Social Media Analytics - CS-EJ5621

Lecture 3

Aqdas Malik (aqdas.a.malik@aalto.fi)

25.09.2020

Aalto University
School of Science

FITech
NETWORK UNIVERSITY

Contents of the lecture

- **Course practicalities**

- **API basics**

- **Twitter basics**

- **TAGS setup and operations**

- **Misc. data collection tools**

Aalto University
School of Science

Course practicalities

- **Quiz 2 due by 2359 today**

- **Few still missing Quiz 1**

- **Case project proposals due next week (10%)**

- **Google and Twitter (+developer) accounts**

Social Media analytics framework
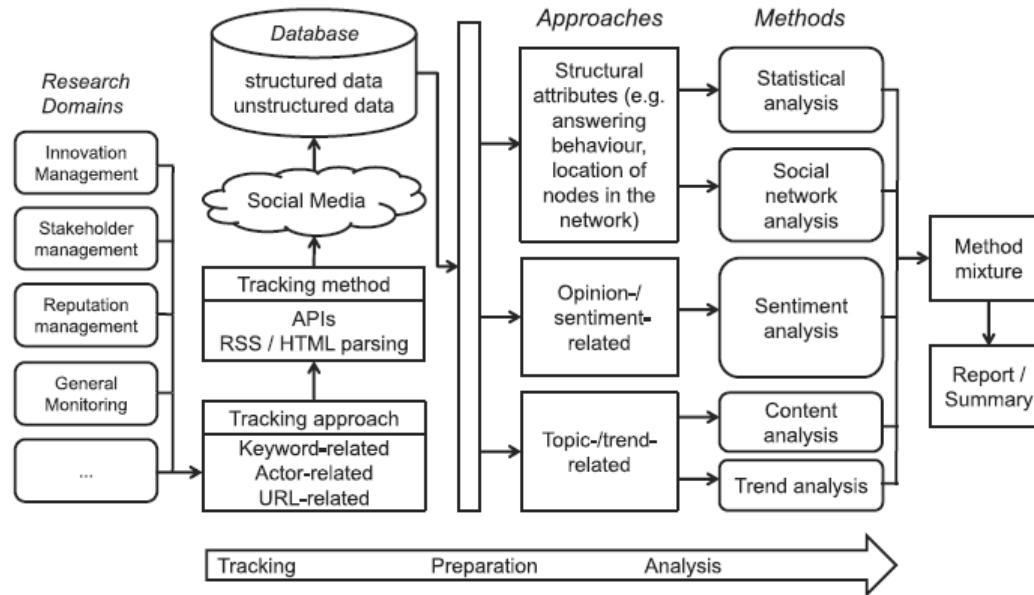
Data collection through APIs - basics

- **APIs (Application programming interfaces) are increasingly used in research**

- **Interface of a computer program that allows the software to "speak" with other software**

- **A data pond made available for public by the company**

- **Enables access to nonpublic Internet environments (requiring authentication)**

- **Not completely open**

- **Few social media grant APIs access for research**

Data collection through APIs - benefits

- **Novel methods & significant opportunities for research**

- **Detailed and systematic access to usage patterns (human-behaviors, communication, activity etc.) that is not possible with traditional research methods (e.g. interviews and surveys)**

- **Nonintrusive and instantaneous**

- **Longitudinal/temporal studies**

Data collection through APIs - challenges

- **Computational skills and resources**
- **Basic know-how of different collection methods (strengths and weaknesses) for selecting the relevant tools**
- **Non-representative sample**
- **Unable to answer - why users are doing something?**
- **Data cleaning (to some extent)**

About Twitter

- **Microblogging**
- **Founded in 2006**
- **280 characters**
- **#, RT, and @**
- **Trending topics**
- **Verified accounts**
- **Protests & Activist movements**
- **Public figures & World leaders**

Key considering before acquiring Twitter data

- **Historical or current**

- **How many tweets?**

- **Complete or sampled**

- **One-time or recurring retrieval**

- **Key attributes (retweets, favorites, geolocation etc.)**
  https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object

- **Technical skills**

- **Resources**

- **Academic or commercial**

- **Analytical skills**
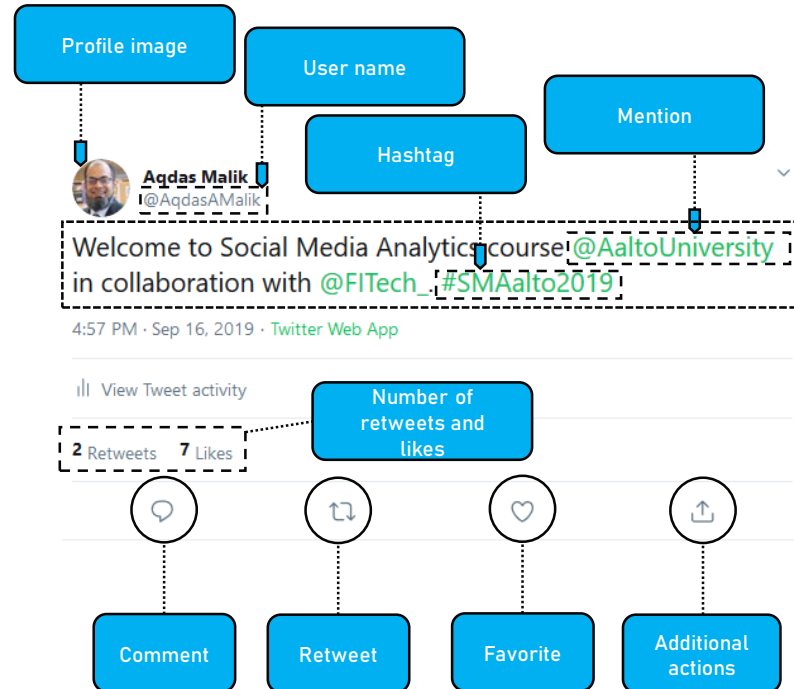
Acquiring Twitter data

- **Write your own code**

- **Copy-pasting from Twitter**

- **Existing Twitter dataset**

- **Purchase from Twitter/services provider (firehose)**

- **Retrieve from public API
  - Software libraries (Tweepy, rtweet, Twitter4J)
  - Web applications (DMI-TCAT, Netlytic)
  - Plugins for popular analytic packages (Atlas.ti, NVIVO, NodeXL, TAGS)**

Limitations of Twitter API

- **Limited access to historical tweets (~3,200 tweets from the last 7-9 days)**

- **Access to current tweets is also limited (~1%)**
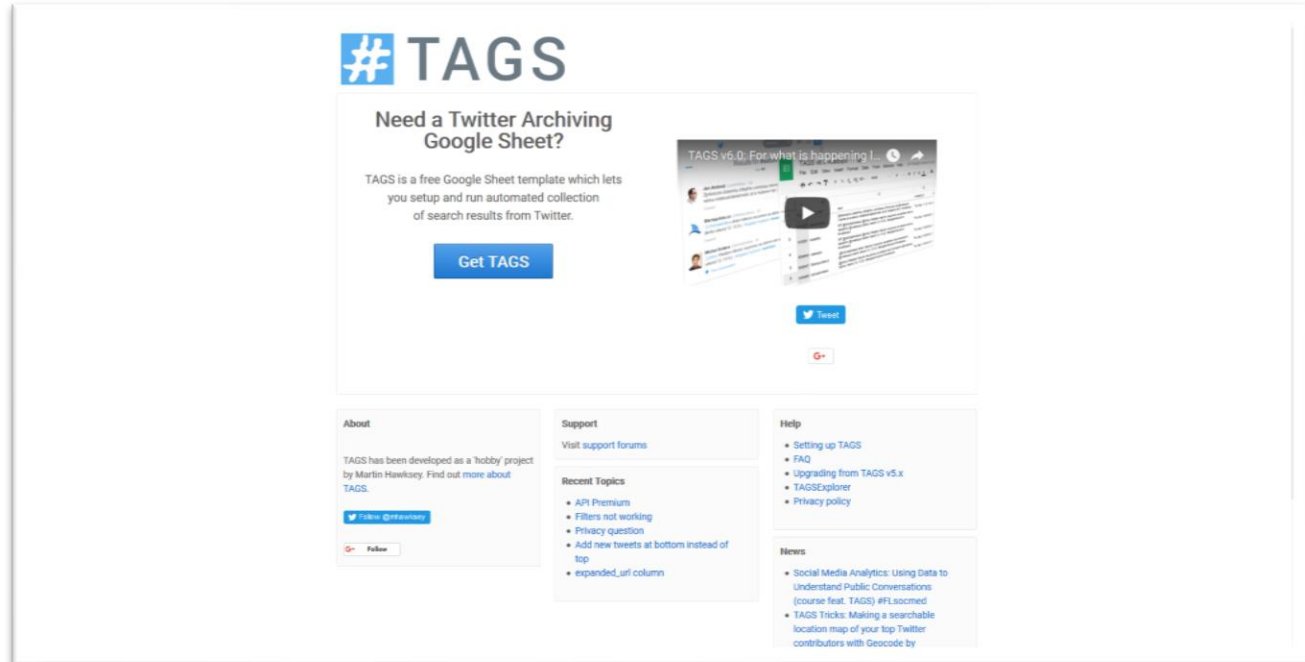
- **How Twitter samples the data is widely unknown**

More info: https://developer.twitter.com/en/docs/basics/things-every-developer-should-know

Anatomy of a Tweet

Profile image

User name

Hashtag

Mention

Aqdas Malik
@AqdasAMalik

Welcome to Social Media Analytics course @AaltoUniversity in collaboration with @FITech_ #SMAalto2019

4:57 PM · Sep 16, 2019 · Twitter Web App

View Tweet activity

2 Retweets   **7** Likes

Number of retweets and likes

Comment

Retweet

Favorite

Additional actions

Aalto University
School of Science

Tags setup

https://tags.hawksey.info/

Publically available datasets

- **Data.World @SocialMediaData (multiple datasets)**
- **2016 United States Presidential Election (280 m tweets)**
- **Winter Olympics 2018 (13 m tweets)**
- **Hurricanes Harvey and Irma (35 m tweets)**
- **#panamapapers (4.9 m tweets)**
- **#WomensMarch (14 m tweets)**
- **4,500 news outlets accounts (26 m tweets)**
- **26 Crises/ calamities (~250K)**

A? Aalto University
School of Science

Other tools for Twitter Data collection

- **Chorus**
- **ContextMiner**
- **COSMOS**
- **DMI-TCAT**
- **Mozdeh**
- **Webometric**

Next lecture – 02.10.2020

- **Native vs. non-native analytics**

- **Metrics for analyzing social media data**

- **Organizing gathered data & cleaning**

- **Detecting bots**

**Aalto University
School of Science**

Thank you