INTERIM PROJECT REPORT

POLITICIAN360

2016

TABLE OF CONTENT

INTRODUCTION

About Politician360

Tab 1

IN THIS SECTION:

- About Project
- Team Member
- Stages of Development

ABOUT PROJECT: -

This project is a part of the final semester for MSC Student NL programme for the year 2015-2016. The vision of this project is to collect and analyze news and social networks to provide a linked information and a rank of Irish politicians by popularity and sentiment Analysis.

TEAM MEMBER: -

There are in total of 5 Members responsible for the completion of the project in the given Time frame

- ARUN KISHORE RAJENDARAN
- MOHD FAISAL RAZA
- RODRIGO SERVIUC PAVEZI
- SOHAIB AHMAD
- SUMITHA DHANASEKHARAN

Duration: - 17-05-2016 to 19-08-2016

STAGES OF DEVELOPMENT: -

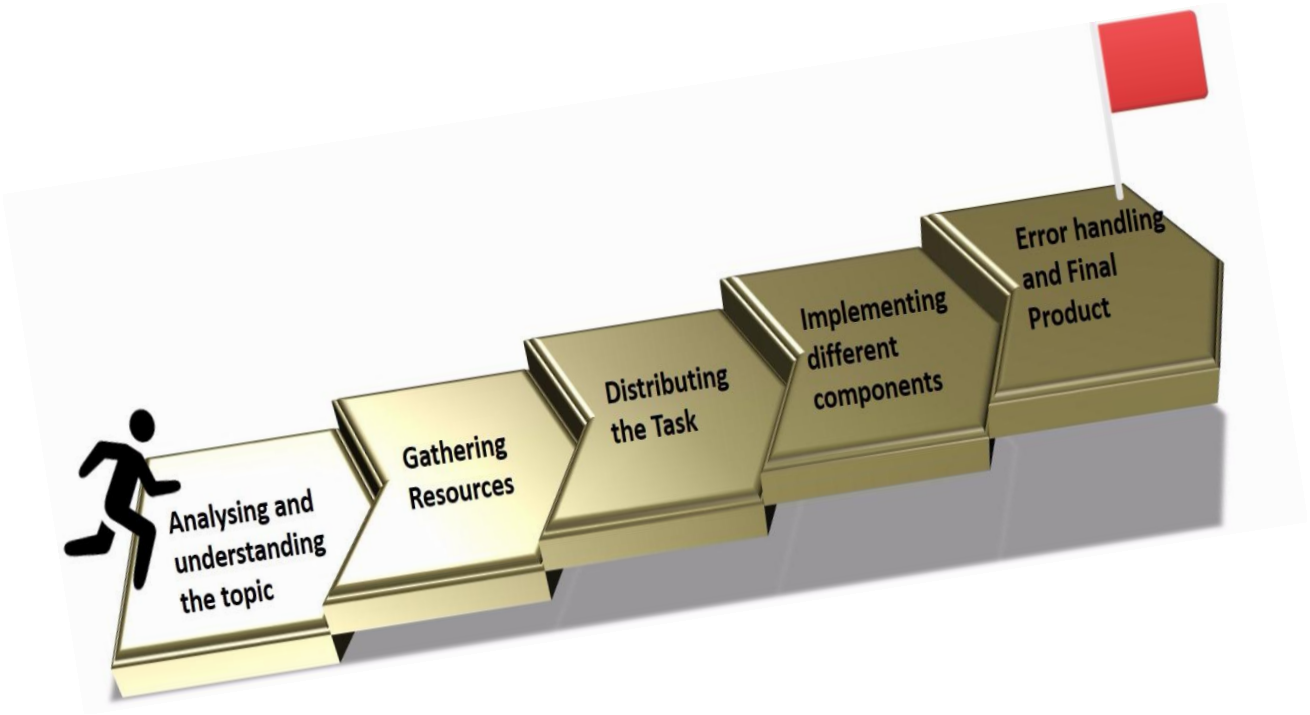

Fig 1.1 Stages of Development

USER SCENARIO

Tab 2

About User

IN THIS SECTION:

- Target user defined
- Importance of user
- What problem is solved for them

Target User: -

We are targeting our solution to Irish voters who do not have time for following all the news streams about politics. Thus, they could be any Irish citizen aging from 18 years old. To be aware of what is happening in the government and what each individual politician is doing currently, a voter would normally need to keep up to date with news and perhaps with social network streams. We can say that this is a very demanding task nowadays as people are overwhelmed by information from any kind. They could also access the public information from the government websites such as debates and votes. There are so many sources of information which makes it harder for a voter to have an overall picture of what is happening in politics. Knowing what an individual politician is interested or involved in while he/she is in the government is also not straightforward. Another fact is that information from a single source might not be reliable as it tends to be biased.

Importance of User: -

Voters are responsible for the election of their politicians, thus, they should know about the person to whom they are electing as their representative. They should also be aware of what their representatives are doing during their mandates, especially if they are fulfilling their promises or not. For instance, a voter could choose a candidate as it promised to work in favor of the environment and renewable sources. Then it would be good if the voter could easily follow the politician actions. Furthermore, if that politician is not working in favor of the environment, the voter could demand him/her to do so. Helping voters with all this can facilitate on the country's progress towards fulfilling the needs of its people like a real democracy where people have consciousness and voice to speak up.

Problem solved for User: -

With all that in mind the main idea behind our application is to link news, tweets and debates data to current Irish Politicians. We believe that this could be an important asset for voters who are willing to know what these source of information is saying about the politicians. It could be interesting to voters to know which are the most popular politicians based on the number of news and tweets related to them. It could also be important to know what is the sentiment attributed to these news and tweets, negative or positive. Moreover, it could be helpful to know what topics the politicians are more active on debates. Furthermore, all this information could be implemented in a form of a rank where voters could compare politicians by popularity, sentiment analysis or debate participation.

Studies have been conducted showing that people that consume more news media tend to be more civically and politically engaged [1]. Other studies show that not everyone who has the right to participate on politics will exercise it [2]. These two facts should help us to understand our potential users so we can build a solution that will provide more useful and summarized information for those already politically engaged, and ultimately increase the number of those who are not.

TECHNICAL REPORT

Technical description

IN THIS SECTION:

- Why does the System exist?
- Core Technical Problem defined
- Reviewing other similar system

Why Does System Exist?

The main problem we are willing to solve is to bring information from different sources into one place where our users can have it in an interesting, meaningful and easy way to consume. Our solution is not just for gathering information about the overall politics, it is meant to link this information to each individual politician. Thus, users can compare and keep track of the individual politician actions in the government. In this matter, our system has been designed to use this linked information in an interesting way which will be the rank of politicians.

Core Technical problem?

We can represent our core technical problem through the below image. Our solution will have to collect, filter, extract, analyse and present the data from different sources. There are a couple of technical challenges involved on building that. They can be listed as follow:

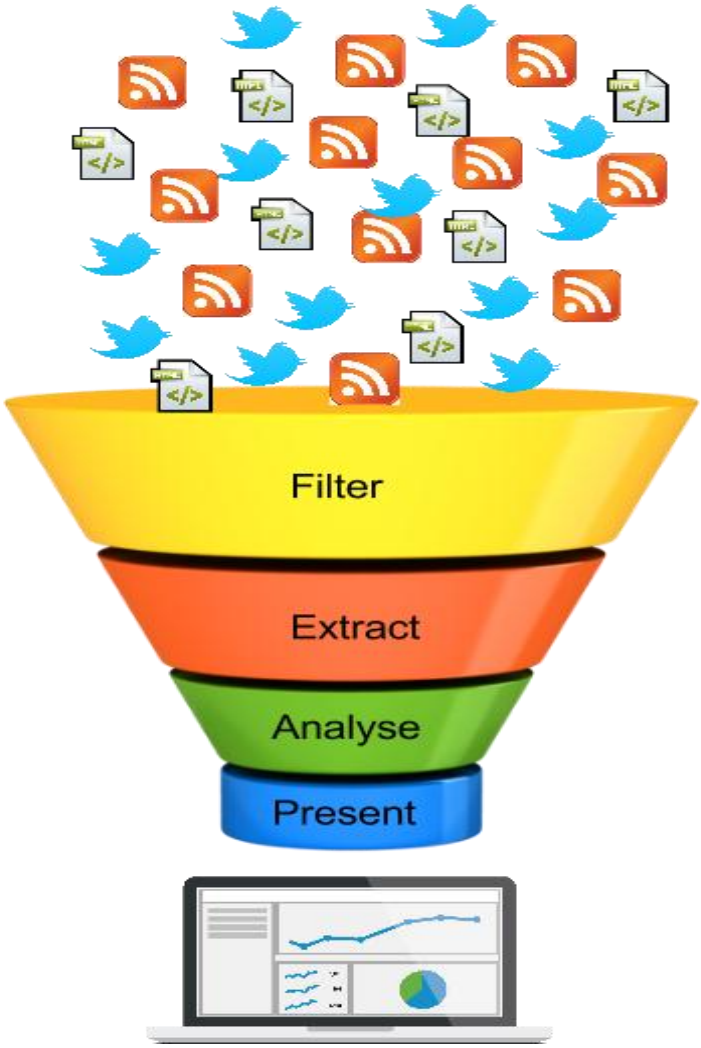

Figure 3.1 Technical problem Diagram

- Data collection - it might be difficult to get data enough for all the politicians entities from the current government, especially if the politician is not that popular.
- Data filtering - we will need to make sure the data is related to the right entity and also to politics.
- Data extraction - we will have to extract the important text from the data ignoring link urls and other non-meaningful texts.
- Data analysis - we will have to make sure the sentiment analysis algorithm has a reasonable level of correctness.
- Data presentation - we will need to make sure we present the summarised data in a intuitive way where users can understand it.

Review Other Similar system: -

Based on our research for other systems, we could not find any that would have the same features we have on our system such as the politicians rank and sentiment analyses. Some systems have gathered and linked data from news and tweets but were targeting the election period while we are targeting the current elected politicians. As some examples we have.

- Insight4News which has linked data to candidates but has not implemented any kind of ranking or sentiment analyses. It has a list of candidates and for each candidate it shows some basic information, related news articles and tweets. It has also implemented a count for party mentions on public tweets, some stats such as the gender of the candidates and a counter of number of articles and tweets per party [3].

- Politwoops which was created to record, store and publish deleted tweets from candidate accounts. Its main purpose is to keep track of deleted tweets so politicians cannot use the trick of making some statement on their twitter account then try to mislead people saying they have never said that by deleting the tweet. It is very useful tool nowadays as social networks have been hugely used by politicians to communicate directly with voters. Nevertheless, it does not have anything similar to our solution [4].

- GE16 Twitter Counter which was created to track and monitor twitter accounts of candidates and parties during the Irish Election of 2016. It also looks at their online engagement with voters. The system analyse tweets but does not have any sentiment analysis as our solution will have [5].

- The study carried out on [6] was able to analyse RSS feeds sentimentally during the US presidential election of 2008. The system was more focused on the visualization of the analysis in which every RSS news with its respective sentiment was plot in a way the position of a 2D square figure would represent a sentiment score. Moreover, it represented different characteristics, extract from the news, using coloring, shape and opacity. Although it is an interesting study, the system did not have the same representation of the sentiment analysis we are proposing on our solution.

- Another study described on [7] was applying sentiment analysis to tweets during the US presidential election of 2016. They have collect tweets through the search api using some keywords and the full name of the candidates. Then, they have chosen to only analyse the tweets containing emoji's and multi classified them into these

categories: happy, sad, angry, laughter and scared. For the visualization they generated a map for each candidate which showed the predominant sentiment on each US state. Their solution has similarities with what we are going to build on our system regarding the sentiment analysis.

Other systems have gathered raw data about politics. That makes them a good tool for researchers but not that much for ordinary voters as there is too much information which is not summarized anyhow. For instance, we have

- KildareStreet.com which is a searchable archive of everything that's been said in the Dáil and all written parliamentary questions since January 2004, everything in the Seanad since September 2002, and all Committee meetings since September 2012 [8].

- Oireachtas.ie which has data about debates and voting records for the Dáil [9].

TECHNICAL SOLUTION

Technical solution

Tab 4

IN THIS SECTION:

- What does System do?
- System Design
- Front end Technologies
- Back End Technologies
- Data Sources

What Does System do?

For solving our main problem, we are building a web app with three main sections (see mock-ups below):

- Home page - where users will land into the app and have direct access to the latest news and tweets related to the politicians, and a Top 10 rank of the 10 most popular politicians.

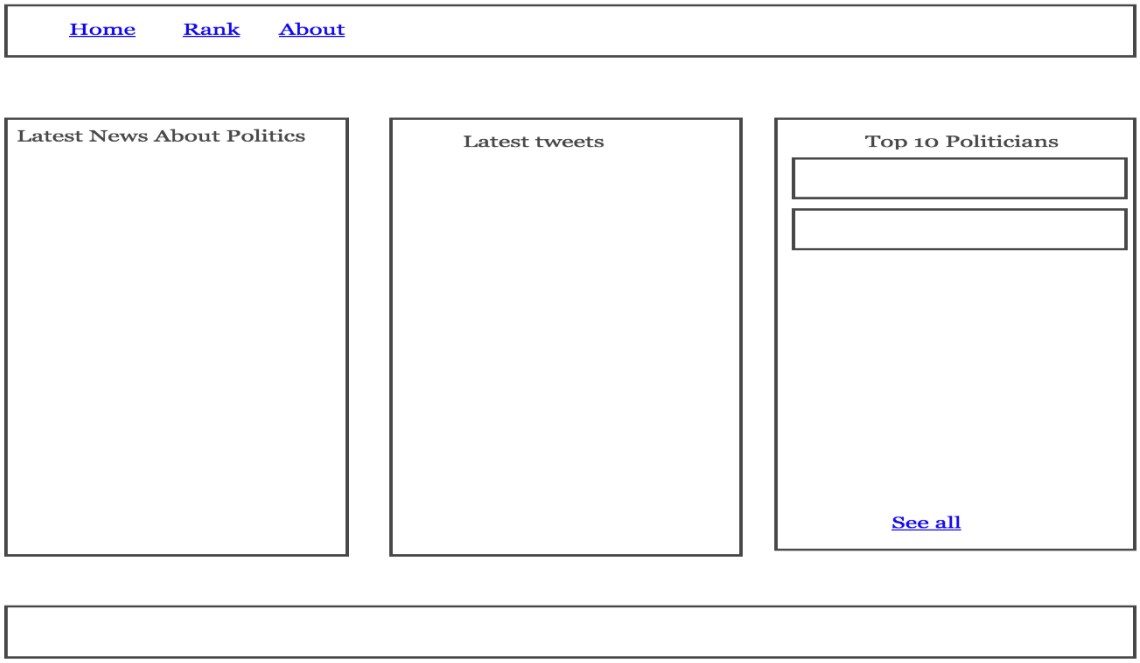

Fig 4.1 Home page Mock up

- Rank page - where users will see a list of all the politicians ranked by either
  - popularity (number of news and tweets related to the politician).
  - positive sentiment (number of positive news and tweets).
  - negative sentiment (number of negative news and tweets).
  - debate participation (number of times a politician participates on a Dáil debate).

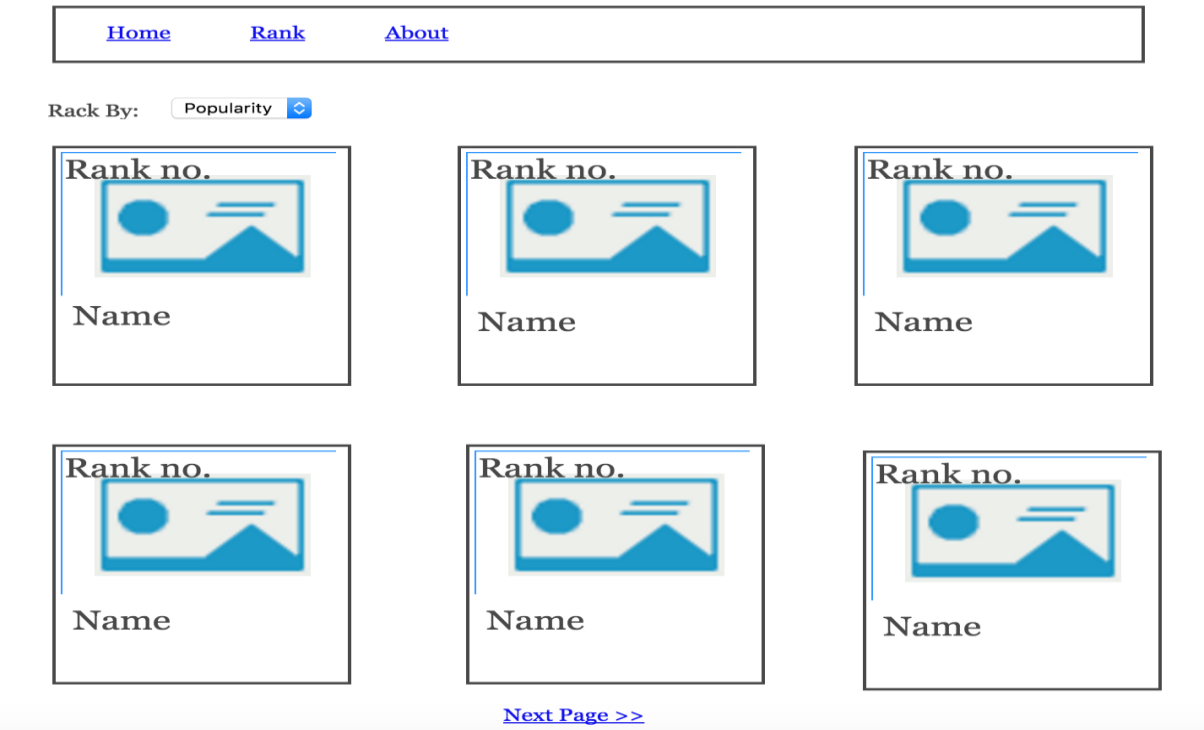

Fig 4.2 rank Page

- Politician page - where user will have access to
  - basic details about the politicians such as name, photo, party, constituency and contact details.
  - counters such as number of news articles, number of tweets, number of participations on debates, number of positive news articles and tweets and number of negative news articles and tweets.
  - list of news articles related to the politician.
  - list of tweets related to the politician.
  - a chart showing what topics the politician participates more on debates.

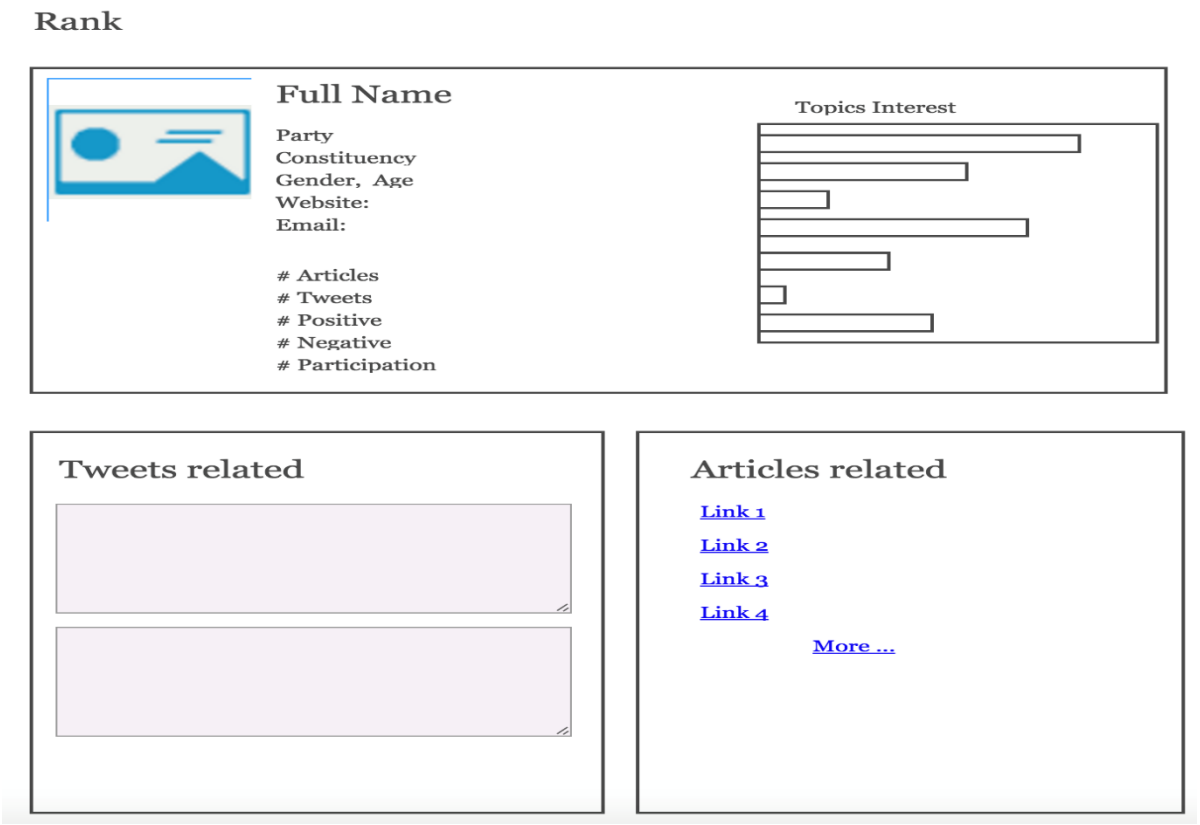

Fig 4.3 Politician page

System Design: -

We are building our system using Docker which allow us to separate it into containers. For instance, we will have a container for the front end, another for the backend and another for the database. All the different containers will be managed by Docker compose. There will be differences when deploying to production as the system will use more reliable web servers such as nginx for the front-end and the gunicorn for the back-end. The project has a couple of scripts which are responsible for building or managing the containers. Everything works really well and there is no need for setup each environment from scratch. Also, it is very easy to deploy to production so we are willing to setup a continuous delivery process.

System Diagram

The following diagram show a high level design of the entire system with the communication between the parts and the collection of the data from the data sources.

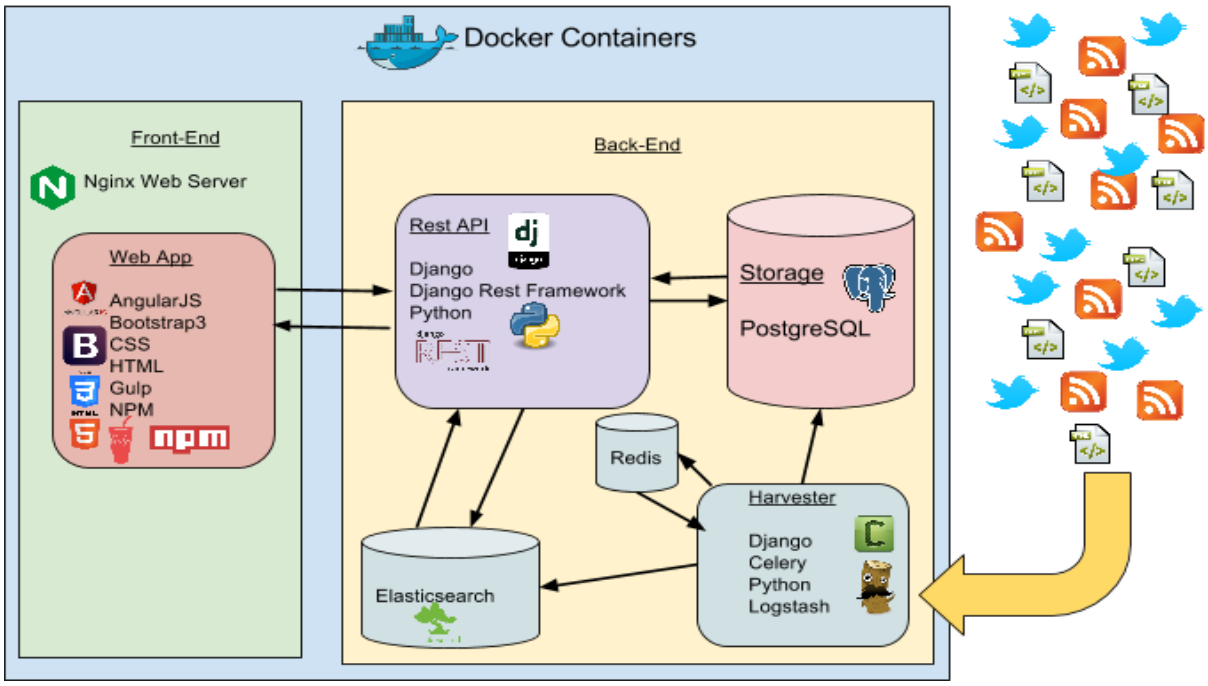

Fig 4.4 System Diagram

Basically, our front-end app will communicate with our back-end using JSON through a rest api which will hold the business logic of the system. This rest api will also consume the data from our database and the index system. Both will be populated by a harvester system which will be comprised of 3 modules

- Rss collector.
- Tweets collector.
- Html scraper.

The harvester jobs will run as often as possible through scheduled tasks so we keep the database and index up to date. We will make sure we respect the twitter api limits for instance.

Once the data is collected, there will be a job to extract and analyse the data. The extraction process will make sure the text is good for the sentiment analysis algorithm. Then this algorithm will run for each tweet and new article storing its sentiment on the database. A table on the database will keep the relation between the data collected, the politician entity and the sentiment analysis result. With that, we can build the counters used on the rank and on the politicians' page.

Front End Technologies: -

The front-end will be implemented using the following technologies

- JavaScript, HTML and CSS
- Bootstrap - UI front-end framework
- Angular-JS - a JavaScript MVW framework
- NPM - JavaScript package manager
- Bower - web app dependency manager
- Gulp - build task manager

All this tools and libraries will help us to develop a responsive and organized web app.

Back End Technologies: -

The back-end will be implemented using the following technologies

- Python
- Django with Restful framework
- PostgreSQL
- Celery
- Redis
- Log stash
- Elastic search

Data Sources: -

As a start point we have a list of elected politicians from 2016 elections. This was used on the Insight4News application. The system will then use that list to get other info about the politician from the IEOD (Irish Election Open Database) candidates API.
For each politician we will need information from news articles and twitter. We have considered the main Irish newspapers which cover politics. The list was taking from FUJO (The Institute for Future Media & Journalism)

- RTE NEWS
- IRISH EXAMINER
- THE IRISH TIMES
- IRISH DEPENDENT
- THE JOURNAL.IE
- NEWSTALK

For each newspaper the system will read their RSS and look for news that are related to any of the politicians. When an item is found it will be saved on the index system.

Moreover, we will also use Twitter data through the streaming and search APIs. As the search API only accepts 150 requests every 15 minutes. We will need at least two twitter accounts for searching tweets for all the elected Irish Politicians (218). The idea is to make 218 search requests every 15 minutes. All tweets related to a politician are going to be indexed so we can easily mine them.

Furthermore, for the analyses of the politician participation on debates and the politician's topic interest chart, we will scrape the Oireachtas.ie website collecting all the information regarding the debates. There we can get who has participated on a debate and within which topic.

EVALUATION STAGE

Evaluation

IN THIS SECTION:

- Meaning of success for the system
- Measures to Evaluate the system

Evaluation Defined: -

The success of the system could be, if our system can provide with maximum valid information to our user in minimum (say less than 10 minutes) time than they will get anywhere else.

The system is not giving any conflict issue in the name of the person means there is no Ambiguity in the name of the politician

The system is able to provide the user with the trustful information in the given time frame

Evaluation Strategy: -

After completion of every MVP there is a day assigned as shown in the fig6.2 Future plan for project Management diagram the system will be tested with respect to that current MVP and after that we will move onto the next MVP and same process is obtained for Evaluation at each MVP stage.

After the completion of the MVP's we will integrate the system and then will Evaluate our system as a whole using the following procedure

We can count the number of users who visited our site on daily basis and can have the monthly chart to see our website progress on the internet.

We are developing a website and for its evaluation we need to perform few structured questions with the users to get the feedback about it. By structured questions we mean to say that we will prepare some question based on our website and the users can answer it after visiting the website.

Question that will form the part of evaluation:

1. How difficult it was to find the information you were looking for on our site?
2. Were you able to find the information you were searching and how satisfactory it was?
3. The option for filtering the list of politician with different types (popularity, number of tweets and articles, articles about corruption) was useful or not?
4. Did you tried the comparison between politicians to know more about them?

We will also promote our website on social media and other platforms and will try to gather the report of evaluation based on the result from the user.

MANAGEMENT

Project Management & Conclusion

IN THIS SECTION:

- Define Project Management Strategy
- Challenges
- Future Plan

Project Management Strategy

Following the concepts of MVP, we have divided our project into 3 main MPVs. One for collecting and listing the news articles and tweets for each politician. Another for implementing the sentiment analysis and the rank. Finally, one more for implementing the debate data collection, counter of participations on debates and chart for topic interests on debates. See below how it looks on our trello board.

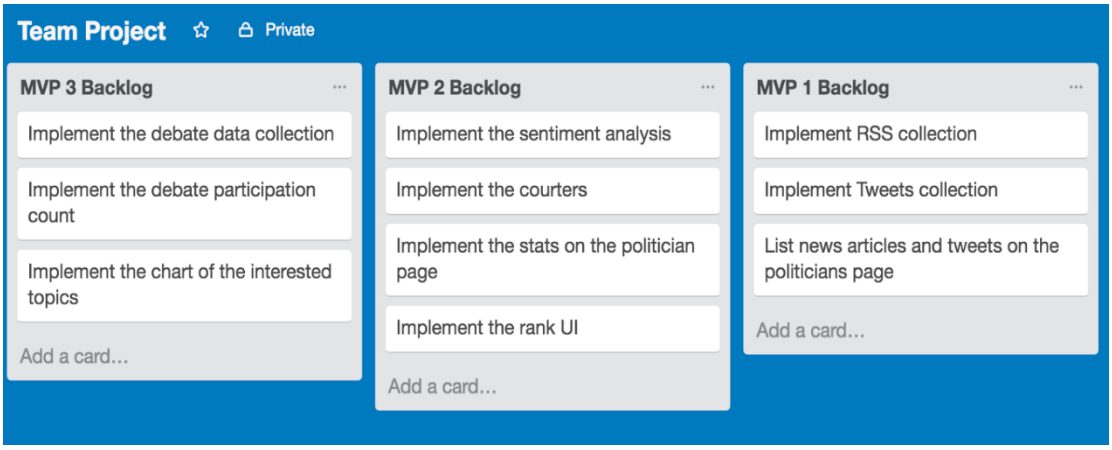

Fig 6.1 Current Tasks shown

What is your project management strategy?

As team we work together on every aspect of the project but explicitly we have assigned each one of us some role to play when it comes to project management. It is always better that to assign each work to an individual so that the work flow is maintained and someone is responsible for it.

For this project we are following the Scrum methology and to keep track of our work we do a 15 minutes' standup scrum daily on google hangouts. It helps everyone to know what progress have been made of the work assigned to each of us. If someone is strcuked in some work then it is the best place to put up the issue in front of the team so that best possible solution can be reached.

Trello is one of the application we are using to keep track of our sprint backlogs, assigned word, blocked work and done work so far. It helps to bind the team together with all the work assigned to each and its due date so that no one misses the target date.

Slack is our application for communicating as a team. It is really helpful as it is now embedded in trello as well keeping every one updated of any situation.

Challenges: -

The biggest challenge that we are currently face was to cope up with time and submit the report and cope up with the work on time due to some un avoidable circumstances which have been sorted out and we are now coming back on time and trying to cope up with the times loss which we have faced and finish everything on schedules.

Future Plan: -

To achieve the success in the given time frame we have set up a deadline for the remaining task so as to deliver and evaluate the system well before in advance before final deadline. Following Time diagram explains our Planning for the Upcoming time frame.

	Politcian 360	Start Date	End Date	Time
2	**MVP PLAN**			
3	**MVP 1**	**25-Jun**	**4-Jul**	**11**
4	Implement RSS Collection	25-Jun	29-Jun	4
5	Implement Tweets Collection	29-Jun	1-Jul	2
6	List news articles and tweets on the politicians page	2-Jul	3-Jul	1
7	Evaluation 1	4-Jul		
8	**MVP 2**	**5-Jul**	**12-Jul**	7
9	Implement the sentiment Analysis	5-Jul	7-Jul	2
10	Implement the Courters	7-Jul	8-Jul	1
11	Implement the stats of politician page	8-Jul	9-Jul	1
12	Implement the Rank UI	9-Jul	11-Jul	2
13	Evaluation 2	12-Jul		
14	**MVP 3**	**13-Jul**	**20-Jul**	7
15	Implement the debate data collection	13-Jul	14-Jul	1
16	Implement the debate participation count	15-Jul	16-Jul	1
17	Implement chart of interested topic	17-Jul	19-Jul	1
18	Evaluate 3	19-Jul	20-Jul	1

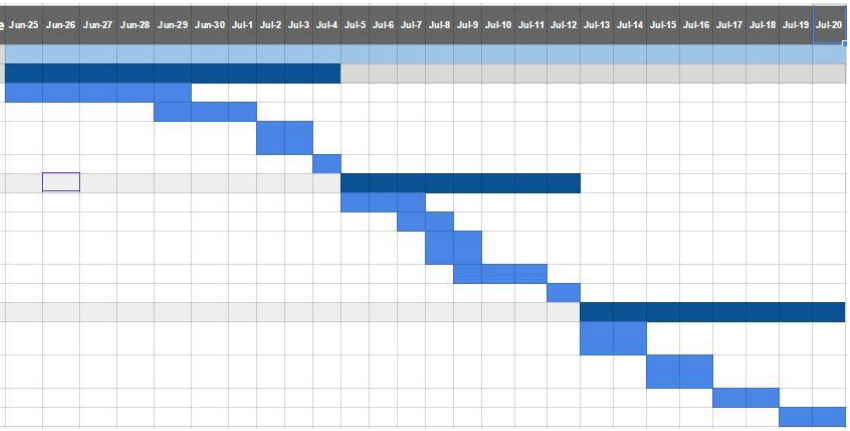

Fig 6.2 Future Plan

References: -

[1] - *How does social media use influence political participation and civic engagement? A meta-analysis*, http://journalistsresource.org/studies/politics/digital-democracy/social-media-influence-politics-participation-engagement-meta-analysis

[2] - *Political Participation,* http://www.encyclopedia.com/topic/Political_Participation.aspx

[3] - http://insight4news.ucd.ie/

[4] - *Introducing Politwoops for the Irish Election,* https://medium.com/@Storyful/introducing-politwoops-for-the-irish-election-cd0705ae328#.p3bwr0e2k

[5] - http://mlg.ucd.ie/ge16/

[6] - *Visual Sentiment Analysis of RSS News Feeds Featuring the US Presidential Election in 2008, http://ceur-ws.org/Vol-443/paper7.pdf*

[7] - *Analysing Twitter Sentiment of the 2016 Presidential Candidates,* http://web.stanford.edu/~jesszhao/files/twitterSentiment.pdf