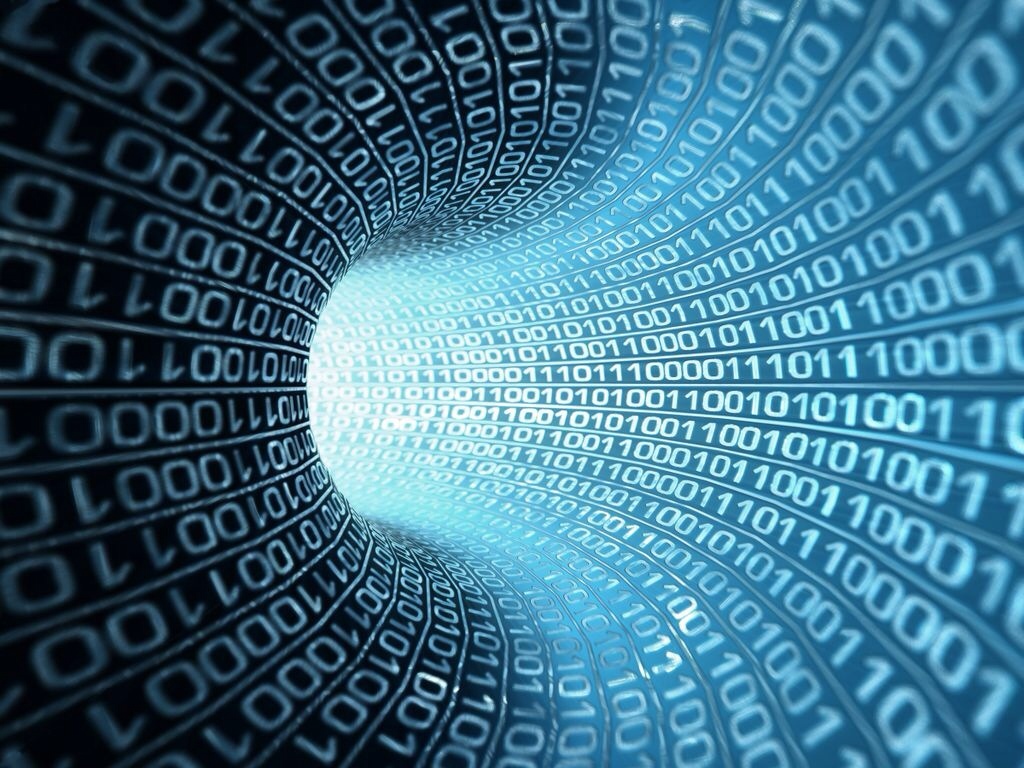
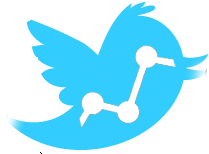
**Knowledge Discovery Management Project**

**Social Trend Analysis**

****

**Team:**

**Ashok Rudraraju,**

**Aditya Deshpande,**

**Tej Kiran Meka,**

**Mahesh Vemula.**

**Introduction:**

Social networking sites like twitter and Facebook has become unprecedented platforms for this generation to express their views and emotions without any fear or bias. We can see people tweeting to support a cause or express their views on a political party or the government. It is not uncommon to find deep discussions about a company’s performance or its credibility on Facebook or twitter. Celebrities, businessmen, political personalities and perhaps the most important of them all the common people participate in such discussions. Every company, every university and almost every country has a twitter account and Facebook page.

The information that is produced by these online accounts is almost priceless. It hides with in itself many hidden patterns, like the sentiments of people of a country or the general trend about what the people think about business acquisition like the recent acquisition of whatsapp by Facebook. If harnessed and processed properly, there is so much we can learn from the data available on twitter. This idea lead us to come up with our project.

**Project Goal and Objectives**:

We propose to collect data from twitter and analyze it to find hidden patterns like how the people in different parts of the world react towards the new developments in business, politics etc. One of our main objectives is to find the geographic and demographic trends of a single event. We aim to present any hidden trends we find in the data in a very user friendly, attractive and relevant user interface consisting of good visualizations like google maps. One of the basic principles of data analysis is to ask the right questions. The questions we expect to answer from this project are: people from which parts of the world are more aware of an issue?, people from which parts are maybe a little less aware?, etc. and the list can go on, we feel there is a real need for such kind of analysis and that it would help raise and also answer many questions.

If we succeed in our analysis and show how the general public are reacting to a topic we might be able to alter the way many decisions are made in this world.

**Future extensions:** If time permits and we are able to achieve the above mentioned goals, we believe we can take a step ahead and try to predict the tone of the tweets i.e the emotion behind a tweet by using a set of sad and a set of happy words to compare against. This would give us the number of positive tweets and the number of negative tweets on a given topic.

**Related Work:**

Predictions have been made based on social media networks like twitter and Facebook in recent years. J.Bollen, H. Mao, and X. Zeng predicted the stock market based on twitter mood[2], Asur and Huberman observed that the success of newly released movies could be forecasted by analyzing tweets relevant to movie titles [3]. Wong, S. Sen, and M. Chiang were even able to predict the story of movie based on social media[4].

Our proposed system is different from all the above mentioned projects in two main aspects, firstly we are interested in patterns that might not be very evident and we are planning on dealing with much more relevant and important topics like the present technological trends, political changes etc. Secondly we are planning on how one issue has different effects on different geographic locations and on different demographic.

**Proposed System**

**1) Requirement Specification**: We plan to represent the analysis and data in very attractive maps and easy to understand charts and graphs in a web based UI, thus the following are our main requirements.

a) A rich and user friendly UI (high Priority)

b) The google maps API (high Priority)

c) The twitter API (high Priority)

d) The yahoo Charts API (high Priority)

e) Well-structured web services which will hold our business logic.

**2) Business Requirements:**

a) Find relevant patterns from the dataset based on a given topic. (High)

b) Analyze data based on geographic locations. (High)

c) Analyze data based on demographics. (Medium low)

d) Find tone of the tweets (Medium)

**3) Technological and Architectural requirements:**

a) Hadoop/Mahout/R, Lucene/Solr, Weka etc.

b) Linux, Eclipse, Java, Html 5, Twitter bootstrap etc.

c)Datasets, we have been able to successfully pull the required dataset from twitter in a dynamic manner.

**Framework Specifications:**

The project is bases on big Data and utilizes Hadoop and its tools for data analysis. For user interface web application and android web app is planned to design.

1. System Architecture:

User query inputs recommendations

Web application or android application for user interface

Create web services using lucene/Solr

Mahout analysis on data to obtain relations and recommendations

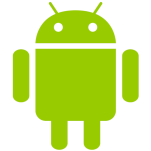
BigData from various sources

Hadoop Platform for data analysis using HDFS and MapReduce



Produce output results in user interface





**Data Sources:**

The source of data is collected from twitter pages using python script.

**Methodologies and Algorithms:**

The data is analyzed using mahout tool. The algorithms we are planning to implement are K-NN clustering and naive Bayes. But it is yet to decide depending which is best suited for it.

**Analytic Tools:**

HBase, Mahout, Solr and Solr rest services, boot strap, etc. There may be more tools will be included in further increments.

**Analytical Tasks:**

In this project the first task is to collect the data and store it using Hadoop file system using map reduce. Then depending on the application we structure the data and run machine learning algorithms on it. Finally we develop user interface where user can access data in graphical view and have some recommendations.

**Domain/Application Specification:**

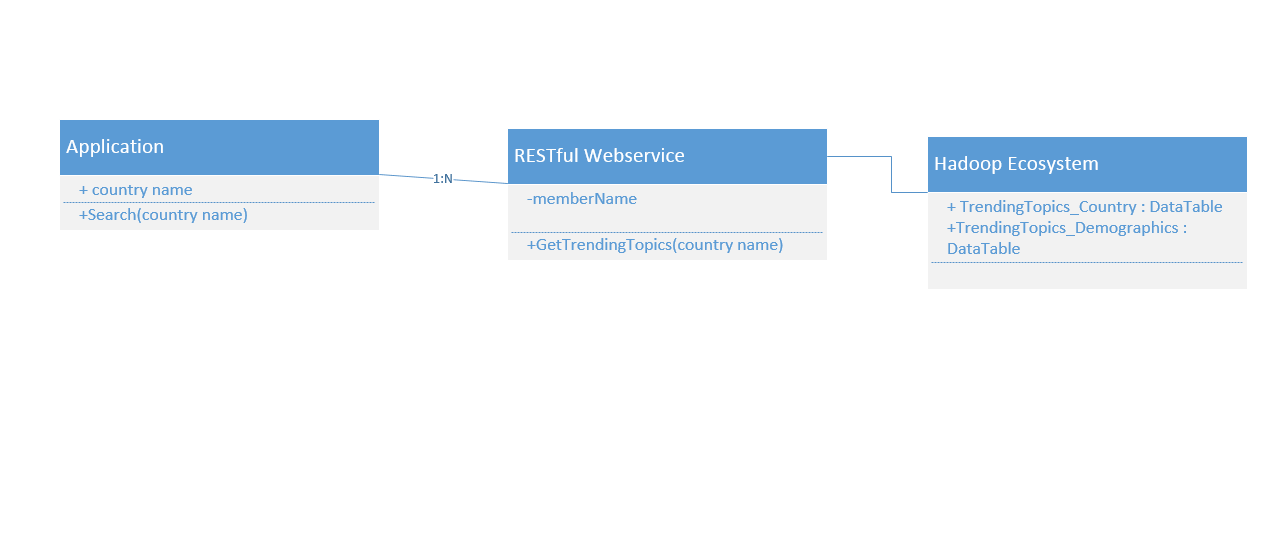
**Existing Applications:** Netflix,Youtube,Amazon,Google Ads

In above mentioned applications would implement classification and recommendation machine learning algorithms to filter the data and present it to the user. These applications would make use of collaborative filtering to filter the results that needs to be displayed to end user.

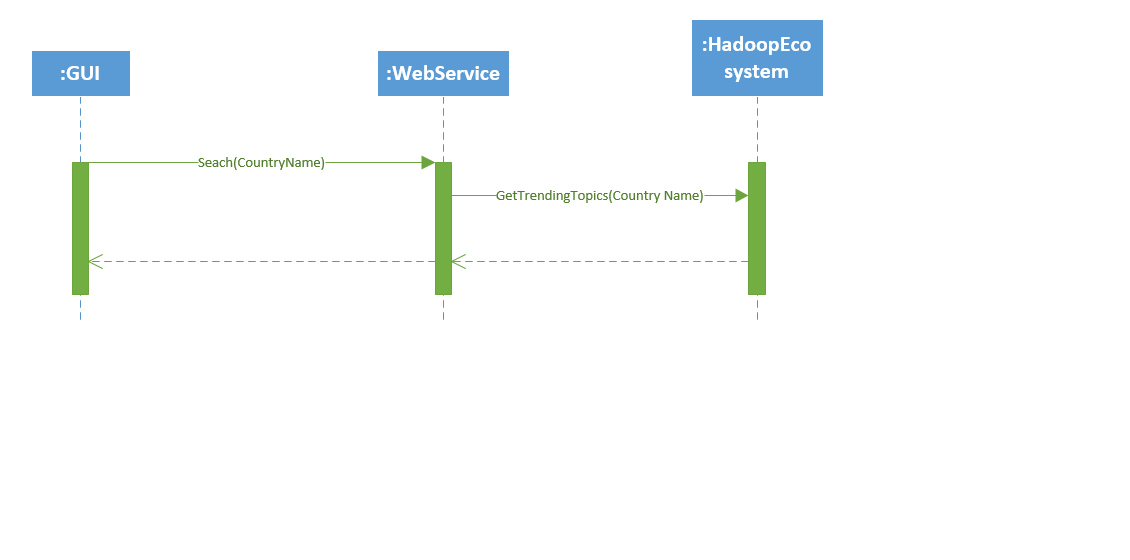
**Features in our application:**

(1)Get Trending topics based on country search

**Class Diagram:**



**Sequence Diagram:**



**Service Specification :**

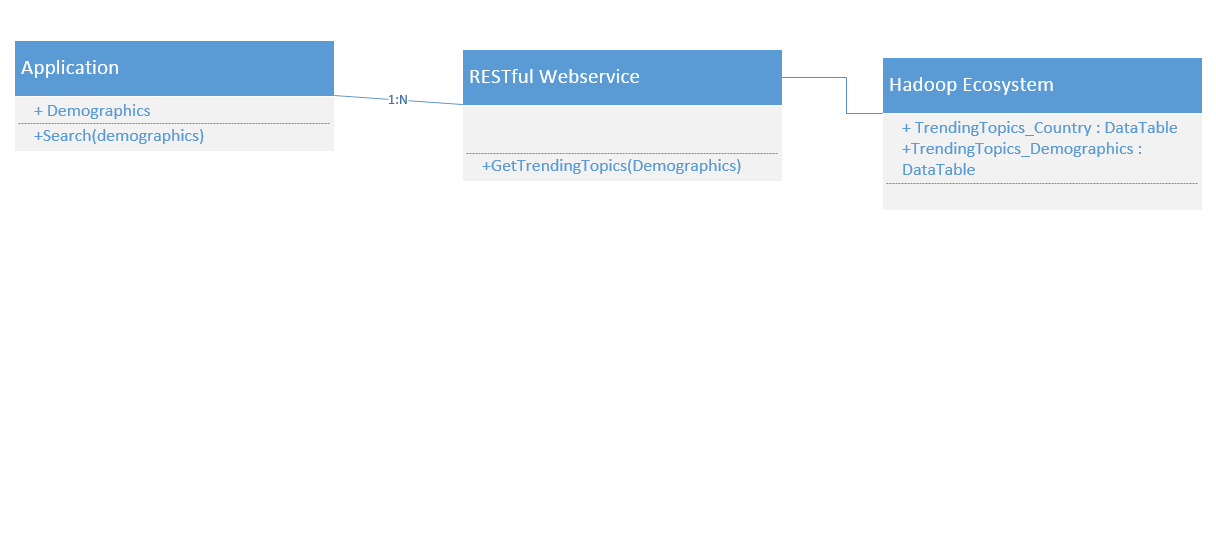
This service would display trending results to end user based on country name.

**Input :** Country Name

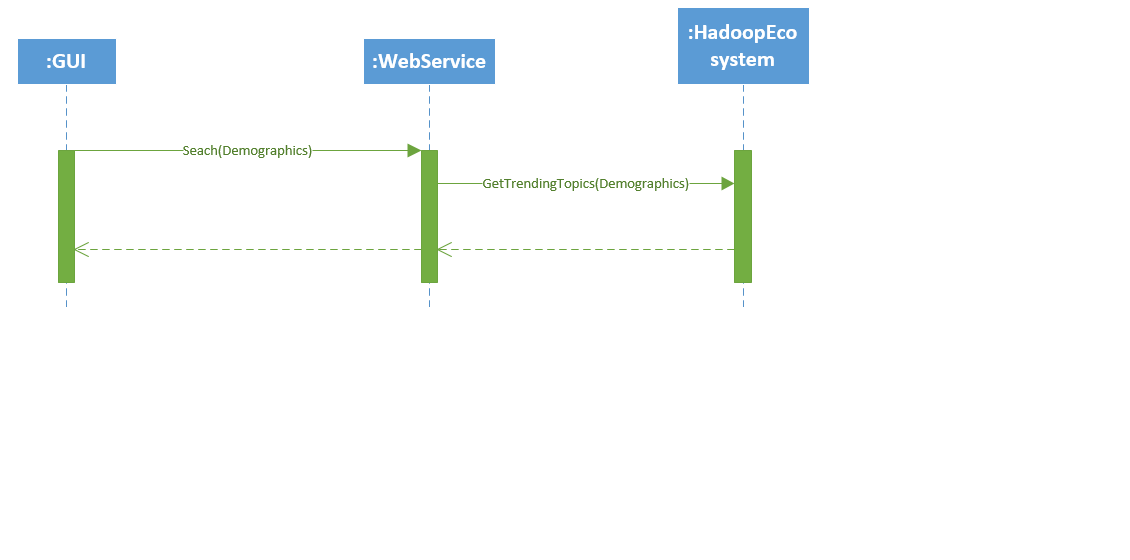
**Output :** Returns set of trending results in JSON format

(2) Get Trending topics based on Demographic search

**Class Diagram**



**Sequence Diagram:**



**Service Specification :**

This service would display trending results to end user based on Demographics.

**Input :** Demographics

**Output :** Returns set of trending results in JSON format

**Activity Diagram for application:**



**Design of Mobile Client:**

Web based approach would be followed in our project. We would be implementing **Twitter Bootstrap** CSS framework for our mobile interface. Mobile Client would consume RESTful Web service with jQuery-Ajax call.

**Features:**

* **Responsive design**: Layout of web pages adjusts dynamically, taking into account the characteristics of the device say mobile,Tablet or desktop
* Contains inbuilt libraries for [CSS](http://en.wikipedia.org/wiki/CSS)-based design templates for [typography](http://en.wikipedia.org/wiki/Typography), forms, buttons, navigation and other interface components
* Contains optional [JavaScript](http://en.wikipedia.org/wiki/JavaScript) extensions
* Compatible with latest versions of all major browsers

**Technologies:**

HTML5, Ajax, jQuery, JavaScript, CSS3

**Plan by Analytical Task:**

**Scenario and Use case specification template:**

**User Search:**

**Scenario:**

In this scenario user will enters his preferences to search for the trending topics

**Use-case specification:**

|  |  |
| --- | --- |
|  | |
| Title | User Preferences |
| Description | User selects his preferences and clicks on the search button in the website |
| Actors | User |
| Initial Status and Pre Conditions | User should have an internet connectivity |
| **Basic Flow** | |
| Step 1 : User selects his preferences like country, area etc  Step 2 : User clicks the search button  Step 3 : Trending topics in the twitter will be displayed on the website | |
| **Post Conditions** | |
| After successful action user can view the trending topics according to his preferences | |
| **Alternative Flows** | |
| * If the incorrect format of data is entered by the user then error message will be returned for the user | |

**Design of Map Reduce Algorithm:**

**Scenario:**

In this scenario Developer will design the Map reduce algorithm for the analysis of the data that was pulled from the twitter.

**Use-case specification:**

|  |  |
| --- | --- |
| **Overview** | |
| Title | Design of Map Reduce Algorithm |
| Description | Developer designs the map reduce algorithm to extract the interesting patterns from the data pulled |
| Actors | Developer |
| Initial Status and Pre Conditions | Developer should have knowledge on design of algorithms |
| **Basic Flow** | |
| Step 1 : Developer designs the map reduce algorithm  Step 2 : Developer execute the map reduce algorithm using cloud era hadoop.  Step 3 : Developer will extract useful patterns out of the data. | |
| **Post Conditions** | |
| After successful execution, developer will extracts the useful patterns. | |
| **Alternative Flows** | |
| * If execution of the algorithm fails then the incorrect or irrelevant patterns are obtained. | |

**Projection of Predictions In the UI:**

**Scenario:**

In this scenario Developers will project the predictions that are driven from the extracted patterns in the UI designed on the website.

**Use-case specification:**

|  |  |
| --- | --- |
|  | |
| Title | Projection of Predictions |
| Description | Developer will Project the predictions driven from the extracted patterns. |
| Actors | Developer |
| Initial Status and Pre Conditions | Developer should run the algorithm to extract patterns and makes predictions out of patterns |
| **Basic Flow** | |
| Step 1 : Developer will extract patterns through the algorithms.  Step 2: Developer will makes predictions out of patterns.  Step 2 : Developer will use the APIs to project the extracted patterns on the UI | |
| **Post Conditions** | |
| After successful action, Predictions will be displayed on the website. | |
| **Alternative Flows** | |
| * Developer has to analyze the patterns clearly to make the accurate predictions. | |

**Testing the services and website:**

|  |  |
| --- | --- |
| **Overview** | |
| Title | Testing of services and website |
| Description | Services implemented in the project will be tested |
| Actors | Developer/Tester |
| Initial Status and Pre Conditions | Developer should access the completed website. |
| **Basic Flow** | |
| Step 1 : Developer/Tester tests all services implemented in the website.  Step 2 : Developer can release the project | |
| **Post Conditions** | |
| After successful Testing, project can be released. | |
| **Alternative Flows** | |
| * Developer/Tester should rectify or report the errors. | |

**Design by Algorithm:**

**Map Reduce Algorithm:**



**Project Timeline, Members and Responsibilities:**

**Increment 1:**

Ashok : Design the script for extraction of data from twitter.

Aditya: Create Login Services for Website.

Tej Kiran : Create Registration Service for the website.

Mahesh: Create Basic UI for Website.

**Increment 2:**

Ashok: Use the script to extract the data from twitter.

Aditya: Design the Map Reduce algorithm for filtering and sorting the data.

Tej Kiran: Design the word count program for analyzing data.

Mahesh: Analyze the data extracted from twitter.

**Increment 3:**

Ashok: Design the required web services for the website.

Aditya: Run Map Reduce algorithm to filter and sort the data.

Tej Kiran: Design detailed UI for the website.

Mahesh: Perform testing to make sure all services till this increment are working.

**Increment 4:**

Ashok: Plot the detected patterns on the UI using API like google maps.

Aditya: Perform the final enhancements to all the services in the website.

Tej Kiran: Host all the services to the VM cloud.

Mahesh: Perform testing on all the services in the website.

**Bibliography:**

[1] B. O’Connor, R. Balasubramanyan, B. R. Routledge,and N. A. Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In Proceedings of the International AAAI Conference on Weblogs and Social Media, pages 122-129, Washinton,DC, May 2010.

[2] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. Journal of Computational Science, 2(1):1-8 March 2011.

[3] S. Asur, and B. A. Huberman. Predicting the Future with Social Media. In IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pages 1:492ˆa˘A ¸S499, Lyon, France, August 2011.

[4] A. Mislove, S. Lehmann, Y. Ahn, J.-P. Onnela, and J.N. Rosenquist. Understanding the Demographics of Twitter Users. In Proceedings of International AAAI Conference on Weblogs and Social Media, pages 554-557, Barcelona, Spain, July 2011.

[5] F. M. F. Wong, S. Sen, and M. Chiang. Why Watching Movie Tweets Won’t Tell the Whole Story? arXiv:1203.4642v1, 2012.

[6] http://www.sentiment140.com

[7] T. Wilson, J.Wiebe, and P. Hoffmann. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In Proceedings of Joint Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing, pages 347-354, Vancouver, Canada, October 2005.

[8] http://www.locomatix.com

[9] N. Jindal, and B. Liu. Opinion Spam and Analysis. In Proceedings of the international conference on Web search and web data mining, pages 219-230, Palo Alto,CA, February 2008.

Class tutorials and syllabus.

http://steinvox.com/blog/wp-content/uploads/2012/10/GoogleAnalyticsDashboard.png

http://icons.iconarchive.com/icons/gakuseisean/radium/256/User-Computer-icon.png

http://i2.wp.com/devleader.ca/wp-content/uploads/2013/10/Android.png?resize=614%2C614