

1 Введение

Платформы микроблогов стали очень популярным способом размещения данных в Сети. В них можно найти сообщения пользователей практически на любую тему, начиная стихийными бедствиями и заканчивая рейтингами музыкальных исполнителей. Правильная обработка доступной информации — нетривиальная задача, которая имеет множество областей применения.

Отслеживание сообщений о стихийных бедствиях в реальном времени поможет вовремя организовать спасательные операции и сохранить жизни людей[1]. Руководствуясь сообщениями пользователей микроблогов можно судить о популярности товаров и вовремя принимать экономически целесообразные решения. Можно делать предположения о рейтингах политиков и эффективности рекламы на основании информации в микроблогах. Помимо перечисленных способов применения доступной информации в микроблоггинговых платформах можно привести множество других.

В данной работе в дальнейшем будет рассматриваться сервис микроблогов Twitter. В нем помимо текстовой информации можно публиковать фото и видео, что так же может быть использовано при анализе. В доступном наборе данных можно проводить анализ разных сущностей, эта работа посвящена выявлению событий среди потоков информации. Поскольку трактовка событий в сообщениях микроблогов может быть субъективной, выделим несколько свойств, которыми должно обладать событие.

2 Описание задачи

Цель данной работы — разработать алгоритм, который по входным данным будет строить набор событий. В качестве данных были использованы сообщения пользователей Twitter с 4 июня 2013 года по 31 июня 2013 года, содержащие в себе хэштег #texas. Прежде чем перейти к описанию искомого алгоритма, рассмотрим некоторые существующие подходы к данной задаче.

3 Полученные результаты

4 Заключение

Список литературы

- [1] Extracting Information Nuggets from Disaster-Related Messages in Social Media, Imran, Elbassuoni, Castillo, Diaz and Meier, 2013.