

Introduction to Text Mining using Twitter, TAGS, and Voyant Tools



Andrew Kulak, M.A.
Department of English
Human Centered Design

What is Text Mining?

- Text mining is a type of data mining
- Data mining refers to the computer-assisted collection and analysis of large bodies of data
- Many tools exist for the collection and analysis of data from various sources, including text-based data
- Data about you is mined every day by government agencies, corporations, and websites

Ethics of Text Mining

- Remember that human beings are at the other end of (many) usernames
- You should not use or share information from text mining for targeting or harassing users
- In your analysis, you should respect the rights of users who have chosen to remain anonymous

Let's Try it Out!

To Prepare

- Have technology ready and a web browser open
- Sign out of all of your Google accounts except the one you will use for this activity
- Sign into the Twitter account you will use for this activity

Plan for Activity

1. Set up TAGS in your Google Drive
2. Determine a keyword and build an archive
3. Use Voyant Tools to analyze trends

1. Set up TAGS in your Google Drive
2. Determine a keyword and build an archive
3. Use Voyant Tools to analyze trends

Set up TAGS

1. Important: Make sure you are only signed into **only one** Google account
2. Access the TAGS website at tags.hawksey.info
3. Get TAGS v6.1 by copying TAGS into your Google Drive from the TAGS website

1. Set up TAGS in your Google Drive
2. Determine a keyword and build an archive
3. Use Voyant Tools to analyze trends

Build an Archive

1. Determine a keyword centered on a controversial topic that you are interested in researching
2. Enter the term in the appropriate box on the TAGS settings page (be sure to check spelling!)
3. To build an archive, connect TAGS to Twitter and select from the menu bar TAGS>Run Now!

1. Set up TAGS in your Google Drive
2. Determine a keyword and build an archive
3. Use Voyant Tools to analyze trends

Analyze Trends

1. Navigate to the "Archive" page of your TAGS sheet
2. Click to select tweet content from the "text" row and copy
3. Access Voyant Tools website at voyant-tools.org, paste tweets into text field, and select "Reveal"

Questions to Consider

- What are the most frequent words used by tweeters discussing your keyword?
- Why do you think these terms are relevant to the topic you selected?
- What terms were unexpected or surprising? Based on the original text, can you infer any connections?

Digging Deeper

- What links and images are being shared? How do they relate to your topic?
- Tip: Do you see among your frequently used terms seemingly random strings of letters and numbers?
- Who are the most frequent participants in your Twitter conversation? Who is associated with these accounts?
- Tip: You can use another Voyant Tools window to analyze other rows from the archive spreadsheet

Keep Exploring

- TAGS can be set to automatically run every hour—build an extended archive of your topic
- You can export data from Voyant Tools to design your own visuals and graphs
- The TAGS explorer and searchable archive are available if you make your TAGS Google Sheet public
- There is a whole universe of tools for other types of text and data mining to experiment with

Assignment

- Identify some hashtags that are frequently used together with your keyword, but that seem to reflect different stances on the topic
- Create different TAGS sheets and build archives for several of the hashtags related to your keyword
- Use Voyant Tools to analyze different hashtags and compare and contrast the conversations

Questions to Consider

- What kinds of claims are being made? What rhetorical moves do tweeters make to support their claims? Which moves are successful? How can you tell?
- Which moves are common among all of the hashtags you are studying? Which moves are unique to specific tags? What does this tell you about the tags and their audiences?
- Do tweets featuring one hashtag seem more credible to you than others? Why?

Follow me on
Twitter @KulakVT

andrewkulak.com



Andrew Kulak, M.A.
Department of English
Human Centered Design