12-2019

Happy toilet: A social analytics approach to the study of public toilet cleanliness

Eugene W. J. CHOY

Winston M. K. HO

Xiaohang LI

Ragini VERMA

Li Jin SIM

See next page for additional authors

Author

Eugene W. J. CHOY, Winston M. K. HO, Xiaohang LI, Ragini VERMA, Li Jin SIM, and Kyong Jin SHIM

Happy Toilet: A Social Analytics Approach to the Study of Public Toilet Cleanliness

Eugene W. J. Choy
School of Information Systems
Singapore Management
University
Singapore
eugene.choy.2016@smu.edu.sg

Winston M. K. Ho
School of Information Systems
Singapore Management
University
Singapore
winston.ho.2016@smu.edu.sg

Xiaohang Li
School of Information Systems
Singapore Management
University
Singapore
xiaohang.li.2016@smu.edu.sg

Ragini Verma
School of Information Systems
Singapore Management
University
Singapore
raginiverma.2016@smu.edu.sg

Li J. Sim
School of Information Systems
Singapore Management
University
Singapore
lijin.sim.2016@smu.edu.sg

Kyong Jin Shim
School of Information Systems
Singapore Management
University
Singapore
kjshim@smu.edu.sg

Abstract—**This study presents a social analytics approach to the study of public toilet cleanliness in Singapore. From popular social media platforms, our system automatically gathers and analyzes relevant public posts that mention about toilet cleanliness in highly frequented locations across the Singapore island – from busy shopping malls to food 'hawker' centers.**

Keywords—sentiment analysis, toilets, social media, analytics

I. INTRODUCTION

Toilet is a basic necessity in everyone's daily life. Unpleasant toilet experiences can negatively impact people's lives. Toilets in public places are shared by many people on a daily basis, and it requires regular cleaning and monitoring. With over 5.6 million local residents and over 18 million tourists from around the globe visiting Singapore annually [1], the government's National Environment Agency (NEA) seeks to ensure the cleanliness of public toilets. In this study, we aim to gather interesting insights such as trending topics and public opinions about public toilets in Singapore.

Restroom Association of Singapore (RAS) was set up in 1998 as a non-profit organization. RAS works closely with NEA to promote cleanliness of public toilets. In 2003, RAS launched the Happy Toilet Program (HTP) as a star-grading initiative for public toilets. It focuses on five areas: Design, Cleanliness, Maintenance, Effectiveness, and Satisfaction. Toilet owners can apply for their toilets to be assessed by RAS. Upon inspection and receiving satisfactory rating, RAS issues a certification with the rating.

Our group met with the RAS team and participated in toilet inspections. RAS's team of volunteers perform manual on-site checks and assessment on various public toilets all around Singapore. With approximately 1,150 public toilets to monitor, it is challenging for the volunteers to identify areas of concern in a speedy manner. Additionally, it is difficult to monitor public toilets on a frequent and on-going basis throughout the year due to resource constraints.

To tackle this challenge of resource constraint, we turned to the World Wide Web. Specifically, we conducted a preliminary research by visiting several popular social media sites, and we discovered that there were lots of digital traces of social conversations about the cleanliness of public toilets around Singapore. They include Reddit, Twitter, Instagram and TripAdvisor. Our solution is an automated system that crawls and analyzes social media mentions of various toilets in Singapore. It performs topic modeling and sentiment analysis. Our system displays analytics insights in a web-based visualization dashboard.

II. SYSTEM OVERVIEW

We received the full list of public toilets registered under the Happy Toilet Program [2]. From toilet inspection sessions, we understand, first-hand, the current rubrics for assessing toilets in Singapore.
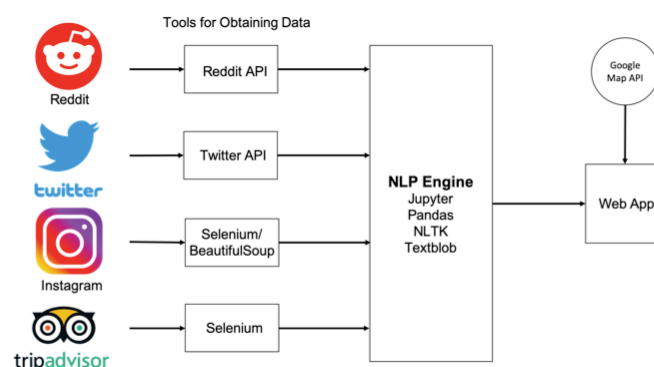

Fig. 1. Public Toilet Cleanliness Study – System Architecture

Our solution (Fig. 1) seeks to understand people's opinions or concerns with regards to public toilets. Our system collects public posts that mention terms related to toilet cleanliness. The collected data include image files of toilets as well as textual data such as hashtags or full sentences.
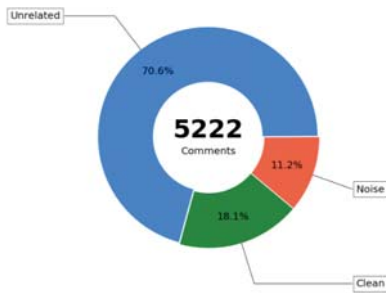
Fig. 2. Reddit Comments – Data Cleaning

Using the Reddit API search endpoint, we retrieved posts that mention 'toilet' from the 'Singapore' sub-reddit in early November 2019. For all the crawled posts, we inspected comments on those posts. Data cleaning reveals that over 70% of the comments do not mention the keywords 'toilet', 'urinal', 'cubicle', 'flush', 'poop', 'urine', and so forth – hence, we removed such comments from the dataset (Fig. 2).

We obtained TripAdvisor data using Python's Selenium library. We crawled all answers to questions mentioning toilet-related keywords. We obtained Instagram posts using Python's Selenium and BeautifulSoup libraries. Crawling was done using certain usernames as well as hashtags – relevant to toilets. Lastly, we crawled tweets via Twitter Premium API endpoint – with toilet-related keywords between 2017 and 2019.
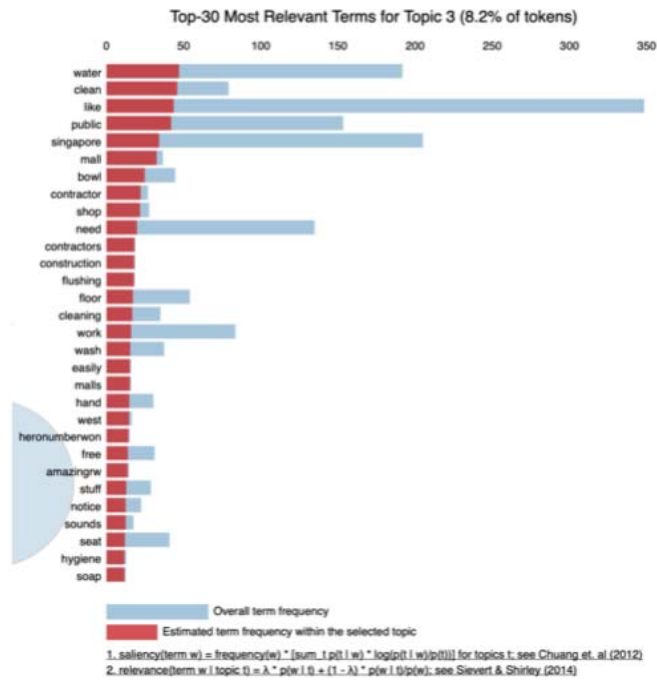

Fig. 3. Reddit Comments – Topic Modeling

We performed Natural Language Processing to analyze textual data to a) assess the overall public sentiment and b) identify specific locations where problem toilets are reported. Our corpus includes major shopping malls and food 'hawker' centers – as they experience high human traffic on a daily basis. We performed topic modeling to identify key topics and conversations that were made based on those relevant topics.

Fig. 3 shows the results of topic modeling on Reddit comments. 20 topics are found, and one topic is closely related to people's opinions about public toilets.

Topic Keyword	Relevant Posts
Water	"There aren't many places in Singapore where I haven't noticed moderately or very clean toilets. Usually, the MRT ones tend to be dirtier than the average mall toilet. Not sure who is the most common culprit." "Considering the humidity in Singapore, I can't imagine what will start to grow with water going everywhere (i.e., mold), if the place isn't kept dry over time"
Clean	Opinions on the toilet cleanness rating tablet: **Positive:** "Actually, i don't rate precisely because i just washed my hands and i don't want to touch that screen because others don't wash their hands after pooing and peeing..." "Exactly. Can't understand why no one else is pointing this out. The toilet isn't really the best place to put a touchscreen device." "Mostly at the airport, to let them know it is a good job keeping it clean. Considering other places with no good review, I will probably skip instead of saying they are bad." **Negative:** "Cleaners are like goalkeepers. Only get noticed if they screw up" "At least clean the toilet bowl first before cleaning the cubical! What's the point of cleaning the cubical first, when you're going to splash dirty water out of the toilet bowl afterwards? It's clear that the company did not think this through enough. Prepare to see sh*t all over the place if this is going ahead with no changes to the procedure." "I don't think it's right for female cleaners to be in male toilets (especially when there are men inside). "

Fig. 4. Relevant Reddit comments with keywords 'water' and 'clean'

Strong Active Sentences Give the Best Accuracy	
"Yep. I always rate the highest whenever i'm at the airport. Their toilets are so clean"	**Positive**, +0.37
"If my legs ever touch a public toilet seat, I'm cutting is off. "	**Negative**, -0.21
"Disgusting behaviour"	**Negative**, -1.0
"Menstrual blood all over the toilet seat. My friend discovered it at work some more, so it was definitely the work of an adult woman seriously wtf."	**Negative**, -0.11
"CCK MRT toilet. The bowl was up to the brim with liquid shit. Wondered how they cleaned that one."	**Negative**, -0.12

Fig. 5. Reddit Comments – Sentiment Analysis (Strong Active Sentences)

Given Reddit's function as a forum in general, most of the comments appear to be candid in nature. The comments contain commonly used acronyms such as ROFL, LOL, and so on. It also contains acronyms or shortened forms of regions (e.g. CCK for Choa Chu Kang) and shopping malls (e.g. Plaza Sing for Plaza Singapura). Our system is configured with a pre-defined list of known acronyms or shortened forms of locations. Further, some comments contain local Singlish terms of different dialects. While not fully comprehensive, we created our own local Singlish corpus based on our manual inspection. Our system is configured to recognize local slangs of importance (e.g. terms indicative of emotions). Additionally, when Reddit users reply to one another, the previous comment is quoted as part of the reply. We removed quoted 'previous' comments so as to remove duplicate comments for sentiment analysis.

Susceptible to False Positive due a Lack of Context	
"Same in this toilet. No brushes or anything provided to clean.	**Positive**, +0.17
"well you are using PUBLIC toilet.. don't expect much."	**Positive**, +0.06
"where can i find a 6 star toilet ??"	**Negative**, -0.03
Weak to Sarcasm	
"Cleaner than your intentions."	**Neutral**, +0.0
"Spreading singapore culture to the natives. Good job. If Peeping Tom was an Olympic sport we'd own the entire podium,then celebrate by filming the female athlete in their own showers."	**Positive**, +0.23
Location detection relies on spelling accuracy of commenter	
"Went in no toilet paper, realised no flush. Bras Bersah."	**Negative**, -0.03

Fig. 6.   Reddit Comments – Sentiment Analysis (Lack of Context & Sarcasm)

Upon data cleaning and transformation, we performed sentiment analysis using Python's TextBlob library. Fig. 5 shows strong active sentences that use terms strongly indicative of emotions, e.g. 'disgusting', 'shit', 'rate the highest', 'wtf'. On the other hand, as shown in Fig. 6, some comments are susceptible to being incorrectly labeled due to lack of context or sarcasm.
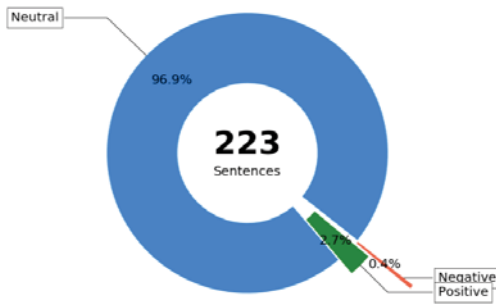

Fig. 7.   TripAdvisor Reviews

Sentence	Score
Positive Sentiment Comments do provide good feedback	
"Yes, it's a good idea to get used to toilet paper. All toilets here are dry toilets. Having said that, for most toilets, the last cubicle, normally, a squat toilet, will have a hand spray or tap. Alternatively you may want to bring a disposable water bottle with water to the toilet. Hope this helps."	**Positive**, +2.0
"the ones at newton circus are very'zen' and lovely:)"	**Positive**, +3.0
False Positives are aplenty due to sarcasm:	
"Anyway, I haven't seen a 'dry toilet' in ages."	**Neutral**, +0.0
"LOL WTF IS THIS SHIT"	**Positive**, +0.03

Fig. 8.   Reddit Posts – Sentiment Analysis

An analysis of TripAdvisor answers reveals that only a fraction of the content include toilet-related keywords (Fig. 7). It is because most of the answers do not only focus on toilets but also other aspects of hotels and restaurants (e.g. price, privacy, transportation, etc.). We inspected the answers manually sentence-by-sentence (Fig. 8). Some sentences appear to be informational while others cannot be labeled properly due to lack of context.
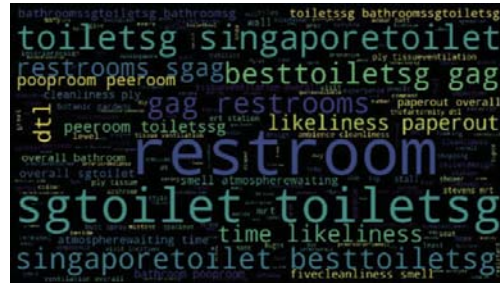

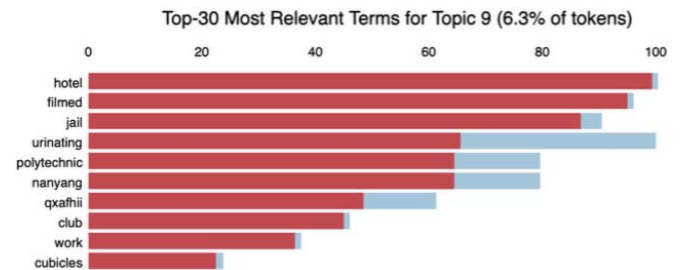Fig. 9.   Instagram Posts – Hashtag Analysis
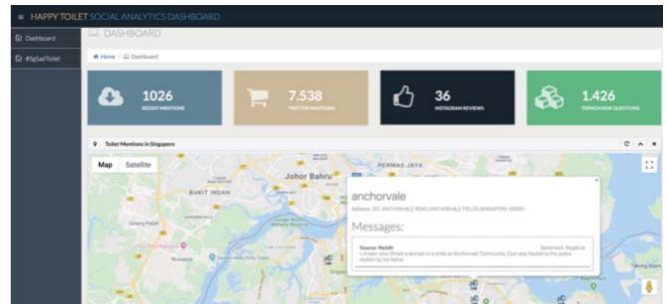

Fig. 10. Twitter Tweets – Topic Modeling

Fig. 11. Dashboard for Live Monitoring of Social Conversations about Toilets

Our analysis of Instagram posts reveals a set of hashtags used by the public (Fig. 9). This is particularly interesting to RAS as they continue to run campaigns to promote clean public toilets. An analysis of Twitter tweets (Fig. 10) shows that there are social conversations about legal issues concerning public toilets such as "… man jailed, … secretly filmed men in toilet cubicles", "…jailed for filming men urinating", etc. Web-based dashboard helps agencies and the public monitor social conversations about public toilets in Singapore (Fig. 11).

III. CONCLUSION & FUTURE DIRECTIONS

In our study, we designed a data-driven solution for listening to social conversations in social media about public toilets in Singapore. With the insights from our analyses, more effective campaigns can be designed to promote active reporting of public toilet cleanliness using common hashtags in social media.

REFERENCES

[1] Singapore Tourism Board. https://www.stb.gov.sg
[2] Loo Map Directory. Retrieved from http://toilet.org.sg/loomapdirectory