SmartBridge

Project report

AI-POWERED-NEWS-APP

Raesah Sheriff
5/20/2020

1. INTRODUCTION

1.1 Overview

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 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.

1.2 Purpose:

The purpose of the project is 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 customer insight on a massive scale and ensure that you don't miss a single conversation.

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.

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

1) **<u>Too much information</u>** – It is difficult for the newsreader to rely on search engines to index, sort and rank externally-produced content that isn't standardized for format. There's still too much information and the organization is distorted by advertising dollars.

2) **<u>News is time-sensitive information</u>** – 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.

3) **<u>News is geographically sensitive information</u>** – The news content has to sort through local to find local news. While GeoRSS is out there, it's still nowhere near being adopted as a widely-used content standard.

4) **<u>Distilling quality news</u>** – 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.

2.2 Proposed solution

To use Watson Discovery Service:

1) It continually crawls into the web to update its Discovery News collection. 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 natural language processing capability represents the first commercialization of IBM Research's Project Debater, billed as an AI system capable of debating humans.

2) The tool is designed to help develop arguments that could ultimately be used to improve machine understanding and help analyze complex human speech.

3) It's still impossible to analyze it manually without any sort of error or bias. Oftentimes, companies with the best intentions find themselves in an insights vacuum. Insights are necessary to inform your decision making and you know that you're lacking them, but don't know how best to get them. The integration of Project Debater technology with Watson includes a sentiment analysis and "summarization" tools along with the ability to cluster data by topic.

4) Sentiment analysis is implemented and is important because companies want their brand being perceived positively, or at least more positively than the brands of competitors. Sentiment analysis, if accurate, can be a very valuable tool for this specific use case.

5) Sentiment analysis provides some answers into what the most important issues are, from the perspective of customer. As sentiment analysis can be automated, decisions can be made based on a significant amount of data rather than plain intuition that isn't always right.

6) The sentiment analyzer can detect "sentiment shifters," so-called because a negative modifier can reverse a speaker's meaning using a phase like "hardly helpful."

7) 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.

8) IBM's Watson Discovery enterprise search platform, the data organizer is aimed at subject experts in specific industries like health care, insurance and manufacturing.

3. THEORITICAL ANALYSIS

3.1 Block diagram
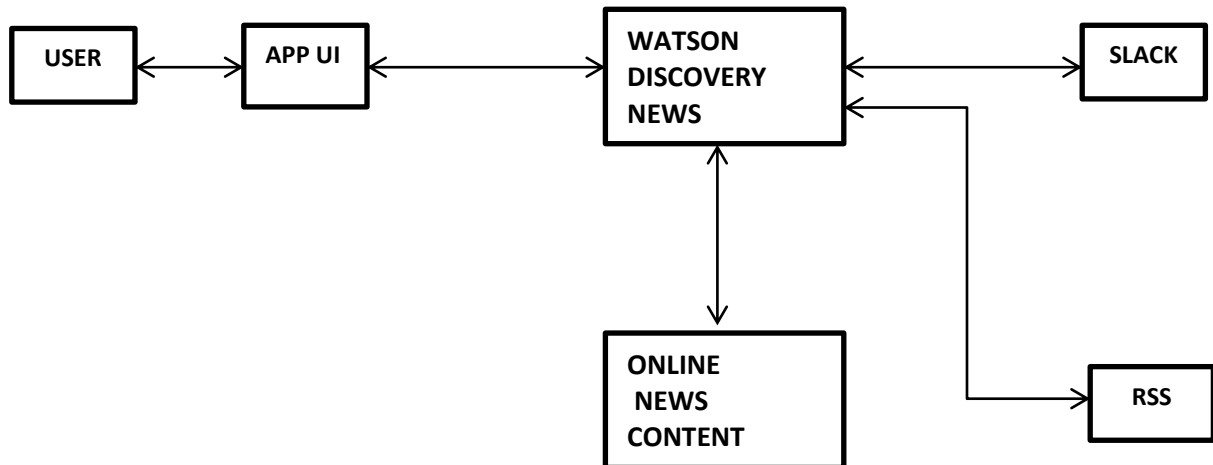
3.2 Hardware / Software designing

- IBM Cloud
- Node Red
- IBM Watson Discovery
- Node.js
- GitHub
- Zoho Writer
- Slack
- Html
- RSS

4. EXPERIMENTAL INVESTIGATIONS

This project used 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. Generally, news stories appearing on a web site were evaluated as less credible than their identical counterparts on paper.

In the new digital era, brands and governments are challenged with how to manage large volumes of data and controversial news and how to address followers and skeptics. This project experimented with cognitical solution that it can deliver immediate insight with great precision and a remarkable level of detail. It offers qualitative data about people's opinions and feelings. The solution processes huge amounts of structured and unstructured data. The process is automated, so it's fast. To compile a similar sentiment analysis report manually would take approximately a week, and by that time, the data would be old. Cognitical include solutions that evaluate sarcasm, satire and emotion.

This project includes a bot, which is a type of Slack App designed to interact with users via conversation, it 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. It can access the same range of APIs and gives the app a face, a name, and a personality, and encouraging users to talk to it.

5. FLOWCHART

Flow
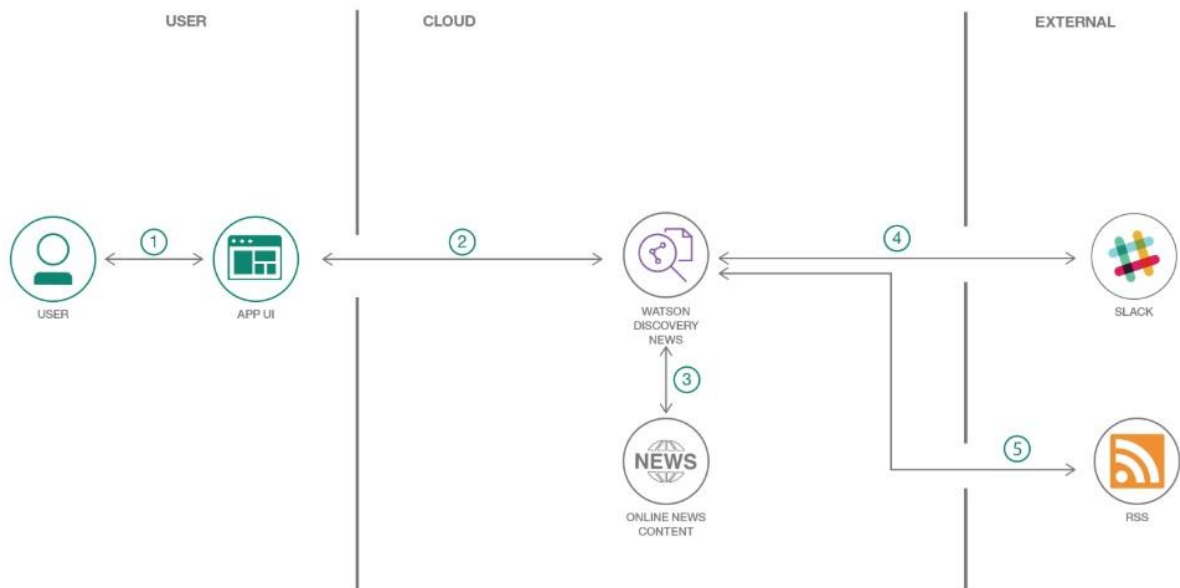
6. RESULT

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

Screenshot 1 (top):

Search Beyond_infy   Skip tutorials

Beyond_infy ˅
raesah_sheriff

@ Mentions & reactions
People   + Add
Apps
Files
↑ Show less

˅ Channels   +
project
team
welcome

˅ Direct messages   +
○ syedanawazahmadi27
+ Invite people

˅ Apps   +
● infintybot

infintybot ☆   ⓘ Details

Messages   About

raesah_sheriff 12:28 AM   Wednesday, May 6th ˅
@infintybot hi
👍 1 😊

infintybot APP 12:28 AM
Hello.

raesah_sheriff 12:28 AM
@infintybot news please

infintybot APP 12:28 AM
Hi there!
What news are you interested in?

raesah_sheriff 12:29 AM
cricket

infintybot APP 12:29 AM
You want me to search for news articles about `cricket` ?

raesah_sheriff 12:29 AM
yes

infintybot APP 12:29 AM
OK searching...

Message infintybot

B I S </> 🔗 ≔ ≔ ≔ ⏱   Aa @ 😊 📎

Screenshot 2 (bottom):

Search Beyond_infy   Skip tutorials

Beyond_infy ˅
raesah_sheriff

@ Mentions & reactions
People   + Add
Apps
Files
↑ Show less

˅ Channels   +
project
team
welcome

˅ Direct messages   +
○ syedanawazahmadi27
+ Invite people

˅ Apps   +
● infintybot

infintybot ☆   ⓘ Details

Messages   About

Hussain Stresses on Importance of Pitch Quality, Laxman Not a Fan of F...   Wednesday, May 6th ˅
Nasser Hussain and VVS Laxman speak on Test Cricket.
Apr 14th (32 kB) ˅

https://www.news18.com/news/buzz/guess-the-cricketer-icc-is-back-with-another-fun-quiz-and-it-is-not-as-easy-as-you-thought-2575547.html (edited)
18 News18
Guess the Cricketer: ICC is Back With Another Fun Quiz and It is Not As Easy As You Thought
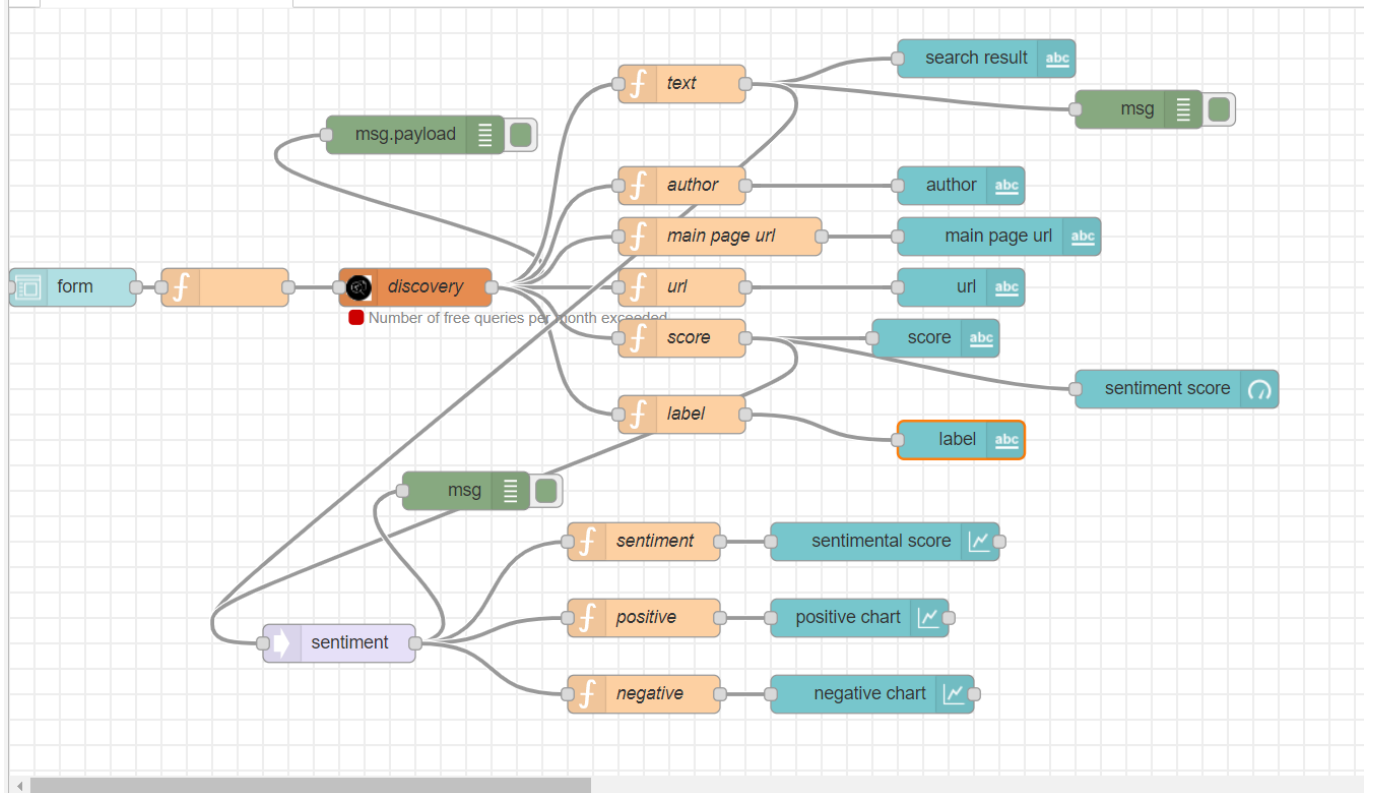The cricket council has hidden a cricketer in an optical illusion and challenged netizens to identify the player.
Apr 13th (51 kB) ˅

Message infintybot

B I S </> 🔗 ≔ ≔ ≔ ⏱   Aa @ 😊 📎

Flow 2

form → f → discovery
🔴 Number of free queries per month exceeded

msg.payload

f text → search result abc
→ msg

f author → author abc

f main page url → main page url abc

f url → url abc

f score → score abc → sentiment score

f label → label abc

msg

f sentiment → sentimental score
f positive → positive chart
sentiment
f negative → negative chart

IBM

Enter the input *

covid-19

SUBMIT	CANCEL

search result

) Nucleocapsid (N) (SARS-CoV-2; COVID-19) Research and Development on COVID-19 Diagnostics and Medtech The Need to Reinvent Ventilators COVID-19 Clinical Trial Landscape Chapter 4 Economic Impact of COVID-19 Impact of Coronavirus on the Global Economy Government Expenditures on COVID-19 Predictions on the Global Economy Quick Recovery Chapter 5 Impact of
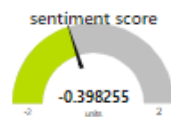
author

main page url

url https://www.marketresearch.com/BCC-Research-v374/COVID-Fallout-BCC-Research-Economic-13228708/
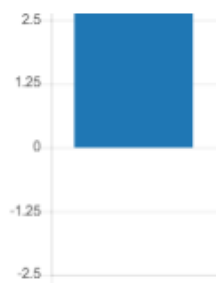
score   -0.398255

label   negative

sentiment score

-0.398255
units

sentimental score   positive chart   negative chart

7. ADVANTAGES & DISADVANTAGES

ADVANTAGES

- By querying and manipulating enriched data, you can build a more insightful search interface. This code pattern provides a Node.js app built on the Watson Discovery Service that does just that. The pattern demonstrates how you can use individual out-of-the-box UI components to extract and visualize the enriched data provided by the Discovery analytics engine.

- Watson Discovery News is a service where Watson will automatically add news articles to a collection of documents that you can then query. New documents are added every day, which is great because it removes the legwork of having to gather these sources yourself.

- It uses data science and creates visualizations which lead to uncovering anomalies in the data and creating alerts when key events occur. The news provided is time-sensitive and hence generates the latest news.

- This project uses Watson Studio and Watson Discovery News to query articles that are linked to the location you reside in.

The main benefit of using the Watson Discovery Service in the project is its powerful analytics engine that provides cognitive enrichments and insights into your data. The app in this code pattern provides examples of how to showcase these enrichments through the use of filters, lists, and graphs. The key enrichments are:

1) **Entities:** People, companies, organizations, cities, and more.

2) **Categories:** Classification of the data into a hierarchy of categories up to 5 levels deep.

3) **Concepts:** Identified general concepts that aren't necessarily referenced in the data.

4) **Keywords:** Important topics typically used to index or search the data.

5) **Sentiment:** The overall positive or negative sentiment of each document.

6) The app uses standard search UI components such as filter list, 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.

- 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 slack bot created listens to messages after specifying the keywords you want to listen to and after specifying whether those keywords are directed to the bot user or the Slack channel. This process leads to the generation of news using an interactive method making the user experience a new way of searching news.

DISADVANTAGES:

- With short sentences and pieces of text, there might not be enough contexts for a reliable sentiment analysis. So, automated sentiment analysis tools do a really great job of analyzing text for opinion and attitude, but they're not perfect.

- Like other APIs and integrations, bot users are free. Unlike regular users, the actions they can perform are somewhat limited. For workspaces on the Free plan, each bot user counts as a separate integration.

- Application Programming Interface (API) is not stable; it keeps on changing at frequent intervals and does not remain stable.

- 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.

- Bots are installed with the motive to speed-up the response and improve customer interaction. However, due to limited data-availability and time required for self-updating, this process may appear time-taking and

expensive. Therefore, in place of attending several customers at a time, Bots appear confused about how to communicate with people.

8. APPLICATIONS

- 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.
- Market Research
  The way market researchers are having a dedicated survey app, many research firms are developing Chatbots to have personalized, engaging conversational surveys with respondents to improve the experience and increase completion rates.
- 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.
- 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.
- 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 code pattern includes Watson Discovery series. It gets the customer sentiment insights from product reviews.

10. FUTURE SCOPE

Improve Querying of news using IBM Watson:

- Provide Discovery instance with training data with more training data, the service instance becomes more accurate and sophisticated in the ordering of results it returns.

- Once trained, Continuous Relevancy Training is used to influence the results of a natural_language_query when using an environment-level query. Continuous Relevancy Training can be used at query time by running a multi-collection natural_language_query across all collections in your environment

- You can monitor and track usage of your Discovery instance and use this data to help you understand and improve your applications. The Events API can be used to create log entries that are associated with specific natural language queries and actions. For example, you can record

which documents in a results set were "clicked" by a user, and when that click occurred.

Making more complex bots

In the current project, a lot of assumptions of simplicity were made. For a real bot in production, some of these assumptions would break the behavior of the bot. To improve:

- Tracking conversations
  Involve tracking the beginning of a conversation, the participants involved, and the progress through the flow.

- Threaded messages
  Be aware that a user might choose to reply to your bot's messages in a thread rather than at the channel-level, add some extra logic to ensure that your bot responds to the user in the relevant location.

- Variations in phrasing
  Bots must broaden its understanding of natural language queries to capture a wider range of potential trigger phrases.

- 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 like the Botkit Community.

11. BIBLIOGRAPHY

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

https://github.com/

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

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

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

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

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

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

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

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

Appendix

 a) Source Code

https://eu-gb.git.cloud.ibm.com/160417733004/NodeREDAJAHK.git