



Conference Rating

By Ankur Gupta

San Jose State University, India

Abstract- Conf Rate is a website designed to gather tweets regarding various conferences occurring around the corners of world and applying an algorithm to rate them on the sentimental analysis of tweets. When we talk of technolog stack implementation, Conf Rate uses Twython; a python based twitter API which connects to twitter on oauth (open authentication) and offers functionality like twitter search. The technique is to use the real time data from the twitter; commonly known as “tweets”. This information is categorized into positive or negative on the basis of the “tweet sentiments”. The Natural Language Toolkit (NLTK) is a powerful tool in Python which identifies tweet sentiments using Bayes classifier. The use of Django 1.3.1 web framework (a python framework) provides a lightweight web server with MVC framework which easily integrates with our python scripts. For designing user interface, html5 using jquery has been used. Finally, the application is deployed on the cloud using AWS (Amazon web services) Elastic Cloud Computer or commonly known as EC2, providing more profitability and reliable browsing and dependability from market perspective.

GJCST-E Classification : G.1.2



Strictly as per the compliance and regulations of:



RESEARCH | DIVERSITY | ETHICS

Conference Rating

Ankur Gupta

Abstract- Conf Rate is a website designed to gather tweets regarding various conferences occurring around the corners of world and applying an algorithm to rate them on the sentimental analysis of tweets. When we talk of technology stack implementation, Conf Rate uses Twython; a python based twitter API which connects to twitter on oauth (open authentication) and offers functionality like twitter search. The technique is to use the real time data from the twitter; commonly known as "tweets". This information is categorized into positive or negative on the basis of the "tweet sentiments". The Natural Language Toolkit (NLTK) is a powerful tool in Python which identifies tweet sentiments using Bayes classifier. The use of Django 1.3.1 web framework (a python framework) provides a lightweight web server with MVC framework which easily integrates with our python scripts. For designing user interface, html5 using jquery has been used. Finally, the application is deployed on the cloud using AWS (Amazon web services) Elastic Cloud Computer or commonly known as EC2, providing more profitability and reliable browsing and dependability from market perspective.

I. INTRODUCTION

Conferences happen on a day to day basis; some are good enough and some are not. As a novice in a particular industry, it becomes difficult to decide at times if it will be profitable enough to attend a particular conference or not? At times, it might happen that there is a tradeoff between the conferences you have to choose to attend. Consequently, nothing is gained by attending conferences that do not conform to their objectives. This offers the user a wide range of choices to choose from. To make a wise choice, the user needs to know which of those conferences are worth attending. This gives rise to the need for a tool that can fetch the reviews about the conferences from the top rated social networking sites such as Twitter. Thus, our goal is to design an application or a website that can assist people in getting conference rating along with visualizing the data gathered on the basis of tweets.

People or professionals attend conferences like IBM Information on demand, IBM Digital Experiences, Oracle World, Dream Force etc and provide their feedback on Twitter. Some may post their experience or provide a feedback about that conference. So the sole idea of our website is to offer users, the ratings of the conferences so that they are acquiesced with that particular conference in a much better way and hence plan their schedule accordingly. The user obtains the reviews in the form of rating such as Must Go, Go, Fair and Useless.

Author: San Jose State University, Shastri Nagar, Delhi, India.
e-mail: anky689@gmail.com

Conference with a very negative feedback will be rated as a useless conference.

Please note that we do not recommend any conference as a useless conference, it's our algorithm that suggests the viability of a particular conference not to be so good to attend only on the basis of twitter sentiment analysis. While at the same time, if a user search for a conference that does not exists, system may prompt an error message saying that you are looking for a conference that doesn't exists as no data is found for that particular occurrence. Added, if a user simply clicks on analyze button without entering any data, it redirects to a page saying error.

While the concept of Conf Rate is not entirely unique, the edge that has been added to it gives the entire concept a new and upgraded view as use of natural language toolkit with libraries such as stop words corpus, words frequency, classifier etc. that could easily be used to analyze the sentiments. Although, many work has been done so far around on twitter sentiment analysis but

When it comes to conference rating, there is no consolidated single tool or website to help you do so. Targeted users for this particular application are professionals, students, freelancers and people who are interested to attend to the conferences.

II. MOTIVATION

Working marathon hours is a part of most of the professional's schedule and many technical conferences are organized on a day to day basis. Say for example, I am new to a particular industry with avid interest in automobile and want to know if an upcoming conference by General Motors would be worth attending as it's a tradeoff between my working day and conference timings. So, an application that can help me get better reviews of the conference would save me from a tiresome day go wasted due to a bad decision. We do not require our users to end up wasting their valuable time in attending the events that do not meet their target. This gives rise to the need for a tool that would help users utilize their time artfully and judiciously. This application gathers user data and gives an analysis report on same while rating the conference. After analyzing all the essentials, we designed a prototype (figure1) for the application. The most vital components of the UI were added to this prototype. The foremost task was to come up with an appropriate name for the application. Next, a search bar was added to help the user look up for conferences. The ultimate goal of this

application is to display the ratings of conference based on the user feedback on twitter; thus, a "Rating" button was added to the prototype.



Figure 1 : Application Prototype

III. IMPLEMENTATION OVERVIEW

The project work as carried out in the form of sessions wherein each session, we worked on different aspects of the application such as requirement gathering, re-designing the user-interface, team management etc.

a) Session Summary

1. Session 1 & Session 2

These sessions were unofficial as during these sessions the team members got to know each other and gathered some background information on the technical aspects of the application to be developed. The sessions that followed, mainly comprised of development and implementation.

2. Session 3: Requirement Gathering

In order to identify an appropriate language for the development of Conf Rate, the team did a lot of research on all potential technologies that could be applied. The major task was to come up with a language that is widely used and recognized. Python dominated the list of all possible alternatives. Python's open source license makes it free to use; so anyone can install it. Twython is twitter's Python API that connects it on Oauth (open authentication). Oauth generates a security token which connects it to twitter. Django 1.3.1 has been used as the framework to run python because of its ability to simplify and speed up the fabrication of compound applications. It provides a lightweight web server with MVC framework which could easily integrate with our python script.

3. Session 4: Planning User-Interface

We decided to re-design the UI (figure 2) to make it look more appropriate and authentic.



Figure 2 : Re-designed User Interface

Besides having a search button, the new UI has a well-designed homepage. This homepage has added features such as Information about the Standards, Principles, Experience and Reviews. Most importantly, it provides information about the upcoming conferences. This helps the user to get updates about almost all the top-notch conferences.

Further, this UI has been re-designed with the help of the following tools:

i. HTML 5

The front end of our application is implemented in HTML5 as it is the latest version and has more features than the previous one. We have used new header and footer tags available in HTML5. Also we have incorporated the favicon using link tag. Html5 has added better functionality to our web pages.

ii. Javascript

JavaScript is used in our project to provide basic client side validation. Since the processing is done on the client's browser. It results in less load on the server. Javascript and JQuery (a Javascript library) has been used extensively in our project as it provides a very good support for representation at the front end.

iii. CSS

CSS is used throughout the presentation layer along with HTML5 to provide styling support and to make the web pages more presentable. We have incorporated various visual display effects using Css. For example, on the on hover event of the home button, the background color of the button changes because we have changed the Css styling properties on occurrence of that event.

iv. JQuery

Jquery has been added as an added functionality to Html5 pages to make our web pages more responsive. Along with that, it has added application scrolling facility to our application.

v. Highcharts

Highcharts is a library used to display dynamic and user friendly charts in HTML pages. Highcharts supports HTML5 and is open source. It is a jquery based library hence requires javascript to be enabled in the user's browser. It has options to display various different chart types like area charts, bar and column charts, pie charts and scatter and bubble charts. We have used pie and column charts (figure 3) in our project. The pie chart is used to display the percentage of positive and negative tweets and the column chart is used to display the sentiment of the last five tweets.

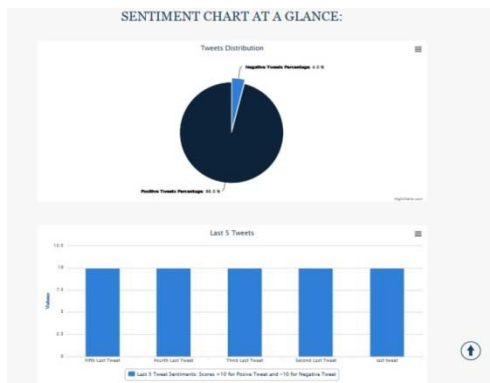


Figure 3 : Pie Chart and Column Chart

vi. Datatables

Data Tables are a flexible jquery based javascript library. They require javascript to be enabled in the client's browser. They also provides various features like sorting data, filtering data on the go, hidden columns, flexible data width and state saving. Besides being trivial to implement, these give a very good look & feel to the front end of an application. Various styling options like adding custom classes, Themeroller support and Theme Forest themes support which are pre-built HTML and CSS templates are supported by Data Tables. We have used Data Tables (figure 4) to display the tweets we are fetching from the twitter api so the end user can also view the tweets.

Figure 4 : Data Table displaying the Tweets

4. Session 5: Execution

The web framework was implemented using Django. It's written in python and provides MVC architecture. We built some responsive templates in HTML5, CSS etc in our view component so that the user could enter a conference name. The model part of the web server remains empty as there's no database required for our implementation.

In order to connect to twitter to search for the conference name entered by the user, Twython has been used. We used naive Bayes classifier based on Bayes' theorem to classify the tweets as positive or negative at the run time. This classifier can be easily implemented with the use of Python external library and Natural Language toolkit.

The algorithm works in the following manner. The classifier needs to be trained prior to analyzing the sentence. We created two lists of manually classified tweets; positive and negative and tagged each of the sentence with the sentiment it carries. The next step was to extract each of the word in the real time tweets except the stop words that don't carry any sentiment. Then, we defined feature extractor that compares our dictionary of manually classified words (in order of its frequency) with the real time tweet words. Finally, we trained our classifier on the training set in order to create a ready classifier. It detects the sentiment on the basis of the association of positive/ negative words learnt on the training set. We also used data table and High chart to build dynamic interactive data representation of our sentiment.

5. Session 6: Deployment

Project has been deployed on Amazon EC2 (Elastic Compute Cloud). Elastic Compute Cloud is a web service offered by Amazon at a minimal cost. We created an instance for our project and established its configuration settings including inbound security for assigning a particular port number for instance to run. To SSH the Amazon web server, we used WINSXP and ran the putty. A new session was opened with instance security key and got launched as a ubuntu user. Thereafter, python, twython, django and nltk were installed on our instance through couple of commands.

We have used screen command to nohup our server and have it run forever in background. We have deployed our website on port number 8000 and following is the url for our website which is already deployed on Amazon EC2 web service:

<http://ec2-54-215-198-104.us-west-1.compute.amazonaws.com:8000/>

After deployment of the project on the cloud, the images depicting the working of this application are as follows:

Figure 5 shows that the user looks up for conferences and the results related to that particular conference are displayed dynamically.



Figure 5 : Conference Search

Next, the user will hit the “Analyze” button and the application will fetch all the tweets related to that conference. These tweets undergo sentiment analysis and a result (figure 6) is displayed by the application. This result is generated in terms of: Must Go, Go, Fair and Useless where Must Go and Go convey a positive feedback, Fair indicates a neutral response and Useless implies that the conference majorly has negative feedback.

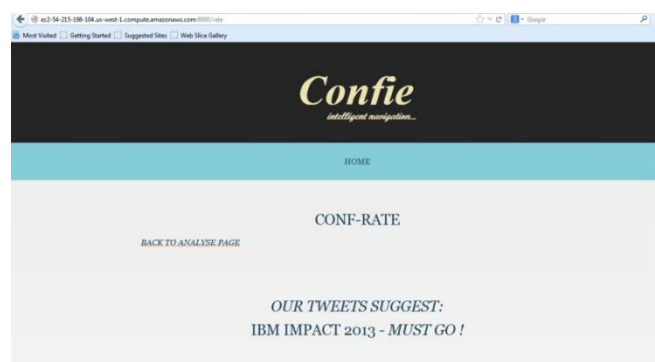


Figure 6 : Rating of the Conference

6. Session 7: Report

The last and one of the most crucial part of the project development was the project manual. The manual follows the IEEE format and describes all the essential aspects of the project in a simplified and easy to understand manner. This report covers all the technical and non-technical facets of the tool.

IV. CHALLENGES

Building an application of this potential is not an easy task. Many challenges were faced during this. Among them, the prime challenges were:

The hardest task of all was to develop and implement an algorithm that supports python and is capable of carrying out sentiment analysis.

This issue was solved by using python's natural language toolkit whose functions support the Bayes classifier.

Another challenge was to run the script on a local web server in order to be responsive from the client front end. This task was simplified by the use of Django which is easy to use framework.

Also, it is a tedious task to enable execution of interactive HTML5, CSS and images on the view component of Django. This problem was solved by defining a static path on the server side.

After successfully executing the tool, its deployment was another challenge we faced. We deployed the tool on Amazon web services. The big task was to ensure that the process runs forever on the ec2 instance. The best possible solution to this problem was to make the server run in daemon mode using a screen that would enable a virtual terminal for the processes to pull off in the background.

V. FUTURE IMPLEMENTATIONS

One of the major challenges of this tool is to fetch the upcoming conferences dynamically. Automatic updating of the conferences that are to be held is a paramount challenge which makes use of cron jobs.

Secondly, data can be gathered and mined in a more optimized manner. We may gather data from yelp for example and use it further to generate ratings. Basically our idea is to integrate website or engines with huge data sets and provide most customized reviews. Opinions from different websites can help give better conclusions about the conferences. More reviews can help analyze sentiments in a better way.

VI. CONCLUSIONS

We have named our website as Confie-intelligent navigation. As the name suggests, we are helping our users making an intelligent navigation further through an optimized search for conferences.

We built Conf Rate, which is a python based application that analyses the feedback given by the people who attend conferences. In short, following describe the tool in the best possible manner.

The feedback given by various people who attend conferences is gathered from twitter in form of “tweets”. These tweets are analyzed using an algorithm and then they are categorized as “positive” or “negative”.

All the tweets related to the search query is fetched at that point of time. Users can have a look at all the tweets that are being tweeted by the people.

Next, the user can view a bar chart that shows the last five recent tweets whether, negative or positive. The pie chart gives an overall percentage of all positive and negative tweets gathered. The user gets the feedback of the conference in term of Must Go, Go, Fair and Useless.

REFERENCES RÉFÉRENCES REFERENCIAS

1. [http://en.wikipedia.org/wiki/Python_\(programming_language\)](http://en.wikipedia.org/wiki/Python_(programming_language))
2. <http://www.python.org/>
3. <http://aws.amazon.com/application-hosting/>
4. http://en.wikipedia.org/wiki/HTML5#Differences_from_HTML.C2.A04.01_and_XHTML.C2.A01.x
5. [http://en.wikipedia.org/wiki/Django_\(web_framework\)](http://en.wikipedia.org/wiki/Django_(web_framework))
6. <http://www.w3schools.com>
7. <http://www.highcharts.com/>
8. <http://datatables.net/>



This page is intentionally left blank