

The Twitter Connection [Working Title]

Gabe Benjamin, Cole Harbeck, and Adrian Wan
Swarthmore College
{gbenjam1, charbec1, awan1}@cs.swarthmore.edu

1. PROJECT PROPOSAL

1.1 Problem

A description of the problem you plan to solve and the motivations for doing so (i.e., why this problem is interesting/important).

Twitter is a social media platform with 284 million users active per month, tweeting approximately 500 million tweets per day. [2] At a maximum of 140 characters per tweet and about 1 byte per character¹ this represents an informational flow of over 70GB of text per day. An average book is about 2MB in size [1], so Twitter users are collectively writing about 35,000 books a day.

With this wealth of textual information that is often supplemented with geo-location data and content-connecting hashtags, it is no surprise that a multitude of tools have arisen to harvest the information encoded in tweets. Fabric, Twitter's API [3], has been used to build visualizations of tweet locations (NEED CITATION), NEED LIST OF OTHER EXISTING PRODUCTS (GABE).

Existing products that leverage the informational wealth of Twitter are mostly interested in geolocation, and are either purely historical or purely live. We seek to move away from this problem space by creating a product that:

- Uses both historical and live-stream data
- Focuses on relationships between tweets, users, and content, rather than on just their location.

blah

1.2 Goals

The goals you have for the project. What constitutes success and how will you evaluate it?

Our goal for this project is to have a fully functional visualization depicting the popularity of tweets (based on re-tweets) and the connectivity of tweets being currently

¹Issues arise with different encodings, but we ignore these for the sake of our back-of-the-envelope calculation

tweeted (based on @replies and hashtags). We will also keep a small (several week long) history of tweets instead of creating a new graph of tweets starting from when the program is initialized.

1.3 Hypothesis

Your hypothesis: given the work you intend to do, what are the results you expect to see? How does this work help to solve the problem?

1.4 Environment

Characterize environment you intend to operate in. Does your project operate on Amazon Web Services? on the general Internet? in a data center?

We intend to operate using Amazon Web Services. TWITCON(WT) consists of an always-on EC2 instance

2. REFERENCES

- [1] L. Daly. How big is the average epub book?
- [2] Twitter Inc. About Twitter, Inc., 2014.
- [3] Twitter, Inc. Twitter developers, 2014.