Harry Scells
Craig Craci
28 January 2015

Candidate Topics For A Research/Programming Project

Predictive analysis of the weather using twitter

This project aims to use geographic data from twitter to predict the weather based on the content of tweets. Using the geo-data attached to every tweet and lexical analysis of the content of the tweet, it can be mapped to a graph which, with enough training, should be able to determine the weather. With enough user input and data, it should be possible to produce a geographical map with the predicted outside conditions. This project will involve artificial intelligence to analyse text, attempt to determine if the text contains weather conditions, make a generalisation against users tweeting about the same content, and map it to a geographical area.

Analysis of mass amounts of social media is in the realm of big data, which is an emerging area of research and can provide some insight into crowd sourcing information. There have already been projects aiming to make predictive analysis of tweets, including predicting earthquakes and typhoons. The outcome, like similar projects of this kind, should be that of getting results before existing technologies and systems can provide them. It can also benefit people whom are at risk of tornados and snow storms in North America.

The results of the project will be evaluated against current weather conditions. This includes overlaying both datasets above each other to determine how close the predictive analysis is to already established and trusted methods of weather prediction. The text analysis, prediction and mapping will all be implemented in lisp, with the networking in python. The resulting output will be a graph plotting the relationship between the temperature (hot or cold which could infer a heatwave or snow or rain) against the original location of the tweet (east coast to west coast America). The output will also aim to provide a geographical map with overlays, however this can be done well into the end of the project.

Sentiment analysis of bias in Wikipedia articles

This project aims to determine if a given Wikipedia article contains bias or is prejudice or if the article is not written in correct encyclopaedic language. This is useful for determining if the person that corrected or wrote the article has any personal judgements against the topic and can even be applied to other textual documents, including scholarly articles and even news and blog reports.

The results of the analysis will offer where the offence occurs in the text and an overall score, to provide a high level overview of the document. The score will be calculated by analysing combinations of phrases and words, and basing a score off training data. The score will be on a 10-point scale, where 10 is highly bias and subjective, and 0, where there is no bias and the text s purely factual.

The artificial intelligence will be written in lisp, with minor networking aspects written in python (downloading a webpage and extracting text). The sentiment analysis will be performed using natural learning processes in which a machine is trained using a word database against known articles which contain bias. The implementation of this from a user perspective will provide an interface in which to check if a web page is reliable by simply typing in the address. The first version will only look at Wikipedia articles, and if the learning process can be generalised, then it should be able to be applied to any web document.

The difference this project will make can impact report writing as well as journalism. It can be used to find reliable sources of web documents and can even be used to see if interested parties are modifying articles on Wikipedia which can negatively impact general consensus of the topic being edited. The resulting program will effectively be able to determine the issues of bias and related, unwanted artefacts in Wikipedia, and articles from the web.