Sub-story evolution tracking on social microblogs

Social media has now become the de-facto information source on real world events. The challenge of social microblogs, however, due to the high volume and velocity nature of social media streams, is in how to follow all posts pertaining to a given event over time – a task referred to as *story detection.* Nevertheless, a global social event normally consists of many underlying and subtle constituent parts which lead to different aspects of social effects (aka. Sub-stories or sub-events). The sub-event tracking task have gained much attention in recent related work. [1,2] propose techniques based on topic-modeling, in specific, leveraging *hashtags*[1] or *hierarchical LDA* [2] for sub-event modeling. [2,3] propose graph-based clustering method for sub-event detection. In this work, we go beyond the original task in attempting to detect sub-stories at early stage, before it bursts. The task is of importance in many critic situations.

This work is relevant for project or master thesis.

Prerequisites for students who are interested in the topic:
- Good programming skills with either Java or Python
- Knowledgeable about basic machine learning and data mining models
- Hand-ons experience with sklearn[1] is preferable

What you will get:
- Cool topic for a master project / thesis
- Machine Learning skill set
- Structured thinking[2]
- Fun

If you are interested, please contact: Tu Nguyen: tunguyen (at) l3s (dot) de.

References
[1] Srijith, P. K., et al. "Sub-story detection in Twitter with hierarchical Dirichlet processes." *Information Processing & Management* 53.4 (2017): 989-1003.
[2] Xing, Chen, et al. "Hashtag-based sub-event discovery using mutually generative LDA in Twitter." *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.
[3] Meladianos, Polykarpos, et al. "Degeneracy-Based Real-Time Sub-Event Detection in Twitter Stream." *ICWSM*. 2015.

---

1 http://scikit-learn.org/stable/
2 Structured thinking is a process of putting a framework to an unstructured problem.