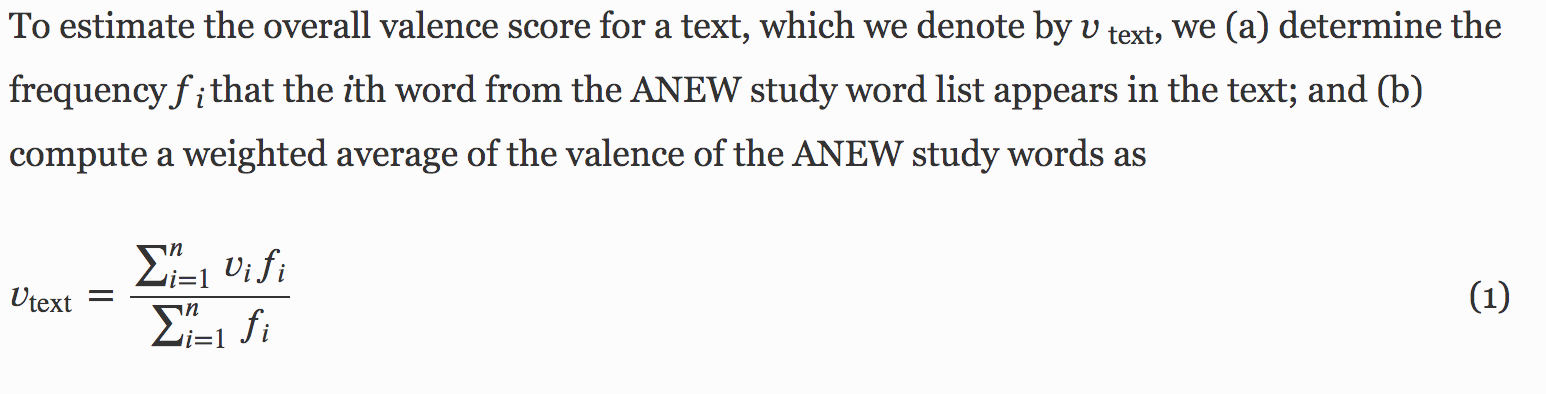
**Twitter Data Mining, Sentiment Keywords Literature Review**

**Dodds, P.S., Danforth, C.M. Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents. J Happiness Stud 11, 441–456 (2010).** [**https://doi.org/10.1007/s10902-009-9150-9**](https://doi.org/10.1007/s10902-009-9150-9)

* The study authors drew directly on the Affective Norms for English Words (ANEW) study (Bradley and Lang 1999) for determining the “happiness” level of individual words
* The following figure from the article represents how the authors scored a given text
* “Since our method does not account for the meaning of words in combination, it is suitable only for large-scale texts. We argue that the results from even sophisticated natural language parsing algorithms (Riloff and Wiebe 2003) cannot be entirely trusted for small-scale texts, as individual expression is simply too variable (Lee 2004) and must therefore be viewed over long time scales (or equivalently via large-scale texts)” This could be an issue for us, as tweets are limited to 240 characters.

**Lui, et al. The Tweets They are a-Changin’: Evolution of Twitter Users and Behavior**

[**https://mislove.org/publications/Profiles-ICWSM.pdf**](https://mislove.org/publications/Profiles-ICWSM.pdf)

* This article generally looks at how Twitter data has changed over time.
* However, I think one of the most relevant takeaways for us is that the study authors used existing Twitter data sets to track change over time (for example, the *Crawl* dataset and the *Gardenhose* dataset). Is there potentially an existing set we can pull from? Of course, this may cause sampling issues, but it could be a way to get the data pre-COVID without a backdoor work-around.

**Quynh C. Nguyen, et al. Geotagged US Tweets as Predictors of County-Level Health Outcomes, 2015–2016, American Journal of Public Health 107, 1776\_1782,** [**https://doi.org/10.2105/AJPH.2017.303993**](https://doi.org/10.2105/AJPH.2017.303993)

* In this study, the authors used Twitter’s streaming application programming interface to collect a random 1% subset of publically available geo-located tweets in the United States.
* This article may provide a useful model on collecting geo-tagged information, though it is worth noting this was county-level, as opposed to country-level

**Chi Y. Bahk, Melissa Cumming, et al. Publicly Available Online Tool Facilitates Real-Time Monitoring Of Vaccine Conversations And Sentiments. Feb 2016,** [**https://doi.org/10.1377/hlthaff.2015.1092**](https://doi.org/10.1377/hlthaff.2015.1092)

* This article analyzed sentiment towards vaccines through social media with a publcially available platform for monitoring vaccination-related content, called the Vaccine Sentimeter. I imagine we’re aiming for a similar goal related to sentiment and COVID-19.
* From this, I gathered it’s important to specify our scope. Are we looking at happiness in general before/during COVID, are we looking at sentiment directed at public health interventions (ex: social distancing regulations/mask mandates)?

**Entity-centric topic-oriented opinion summarization in twitter**

**Meng et al, Aug 2012,** [**https://doi.org/10.1145/2339530.2339592**](https://doi.org/10.1145/2339530.2339592)

* This article studies the problem of opinion summarization for users in Twitter with an entity-centric topic-based opinion summarization framework.
* We need to consider how we are summarizing sentiment, etc.

**A twitter recruitment intelligent system: association rule mining for smoking cessation**

**Ahmed Abdeen Hamed, 2014,** [**https://doi.org/10.1007/s13278-014-0212-6**](https://doi.org/10.1007/s13278-014-0212-6)

* This article presents the study authors’ approach to digial recruitment, presents the algorithms and data mining approaches using Twitter data, and describes their system and results.
* They created an algorithm that identifies potential recruits for a smoking cessation program
* This article raises questions about digital recruitment. What do we need to keep in mind when scraping these tweets for research purposes?
* The article also raises questions about the type of Twitter data being taken into account (original posts, retweets, replies, etc.)

**Ahmed A. Hamed ,An Exploratory Analysis of Twitter KeywordHashtag Networks and Knowledge Discovery Applications, 2014,** [**https://scholarworks.uvm.edu/cgi/viewcontent.cgi?article=1324&context=graddis**](https://scholarworks.uvm.edu/cgi/viewcontent.cgi?article=1324&context=graddis)

* This dissertation provides a good overview of existing literature (some of which I have discussed here)
* This appears to be the dissertation of the author of the above Twitter recruitment intelligence article. This document goes into more detail about the background, etc.
* The 4th chapter provides a discussion of hashtag networks in biomedical web science. It may be worth exploring the *HashnetMiner* algorithm discussed by the author. Hashtags such as #stayathome could be a basis for different associations.