

# Comparing User-Generated Comments from Multiple Sources for a Specific Hotel

Joseph Cilli, Lei Li, Chao Shen  
School of Computer Science  
Florida International University  
cillij@fiu.edu, {lli003, cshen001}@cs.fiu.edu

## Abstract

User-generated comments and reviews for hotels on the web are an important information source for hoteliers and for travel planners. In addition, understanding and reacting to these comments is important for quality control. We present a system that collects such comments from multiple sources on the web, compares source data, ranks and generates visual output for the consumer and/or hotel manager to react upon. The system ranks and compares hotel reviews from multiple user-generated sites. In addition, provides output comparison of individual features of a hotel, where applicable.

## Keywords

Data analysis, Opinion mining, Hotel reviews, Data visualization, Summarizing differences, Rank comparison

## 1. Introduction

With the increased use of Online Travel Agents (OTA), and the recent growth of both social media applications and mobile technology, consumers can easily research hotel properties and read customer reviews. In addition, a 2010 eTRAK study examined the top 30 hotel brands and concluded that 57% of reservations were made online. This was a 19% increase from 2006 (TravelClick, 2011). A recent ComScore study

found that 37% of smartphone users had accessed travel related sites from their phones for research; whereas, one in five actually made a reservation with their mobile device (ComScore, 2012; Thomson, 2012). Projections for 2012 indicated there would be over 117 million online travel researchers and over 98 million reservations will be made online while traditional Global Distribution System reservations continued to decline to its lowest point. One of the most well-known user-generated comment sites is TripAdvisor.com (O'Connor, 2010), with nearly 30 million visitors monthly (Gretzel, 2007). Most of this can be directly attributed to consumer online interaction (Hockenson, 2012; Sarkov, 2011). The growth and wide use of social media and mobile web applications fostered this evolution. Although many hoteliers began marketing campaigns across multiple channels in an effort to produce greater returns, the majority of consumers are less willing to believe advertising and more willing to trust peer reviews. 90% of consumers trust peer recommendations, while 70% trust online comments left by strangers; whereas, only 14% trust advertisements (NielsonWire, 2009; Weber, 2007). (Gretzel, 2008) determined that 97.7% of TripAdvisor.com users were influenced by other traveler comments.

The purpose of this study is to compare user-generated, hotel rankings from individual sources.

This comparison is initially based on word distribution and sentiment score. In addition, individual features and attributes are compared, as well as, association is measured for each feature from the different sources using Kendall's Tau coefficient. Individual comments are analyzed to generate feature sentiment score and attribute score, which can be the basis for a summative report defining quantitative measures of products, features, and services of hotels based on user-generated comments written in online communities. Ultimately, we are identifying the statistical significance of user-generated comments taken from multiple sources for the same hotel. The sentiment word score ranking can also be generated for a specific time range. In addition, we will also provide results of individual hotel features (star ranking) taken from multiple sources.

## 2. Significance

This study presents both practical and academic significance. Hoteliers gain insight into specific aspects of the hotel that are receiving both positive and negative comments from specific sites and may act on these comments. Being able to quickly identify, by source, an area that requires improvement in order to provide a better customer experience would be significant to hoteliers. Similarly, consumers can quickly evaluate a hotel based on the aggregation and summary of user-generated comments. From an academic standpoint, many studies have compared data groups; however, this is the first study to analyze multiple data sources in an effort to determine association for a specific property. This can assist in defining correlation amongst different sources, as well as, identify outliers. A significant number of outliers may be caused by opinion spam; however; for the purposes of this study, we do not evaluate user-generated comments for truthfulness.

## 3. Related Work

Mining user-generated comments for information has received increasing attention in recent years. Many research studies focus on consumer products. Movie reviews, Youtube.com reviews, and Twitter are also popular due to the large datasets that are publicly available. This area of study can range from identifying polarity (Wilson, 2005), sentiment analysis of features (Pang, 2008), product feature extraction with manually generated rules (Kasper, 2011), or the use of machine learning techniques for product feature extraction (Popescu, 2005). An analysis identifying areas of concern for hoteliers based on user-generated comments was defined in (Barreda, 2013). This study is a managerial oriented approach to assist in business decisions related to brand image. A similar approach to ours is the use of linguistic preprocessing used in (Aciar, 2009) with the use of SentiWordNet (Esuli, 2006); however, this work involved the context of the user-generated comment to segment reviews by type (i.e., business traveler comment versus vacation traveler comment).

The aforementioned research focuses on many aspects of user-generated comments, most of which are extracted from a single source. Ideally, hoteliers should be able to use the information gained in these studies. (Barreda, 2013) provides a summative analysis. Similarly, (Hu, 2004) also provides a summary. These are closely related to our work in that they provide a summative report to easily identify problem areas of a hotel, as identified by a specific source. Neither examine statistical consistencies from multiple sources for the same hotel, nor feature ranking as defined by sentiment score star ratings.

## 4. Research Questions

This study was driven by the following research considerations:

1. The aggregation and comparison of hotel features (star ratings) from multiple sources.
2. The development of a feature ranking system based on sentiment analysis of individual reviews.
3. The creation of a graphical summative report of feature rankings used to compare sources of overall comments, attribute score, and sentiment score.
4. The comparison of hotel overall ranking from individual sources, looking for discrepancies.
5. The comparison based on sentiment, word distribution.
6. Ranking individual features based on review content.
7. Generate summaries for each hotel, identifying differences.
8. The comparison star rating of attributes.

## 5. Data Corpus

This study involved the analysis of user-generated comments on four web-based sites for hotels: TripAdvisor.com, Hotels.com, Yelp.com, and Orbitz.com. All four sites offer consumers a review instrument in the form of comments and ratings. Similar to consumer retail products, consumers can write reviews based on their experience and have the option to provide an overall (star) rating, 1=worst, 5=best. We then identified twenty U.S. hotels that had user-generated reviews in all four sites over the same time period. No less than 100 reviews (no more than 150) were extracted using an automated web crawler for a period ranging from 2008 to 2012. The reviews provided both positive and negative comments and were used for the primary analysis of sentiment word scoring. The data collected was user-generated reviews for: a particular hotel, hotel name, website name, user-generated title of comment, star ranking, if applicable, and review date. The individual star (feature) rankings focused on seven attributes: Amenities, Value, Cleanliness, Service, Location, Room Quality, Sleep Quality. There was also an overall star

ranking, in addition to the seven attributes. All the sites provided an overall star rating for each hotel; however, individual attribute (star) ratings were not available on every site. In addition, these attributes varied based on the source. For example, TripAdvisor.com was the only site to have a star rating for Sleep Quality; whereas, Yelp.com and Hotels.com only offered an overall star rating and no individual attribute rating. Similarly, Orbitz.com provided a star rating for Amenities. As such, some attributes are not comparable based on attribute rating and therefore can be ignored.

## 6. Methodology

The primary goal of this study was to provide hoteliers and travel consumers with an understanding of varying user-generated comments on different websites, and to determine the consistency of the comments across multiple websites.

In this study, the authors collected data from four publicly available websites designed to capture user-generated comments for hotels. The collected data was inserted into an SQL database. The user-generated comments were analyzed, which used SentiWordNet to compare synsets and provide a sentiment word score. Similarly, the user comments were analyzed to provide a sentiment score for each feature (star rating). This was done without regard for context.

In addition, we provided an attribute rating score based solely on the star ratings provided by the users for the sites that provided such information.

Later, Kendall's Tau rank correlation was applied to all the datasets to determine a rank coefficient. The resulting information was also stored in a database. A simplistic web-based interaction was created using Highcharts.com to provide the end-user with the ability to visually compare multiple dimensions of data.

## 7. Experimental Results

In general, the ranked results of user-generated comments based on sentiment word score is statistically significant amongst all four sources. Figure 1 shows hotel rankings, from all four sites, based on the aggregate sentiment score. The graph shows a fairly noteworthy correlation of some of the site rankings. Orbitz.com and Hotels.com had the highest rank correlation for the 20 hotels (.51); whereas, TripAdvisor.com and Hotels.com had the lowest rank correlation (.11) for aggregate sentiment word score.

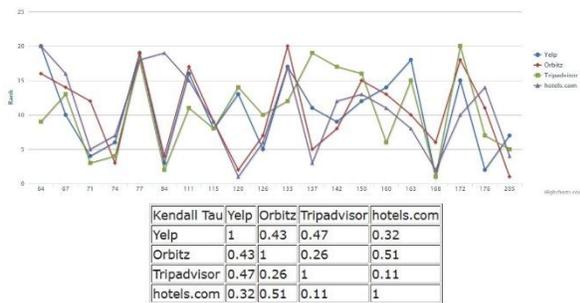


Fig.1. Hotel rankings based on the aggregate sentiment score.

Figure 2 shows the hotel rankings, from all four sites, based on the average sentiment score. Yelp.com and Orbitz.com had the highest rank correlation (.61); whereas, TripAdvisor.com and Hotels.com had the lowest rank correlation (.11) for the average sentiment word score.

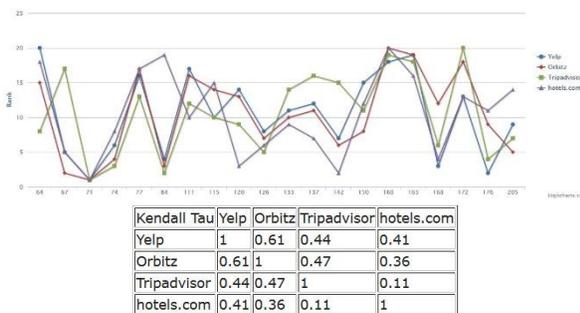


Fig.2. Hotel rankings based on the average sentiment score.

The use of the sentiment score for individual hotel features was used to generate a rank coefficient as well. This was performed on six attributes. Table 1 outlines the results for Room Quality. In addition, the interactive web-based tool provides the words in the user comments that were analyzed

for the attribute. Figure 3 shows the representation used to define the sentiment score for Room Quality.

Room Quality	Yelp	Orbitz	TripAdvisor	Hotels.com
Yelp	1	.48	.38	.36
Orbitz	.48	1	.62	.12
TripAdvisor	.38	.62	1	.14
Hotel.com	.36	.12	.14	1

Table.1. Individual hotel feature (Room Quality) rankings based on sentiment score.

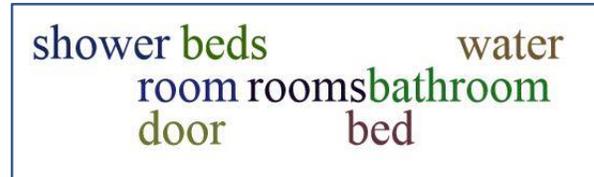


Fig.3. Individual hotel feature (Room Quality) words used to generate the sentiment score.

Similar to the process used to determine the individual attribute sentiment score, we then calculated the individual attribute rating score based solely on the star rating system and generated the rank coefficient. This was done for six individual attributes and only for Orbitz.com and TripAdvisor.com, as they were the only sites that provided star ratings for these attributes. Table 2 shows the rank coefficient for each attribute when comparing the attribute rating of the two sites referenced above.

Attribute Rating	Kendall Tau
Cleanliness	.68
Service	.42
Location	.56
Value	.51
Room Quality	.64

Table 2. Individual hotel feature rankings based on attribute score for Orbitz.com and TripAdvisor.com.

As shown in Tables 1 and 2, it is clear that ranking between Orbitz.com and TripAdvisor.com for Room Quality has a higher statistical significance when using attribute rating versus sentiment score. This seems consistent with most attributes. When focusing on the coefficient scores from TripAdvisor.com and Orbitz.com, the sentiment score (.62) related to Room Quality is similar to the attribute rating score (.64); although this occurrence does not seem to be the norm. This

may be an indication that the results of the sentiment analysis coincide with the stated user star rating, and thereby demonstrate that user-generated comments mirror their numerical star rating for individual hotel features. Table 3 shows the rank coefficient for each individual attribute and the corresponding sentiment score for TripAdvisor.com and Orbitz.com.

Attribute	Attribute Rating Score	Attribute Sentiment Score
Service	.42	.43
Location	.56	.52
Value	.51	-.084
Room Quality	.64	.62

Table 3. Individual hotel feature rankings based on attribute score and sentiment score for Orbitz.com and TripAdvisor.com.

Similar to the Room Quality score consistency, Service and Location have comparable consistencies in rating and sentiment score. Value showed drastically different results in this comparison. This may be attributed to user comments not being consistent with their star rating. Another explanation would be not enough information was available in the comments to provide an accurate sentiment score for Value across all four sites. This seems evident in that the sentiment analysis was only able to extract two terms (experience, price) when calculating sentiment score for Value. Additionally, the rank coefficients for the sentiment score for Value were substantially lower than the other attributes.

## 8. Conclusion and Future Work

Due to the intense growth of Web2.0 technologies, specifically social media, user-generated comment sites are rapidly growing with high consumer reliance. Literature shows that user-generated comments are increasing in the travel industry (O’Conner, 2010). As such, hoteliers need to understand the impact user-generated comments can have on their business, both positive and negative. In addition, hoteliers need to understand that there are multiple sites and opportunities for consumers to provide comments and that hotel

rankings may differ due to these dissociated sites. Similarly, consumers can also benefit from aggregating user reviews from multiple sources thereby assisting in their decision making.

The results indicate that there exists a statistical significance when comparing hotel rankings from multiple user-generated comment sites. Although the rank coefficients for most were not high, we were able to show a positive corollary relationship amongst the data sources. In general, this correlation was statistically higher when comparing the overall sentiment word score with the individual attribute sentiment scores. In most instances, the correlation was statistically higher when comparing the attribute rating score with both the sentiment word score and the individual attribute sentiment scores. Areas that could have a significant effect on results would include the time range and opinion spam. Some of the variation in the results may be credited to the fact that one site may produce more current comments and therefore does not give an accurate comparison based on the other sites’ older comments. This may be especially apparent when dealing with hotel renovations. For example, TripAdvisor.com may provide 100 comments for one hotel over a period of two weeks; whereas, Hotels.com can only provide 100 comments over a six month period for the same hotel. In this example, comparing comments from six months ago against more recent comments could provide skewed results. This can be managed by limiting the time range of comments, assuming all sites provide an adequate number of reviews in that time period.

Additional work for this area would be to build upon the individual attribute sentiment score model and based on this information, construct a formal star rating. This is particularly obvious for sites that do not already provide a star rating system. Other work may include the automatic generation of a summative report identifying operational areas of concern for a hotel based on

user-generated comments and/or star reports from multiple data sources. Other areas of study to further this research would include the development of a framework to create hotel profiles based on the aggregation of user-generated comments. Furthermore, the development of a framework to create user profiles based on individual user-generated comments. This would automatically generate a user profile based solely on that user's generated comments. Such systems would lend themselves to a recommender application to match user profiles with hotel profiles.

## References

- Aciar, S. (2009). Mining context information from consumer's reviews. In Proceedings of the Context-Aware Recommender Systems (CARS) Workshop.
- Barreda, A. and Bilgihan, A. (2013). An Analysis of User-Generated Content for Hotel Experiences. *Journal of Hospitality and Tourism Technology*, vol. 4, no. 3, pp. 263-280.
- comScore. (2012). Smarter Travel: Nearly 2 in 5 Smartphone Users Accessed Travel Content from their Device in July. Retrieved from: [http://www.comscore.com/Press\\_Events/Press\\_Releases/2012/8/2\\_in\\_5\\_Smartphone\\_Users\\_Accessed\\_Travel\\_Content](http://www.comscore.com/Press_Events/Press_Releases/2012/8/2_in_5_Smartphone_Users_Accessed_Travel_Content).
- Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC-06).
- Gretzel, U. and Yoo, K.H. (2008). Use and impact of online travel reviews. *Information and Communication Technologies in Tourism 2012*, pp. 35-46.
- Gretzel, U., Hyan-Yoo, K. and Purifoy, M. (2007). Online Travel Review Study: The Role and Impact of Online Travel Reviews, Laboratory for Intelligent Systems in Tourism.
- Hockenson, L. (2012, February 21). The Evolution of Online Travel. Retrieved from: <http://mashable.com/2012/02/21/online-travel-infographic/>.
- Kasper, W. and Vela, M. (2011). Sentiment Analysis for Hotel Reviews. In Proceedings of the Computational Linguistics-Applications Conference.
- NielsonWire. (2009, July 7). Global Advertising: Consumers Trust Real Friends and Virtual Strangers the Most. Retrieved from: <http://blog.nielsen.com/nielsenwire/consumer/global-advertising-consumers-trust-real-friends-and-virtual-strangers-the-most/>.
- O'Connor, P. (2010). Managing a hotel's image on TripAdvisor. *Journal of Hospitality Marketing & Management*, vol. 19 no. 7, pp. 754-772.
- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1-135.
- Popescu, A. and Etzioni, O. (2005). Extracting product features and opinions from reviews. In Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP).
- Starkov, M. Book Now!. (2011, April). *Special Report: Distribution*, 24-28. Retrieved from: <http://www.hotelsmag.com>.
- TravelClick. (2011). eTRAK Benchmark Report. Hotel Bookings by Channel. Retrieved from: <http://www.travelclick.com/>

information-center/bookings-by-  
channel.cfm.

Weber, L. (2007). Marketing to the Social Web.  
Wiley Publishing.

Wilson, T., Wiebe, J. and Hoffmann, P. (2005).  
Recognizing contextual polarity: An  
exploration of features for phrase-level  
sentiment analysis. Computational  
Linguistics.