Exploring sentimental values across vast amounts of data

Ryan Ang
Arizona State University
Tempe, Arizona
rang1@asu.edu

Kushal Reddy Papakannu
Arizona State University
Tempe, Arizona
kapapkan@asu.edu

Vishal Kumar
Arizona State University
Tempe, Arizona
vkumar47@asu.edu

Aditya Shah
Arizona State Univeristy
Tempe, Arizona
aditya@asu.edu

Tushar Kapoor
Arizona State University
Tempe, Arizona
tkapoor1@asu.edu

ABSTRACT

Social media is growing in popularity as more people become connected to the Internet. The benefits are endless; we can find friends from high school, we can communicate to relatives across vast distances, we can watch sports around the world and we can tune into world events as the happen at this very second. All this global chatter makes us wonder, could we find the emotions in all this traffic? Are there areas where people say happiness can be found at this place, or are there areas where people would warn others not to visit that place. In this paper, we discuss finding sentiment values in hotel reviews.

ACM Reference Format:
Ryan Ang, Kushal Reddy Papakannu, Vishal Kumar, Aditya Shah, and Tushar Kapoor. 2019. Exploring sentimental values across vast amounts of data. In *Proceedings of CSE 578: Data Visualization (CSE 578).* ASU, Tempe, AZ, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

1.1 Sentiment Analysis

While the invention of the Internet and World Wide Web is not new, social media has allowed people to share information faster than the founders could ever predicted. In the past decade, platforms for all sorts of media have sprung up, notably *Youtube*, *Facebook*, *Instagram* and *Twitter*. *Twitter* is particularly interesting due to the fact people can "Tweet" out live updates wherever they are. From 126 million tweets a day[31], analysis of these tweet could provide parties with some interesting relations.

Sentiment analysis is determining if a word or a phrase is a positive, negative, or neutral attitude, also known as 'polarity'. In 2011, Agarwal et al. noted that due to Twitter character limit, users tend to "shorten" their expressions, either by using acronyms (LOL) or emoticons (:O). After preprocessing the text back into full english words, Agarwal and his team compared a unigram model to a tree based model. The unigram model contains 10,000 features pre- built from others whereas the tree based allows the words or phrases to be in various "categories." By considering all possible sub-trees, Agarwal aims to find any correlations between features (defined in the unigram model) and categories of features. The accuracy for the tree alone was 73.93% compared to unigram's 71.35% [2]

1.2 Sentiment in reviews

Not only has the Internet provided us with social interactions across vast distances, we can also review products and services. Two big websites are *Amazon*, an online retailer, and *TripAdvisor*, a user-supplied database of reviews for hotels and restaurants. On a dataset of 20,000 products with over 5 million reviews from *Amazon*, Fang and Zhan mention that normal text analysis is not enough to determine how positive or negative. Their research has determined that machine learning increases the F1 score, where a high F1 means low false positives and false negatives. Overall, Support Vector Machines (SVM) provides the best algorithm to determine the polarity of a review. [9]

Another method to provide a better accuracy for sentiment analysis on large pieces of text is using Latent Dirichlet allocation (LDA) with naive bayes. Unlike SVMs, which provide a simple linear classification, Karim and Das use LDA to model the review as "topics" and "documents". The idea behind LDA is to categorize each "document" into n number of "topics" and find similar clusters to make generalizations. Karim compared the results of LDA with a rule-based tagger, called SentiWordNet, which takes a single word and determines if it positive and negative. Using a dataset of 27,000 movie reviews, the accuracy of SentiWordNet was 59% while LDA was 75%. [17]

1.3 Web applications

When the Internet first started, information was downloaded into static file. However, with the inclusion of Web 2.0, web sites started to include dynamic content, components that would update depending on the time and day. Social media is a major player in Web 2.0. Humans naturally fall into groups they feel familiar with and through the Internet, millions use Instant Messaging to communicate. In 2010, Correa et al. has determined three major personalities that utilize social networking sites (SNS) the most[7]:

- Extraversion. Those who extraverted are likely to use social media, which makes sense since these types want to chat.

CSE 578, Apr 2019, ASU, Tempe, AZ, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

- Neuroticism. The neuroticism personality is related to anxiety, worry, fear, anger, and loneliness. Women are found to be more likely to use social media compared to men in this area.
- Openness. Like extraverts, those who are open to experience will be engaged in interacting with the web applications and are curious about experimenting with the applications. We plan to create an interactive application that appeals to these types of people, which is discussed later.

1.4 Designing effective web applications

Web applications are now essential in today's world, they are well-planned strategies for various industries which help them achieve their goals while providing the same level of comfort to the consumer. However, the development phase of a web application differs from that of a traditional systems [13], and takes a lot of time in using traditional development cycle model. The development of Web-based business applications is broken into three stages called 'tiers' [10]. The three stages/tiers are basically the presentation layer, application layer and storage layer:

- Presentation layer: This layer is mostly concerned with displaying information to the user in the web browser. There are 4 different technologies that are used during the development of this, they are namely Script based, Plugin based, Browser based and Web-based desktop technologies.
- Application Layer: This layer basically hosts the application logic. PHP, PERL, ColdFusion, ASP.NET are some of the technologies that are used in this layer.
- Storage Layer: This layer handles the database connection.

Depending on the budget, scalability, security and other factors industries make their choices among the given options[18].

The architecture of a web application basically defines how the application layer interacts with the database, which comes down to the 'piece of code' which has been parsed to the browser. A web application has 2 parts, shown in Figure 1:

- Server Side: The code running on the server which responds to HTTP requests.
- Client Side: The code parsed in the browser and responds to user input.

Depending on the need, the development team has their options, but Stringfellow advises that an effective web application should have the following features for the satisfaction of the user as well as host as they affect both application and logic layer[25]:

- Solves problems consistently and uniformly
- Is as simple as possible
- Offers fast response times
- Utilizes security standards to reduce the chance of malicious penetrations
- Does not have a single point of failure
- Errors logged in a user-friendly way

Choosing technologies also affect the working of the application/logic layer. The Business Logic Layer contains the determinant part of the application logic. Adamko from University of Debrecen recommends that this layer include[1]:

- performing all required calculations and validations

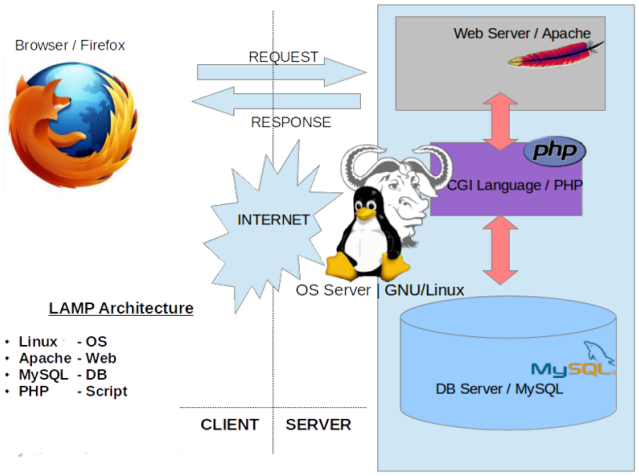

Figure 1: A common web application architecture. Source: Wikipedia

- managing workflow
  (1) state management: to keep track of application execution
  (2) session management: to distinguish among application instances
  (3) user identification
  (4) service access: to provide application services in a consistent way
- managing all data access for the presentation layer

The Business Logic Layer is generally implemented inside an Application server (like Microsoft Transaction Server, Oracle Application Server, or IBM WebSphere). The Application server generally automates a number of services like transactions, security, persistence, connection pooling, messaging and name services[1].

2 RELATED WORK

One of the easiest examples of sentiment analysis is opinion mining. Bjorkelund et al. defines opinion mining as a way to rank a hotel by the reviews instead of using a set number of ratings, such as 1-star to 5-star. It is also beneficial to show the history of such reviews to determine if there is a positive trend for the user. To accomplish this task, Bjorkelund and his team display the polarities (positive and negative) in a form of spatial aspects, or a map. Their interactive application shows the locations of each hotel with heat maps where a larger circle means more reviews and the color green is related to "good" whilst red is "bad." [6]

Sharma and his team has also performed a spatial visualization on sentiments. Using *Twitter's* Application Programmer Interface (API), they ask the user to enter a keyword and pulls the mean and weight polarity of each Tweet. Location metadata in the Tweet is used to display a world map of the polarities for the specified keyword. Another interaction for the user is "drill-down" in which when a country is clicked, the application will zoom in on the country and display individual states with sentiment values. [23]

In 2013, Duan et al. also performed a study on hotel service quality through the reviews. The five dimensions in which a service can be quantified are:

- TANGIBLES. The appearance of physical facilities, equipment and employees
- RELIABILITY. The ability to perform the guaranteed service consistently and correctly
- RESPONSIVENESS. The willingness to assist customers and provide quick and easy service
- ASSURANCE. The confidence in the customers that the employees are sympathetic and their ability to encourage trust
- EMPATHY. Like assurance, this lets the customers know the level of caring and personalized experience are expected of the employees

Duan's research has determined that from the reviews gathered from 1999 to 2011, the TANGIBLES category are the deciding factor when determining sentiment. If a hotel looks nice, then the user will have a better chance of leaving a positive review. His team also sees that having a good experience in RELIABILITY and ASSURANCE leads to a positive review whereas having a bad experience in RESPONSIVENESS will lead to a negative review. [8]

Similar to opinion mining, Barbu and Ziegler discuss how to make a hotel recommender system through personalized user experience. Barbu first asks for features a user would like ("I prefer to sleep on a soft mattress") and compares these preferences to the text of reviews ("I absolutely loved this hotel. The mattress was clean and comfy"). Once the user model is built, the visualization presents how the recommender computes closely related hotels using sankey, a type of flow diagram. These flows can show the user which preferences was used to recommend the list of hotels, show what other users think about the same preference, and show groups of users(young couples or business travelers) with the same preference. For each feature, the sankey displays green for positive and red for negative sentiments. The blend of colors gives the user the ability to see how much of each feature is positive for a specific hotel. When the reviews themselves are displayed, the application highlights which phrases have a positive or negative polarity. [4]

3 MOTIVATION

"Imagine a world in which every single human being can freely share in the sum of all knowledge." [11] One of the biggest websites on the Internet is Wikipedia, a free and open encyclopedia. Backed by the Wikimedia Foundation, Wikipedia aims to share as much information as possible. Due to this success, we aim to build an encyclopedia-like of hotels, thus we named our project, WikiHotels.

To be easy as possible, we plan on not requiring a user account to be created. We aim to educate viewers on which hotels are the best in a certain city using the best design philosophies. Our TripAdivsor dataset contains over 700,000 reviews and the challenge is how should we present this information. Wikipedia is a text and picture based website, so the audience is expected to read in detail. But, what if our users would like a quick glance at a city they are staying at? Designing a good visualization for such a large dataset raises several key questions when developing the application:

- Are the review of late reviewers biased compared to instant reviewers?
- How similar are the reviews from each country?
- Can we accurately determine the positive and negative sentiments of each review?
- Are hotels paying attention to the reviews and trying to improve?
- How effective is a keyword network visualization?

4 SYSTEM DESIGN

4.1 Interactive visualization

There are various reasons to use interactive visualizations over static content. One of the main reasons is because information is absorbed easily and at a faster rate. We have a lot of data and just plotting them on simple line charts could have been enough but adding an interactive component makes it more visually appealing and easier to comprehend. Not only this but a well designed interactive visualization helps attract the users attention and maintain it. If a user looks at a website and sees a cluster of dull graphs full of numbers it is highly likely that the user will not be visiting or using the site often.

Using the visualization mantra as described in Shneiderman, "Overview first, Zoom and filter, then details-on-demand."[24] We want to make a word cloud using this mantra on our word cloud. Initially the user will be able to see the whole word cloud of reviews of the hotels. The user will be able to zoom into the cloud to see the smaller words or filter out positive or negative words. Finally the user will be able to click on a word in the cloud to see more details about it, such as the number of times that word occurred in the reviews and preview of the reviews in which it occurred.

4.2 Vision and color

We plan to use color sequences to display graphs and positive to negative scales. Using color sequence techniques such as the one described by Evan Oliveri and Colin Ware will help us get a smoother color distribution and equalise difference between the scales[22]. According to various researchers if the color sequence increases in luminesce it is easier on the human vision as humans perceive colors via a "luminesce channel" instead of a "chromatic channel". Figure 2 shows Oliveri's and Ware's Java application to select a color sequence. [32]

4.3 Maps and cartography

As our target audience is not limited to any certain region, we need to deliver information in a language that everyone can understand. Maps are one of the ways of communication that everyone can understand without actually speaking or listening to. Automatic translation of text has not reached its full potential and has many drawbacks. Thus, for information which can be perceived visually (like surroundings, directions e.t.c.) maps are the best choice.

All the information about one place cannot be put in a single map. Different aspects of a place can be shown in different types of maps. Cartography is the art science and technique of map reproduction. It is characterized as a refining process between the original source material and the production putting into consideration objective of the map being a meaningful picture of reality. It focuses on the production of maps, including the construction of projections, design, compilation, drafting, and reproduction. Cartography builds on the

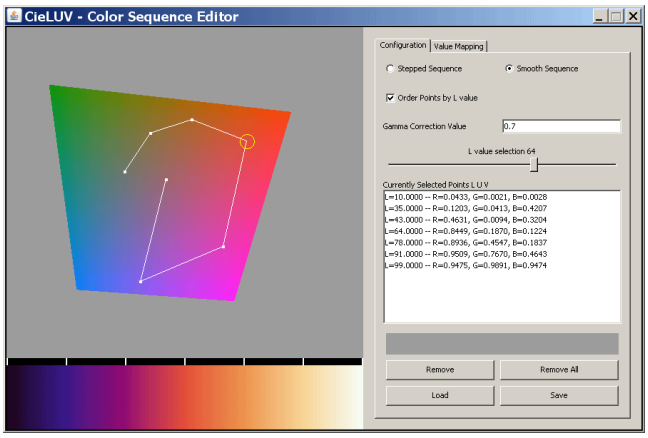

Figure 2: Oliveri's Implementation of CieLUV

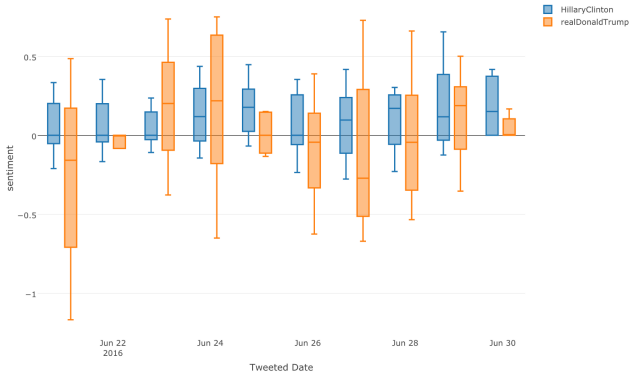

Figure 3: Clinton and Trump tweets over time

premise that reality can be modeled in ways that communicate spatial information effectively [30]. We also wanted our map to stand out so that the user can perceive information as quickly as possible, one of the ways was to color different regions(circle) with different colors.

4.4 Perception & Attention

A user looks at a webpage for a few seconds and forms their opinion in that quick look. There are a couple of things we will improve so as to leave better first impression on the users as well as improve the usability of the website.[29] A search bar for the hotels will be added that auto fills the name of the hotels, this makes it easier for the users who may not remember the full name of the hotel they are trying to look up. Along with this we want the the transition between the cities quick and smooth i.e. the map will zoom out move to target city and zoom into it. Everytime the user clicks on a hotel pin on the map, we plan to add a popup that shows all the graphs at once.

4.5 Time Series & layering

Being able to interpret large amounts of data as a visualization reduces the cognitive load of human perception.[19] We want to show the changes in the average sentiment of the reviews over time for each hotel and the most effective way to do that without overloading the user with information is to use time series, something similar to what is shown in the following picture. The graph will show sentiment values in a range from highly positive to highly negative for all the months from which we have data. Nishida shows how he looks at Twitter over time in Figure 3 [21].

4.6 Social Network Visualization

Visual representations of social networks are important to understand network data and convey the result of the analysis. "images of social networks have provided investigators with new insights about network structure and have helped them communicate those insights to others" [12]. As we want to build a Hotel recommendation system, there can be scenarios where a user may want to know whether any of his friends have visited a particular hotel. One way

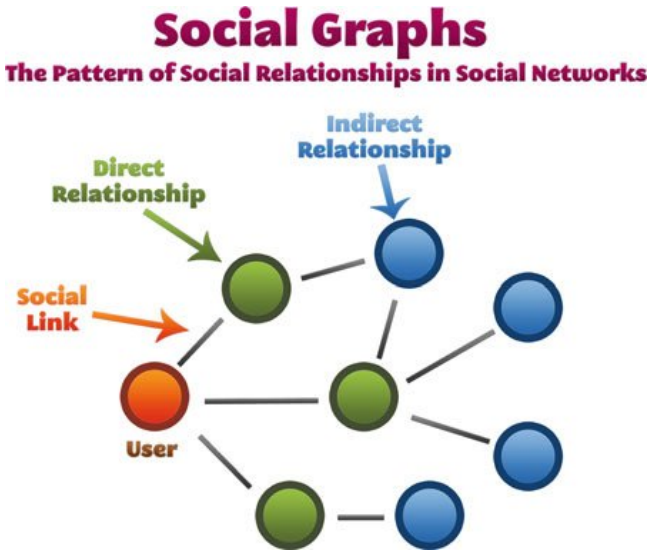

Figure 4: Social relationship in social networks

to show this information can be visually representing photographs of the users' friends when they hover over a hotel on the map or list, this can be an additional feature if the user has a WikiHotels account. Falahi demostrates how a social network works in Figure 4 [3].

Recommending something to a user can be improved when the system looks at the friend circle of that user. In fact, this is how many recommendation systems work. Nayebzadeh et al. mentions "data mining approaches with collaborative filtering model is getting more attention these days to receive a better accuracy. Analyzing the data generated by users within social networks has several practical applications that can be used to develop recommendation systems"[20]. A hotel recommendation to a user may prove to be helpful if the user is not able to make a proper choice.

Figure 5: A tag cloud of words related to Wikipedia. Source:Wikipedia

4.7 Machine Learning & Visualization

As WikiHotels' Business executives, we want to know whether using a specific service or technology will improve our business, provide for better customer experience, and generate operational efficiencies such as speed, cost savings, and greater precision or not. Machine learning today is not like machine learning of the past; many machines learning algorithms like Neural Networks are still a black box but perform really good when applied on a specific task. We ourselves can use machine learning algorithms for several purposes like creating a prediction system for "How frequently we should send emails, news to our registered users." This can be done by using previous data telling how many users subscribe or un-subscribe on receiving which kind of email (advertisement, suggestion, or payment receipt).

While data visualization is used for visualizing current data, machine learning is used for predicting data. We can merge both and instead of predicting data, we can predict which technology to use. This can be achieved by training a machine learning algorithm that learns your productivity graph of past years and when given a specific curve(graph) can predict which technology should be used for getting that curve. For example, Google Quick Draw recognizes what you have drawn in the given canvas and tell its most accurate decision [14].

4.8 Text Visualization

Data visualization expert Edward Tufte explains the importance of words in his book, *The Visual Display of Quantitative Information, Second Edition*. He writes, "Words and pictures belong together. Viewers need the help that words can provide."[28] There needs to be a way in which we can provide a gist about a hotel, whether it's about hotel services, amenities or some other data from reviews. One of the most common ways to represent most frequent info is to use, word/tag cloud, which doesn't provide complete information but gives the basic idea in just a look. Another thing to use can be a word tree.

All these text visualization techniques are useful but used incorrectly or in the wrong place may give negative results. If we use it we need to take care context and semantics, if a user if confused the presented form of text visualization should provide relevant context to aid understanding[15]. Hsiao recommends an effective text visualization should contain the following[16]:

- content linguistic structures
- content semantics
- content similarities
- content connections
- content evolution
- improve text searches
- innovations

4.9 Visual Analytics & Dashboard

Visual analytics is "the science of analytical reasoning facilitated by interactive visual interfaces"[27]. Basically, it's a practice identifying trends, patterns, and relationship in the data which explains for certain event that has occurred. For instance, recall the designs and pattern you see drawn around a player(on a digital screen) during a match of soccer/ baseball. These curves/patterns basically tell us how that player is playing and other info. That's visual analytics. Today's visual analytics of sports-related data is widely used in research[5].

Dashboards are of critical importance in today's business world. And to compete with our rivals, WikiHotels need one too. "A dashboard is a collection of several views, letting you compare a variety of data simultaneously"[26]. Basically they are an 'Engine-check' of our business. We can use a dashboard telling us following info about our WikiHotels:

- A line graph showing how many users and registering/un-registering any day.
- A bar graph showing how many emails you received (category wise).
- A pie chart showing how many people actually booked a hotel out all visited uses.
- A similar pie chart but just for registered users.

4.10 Implementation

As a course semester project, we implemented our prototype for about 8 weeks. We built the prototype to answer some of the questions mentioned in the Motivation section above. We pre-processed our data by running each review text into SentiWordNet to find the sentiment score and store it in the database for fast lookup. All of our initial classification relies on the sentiment value below 0 for negative, above 0 for positive, and neutral otherwise.

Figure 6 shows a map of the selected city from the dropdown menu. When a hotel is selected, a tree map of most frequent words is shown to the right. Scrolling down, the rating history of the hotel is displayed, Figure 7. The number of reviews between those who review within 30 days and after 30 days is also part of the Figure. Lastly, we compare the sentiments between countries in our world map view, as seen in Figure 8.

5 METHODOLOGY

Our design is roughly nine different design concepts and is pretty elaborate. We believe that following these ideas will help us answer our research questions mentioned in the Motivation section. Our main color palette is green for positive and red for negative, but we will include an option to change the colors if the reader can not see the color gradient. Overall, every visualization is interactive as we want those who are curious to click every option.

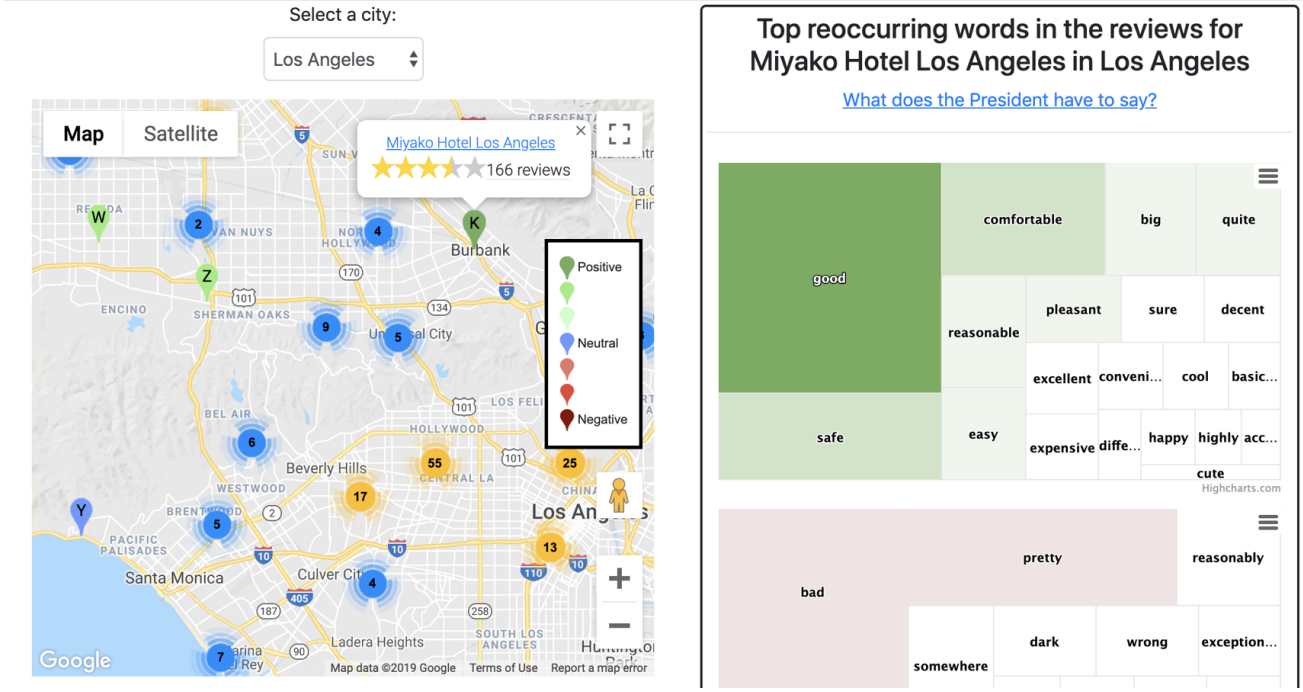

Figure 6: Overview of Los Angeles and tree map

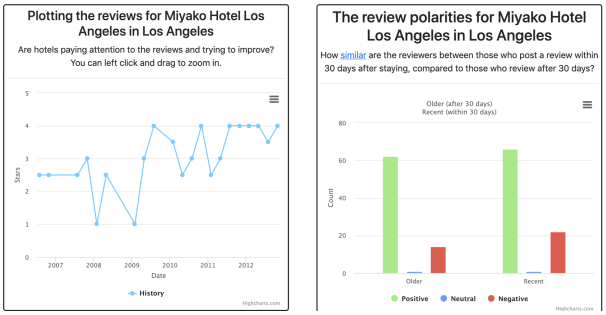

Figure 7: History and comparing the number of reviews

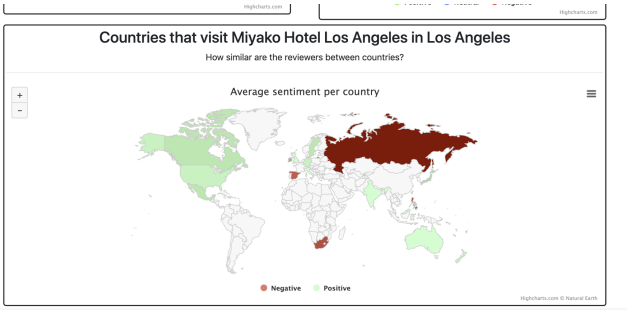

Figure 8: Comparing sentiments across countries

Our first question, "Are the review of late reviewers biased compared to instant reviewers?" can be solved using social network visualization. We classify "instant reviewers" as a person who reviewed within 30 days of their stay date, "late reviewers" after 30 days. Through a network of common tags, we can determine if either of these two groups have similar words or phrases. If one of the groups has no related words, this group might be trying to persuade the reader to second guess their hotel choice.

A maps and cartography solution helps us with our second question "How similar are the reviews from each country?" By definition, a country is a place defined by political geography. To visualize all the dimensions between these areas, we would need a world map. We plan to show the average sentiment, number of reviews, and top phrase on each country to help the reader.

The third question is "Can we accurately determine the positive and negative sentiments of each review?" This one is difficult as manual checking of all reviews is near impossible. The objective is to use machine learning to figure out which review is positive or negative. By training a small set for the machine, our manual review can determine if the algorithm was accurate or not.

Our fourth question is similar to other websites; display a line chart as a time series of 1 to 5 star ratings. This accomplishes the question "Are hotels paying attention to the reviews and trying to improve?" If a hotel does not like its 1 star rating, it will try to change management and keep customers happy for a better review.

Lastly, "How effective is a keyword network visualization?" Network visualizations provide us with how closely are two nodes connected? To reach our goal, we aim to have a hotel recommender in our application. When a user creates an account, adds his or her

preferences, connects with their peers, and starts to "favorite" hotels, the application will start to build a user profile. We then gather common keywords from both the hotels and the user preferences. A network is displayed on the "recommended hotels" page to the user connecting him or her to their friends.

6 EVALUATION PLAN

This section of the report deals with the plan followed to design and develop the project. The development of this project was to determine and understand how people can search easily and in a better fashion for hotels in any particular area. To understand how to justify the difference between a good hotel and a bad hotel, we looked and explored many websites related to travel that have hotel booking as an option. One common thing in most of the websites was that none of them gave a broader outlook on the area selected by the user for considering a hotel in a particular region.

WikiHotels is the idea we created and developed it with the cons of many websites present today so that we can truly justify and deliver what the people want. Making a website interactive is one of the key things that could be done to attract more user base. But, there's always a limit on how much interactive or loaded the website is. We found a perfect balance between text and interactive visualization to be put on the website for display. Thus, the evaluation plan holds both process and product evaluation accountable for determining how effective the visualization was while delivering it to the customer and whether the deliverables were accomplished or not.

The guidelines of the project evaluation are as following:

(1) Did the project implementation face any problems while executing the plan for its design?
(2) Were the deliverables able to accomplish the desired satisfaction from customer?
(3) What was the impact of the visualization built into the project on the customer?

Evaluation took place in two stages:

(1) Process evaluation wherein we gathered the information required to make WikiHotels successful and strategies for the same and analyze and assess the impact WikiHotels had on the customer. It qualitatively and quantitatively assesses the outcome of WikiHotels.
(2) Product evaluation where the focus is to measure the final outcomes against the goals and objectives that were preset for the project. As we go about the project, many changes come along the way compared to the original plan. Such changes are analyzed and then used to determine how effective it would turn out to be in the end.

7 DISCUSSIONS AND CONCLUSIONS

An interactive web application with underlying sentiment analysis has been designed and prototyped for the course objectives. Many interesting topics were discovered by the team, such as how sentiment analysis is helpful, what is the importance of web applications, and previous work on displaying sentiment values on similar datasets. Due to the goal of building an interactive visualization, we needed to understand a handful of design philosophies: Perception & Attention, Vision & color, Maps & cartography, Social Network Visualization, and Text Visualization. In the end, WikiHotels will be good application that helps users to understand which hotels are the best in a certain city.

This application was given to 20 Computer Science students. Results of the prototype showed that most of the students wanted to know if the reviews were fake and compared it to Amazon. The second most requested option is to have a search bar or some kind of filter. As we did not spend much time on it, our first version displayed all hotels in a user-picked city, leading the students complaining that they forgot where they were. In relation to a filter, some would like a list of hotels in the area the map was focused on. Interestingly, those who were male were likely to suggest functional improvements while the majority of the females gave aesthetics feedback.

8 FUTURE WORK

While working on the implementation, our beta testers and classmates mentioned some possible improvements for the next version of WikiHotels. They are as follows:

- The Country graph can be evaluated by adding weights to the reviews of a country based on the percentage of total reviewers. This will eliminate scenarios where there are very few reviewers from a country compared to the rest of the world but still has high impact on the reviews for that country.
- Displaying number of reviews for Countries also help people decide whether to follow the lead or not. One bad review does not mean the whole country thinks/feels that way
- Heatmaps based on sentimental analysis would give a higher accuracy and better judgment for people who want to navigate to the best spots instantaneously.
- The comparative study between old and recent review can only be assumed to be bias. If the person really wanted to leave a genuine review then, he/she will leave a genuine review even if its older. So it can be eliminated to show most genuine reviews.
- Side by side comparison of different hotels will give a better understanding of the hotels to the viewer and help him decide on the best one according to his likings.

REFERENCES

[1] Attila Adamko. 2014. Layered Architecture for Web Applications. https://gyires.inf.unideb.hu/GyBITT/08/ch04.html
[2] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)*. 30–38.
[3] Kanna Al Falahi, Nikolaos Mavridis, and Yacine Atif. 2012. Social networks and recommender systems: a world of current and future synergies. In *Computational Social Networks*. Springer, 445–465.
[4] Catalin-Mihai Barbu and Jürgen Ziegler. [n. d.]. Designing Interactive Visualizations of Personalized Review Data for a Hotel Recommender System. *RecTour 2018* ([n. d.]).
[5] R Basole, E Clarkson, A Cox, C Healey, and J Stasko. 2013. CS (Organizers). *First IEEE visworkshop on sports data visualization* (2013).
[6] Eivind Bjørkelund, Thomas H Burnett, and Kjetil Nørvåg. 2012. A study of opinion mining and visualization of hotel reviews. In *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services*. ACM, 229–238.
[7] Teresa Correa, Amber Willard Hinsley, and Homero Gil De Zuniga. 2010. Who interacts on the Web?: The intersection of usersâĂŹ personality and social media use. *Computers in human behavior* 26, 2 (2010), 247–253.

[8] Wenjing Duan, Qing Cao, Yang Yu, and Stuart Levy. 2013. Mining online user-generated content: using sentiment analysis technique to study hotel service quality. In *2013 46th Hawaii International Conference on System Sciences*. IEEE, 3119–3128.

[9] Xing Fang and Justin Zhan. 2015. Sentiment analysis using product review data. *Journal of Big Data* 2, 1 (2015), 5.

[10] Caesar Fernandes. 2003. Web Application Development. http://www.sitepoint.com/development-guide-success/

[11] Wikimedia Foundation. [n. d.]. Wikimedia Foundation. https://wikimediafoundation.org/

[12] Linton C Freeman. 2000. Visualizing social networks. *Journal of social structure* 1, 1 (2000), 4.

[13] H-W Gellersen and Martin Gaedke. 1999. Object-oriented web application development. *IEEE Internet Computing* 3, 1 (1999), 60–68.

[14] Google. [n. d.]. Quick, Draw! https://quickdraw.withgoogle.com/

[15] Florian Heimerl, Steffen Lohmann, Simon Lange, and Thomas Ertl. 2014. Word cloud explorer: Text analytics based on word clouds. In *2014 47th Hawaii International Conference on System Sciences*. IEEE, 1833–1842.

[16] Sharon Hsiao. [n. d.]. CSE578 - DataViz 2019 Spring. http://www.public.asu.edu/~ihsiao1/slides/cse591/L8-text.html#/3/2

[17] Mirsa Karim and Smija Das. 2018. Sentiment Analysis on Textual Reviews. In *IOP Conference Series: Materials Science and Engineering*, Vol. 396. IOP Publishing, 012020.

[18] Asha Mandava and Solomon Antony. 2012. A review and analysis of technologies for developing web applications.

[19] Peter McLachlan, Tamara Munzner, Eleftherios Koutsofios, and Stephen North. 2008. LiveRAC: interactive visual exploration of system management time-series data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1483–1492.

[20] Maryam Nayebzadeh, Akbar Moazzam, Amir Mohammad Saba, Hadi Abdolrahimpour, and Elham Shahab. 2017. An investigation on social network recommender systems and collaborative filtering techniques. *arXiv preprint arXiv:1708.00417* (2017).

[21] Kan Nishida. [n. d.]. Twitter Sentiment Analysis. https://blog.exploratory.io/twitter-sentiment-analysis-scoring-by-sentence-b4d455de3560

[22] Evan Oliveri and Colin Ware. [n. d.]. Color Sequence Editor Using CIEluv Uniform Color Space. http://ccom.unh.edu/vislab/projects/ColorSequenceEditor.html

[23] Nitesh Sharma, Rachit Pabreja, Ussama Yaqub, Vijayalakshmi Atluri, Soon Chun, and Jaideep Vaidya. 2018. Web-based application for sentiment analysis of live tweets. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*. ACM, 120.

[24] Ben Shneiderman. 2003. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*. Elsevier, 364–371.

[25] Angela Stringfellow. 2017. What is Web Application Architecture? https://stackify.com/web-application-architecture/

[26] Tableau. [n. d.]. Dashboards. https://onlinehelp.tableau.com/current/pro/desktop/en-us/dashboards.htm

[27] James J Thomas and Kristin A Cook. 2005. Illuminating the path: The R&D agenda for visual analytics national visualization and analytics center. In *National Visualization and Analytics Center-US*. Department of Homeland Security.

[28] Edward R Tufte. 2001. *The visual display of quantitative information*. Vol. 2. Graphics press Cheshire, CT.

[29] Martin Varela, Toni Mäki, Lea Skorin-Kapov, and Tobias Hoßfeld. 2013. Towards an understanding of visual appeal in website design. In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 70–75.

[30] Ismail Wadembere. [n. d.]. Introduction to Cartography. https://sites.google.com/site/wadembere/cartography/cartography/introduction-to-cartography

[31] Kurt Wagner. 2019. Twitter finally shared how big its daily user base is. https://www.recode.net/2019/2/7/18215204/twitter-daily-active-users-dau-snapchat-q4-earnings

[32] Colin Ware. 1988. Color sequences for univariate maps: Theory, experiments and principles. *IEEE Computer Graphics and Applications* 8, 5 (1988), 41–49.