Spring | 2012

Evaluating Corporate Social Responsibility Practices

A Case Study of Using Twitter for Measuring Public Perception towards CSR

Saghar Iranzadeh Tamaddon

Chloe Reynolds

School Of Information

University of California Berkeley

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Corporate Social Responsibility (CSR) is a self-regulating mechanism established in a company in order to help it embrace responsibility for the company's actions and encourage a positive impact through its activities on the environment, consumers, employees, communities, and all other members of the public sphere who all can be considered as stakeholders. Various business activities fall under CSR umbrella. It is widely accepted that CSR adheres to a business bottom-line and stakeholders' interests but with no formal act of legislation. ISO 26000 is an effort to offer guidance on socially responsible behavior and possible actions; it does not contain requirements and, therefore, in contrast to ISO management system standards, is not certifiable (Wikipedia 2012).

Considering CSR broad definition, evaluating companies in respect to their CSR practices is a difficult task. The main reason lies with the challenges of defining a unique fits-for-all CSR policy, since CSR pertains the values that matter to business stakeholders who might have different preferences.

Companies need to take a variety of actions and measurements as a means to assess their CSR actions' effectiveness. One direction is to understanding stakeholders' perceptions. This can be accomplished through listening to their dialogues on social media platforms. People tend to express themselves more freely and with less researcher-bias in social media compared to other communication media. Several cognitive steps take place before a person expresses herself in words. Therefore the communicated message may not express the full gamut of intended meaning.

The "Echolytics" project is an experiment of analyzing publicly communicated opinions via social media in respect to social and environmental responsibilities of corporations. The purpose is to discover whether it is feasible to create a reliable and meaningful score that reflects social media users' perception about the CSR practices of a company by mining social media data.

For the first phase of our project, we decided to choose Twitter as the main source of social media data. Twitter tends to host high volume of publicly communicated messages and is widely accepted as a buzz platform. Tweeters do not need to be a member of a blog or a fan of a page in order to share their thoughts with the community. Tweets are usually short and tend to capture moment-by-moment feelings of the tweeter.

For a period of two months, we gathered tweets that contain one or more of 6 brand names: Clorox, Dettol, Lysol, Panera Bread, Chipotle, McDonald's (including aliases, such as mickey d's for McDonald's). We used Natural Language Processing (NLP) and data analysis techniques to classify a tweeter's perception towards the companies' social responsibility. Classification was done according to the sentiment of a tweet and also its relatedness to the CSR context. We ranked companies based on the volume of positive and negative tweets that were flagged as CSR-related tweets. The ranking delivered an absolute score for each company and also placed them in one quadrant of CSR BUZZ and Sentiment charts. The former score helps our audience understand a brand's performance over time. The latter placement compares a brand against their peers in terms of number of tweets, "Buzz," and expert versus social media perception ratings, or "Goodness." Our Echo score is a subjective indicator rather than an objective one.

In conclusion, we found that it is feasible to track and measure tweeters' judgments about a specific brand. Our findings spoke to how brands' CSR were perceived and cascaded in Twitter. We propose this roadmap as a feasible methodology for future work that combines computer and social science to evaluate socially perceived CSR of any brand, industry and corporation in any social media platform.

KEYWORDS

Corporate Social Responsibility (CSR); Natural Language Processing (NLP); Sentiment Analysis; Social Media

INTRODUCTION

In the era of social media, companies can listen to digital voices of their stakeholders all around the Internet. This digital voice, most often, is the echo of companies' decisions, actions, and performance cascaded through online communities. Many businesses around the world have already started using social media as the next source of market research. Our users' interview shows that many companies, however, struggles to devise valid question in the context of social media analytics. While most part of social media data analysis is around customer acquisition and retention, our analysis is focused on people's perception regarding CSR-related behavior of companies.

Imagine that you are the head of CSR at McDonald's and you need to find an answer for the following questions in order to recognize your stakeholder's perception in social media (i.e. Twitter):

1. How much do people talk about our CSR activities?
2. What is the general sentiment (i.e. positive or negative) of CSR-related tweets?
3. What is the "mood" of people on Twitter about our CSR activities? How does this mood change over time? What does the change correlate with?
4. What are the mostly-talked topics of CSR-related tweets?
5. What CSR topics correlate with negative and positive tweets? (You then might decide to foster the areas that are perceived positively and redesign the business processes that are perceived negatively)
6. What CSR-related tweets are retweeted more? How viral is a CSR-related topic?
7. Are we doing better or worse than our competitors on these dimensions?
8. If we launch a new campaign, what people will think about it on social media?

The list of question can keep going but the main logic behind such questions is how social media analytics can possibly help a company such as McDonald's make better decisions relating to its CSR principles.

CSR is a self-regulated policy and many argue that companies do it as an attempt to mislead their stakeholders. Analyzing spoken thoughts of people who tend to be real stakeholders speaking freely and unbiased on social media can tremendously help watchdog institutions and socially responsible investors to compare the self-reported documents of companies with public opinion.

In this experiment, we conducted a feasibility study by collecting, classifying, and quantifying Twitter data for understanding tweeters' perception toward a brand's CSR. We initially started our idea with the question of what are the possible ways of understanding people's judgments in social media about a brand's doing-good behavior. Our Twitter experiment is to determine the proper methodology and feasibility of addressing our question.

RESEARCH TOPIC

Our experiment is designed to tackle the question of "how a brand's CSR activities are perceived in social media". In order for us to answer such question, we need to define CSR and CSR taxonomy. Then we need to determine if, how much, and how often people talk about a company's CSR in online media – and what they say. Extracting such information yields a scoring platform that not only answers the primary question, but also compares various companies with their competitors and with their industry average.

Company rankings may encourage and promote socially responsible business practices, environmental stewardship, consumer protection, human rights, and diversity, as part of the doing-well-and-doing-good movement.

For this particular experiment we chose to analyze two months of Twitter data for 6 brands in 2 industries – fast food and cleaning products.

USER GROUPS

We envision to serve variety of users: socially responsible investors, consumers, consumer influencers, socially responsible corporations, watchdog institutions, etc. Our user groups have both overlapping and separate needs. Each piece of data might need to be visualized and delivered differently according to each group's interests. Our future works consists of a comprehensive user study and usability testing hence we can customize our findings based on user's needs.

We have discovered the following needs according to a few interviews that we did with limited users.

INVESTORS

Socially Responsible Investment (SRI) principles already attracted a lot of socially responsible investors. The UN also has developed the Principles for Responsible Investment as guidelines for investing entities. We believe that SR investors would like to see the following information in our analysis:

1. Standard CSR scores (charity donations, toxic produced, labor/executive diversity, etc.)
2. Public perception (positive, negative or neutral feeling of people toward CSR activities of a brand)
3. Trends over time (monthly rather than short periods like daily).

WATCHDOGS

CSR is a self-regulated principle and some believe that it is just a window-dressing attempt. Watchdog institutions want to see whether and how a company's action in CSR are perceived and interpreted by the public. Considering the penetration of social media and users engagement, such platform is deemed an invaluable source of unheard buzz. Watchdog entities are interested in knowing:

1. Buzzwords and topics
2. The volume and type of the sentiment about a company
3. Changes of public feelings over time
4. What might have likely happened that correlates with any major change in people's opinion or buzzwords

CORPORATIONS

Socially responsible corporations are interested in knowing the impact of their actions. One approach is analyzing spoken words of people that are related to the context and making sense of such data. Other companies, on the other hand, that do not follow CSR practices or postpone embedding such practices in their business core processes can be encouraged to take an action before it becomes too late. Our analysis can yield the following findings for this user group:

1. Public opinion regarding and reaction to corporate behavior
2. Volume and type of sentiment
3. Their CSR perception status relative to others inside and outside of their industry
4. An informative corporate dashboard for corporate top managers including analysis for each of their different brands
5. Data analysis over any period of time (i.e. capturing public reaction to a specific campaign or event)

CONSTRAINTS

Social media is hosting more and more data from all over the world every moment. This data, most often in text format, encompasses invaluable information that often is hard to extract and make sense of. The art of social media analytics is filtering out the rubbish information, capturing the sensible, and interpreting it according to the context. On the other hand, CSR by itself is an arguable concept and there is no unique global standard that delivers a fits-for-all CSR policy. These two challenges together make the task of analyzing CSR-ness of social media postings even more difficult. Our methodology tries to define the steps required for overcoming these challenges by carefully examining all human/machine-oriented steps.

Additionally, several general constraints limited the accuracy of our scoring system.

1. Understanding the intended meaning of a tweeter's tweet
2. Making sure that the brand name used in the tweets refers to the brand
3. Recognizing whether the tweeter talks about the brand or company and not using its name for another purpose (example tweet: "you're cleaner than Clorox!")

Note here that steps 2 and 3 contain some room for two types of precision error: 1) We may have omitted some aliases, 2) We may have false negatives, such as a search for tweets containing the word "McDonald's" would return a tweet like "Sarah McDonald's new car is sweet!' Depending on the wording of the tweet, it may be difficult to conclusively decipher if a specific tweet is a false-positive or a true-positive.

4. Extracting the meaning within the context
5. The low volume of CSR tweets compared to the general volume of tweets that made our training-set small.

In addition to the general constraints, our research team had a few administrative restraints:

1. Time
2. Budget
3. Human resource

METHODOLOGY

Our methodology is summarized in "Figure 1 Methodology: Twitter Data Classification (Sentiment and CSR)".

1. We first chose 9 brands from 3 industries and we later narrowed the scope to 6 companies from 2 industries. We carefully studied those industries, brands, and their current existing CSR and expert rating (GoodGuide for this experiment). We interviewed a few experts to understand the overall needs. GoodGuide objective rating of each brand was saved for further comparison analysis.
2. Next we created a list of Twitter search queries based on these brands' names and some variants of their names. For example, in the case of McDonald's, we ran case-insensitive searches for McDonald's, Mackey-D's, Mickey-D's, and golden arches. The gathered list of aliases was then pre-tested in Twitter's search feature to ensure the terms were fruitful (actually used by tweeters) and did not result in too many false-positives. Additionally, we used Twitter's search function to gather the handles of brands (Appendix 1), which we then added to the list of daily search queries. For simplicity, we limited the tweets to English-language only. The complete list of search terms is in Appendix 1.
3. We then created a Python code to access the Twitter Search API and run these queries daily between 01.01.12 and 02.29.12. Gathered tweets were then saved in a PostgreSQL database.
4. In the next step, we attempted to classify which tweets were related to CSR. In order to do this, we needed to create a training-set. We used crowdsourcing mechanism to flag a portion of tweets with their sentiment and whether or not they are CSR-related. Since these are subjective tasks, we assigned each task to multiple individuals and used the agreed-on answers as the right category. and we adjusted our NLP Algorithms.

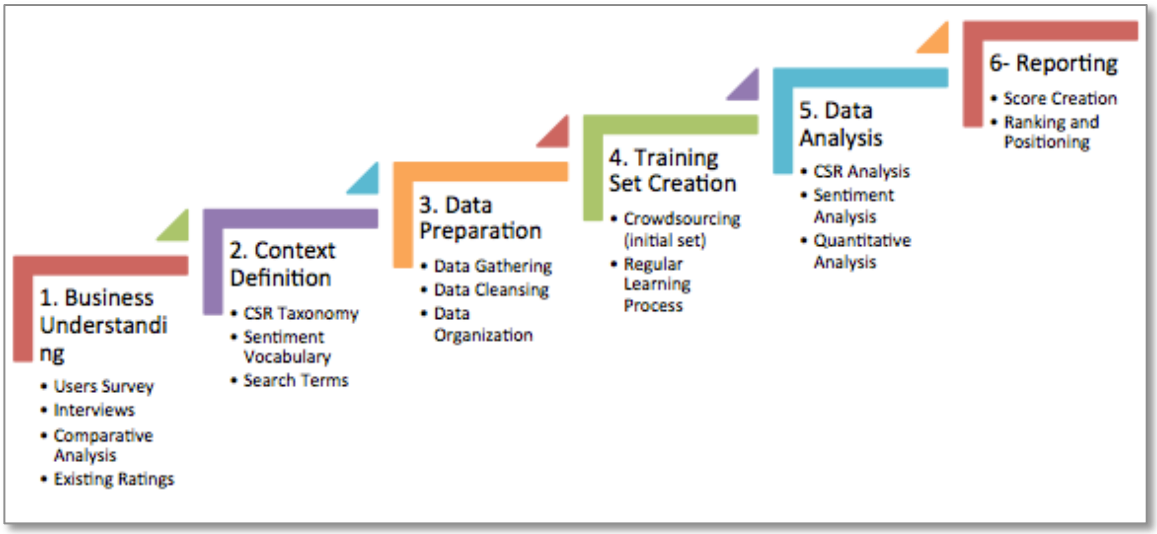


Figure 1 Methodology: Twitter Data Classification (Sentiment and CSR)

5. With a training set in hand, it was then possible to classify the yet-unclassified tweets into:
    a. those which do or do not refer to the brand (eliminating false positive tweets),
    b. those which do or do not discuss CSR,
    c. and what the sentiment of each tweet is

From this point forward, references to positive CSR tweets for example refers to tweets that were classified as tweets that do discuss CSR and have a positive sentiment toward the brand of interest. Tweets that were classified as not referring to the brand of interest were excluded from CSR and Sentiment analyses.

Note that the sentiment of the tweet may be different than the sentiment toward the brand of interest mentioned in the tweet. It was the sentiment toward the brand that we classified.

Classification was the most important analysis we undertook because it was the foundation of all ensuing analyses. Once classified, we did quantitative analysis to quantify the volume and percentage of CSR-classified and positive/negative tweets for each brand.

6. Later stage analyses included finding the most common terms -"tokens" and "bigrams"-and the most common hashtags of positive and negative CSR-classified tweets.

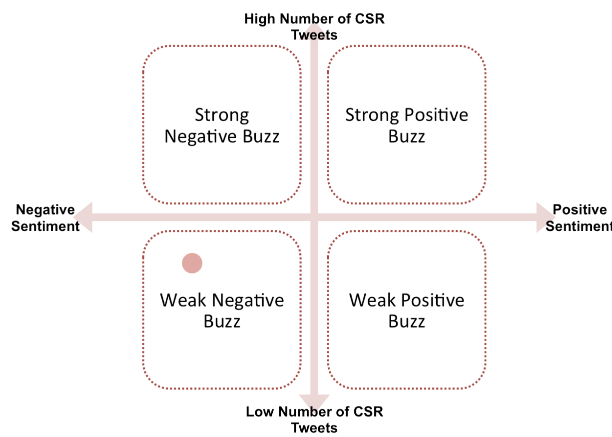


Figure 2 CSR Buzz: CSR-related Buzz in Social Media

The final stage was ranking and scoring each brand according to the quantitative analysis of classified tweets

The Echo score – an absolute score - is percentage of Positive CSR Tweets to Negative CSR Tweets. The ranking number is a relative score that places a brand among all rated brands in terms of highest portion of positive versus negative CSR Tweets.

We also positioned each company to a quadrant in 2 charts.

“CSR Buzz” chart -Figure 2 CSR Buzz: CSR-related Buzz in Social Media- positions a company according to the volume and sentiment of CSR-classified tweets of a company.

"Goodness" chart -Figure 3 Goodness: Social Media CSR Perception- positions a company according to its GoodGuide objective rating vs. our subjective rating.

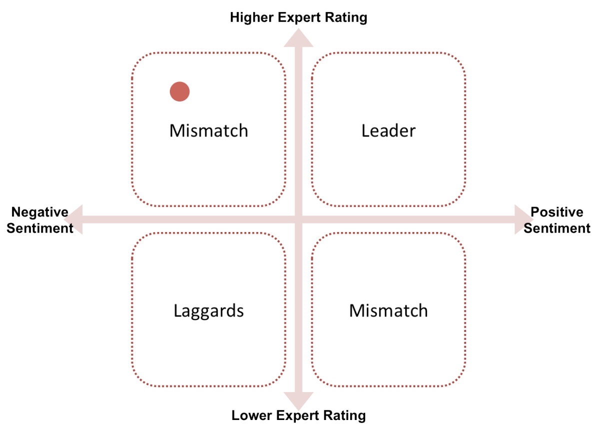


Figure 3 Goodness: Social Media CSR Perception

FINDINGS

Running all the steps described on the methodology section led us to ranking and positioning each company according to its data. Here is a quick stat of each brand. The full report is available at our website www.echolytics.org.

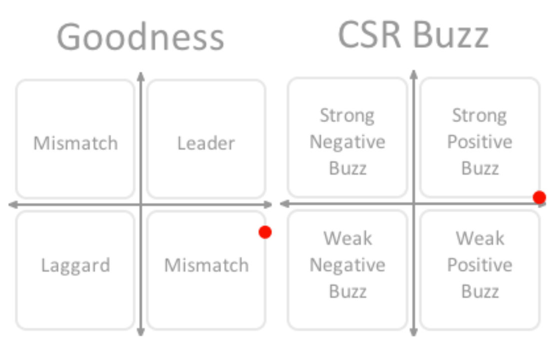
CSR-related tweets were only about 0.05% to 2% of all tweets, according to manual ratings of a subset of tweets. For the 3 cleaning product brands, not enough CSR-related tweets were identified manually on which to create an NLP training set. On the other hand, the number of CSR Tweets identified in the food industry were 7% or over of all the manually rated tweets. Because so few tweets were CSR-related, we brainstormed CSR keywords and tested the accuracy of how often a tweet containing each keyword was actually CSR-Related. Those keywords that yielded over 90% accuracy of correctly predicting CSR-relatedness were used to flag yet-unclassified tweets as CSR-related. After using classified and new key terms, percentage of CSR-related tweets increased. The accuracy of our classifier is based on a 70% training-set and 30% test-set of pseudo-randomly selection of the manually and crowdsource-rated tweets. Also, the size of the training sets used varied by brand.

Having applied our methodology on the experiment data, we ranked and scored the selected brands. The summary of our data analysis findings is below. However the complete report is available online on our website (www.echolytics.org).

PANERA BREAD CAFÉ

Brand: Panera Bread Cafe
Industry: Fast Food
Ranking: 1
Echo Score: 90

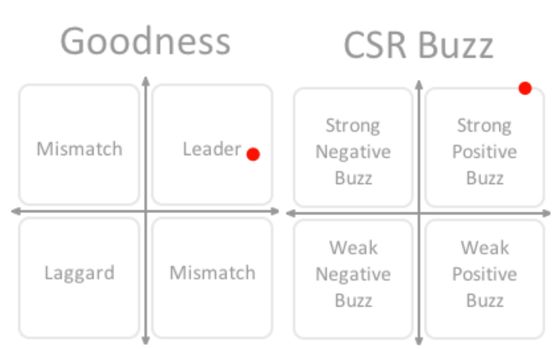
Position in Quadrants:



MCDONALD'S

Brand: McDonald's
Industry: Fast Food
Ranking: 2
Echo Score: 87

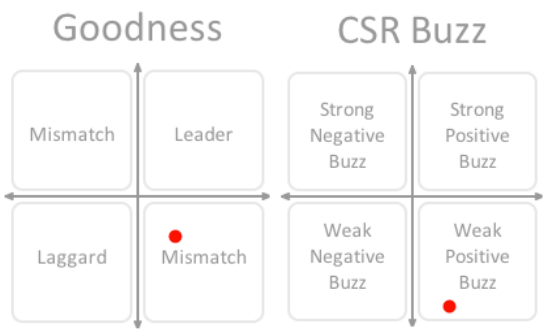
Position in Quadrants:



CHIPOTLE

Brand: Chipotle
Industry: Fast Food
Ranking: 3
Echo Score: 76

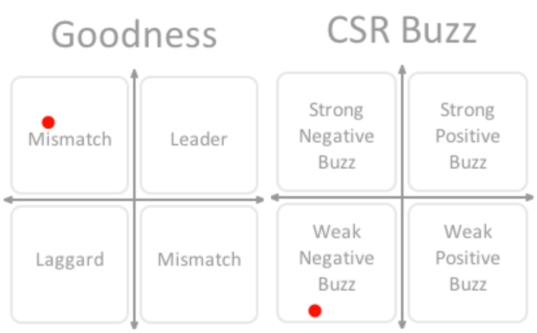
Position in Quadrants:



DETTOL

Brand: Dettol
Industry: Cleaning Products
Ranking: 4
Echo Score: 56

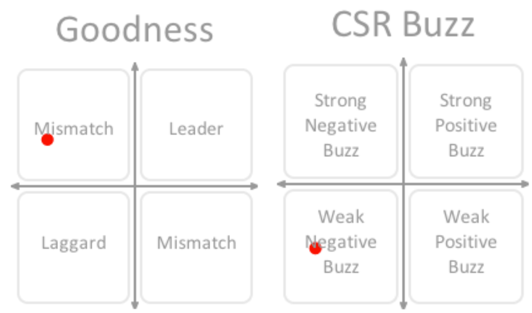
Position in Quadrants:



CLOROX

Brand: Clorox
Industry: Cleaning Products
Ranking: 5
Echo Score: 55

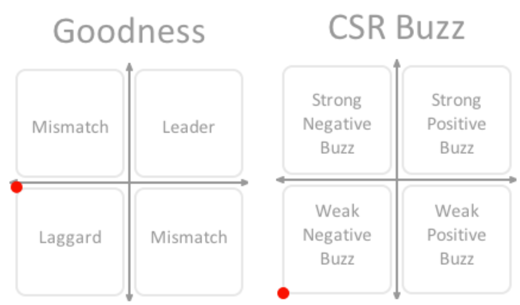
Position in Quadrants:



LYSOL

Brand: Lysol
Industry: Cleaning Products
Ranking: 6
Echo Score: 50

Position in Quadrants:



Other Findings according to Top Terms analysis:

- Website links and retweets were common in positive CSR tweets.
- Retweets were common element of negative CSR tweets.
- PR tweets occupied noticeable amount of retweets, so they can change the scores for a brand. In our future refinement we need to separate out retweets (example "fat city bitch" video link for McDonald's).
- Having looked at the top terms and hot topics, it seems that the opportunities for CSR improvement is different across industries (i.e. producing a safe-to-drink bleach vs. healthy food). We believe it is fair that as a future refinement we separately calculate our score per industry.

Not enough CSR tweets were identified manually on which to create an NLP training-set for the 3 cleaning product brands. The number of CSR Tweets identified in the food brands were 7% or over of all the manually-rated tweets.

Our training-set was relatively small that caused our NLP classifier accuracy to be lower that our expectation. This accuracy percentage is based on a 70% training set and 30% test set of pseudo-randomly selection of the manually and crowdsource-rated tweets. The size of the training sets used varied by company.

DELIVERABLES

Our deliverables for the current experiment is a package including the following items:

DATA COLLECTION INFRASTRUCTURE

- Python code to collect tweets through the Twitter Search API
- Python code to remove duplicates
- PostgreSQL database
- Lots of SQL queries
- Data cleaning and organizing procedures
- Python Unicode cleansing code

TRAINING-SET INFRASTRUCTURE

- Procedures and data preparation instructions for crowdsourcing platform
- Manual spot-check verification and refinement
- Manual CSR keyword prediction - per company
- Manual CSR keyword prediction - per industry

DATA ANALYSIS INFRASTRUCTURE

- Python (natural language) NLP Classification code
- Python Tokenization code
- Python Bigram code
- Python Part of Speech Tagging code
- Python Chunking code
- Python Frequency Counting code
- Definition of our scoring/ranking system
- Ranks and scores for 6 pre-selected companies
- Future: Sentiment prediction NLP code
- Future: Find or create a tweet corpus for POS tagging

REPORTING INFRASTRUCTURE

- Website (www.echolytics.org)
- Information visualization charts

FUTURE WORK AND REFINEMENTS

We have identified several items as the future work and refinements based on our experiment. All is listed in Appendix 2.

CONCLUSION

We have defined a CSR score for 6 companies in 2 industries according to public opinion expressed in Twitter. The results of our efforts confirmed that it is feasible to use social media data to create a quantitative CSR score according to the buzz. We believe that after applying future refinements, our approach can be used to collect and analyze the public opinion from any social media network towards any company or industry's CSR actions. According to such analysis, companies and industries will be scored and ranked against each other. Some parts of our approach need human intelligence until our CSR data for each industry and company reaches to acceptable limit of machine-learning accuracy.

ACKNOWLEDGEMENTS

We thank Iris Cheung for partial website development, Mona Ghandi for logo design, Anthony de Jesus for video direction and production, David Shamma for tweet analysis consulting, the Beyond the Arc team for analysis consulting, and Aron Cramer and David Challis for providing business need insights.

APPENDIX 1 – TWITTER SEARCH TERMS

Brand	Search Term
Clorox	'Clorox', 'from:@Clorox
McDonald's	'McDonald\'s', 'golden arches', 'mickey-d\'s OR mackey-d\'s', 'from:@McDonalds OR from:@McDonaldsCorp'
Panera Bread Cafe	'Panera',
Chipotle	'Chipotle -sauce', 'Chipotle AND sauce', 'from:@ChipotleTweets OR from:@ChipotleMedia'
Dettol	'Dettol', 'from:@Dettol'
Lysol	'Lysol', 'from:@Lysol'

APPENDIX 2 – FUTURE REFINEMENTS

- Separating tweeters into organizational vs. individuals categories and compare all findings by those classification
- Separating tweets into retweets vs. non-retweets to see if any findings differ across these categories
- Automating all parts of the system that do not need human intelligence (i.e. scraping tweets, updating database, cleaning data, running classification and NLP codes, updating the website, posting crowdsourcing tasks)
- Identifying influencers and clustering of users and "bridges" according to social network theory
- Identifying networks of hot discussions
- Building a taxonomy of terms related to one incident or one theme, which may be iteratively updated as news
- Using a thesaurus-like tool for better NLP
- User needs analysis and usability testing of our service
- Identifying various personas and improving our website user experience
- Creating information dashboard per persona
- Clean data dynamically (with some scripts and some alerts)
- Refine keyword lists. Refine process for refining keyword lists (sites to check, searches to run)
- Improving CSR identification (improving CSR definition based on each industry and brand, using larger and quality-checked traning-sets)
- More NLP procedures to perform such as Part Of Speech (POS) tagging, chunking and narrowing top terms to certain parts of speech or chunks- specifically noun and adjective phrases
- Refining keyword lists and related processes
- Categorizing ranked brands for each industry

BIBLIOGRAPHY

Adam Marcus, Michael S. Bernstein, Osama Badar, David R. Karger, Samuel Madden, Robert C. Miller. *twitInfo.* http://twitinfo.csail.mit.edu/detail/31/ (accessed 05 01, 2012).

Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. *Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments.* Pittsburgh, PA: School of Computer Science, Carnegie Mellon Univeristy,.

Sites, Google. *Natural Language Toolkit.* http://www.nltk.org/.

Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python.* O'Reilly Media, 2009.

Wikipedia. "Corporate social responsibility." *Wikipedia.* Apr 28, 2012. http://en.wikipedia.org/wiki/Corporate_social_responsibility (accessed May 2, 2012).