Risk assessment   **Protection of human health**

Alert Center for foodborne diseases in Switzerland

Jacky Casas, Laurent Zufferey, Omar Abou Khaled
Haute Ecole d'Ingénierie et d'Architecture Fribourg, Boulevard de Pérolles 80, 1700 Fribourg

Keywords
Epidemic detection, food intoxication, food safety, tweets classification, neural networks, tweets localization

Aim of the study
This project was performed to: (i) analyze a known epidemic and the repercussions on social networks like Twitter; (ii) develop a platform that collects tweets based on keywords, localizes the tweets with different approaches and classifies the tweets according to their relevance; (iii) develop a web platform showing statistics about the data collected and the potential threats to the Swiss population.

Material and Methods
In a first step, an analysis has been made on a known epidemic (Le Locle 2015) to find out if a social media like Twitter can indicate a problem. Tweets related to this epidemic were found and they made it possible to draw a graph showing the evolution of this epidemic over time.
The project has then been taken further and a platform was built in different modules. The first module extracts in real time the tweets that contain specific keywords related to food intoxication. Another module uses different algorithms to localize, as precisely as possible, each and every tweet collected. The third module is a machine learning model trained on hand labelled tweets. Its goal is to classify the tweets by their relevance: either related to food intoxication or not related. These three modules work separately and are not tied together in order to give more flexibility to the whole system. Each module is developed in Python.
And finally, the third part was to visualize and to understand the augmented data collected. The visualization dashboard is a webpage developed with Vue.js, a Javascript framework, and the data is loaded on the interface via a homemade API developed in Python with Flask. All the modules communicate with a MongoDB database. The dashboard contains many charts and maps to visualize the tweets but also has many features like the possibility to modify the keywords or setting threshold to receive alerts by email.

Results and significance
The real-time collection of tweets is operational, with around 75'000 tweets collected every day (containing at least one keyword from the given list). The localization algorithms are able to find information on about 80% of them (i.e. time zone, country) and localize precisely almost 30% of them (i.e. city, geolocalization). Each day, around 1'000 tweets are localized in Switzerland. As for the classification process, between 400 and 500 tweets are considered as relevant every day. The score of the machine learning model is 89%, even if it is sometimes hard for a human to say if it's relevant or not. The average quantity of Swiss relevant tweets per day is equal to 4.2 on a duration of 30 days.
A dashboard allows the supervisors to have access to numerous representation of the data collected: general statistics, maps with pin point of every tweet (relevant or with a certain keyword), tweet feed with link to Twitter and to the tweet's author, graphs with keywords statistics. It also allows the supervisors to manually set a tweet to relevant (or set a tweet to non-relevant in case of false-positive). An email alert will be sent to every registered supervisor if anything suspicious occurs.

Publications, posters and presentations
None (yet)

Project *17AC - Alert Center for foodborne diseases in Switzerland*

Project duration *6 months (November 2017 – April 2018)*