Twitter is one of the most popular microblogging services in the world. Half billion tweets are generating every day by its users [1]. Messages posted on Twitter have been reporting everything from daily life stories to the latest local and global news and events. Monitoring and analysing this rich and continuous user‐generated content can yield to valuable information, allowing users and organizations to obtain actionable knowledge. This information can be accessible via Twitter’s public APIs. In order to do so, scientists proposed many different techniques for event-detection approaches that monitored Twitter’s events. Monitoring a set of messages in a specific period of time that are overlapped in vocabulary might lead to a common event. This titled as event stream.

large amount of meaningless messages and polluted content, can be obtain during Twitter stream, which negatively affect the detection performance. To overcome this challenge, instead of using the random sampled data in Twitter’s API, retrieving tweets can be matched by specific condition and concept of that event. This matter can be achieved by collecting some features from the event, like, special keywords, #hashtags, @mentions, and URL links. For example, Starbird and Palen collected information about the 2011 Egyptian uprising by using the keywords“egypt, #egypt, #jan25” , Nichols et al. collected sport related tweets using keywords “worldcup” and “wc2010” . However, the set of predeﬁned keywords is subjective and can easily lead to incomplete data but, the main drawback is, researchers need to know the event in advance. In [2] the objective is to generalize the solution for different kind of event without any prior knowledge.

Burst Detection approaches are the most common practices used for Twitter sub-event detection. This is a way of identifying periods of time in which some event is unusually popular. In other words, it can be used to identify fads, or “**bursts**,” of events over time. To measure popularity, different statistical tactics might be tested. For example, Earle at el. compare the short-term-average of tweets volume to the long-term-average for identify possible earthquakes, as well as the sport events and political affairs. In another example, ESA (The Emergency Situation Awareness) burst detector uses a binomial model B(p, N) to generate an expected distribution of feature occurrences in a given time window. Some other researchers propose modelling the event stream as a mixture of multiple topic models. However, this technique reached good result but a priori knowledge regarding the number of sub-event to be detected and thus cannot be applied to the real-time sub-event detection task.

Sub-events detection by Real-time microblog monitoring (STRIM)

framework

In [2] the focus is to achieve higher accuracy on sub-event detection by increasing the coverage of the event content. By collecting, detecting and then extracting the most descriptive event-relevant tweets, the framework automatically formulates descriptions of the sub events.