Harnessing Restaurant reviews made simple with *RestoScraper*

Moharnab Saikia
North Carolina State University
Raleigh,NC
msaikia@ncsu.edu

Pranjal Deka
North Carolina State University
Raleigh,NC
pdeka@ncsu.edu

Rashmi Sandilya
North Carolina State University
Raleigh,NC
rsandil@ncsu.edu

Sindhu Balakrishnan
North Carolina State University
Raleigh,NC
sbalakr2@ncsu.edu

ABSTRACT
Consumer reviews of restaurants are scattered across various websites in the Internet. A typical user, on an average visits more than two websites to get the reviews of the restaurant to dine in. He also spends a considerable amount of time to get an overall idea of the quality of the restaurant by manually comparing the reviews in those websites. On the other hand, users who visit just one or two sites are not satisfied with the quality of the reviews and are often misled by the ratings. In order to gain insights into the problem, we conducted an online survey with a set of relevant questions and collected the opinion of various users. In this paper, we discuss the data collection methods employed and how it helped us to identify the problems faced by the users. A sample solution could be an online tool that uses web scraping [1] to aggregate the reviews from multiple websites and presents the information to the user by normalizing the ratings. Such a tool can save the time and the efforts taken by the users in getting the reviews and provide a hassle-free option to get the aggregated reviews from the most popular websites.

Keywords
Web Scraping, restaurant, online reviews, Aggregated reviews

1. INTRODUCTION
Who does not want to dine at a place that serves the best? Restaurants are at the heart of 21st century life. Beside the basic need for providing food and drink, restaurants have fulfilled the human need for connecting and maintaining social relations. In such a world the importance of restaurant reviews is immensely important.

The world of digitalization has given people, the handy option to use online sources for getting instant reviews. Restaurant review websites are the most used online source to exploit the dining alternatives. Often users come across lot of inconsistencies in the information provided in each of these websites. Each of these websites can give a different review for the same restaurant. It becomes very difficult for the user to make an informed decision. How one could leverage information from multiple review websites and make a more informed decision becomes a key problem.

1.1 Identification of P, U, T, G
P: The problem P is "the differences in the cross website reviews causing user to invest significantly large amount of time to make an informed decision about a restaurant".

U: The target user group U is "consumers who want to get the most out of the online restaurant reviews for making the best choice to dine in".

T: The user group is said to be facing problem in using tool T which in context of our problem is "web browser such as chrome, firefox etc".

G: The goal G is "to get an aggregated review combining all the top 5 visited websites to establish both summarization and correctness of information".

2. DATA COLLECTION METHODS
Data collection on the usage of a certain tool on set of intended user groups helps to identify the problems faced by them better and come up with fix to problem or devise a better tool. We performed two types of data collection to gain a better understanding of the problems faced by the users while collecting restaurant reviews. First we performed an online survey with a specific set of questions and collected the results. For simplicity we limited the number of websites to top 5. We also asked the users to rate the top reviewed sites so that it would be easier for us later to select the websites to aggregate the results while providing fix to the problem. The results from the survey are graphically plotted using various charts so that it will be very intuitive for analysis. Next we performed participant observation studies where we enquired few users about their experience with the usage of various websites in getting reviews for the restaurants.

2.1 Online Survey
In order to perform an online survey, we came up with a set of questions which will help us to better understand the

issues which users face while looking for restaurant reviews. The questions discussed are as follows:

Question 1: How often do you eat in a restaurant?
This question was asked to help us to decide if the user is actually interested in visiting restaurants and if so, his/her reviews will be useful for us. The user had the option to choose very often, often, sometimes, not much and never as a response for this question.

Question 2: Do you read/get reviews before dining in a restaurant or you go about it randomly?
As a response for this question the user could choose either 'get reviews first' or 'go randomly' as a response. This question was intended to help us to better figure out the fraction of the users who give importance to online reviews.

Question 3: What are your sources of reviews?
This question was put forward to help us to know the different ways users use to collect reviews before planning to dine in a restaurant. The options for response were Friends/Family, Newspaper/ Magazines, TV Ads and Review websites. We had also provided a textbox for the user to enter some other type of sources.

Question 4: Do you think the website reviews were helpful?
This was a yes/no question to help us better analyse the credibility the review websites provide to the users.

Question 5: Rate the sites you use the most for restaurant reviews on a scale of 1-5(1:don't use, 5: use the most)
In this question we had listed the top restaurant review websites such as Yelp, Trip Advisor, Google Ratings, Foursquare and Zomato. The users were asked to give a rating in the range of 1-5 starting from the least to the most frequently used websites respectively. Based on the response, we are planning to pick the most frequently used websites to aggregate the reviews.

Question 6: How much time do you spend on average to get reviews for a restaurant?
This question helped us to analyze the time spent by a user while looking for a restaurant review. We had given three options such as less than 5 minutes, 5-15 minutes and more than 15 minutes to be chosen as a response.

Question 7: On an average how many websites do you visit to find the reviews?
The users were presented with four options to answer this question. They are one, two, three and more than three which indicates the number of websites a typical user visits to get a better idea about a restaurant.

Question 8: Would you like to have all the reviews for a restaurant from multiple websites on a single place with aggregated statistics and other info?
This question was put forward to get an idea if the users are likely to use a tool that can aggregate the reviews for them. The response option were be yes/ no or doesn't matter.

Question 9: Tell us one thing which could make the review websites better.
In order to provide the user the option to enter some additional feedback about the review websites, we had included this question that provides a textbox for the user to enter his opinion.

We have created an online survey form that consists of the above questions and used it to collect feedback from various users. The results of the survey and our analysis will be discussed in the further sections.

2.2 Participant Observation

To make data collection method more real and intensive, our team presented the problem of searching the restaurant of choice to the users for a particular location. For simplicity we instructed the participants to limit their search to top 5 websites namely yelp, TripAdvisor, Google ratings, foursquare and Zomato. We closely monitored how each of the participants went about the search and at the end we also interviewed them with some specific questions. Here is a brief outline of task monitoring analysis and response to interview questions.

- *Swagat, an NC state computer science graduate quoted that he needs to spend on an average 10-15 minutes to figure out honest reviews to select a restaurant to dine in at a new location, which is a time consuming effort*
- *Caterine Carl, a senior employee of NC state was seen quite messed up doing the task. "There is a lot of variations in review information presented by these websites. I have no idea which website to follow. It is time consuming indeed", she said.*
- *Deepthi,an employee of Broadcom quoted as saying "Restaurants help employees connect better. Summarizing reviews from multiple website at one place will ease the search for restaurant and will save our time"*
- Out of every 10 participant, around 9 found online reviews very helpful but very misleading at the same time due to dissonance of information presented on multiple websites.
- When asked if a tool or solution which can give an aggregated review will be helpful, the answer was a big 'YES' from almost all the participants.
- A majority of the participants expressed the need of sorting the reviews from negative to positive and very positive reviews in the summarized solution.

3. RESULTS AND ANALYSIS

3.1 Online Survey Analysis

After collecting the data we analysed it to arrive at various conclusions relating to our problem statement.The first three questions were mainly focused at the frequency of the participants in visiting restaurants and their sources that helped them decide on the same. As given in Table 1 it was observed that 75.51% of the users had some sort of medium that helped them decide on a restaurant.

Get Reviews First	75.51%
Go Randomly	24.49%

Table 1: Decision criteria

Also, it was found that out of all the people resorting to reviews for their decision, most of them used website reviews. Figure 1 reflects peoples' opinion in this regard.

The survey also reflected that most people found online review websites to be helpful. This supported our proposal of

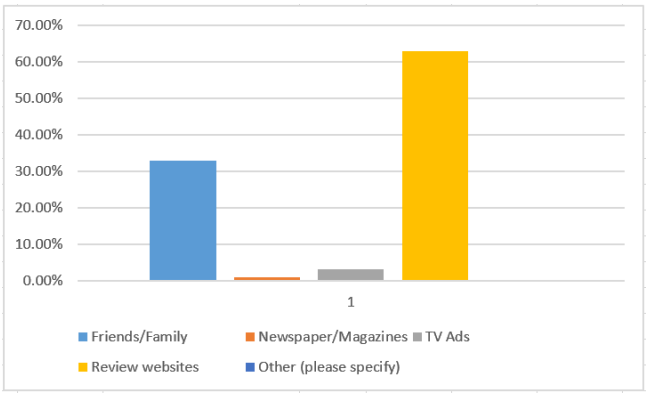

Figure 1: Sources of reviews

enhancing the process as it is one of the most widely used methods. We can refer to Figure 2 for the detailed results.

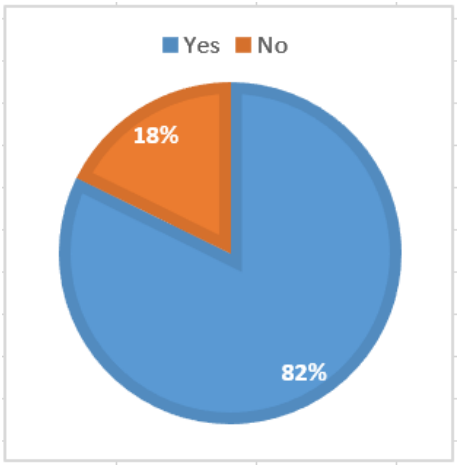

Figure 2: Helpfulness of review websites

We also found out from question no. 5 that Yelp, Google Ratings and Trip Advisor are the top 3 most favorite sites(in the given order) for people when it comes to restaurant reviews.Though we asked for other options too none was provided, which meant that people usually didn't bother to explore beyond the most popular sites. Figure 3 provides the detailed results.

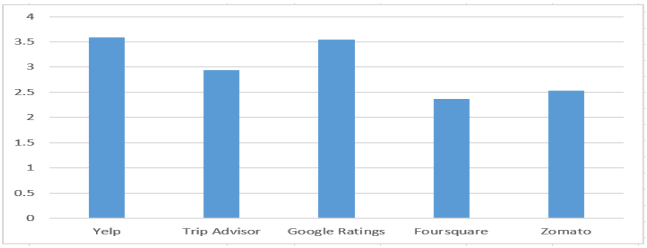

Figure 3: Ratings of common review websites

While searching for reviews, almost 37% of the users spent more then 5 minutes in the process which indicated that the need to optimize the process to help these user-group save a lot of time. Figure 4 presents the detail.

It also came to our focus that more than 70% of the users had to visit multiple websites before they could zero in on a specific restaurant.This can be referred from Figure 5. Aggregating information from multiple sites would make the decision making process for all these users a hassle free process and would help them save effort and time.

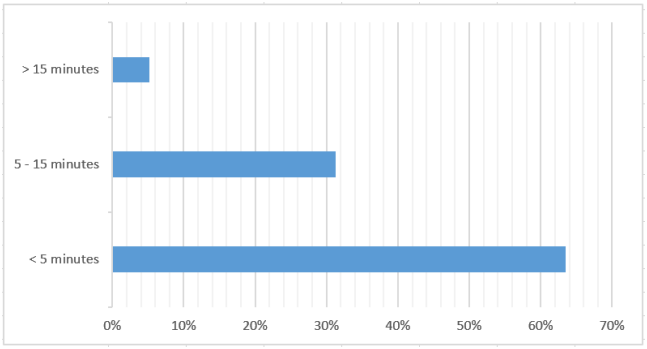

Figure 4: Time taken

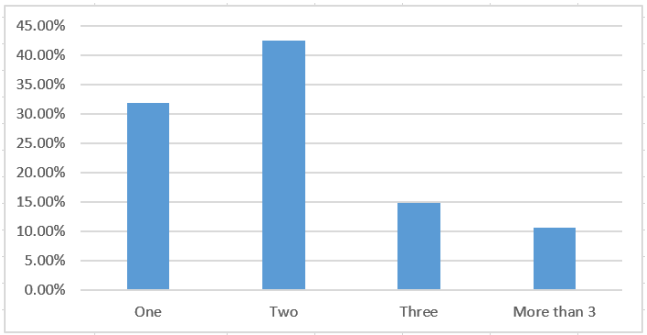

Figure 5: Number of websites visited on an average

As given in Table 2 We could also know from question no. 7 that approximately 70% of the users were supportive of an aggregated website for the reviews.

Yes	72.16%
No	4.12%
Doesn't matter	23.71%

Table 2: Need for an aggregated website

3.2 Participant Study Analysis

The main findings from the Participant studies and interviews primarily focus in four areas where people are facing problem. First being, wastage of time in manual analysis of restaurant review to reach a conclusion. Second problem mentioned was validating the accuracy of review. Third problem is, user wants to see the negative reviews first.This may be attributed to the fact that consumer believes nothing is perfect, so expects to see some negative information also. Lastly it could be analyzed that people are not concerned much about star ratings but more about the review content. The underlying reason can be attributed to the fact

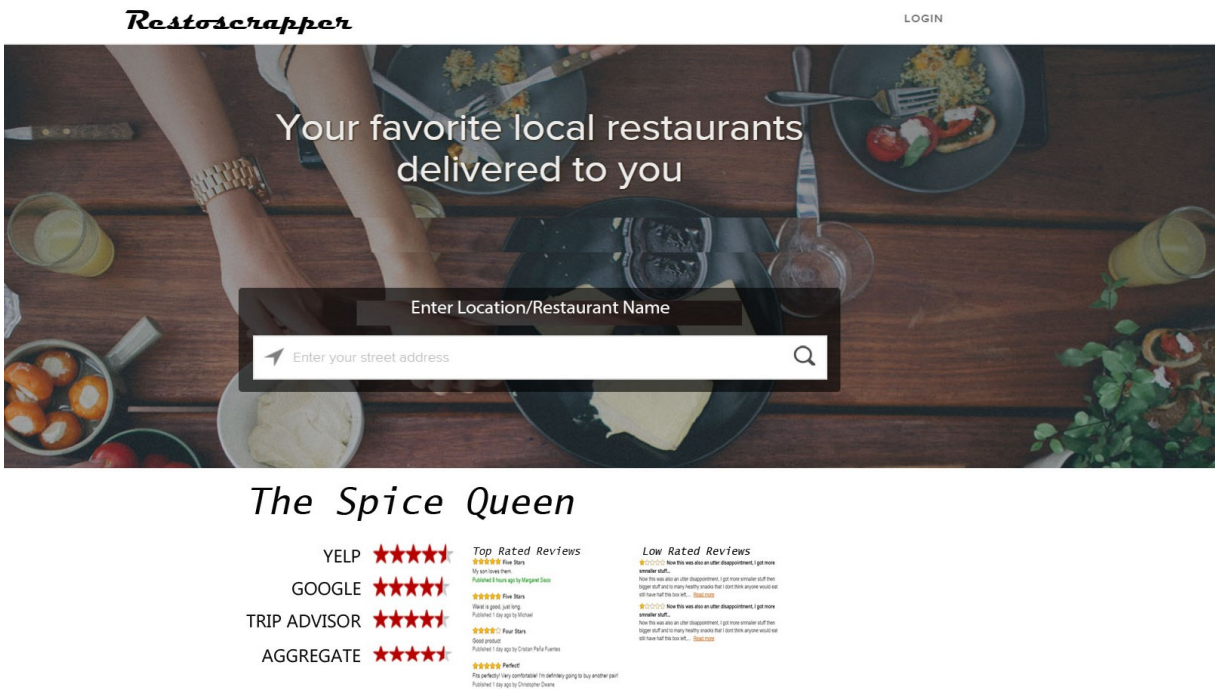

Figure 6: Sample Solution

that there can be increased number of high star ratings with blank content.

4. START UP EXERCISE

As a start up exercise, one of the proposed solutions can be a tool to aggregate the reviews from multiple websites [2] and present a weighted average to the users. Weighted reviews are the reviews that have a perfect balance between rating and content. The goal of the tool may not be limited only to aggregating the reviews from the top 5 websites that we surveyed for but also to identify fraudulent reviews and eliminate those kind of reviews in our calculation of average ratings [3]. More emphasis can be given to those reviews that are highly useful for the users. The design of tool can be a user interface or a plug-in or a full fledged software. The sample output can be one similar to Figure 6.

5. CONCLUSION

Our extensive real time research with online survey and participant study helped us having a closer look at the problem the users are facing in making an informed decision based on website reviews. We can conclude that the obvious choice of the people is to spend a considerable amount of time on website and manually merge and compare and contrast the reviews. The survey and study results not only established the existence of the problem but also threw light on the complexity of the problem.The graphical plot of responses can be used as computational data for presenting the solution for the problem.

A solution to the problem stated in the paper can be an effective tool which can automate the process of decision making for restaurant search based on online reviews from multiple websites. This tool should help user to save time, view aggregated reviews, validate the correctness of reviews and emphasize on negative reviews as well. This would convince the users that our aggregated results are from a reliable source and provide a better view of the quality of the restaurants.

6. REFERENCES

[1] J. I. Fernández-Villamor, J. Blasco-García, C. A. Iglesias, and M. Garijo. A semantic scraping model for web resources - applying linked data to web page screen scraping. In *ICAART 2011 - Proceedings of the 3rd International Conference on Agents and Artificial Intelligence, Volume 2 - Agents, Rome, Italy, January 28-30, 2011*, pages 451–456, 2011.

[2] A. J. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos. Trueview: Harnessing the power of multiple review sites. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 787–797, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.

[3] A. Mukherjee, B. Liu, and N. Glance. Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 191–200, New York, NY, USA, 2012. ACM.