**ATOS 4 group (Social media):**

**Andreas Mihaloianis, Momshad Alvee Dinuri, Chaitanya Agrawal**

**13.03.2015**

**Overview:**

Over the last two weeks we implemented the bag of words algorithm and presented a first prototype of the algorithm live on the server. We fixed all major and minor issues with our code base such as rendering the emoticons, elimination of empty tweet boxes and time rendering. The emoticons rendering worked perfectly on the local-server however the same code did not work on Azure.On top of this we had missing tweet-boxes and wrong time-rendering so were unsure as to the root of the problem. The empty boxes problem was fixed by adding some more functions that cleaned the tweets properly and safely passed data between php, javascript and html. After having the missing boxes problem fixed we were quick to identify that the emoticons could be fixed by updating Azures php setting from 5.4 to the latest 5.6 which we were running on the local-server. The time rendering was fixed by subtracting a constant number form the final output. This was because the time being returned by Azure was of Northern Europe as the database was stored there and so we had to compensate for the time difference.

**Summary of meetings held:**

During the last sprint we have only had one meeting with our client(12th, March), because we had to fully implement the bag of words algorithm before showing it to the client. He has declared himself extremely happy with our flawless progress achieved this sprint. In the meantime, we had 2 internal meetings. In the first one (3rd, March) he have discussed how we should approach the implementation overall. We have divided the work within the team members and settled down the time of the next meeting. With all the work being already done, we have had the second internal meeting on the 10th of March, pulling everything together and deciding on  how we are going to present everything to our client.

**List of tasks completed and estimations:**

We have a fully operational version of the code base on Azure meaning the latest version is accessible by our client at anytime. This version has all the bugs fixed such as appropriate time rendering which takes into account the location of the Database server, Azure server, client machine and twitters server time location. Every tweet is properly displayed now, previously on rare occasions we had an empty box instead of the tweet. this was caused when the tweet contained any of the following characters in any combination ““ ‘ \”. The performance of the project is now ideal (under 1 second) as compared to before (25-30 seconds), this is due our use of local code instead of making REST api calls. We have improved the User-Interface for the control panel. Now the options are evenly spaced out and adjusts depending on the number of available parameters.

**Plan for next two weeks:**

The plan for the following 2 weeks include the following:

* Refine the bag of words algorithm to use percentages rather than fixed values
* Make the list of tweets expandable, so that the user can see more than 30 without reloading the page and earlier tweets as well

We aim to have those finished by the end of this sprint, which is around this time, next week.

**Individual Paragraphs:**

**Andreas:**

During the last 2 weeks I have worked towards the most important part of our project - the development of the sentiment analysis algorithm (bag of words). After we have discovered that the algorithm should significally improve the run-time of our application, we have started implementing the different functions and methods needed to make it run. Along having a solid contribution to this, I have also solved some of the server issues that we had, out of which I should outline the dispay of the wrong time of the tweets. In addition to that, I have written a script to handle some corrupt files which were given as input to our algorithm in order to train it. As usual, keeping contact with our client and arranging internal and external meetings was one of my key responsibilities as a group leader.

**Alvee:**

I primarily worked on implementing the bag of words algorithm and also solving some issues with the current code base. I worked heavily on cleaning the training dataset, creating and training the model for sentiment analysis and writing the code that returned the sentiment. One of the key challenges was to clean the training data as on windows machine each line ends with “CR” and “LF” character and these characters are not shown on most text-editors. Having the data cleaned and stored nicely in an array I serialised the model which was made from the training data. I did this to avoid having to train the model every time we wanted to use it. The serialised array was stored in a file which can now be loaded easily into a php script and used as part of the sentiment analysis. I also worked on solving the missing tweet-boxes and emoticons rendering issues, both of which has now been solved.The missing tweet-boxes was a bit hard to resolve as javascript is embedded alongside php inside html and the use of \ “ ‘ inside strings between these three languages is quite tricky and caused most of the problems. Finally, I implemented the bag of words algorithm and replaced the DatumBox Api and pushed the working project on the server.

**Chaitanya:**

These past two weeks I focused my work on helping clean some portion of the training data, making the control panel’s UI dynamic so that the buttons re-scale according to the number of options on the scrren and fixing the difference between azures server and our local testing server so that we have a deterministic perfomance on both servers. I also gave a helping hand to solve the emoticon rendering issues and missing tweet-boxes.