",
            "#d62728",
            "#ff9896",
            "#9467bd",
            "#c5b0d5"
        ],
        "useOldStyle": false,
        "outputs": 1,
        "x": 700,
        "y": 460,
        "wires":
        [
            []
        ]
    },
    {
        "id": "1cc689aa.34edf6",
        "type": "function",
        "z": "afe6bac3.54c008",
        "name": "sentiment",
        "func": "var
val=[msg.search_results.aggregations[0].results[0].matching_results];\nval.push(msg.sea
rch_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": 500,
        "y": 400,
        "wires":
        [
            [
                "518c773a.00c868"
            ]
        ]
    },
    {
        "id": "8efc2266.52018",
        "type": "ui_text",
        "z": "afe6bac3.54c008",
        "group": "5551d7cf.02f518",
        "order": 4,
        "width": "14",
        "height": "1",
        "name": "-> url",
```

```
        "label": "",
        "format": "URL - <a href={{msg.payload}}>click here</a>",
        "layout": "row-left",
        "x": 690,
        "y": 160,
        "wires": []
    },
    {
        "id": "a5ccfae8.9c7258",
        "type": "ui_text",
        "z": "afe6bac3.54c008",
        "group": "5551d7cf.02f518",
        "order": 2,
        "width": "14",
        "height": "1",
        "name": "-> title",
        "label": "",
        "format": "<h1 style=\"text-transform: uppercase\">{{msg.payload}}</h1>",
        "layout": "col-center",
        "x": 690,
        "y": 40,
        "wires": [],
        "info": "# The Article"
    },
    {
        "id": "b06959eb.dc27c8",
        "type": "function",
        "z": "afe6bac3.54c008",
        "name": "extract input as title",
        "func": "msg.payload=msg.payload.Input\nreturn msg;",
        "outputs": 1,
        "noerr": 0,
        "x": 450,
        "y": 40,
        "wires":
        [
            [
                "a5ccfae8.9c7258"
            ]
        ]
    },
    {
        "id": "5551d7cf.02f518",
        "type": "ui_group",
        "z": "",
        "name": "Get the news!",
        "tab": "d8108a45.772378",
        "order": 1,
        "disp": true,
        "width": "14",
        "collapse": false
    },
    {
```

```
        "id": "d8108a45.772378",
        "type": "ui_tab",
        "z": "",
        "name": "NEWS",
        "icon": "news",
        "disabled": false,
        "hidden": false
    }
]
```