



Monday, 6 February 2017

## Real-time Event Detection with Twitter (Master Thesis Proposal)

Twitter is a popular microblogging service. An estimated 317 million users produce daily more than 500 million messages—also known as tweets—that are broadcast to all of the users' followers. Tweets contain at most 140 characters and cover almost any topic, including personal messages and opinions, celebrity gossip, entertainment, news, and more. Current events are widely discussed on Twitter; for example, around 1.7 M tweets were sent on 7 January 2015 in response to the Charlie Hebdo attacks in Paris. Twitter users often provide live updates in critical situations; for example, users tweeted *'In Brussels Airport. Been evacuated afer [sic] suspected bomb.'* and *'Stampede now. Everyone running'* during the attack at the Bruxelles airport on 22 March 2016. Most tweets can be read even by unregistered users, so Twitter can potentially provide a real-time source of information for detecting newsworthy events before the conventional broadcast media channels.

armasuisse Science + Technology is developing an experimental Social Media Analysis (SMA) system to help detect events such as natural disasters, accidents, political outbreaks, or targeted attacks by analyzing Twitter posts. A key component of the system is a retrieval model for finding relevant Twitter posts, given a topic of interest as expressed by a query. Different models have been proposed in the literature but these models do not perform well for Twitter messages.

The goal of this thesis is therefore to develop a retrieval model that is better suited at finding relevant Tweets for event detection. The objective is to develop an unsupervised learning technique in order to avoid cumbersome training phases for all the possible event types. The focus will lie on query expansion techniques that rely on pseudo-relevance feedback. Such techniques have been widely researched for Web search applications but fail to capture the temporal aspects and text structure encountered in Twitter messages. The student shall therefore enhance such models with a more advanced models and demonstrate the improvement of the state-of-the-art.

This master thesis can be conducted at the the ETH Zürich or in the research labs of armasuisse Science and Technology in Thun. In both cases, the student will get access to a large body of Tweets in a computing cluster to conduct experiments with real-world events.

### Contact at armasuisse S+T:

Dr. Vincent Lenders  
Research Program Manager Cyber and Information  
Federal Department of Defence, Civil Protection and Sports DDPS  
armasuisse Science + Technology  
Feuerwerkerstrasse 39  
CH-3602 Thun

Tel.: +41 (0)58 468 27 68  
E-Mail: [vincent.lenders@armasuisse.ch](mailto:vincent.lenders@armasuisse.ch)  
Web: [www.armasuisse.ch/wt](http://www.armasuisse.ch/wt)