

Twitter Product Referral Bot

Team members:

- Geobio Boo
- Leslie Chang
- Lauren Fratamico

All the members of our group will contribute equally throughout the project. Each member is expected to contribute to coding, research, etc.

Mentor:

- Marcus Phillips

Project Goal:

We are building a referral bot that first finds keywords in tweets where someone is expressing a desire for a product and then sends a reply to the tweeter with a message such as “Here is the top selling [product they were looking for]!”. This will be a subscription only bot. It will only tweet to followers if they express a tweet in a form similar to “I need x” or to people who tweet at the bot with a tweet of that form. The results will be queried from Amazon. Since this bot provides automated responses, we need to be sure that the tweets are valuable for our subscribers so we don’t get marked as spam. The grade we earn will be determined as follows:



Grading Criteria:

A: We finish the bot and get at least one non-spam follower (not including us) or have people interacting with our bot.

B: We finish the bot but nobody follows/uses it.

C: We do not finish the bot in time.

Project timeline with milestones, strategy:

Deadline	Task	Completed?
11/5	Do background research in the areas of Twitter spam detection, Amazon referral links, and Twitter bots in general.	
11/13	Bot is set up and is able to read tweets and reply with a recommendation.	
11/26	Perform any refactoring necessary to have a more successful referral bot.	
12/10	Bot is finished and we have successfully referred a product. Perform analytics: number of referred products sent out, which categories they were in, most successful referred products, if any.	

Literature Review:

Most of our research was done on spam detection, referral efficiency over social networks, and tweet text processing. Our research on spam detection was needed to help our bot avoid being classified as spam by Twitter and avoid replying to spam tweets we receive. Our research on referral efficiency will increase the probability a user clicks and buys the recommended product. Our research on word processing helps us correctly extract the keyword from a tweet and give a relevant product recommendation.

The next steps for literature review is to look into successful product referrals and algorithms for detecting what people really want. Additionally, extra research will go into detecting what the user meant, since the top search result on Amazon often does not match what the user intended.

Spam Detection Research

- Search engines rank tweets based not only on the content of the tweet, but also on how influential the user who posted the tweet is and the more followers the user has, the more influential they are. (1)
- Spammers increase their influence by following popular users. (1)

- The more popular a user is, the more likely they are to reciprocate to spammers' follower requests. (1)
- Many spammers use popular hashtags in their tweets along with a link unrelated to the trending topic. (2)
- Spammers post a higher number of tweets with URLs versus non-spammers. (2)
- Spammers post a higher number of hashtags per tweet versus non-spammers. (2)
- Most spammers have very new accounts. (2)
- Spammers mention many users in their tweets in order to increase their visibility. (3)
- Spammers tend to follow more users than have users following them. (3)
- Spammers who tweet out links to the same site are part of a spam campaign. If the link is not blacklisted, then the spammers may be undetected. (7)
- Tweets that contain a high number of frequently used spam words versus the total number of words are more likely to be spam. (7)

Referral Efficiency

- Tweets with longer lengths are more likely to be clicked. (4)
- Tweets that had linked about 25% of the way through had a higher click rate. (4)
- Tweets with more adverbs and verbs than nouns and adjectives had a higher click rate. (4)
- Weekends and nighttime give you the highest visibility. (4)
- The higher the time gap between tweets, the more likely your link will be clicked. (4)
- Share valuable content in your own voice - personalize the tweets (8)
- use keywords - they are the backbone of content (8)
- connect with followers - in our case, give them more personalized recommendations to better ensure that they will use our service again (8)
- engage your audience - don't just set the twitter account on autopilot (8)
- use hashtags to curate conversation around our bot (8)
- to increase twitter followers (9):
 - encourage retweeting - gets your @username more seen
 - fill in your bio
 - include pictures in tweet
 - get involved with hash tag memes and trending topics
 - track results
 - TwitterCounter (<http://twittercounter.com/>) shows how many new users you're adding per day
 - Qwitter (<http://useqwitter.com/>) will email you when someone unfollows you after a tweet

Word Processing

- Research on detecting questions that are desired to be answered by friends in tweets - related as we are detecting desires and may branch out to more than just the basic "I want X" construction (5)

- people trust their friends to answer questions that are difficult to answer using search engines. they trust their friends to provide tailored, contextual responses. social network like Q&A is becoming more popular
- hard to extract questions from twitter due to the short, informal, nature of tweets
- created a NLP parser for tweets, as sometimes even things that are formatted like questions, are not ones that are meant to be answered - we may run into this problem too
- problems with tweets that make them hard to parse
 - people less concerned with correctness when composing tweets, errors are common
 - garden variety spelling errors can be handled with spelling correction algorithms
 - tweets often have repeated letters "hrmmmmmm", "hahahahahaha" - probably won't have to deal with this in our case as we will be looking for the nouns
 - homophones - shorter syllables generally substituted for longer ones "b4", "2morrow", "sumthing"
 - punctuation - can be random
 - emoticons
- developed a parser that took the above into account - if we run into trouble parsing our tweets, could follow up with them for more direction
- Research on creating a part of speech tagger that is more appropriate for use with tweets (6)
 - poor ability of standard NLP tools on tweets. research focused on re-building the NLP pipeline starting with part-of-speech tagging
 - difficult because 140 character limit - lacks sufficient content to determine what a word is referring to (could be a band, movie, store, company, product)
 - by training their parser on large amounts of unlabeled data, with dictionaries of words in Freebase, and using topic models, they were able to develop a parser that achieves a 25% increase in F1 score over a previous approach
 - tools available at http://github.com/aritter/twitter_nlp with directions on how to run there

Accomplishments:

We have a fully working prototype which currently streams tweets, filters for “I want _____”, and then does natural language processing to get the adjectives/noun that the user is looking for, and then does a search on Amazon, and tweets to the user that they should look at the product we found from Amazon. Currently we use streaming because we have not yet built up our subscribers, and are just prototyping (all current tweets are manually deleted within a few minutes of posting).

Currently, we use the twitter4j library for streaming and posting, OpenNLP for natural language processing, Amazon’s API for product lookups, and bitlyj for shortening URLs.

Additional work will go into improving the matching algorithm, and including more info as to why we selected this item to recommend to the subscriber.

Next Steps:

The next steps for our project is to fine tune our keyword extraction from tweets sent to our Twitter bot. We would like to extend it to handle more complex phrases - infer desires that are not expressed in a basic "I want X" sense. To accomplish this, we may use some of the Twitter NLP taggers that we found during our research to extract the main noun phrases. We also plan to correct for spelling mistakes and handle products that cannot be found on Amazon. In addition, we will also make improvements to our product recommendation algorithm by considering the user's age, gender, etc.

Work Percentage:

	Geobio	Lauren	Leslie
Code	33.3%	33.3%	33.3%
Research	33.3%	33.3%	33.3%
Report	33.3%	33.3%	33.3%

Bibliography:

1. Ghosh Saptarshi, Viswanath Bimal, et al. "Understanding and Combating Link Farming in the Twitter Social Network." *21st International World Wide Web Conference*. Lyon, France. April 2012. <http://www.mpi-sws.org/~farshad/TwitterLinkfarming.pdf>
2. Benevenuto Fabricio, Magno Gabriel, et al. "Detecting Spammers on Twitter." *Seventh Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference*. Redmond, Washington. 2010. <http://ceas.cc/2010/papers/Paper%2021.pdf>
3. M. McCord, M. Chuah. "Spam Detection on Twitter using Traditional Classifiers." *Autonomic and Trusted Computing Conference*. September 2011. http://www.cse.lehigh.edu/~chuah/publications/atc11_spam_camera.pdf
4. Dan Zarrella. "How to Get More Clicks on Twitter." *The Social Media Scientist*. <http://danzarrella.com/infographic-how-to-get-more-clicks-on-twitter.html#>
5. Kyle Dent, Sharoda Paul. "Through the Twitter Glass: Detecting Questions in Micro-Text." AAAI. Palo Alto Research Center. May 2011. www.parc.com/content/attachments/through-twitter-glass.pdf

6. Alan Ritter et al. "Named Entity Recognition in Tweets: An Experimental Study". *EMNLP*. University of Washington. 2011. turing.cs.washington.edu/ritter-emnlp2011-twitter_ner.pdf
7. Zi Chu, Indra Widjaja, et al. "Detecting Social Spam Campaigns on Twitter." *ACNS 2012*. Singapore. June 2012. <http://www.cs.wm.edu/~hnw/paper/acns.pdf>
8. Cindy King. "17 Twitter Marketing Tips From the Pros". *The Social Media Examiner*. October 2011. www.socialmediaexaminer.com/17-twitter-marketing-tips-from-the-pros
9. Kevin Rose. "10 Ways to Increase Your Twitter Followers". *Techcrunch*. January 2009. <http://techcrunch.com/2009/01/25/kevin-rose-10-ways-to-increase-your-twitter-followers/>