Harvesting Online Contents:
an Analysis of Hotel Reviews Websites

Elena Marchiori
Davide Eynard
Alessandro Inversini
Lorenzo Cantoni
Francesco Cerretti

webatelier.net
Faculty of Communication Sciences
University of Lugano, Switzerland
(elena.marchiori; davide.eynard; alessandro.inversini; lorenzo.cantoni)@usi.ch
francesco.cerretti@mail.polimi.it

Abstract

Hotel Reviews Websites (HRWs) are the most used online sources to evaluate accommodation alternatives. However, they often present an overwhelming amount of unstructured or only semi-structured information which is not shared between all the systems and which cannot be easily analyzed in an automatic way. This study aims to automatically analyse hotel evaluations for a given number of Swiss hotels by comparing hotel reviews. Furthermore, the consistency of users' countries of origin in their evaluations has been studied. The results show that there is an overall agreement on considered HRWs and a general consistency among reviewers with different countries of origin.

Keywords: hotel reviews website, evaluation tools, online content harvesting

1. Introduction

During the tourism consumer decision-making process, prospective consumers can access Internet as a platform for booking, and for information and recommendation seeking (Fesenmaier et al., 2010). The amount of tourism related purchases online are numerous and continuously growing. The greatest share of online tourism sales in fact is generated by air travel, followed by hotel bookings, which accounts for 19%, and package tours, rail and car rentals (Marcussen, 2009). Amongst consumers, the preferred criteria for online hotel booking are recommendations from friends and online reviews (Dickinger and Mazanec, 2008). Both criteria represent the most important factors that influence online hotel booking. Hotels Reviews Websites (HRWs) are the most widely used online resources to help with the evaluation of accommodation alternatives (O'Connor, 2008). Chatterjee (2001) noted that consumer reviews and ratings are the most accessible and prevalent form of user-generated-contents (UGCs) available online. HRWs present mainly UGCs, such as reviews of tourism products and services, which are generally perceived as equally or

more trustworthy than those of official websites (Nielsen Global Online Consumer Survey, April, 2009). However, the amount of online data can overwhelm prospective travellers and tourism managers in their decision-making process or their evaluation of hotels online performance, due to the fact that different HRWs can potentially present different types of more or less consistent hotel information. Related work from Schegg and Fux (2010) about a Swiss hotel presence online, shows that small evaluation differences exist between two HRWs such as Tripadvisor and HolidayCheck. The authors suggest further analyzing the reviewers' country of origin as a possible determinant in their evaluation of tourism services and offers. Therefore, the aim of this research is: (i) to compare a group of relevant Swiss hotels' evaluations among HRWs in order to confirm the HRWs tendency to provide similar hotel evaluations; (ii) to understand if hotel evaluations are consistent among different countries; (iii) to investigate whether consumers' country of origin can predict a higher or lower hotel evaluation. The comparison of information belonging to different HRWs is performed on a dataset gathered by scraping data from the Web and aggregating matching information with semi-automatic tools that have been developed ad-hoc for this study. From a research perspective, this study shows how hotel evaluations are consistent among different HRWs systems. This study expands on the findings of Schegg and Fux (2010), further analysing the coherence among the nationalities of the reviewers, and leveraging automated tools to harvest and analyse online information. Contributions to the Internet-based data acquisition are also made by this research, in particular with regard to the information provided by HRWs; which can be an alternative source of marketing research for the forms of assessment analysis on customer evaluation portals. This work should also be apt for all destination marketers who need to monitor the online performance of their hotels and competitors on HRWs, and understand the markets involved on their evaluation.

The paper is organized as follows: the following section presents a literature review about the use of HRWs in travel planning and the creation of valuable knowledge from unstructured and semi-structured sources. Succeeding that, the research approach is presented, showing the instruments used to gather information from the Web, the characteristics of the collected dataset, and the analyses that have been performed over it. The following section shows the results of the analyses; and finally, conclusions are drawn and a plan for possible future work is sketched.

2. Literature Review

eTourism and web2.0

Tourism has been always recognized as an information intensive domain (Sheldon, 1997), where the use of technology is essential for the day to day life and management (Poon, 1993). The advent of the Internet dramatically changed the tourism landscape, giving a real added value to those wise tourism and hospitality managers who were able to exploit the potential of such an instrument (Buhalis, 2003). In fact, tourism and hospitality managers started to use the Internet firstly as a mean of communication and promotion (Buhalis, 2000) and secondly as a selling channel (Werthner and Klein, 1999). As a result, consumers have been more and

more overwhelmed by different content providers and different selling channels (Inversini and Buhalis, 2009). What is clear nowadays is that around a given destination, a tourism online domain does exist (Xiang et al., 2009): information about a given destination or hotel made available from official information sources but also by unofficial ones (Inversini et al., 2009). The different players (being them official or not) are competing to reach the end user's attention offering them both information and purchasing possibilities (Inversini and Buhalis, 2009).

From a traveler perspective, this content can be useful within all the tourism phases (Gretzel et al., 2006): before going to the destination (e.g. image creation, decision making support and purchase), during the stay at the destination (e.g. updated information about events and places) and after the stay at the destination (experience recall and experience sharing). Moreover, the advent of web2.0 (O'Reilly, 2005; Cantoni and Tardini, 2010) as preferred way to express personal feelings and emotions (Inversini et al., 2010; Marchiori et al., 2010) radically reshaped the way tourists are actually browsing for travel information, promoting the so called eWord-of-Mouth (Xiang et Gretzel, 2009).

Use of HRWs in Travel Planning

Internet offers the opportunity to seek information, book, and leave feedback (Buhalis, 2003). eWord-of-Mouth (eWOM) is a major driver for hotel purchase decisions (Dubé and Renaghan, 2000). As Smith et al. (2007) pointed out, in an information intensive situation such as the online environment; consumers actively seek others' opinions as a means of managing perceived risks. Hotels are characterized by different attributes that give information on their quality, namely star category and ranking evaluation. This information is a basis for making purchase decisions (Callan, 1998). HRWs are the online platforms which provide these "other's opinions" in the form of:

- Hotel's Popularity Index: hotel ranking among other hotels on a selected destination; this parameter is usually provided by the system without entering the details of how it is calculated; only in Booking.com the hotel popularity index is a direct consequence of the sum of the hotel's characteristics;
- Rating of hotels specific characteristics (i.e. cleanliness, service, location) given by users;
- Consumer Hotel's Reviews: unstructured text written by past guests regarding their experience in a specific hotel.

The amount of this unstructured or semi-structured information about a given hotel in a HRW can be potentially different in another one. As noticed by O'Connor (2008), in HRWs like Tripadvisor users might leave a comment on a site even if they have booked their hotel room somewhere else. For the providers who cannot afford the task of backing reviews up with transactional data, detecting deception has become an important necessity in order to maintain usefulness and credibility of the content provided in their sites. As affirmed by Shao and Gretzel (2010), looking doesn't mean booking; a prospective consumer can remain online searching for the best offer without taking a decision. In order to be sure of the evaluations they find, they usually compare them across different websites. Popular good hotels on a given tourism destination should be presented on the top rank positions of these HRWs, and the

expectation is to have hotels' evaluation coherence among these websites creating the wisdom of crowd effect (Surowiecki, 2005) which is popular within the social web environment (Kittur and Kraut, 2008).

Knowledge creation from unstructured sources

The success of participative systems such as HRWs relies on concepts like wisdom of crowds and collective intelligence. The first one, introduced by Surowiecki (2005), states that under the right circumstances (diversity of opinion, independence, decentralization, aggregation) groups are often smarter than the smartest people in them. According to Surowiecki, collecting information from every single person in a group (no matter whether the individual is an expert or not), and then calculating an average of all the responses, allows to obtain an answer that is at least as good as the one of the smartest members in the group. Segaran (2007) generalizes the concept of collective intellignece by focusing on the idea of drawing new conclusions by using more or less sophisticated algorithms to combine data collected from many different people. However, as information becomes more and more available, even just locating a particular piece of it becomes a complex task, and trying to combine it with other information is even harder. Stuckenschmidt and van Harmelen (2005) suggest that this problem of information sharing can only be solved by allowing machines to have a better access to the semantics of information. As all the sources accessed within the current project provide information as Web pages, in formats which are semi-structured (as in hotel ratings) or completely unstructured (as in plain text user reviews or hotel name and address), a consistent part of the research is represented by the extraction of useful information out of this data. To perform this operation, ad-hoc Web Data Extraction Systems have been developed. Baumgartner et al. (2009) define a Web Data Extraction System as a software system that automatically and repeatedly extracts data from web pages with changing content, and delivers the extracted data to a database or some other application. The five main functions of these systems are: web interaction, that is the navigation to the Web pages that contain desired information; wrapper generation, where a wrapper (also called scraper) is a software able to extract data from a target Web page and save it into a structured format; scheduling, which manages the repeated execution of scrapers; data transformation, that includes data filtering, conversion, and integration; and delivering, which is returning the resulting data in the format required by an external application. For this specific project, scrapers have been developed following the textual approach model, which considers Web pages as text strings and relies on string pattern matching with regular expressions for data extraction (Muslea et al., 1999).

3. Research Design

Given this context the research focuses on three (and then two) popular HRS, namely Tripadvisor, Booking and Venere. The aim of research is to (i) understand the level of agreement among different HRWs (ii) to study the coherence among different national groups and (iii) to explore the general appreciation of Swiss hotel by guests coming from different countries. The study has been performed on a sample of Swiss

hotels and data was collected in July 2010. The following paragraphs outline the research design used, the data collection and the data analysis.

Sampling and research instrument

The research focuses on Swiss hotels which are present in the three above mentioned Hotel Review Websites with a given number of reviews. In order to compile the sampling of hotels to be analyzed, an investigation of the tourism cities (i.e. using the name of the city as driver to perform a search activity within the websites) in Switzerland was needed. The website of MySwitzerland.com, the Swiss Destination Management Organization, was analyzed and the summer holiday destinations were listed region by region. This preliminary activity detected 197 cities and destinations. This list has then been checked against the three HRWs and all those cities which did not appear in any of them have been removed as well as the destinations that cannot be considered as a "city", such as aggregated destinations (e.g. mountains or lakes). The shortened version contains 170 cities, which represent 86% of all the cities harvested on Myswitzerland.com. What is clear from this preliminary result is that 170 Swiss cities are present in at least one of the three sources analyzed (Tripadvisor, Booking and Venere). Researchers then built a web scraper able to submit all the cities to the HRWs internal search engines and to collect reviews from the websites. Reviews and other relevant indicators were then stored in a MySQL database. A description of the dataset wecollected is shown in table 1: for each source, the total number of cities, hotels, and reviews are shown; aggregate data are also present like the maximum, minimum, and average number of hotels per city, and max, min, and average number of reviews per hotel.

Source	Cities	Hotels	Reviews
Booking	146	1442 (max: 102, min: 1, avg: 9.87)	94102 (max: 820, min: 5, avg: 79.61)
Tripadvisor	162	1818 (max: 115, min: 1, avg: 11.22)	26852 (max: 418, min: 1, avg: 18.58)
Venere	108	639 (max: 54, min: 1, avg: 5.91)	2891 (max: 93, min: 1, avg: 9.00)

Table 1: A summarized view of our dataset.

From this information, it is possible to deduce the importance of the different sources in this specific domain: Venere is the one which has the least amount of hotels and reviews, with an average of 9 reviews per hotel; Tripadvisor is the one which has the most cities and hotels, while Booking has the most reviews. The average number of reviews per hotel in Booking is about four times the amount in Tripadvisor and almost nine times the amount in Venere. This information is even more surprising if we look at the *timespan* covered by these reviews: while Venere and Tripadvisor both have reviews that date back to 2002, the oldest review in Booking dates back to May, 2009. For this reason, we chose to keep for our comparison only the reviews dated after June, 2009. Due to this constraint, the number of reviews from Venere became so small that it was decided to omit them from the following analysis. When reviews were downloaded, they were stored separately according to the source they were derived from. To understand if two reviews coming from different sources were

related to the same physical hotel or not, we needed to *infer* new information from the one we have downloaded. The inference process was actually very simple: as hotel names and addresses were provided in an unstructured format (i.e. not consistent across the three HRWs) we matched two hotel ids *a1* and *a2* if their name or their addresses "loosely matched" each other. The loose match was implemented by using the SQL LIKE comparison operator, with wildcards that allow one string to match if it appears as a part of the other one (i.e. the strings "Hotel Beau Site" and "Boutique Hotel Beau Site Fitness & Spa"). Albeit simple, the double match on name and address allows us to have satisfying results. For instance, for the city of Zurich we were able to download information about 54 hotels and hotel reviews from Venere, 102 from Booking, and 115 from Tripadvisor. The semi-automatic aggregation procedure gave the following results: 36 triple matches (i.e. hotels belonging to all of the sources) were correctly identified, together with 54 double matches (i.e. hotels belonging only to two of the three sources). Hotels which, at the end of the semi-automatic matching process, did not belong to a match were actually present in only one HRW.

Details on collected data

From this linked dataset, 77 hotels were selected satisfying the following conditions: (i) the hotel must have reviews in all of the considered sources (reduced to only Tripadvisor and Booking, due to the small number of reviews available in Venere); (ii) the total number of reviews per hotel must be greater than 200. The information available for each hotel follows: (i) hotel rating (normalized on a 1-10 scale);(ii) number of reviews per hotel (timeframe: June 2009-July 2010); (iii) reviewer information (i.e. name/nickname and country of origin). Figure 1 shows an example of the data harvested from the two HRWs and table 2 shows the quantity of information scraped during the research.
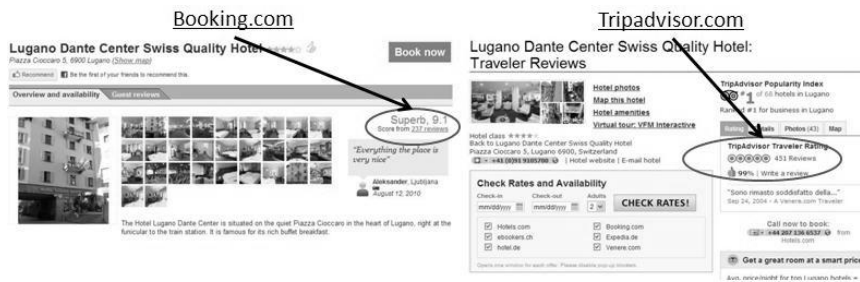

Figure 1: An example of evaluation data scraped from Tripadvisor and Booking.

	Booking	Tripadvisor
# REVIEWS	19434	2630
# VALID EVALUATIONS	19434	2630
# VALID COMMENTS	15656	2541
# VALID AUTHORS ' NAMES	12820	2019
# VALID NATIONS OF ORIGIN	18804	1993

Table 2: Quantitative details of scraped information

Table 2 shows that all the reviews harvested on Booking (19434) and on Tripadvisor (2630) presented a valid evaluation (i.e. the number corresponding to the hotel rating), but not all presented a valid comment (in terms of text): 20% of hotel reviews in booking, and 3.4% in Tripadvisor do not present any text. Author name is present only in 65% of the reviews in Booking and 75% in Tripadvisor. Finally, it has been noticed that the countries of origin were present in 96% of the reviews in Booking and 75% in Tripadvisor.

<u>Data Analysis</u>

After the first phase of scraping, hotels matching and selection of the top 77 hotels, the work followed the objectives of the study:
i) two tables for each HRW, containing the list of hotels and for each one the evaluation given by the reviews of the two booking systems per month were created. The weighted average of each hotel's reviews in the chosen timeframe and their standard deviation for the two HRWs was calculated;
ii) the work on the country of origin presented more problems. In fact, only Booking provides a standard way to choose one's nation, while in Tripadvisor a reviewer can fill the country of origin field with any value. As a result, while the extraction of countries from Booking was trivial, in Tripadvisor a lot of abbreviations, different spellings for the same place, or even city names instead of countries were found. To solve this problem we implemented a semi-automatic tool which allowed us to cluster reviews under a chosen name. This step allowed us to study reviews according to their country of origin. A limitation of this approach is the fact that many reviewers, especially in Tripadvisor, did not specify their country of origin at all. For all those reviews which could not be definitely geographically located (6% of the total), we were able to automatically identify the language from the review text.

4. Results

Among the 22.064 reviews about the 77 Swiss hotels, Booking was the HRW which presented more reviews with respect to Tripadvisor (19,434 against 2,630) for the period between June 2009 and July 2010. The average of the users' evaluation per each hotel was compared between the two HRWs. Results show how the level of evaluation agreement between HRWs was significant (Table 3). Considering a threshold of 1 for the standard deviation (as the hotel evaluations on HRWs were from point 1 to 10 without half points), only 2.06% of the hotels did not show an evaluation agreement. Moreover, a threshold of 0.5 for the standard deviation confirmed an agreement on the evaluation for almost 80% of the hotels.

	%hotels (SD>0.5)	%hotels (SD>1)
Agreement	79.03	97.04
Non-agreement	20.07	2.96

Table 3: Agreement on hotel evaluations between Tripadvisor and Booking.

Each review, of the 22.064 chosen, was associated to a hotel evaluation, and it was possible to identify the country of origin of most of their authors.

Results show that 69.7% of the hotel evaluations got a nationality attribution. 6% were attributed to "none" due to the fact that for those reviews no country of origin had been specified. 24% of the evaluations were grouped as "other" due to the fact that no country within this group reached the frequency of 2% (i.e. minimum threshold to be considerate relevant for the study). Therefore, only nations with greater than or equal to 2% of presence (n=10) were taken into consideration for the actual analysis: Switzerland (24%), France (8%), Italy (7%), UK (7%), Germany (6%), USA (5%), Netherlands (5%), Spain (4%), Russia (2%), and Belgium (2%), represented the main audiences around Swiss hotels on HRWs.
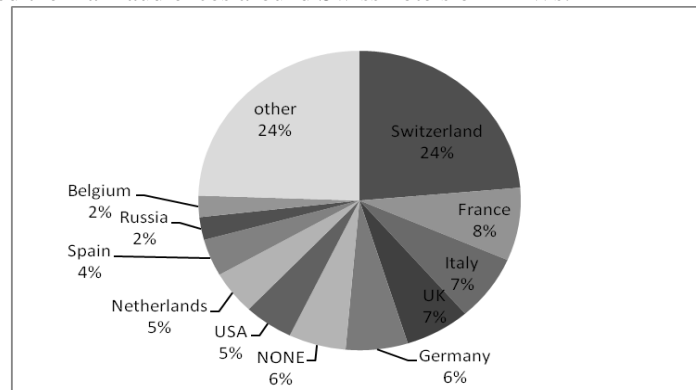

Figure 2: Review nationalities for the Swiss hotels case study.

The standard deviation among evaluations per hotel and country of origin were calculated. Then, the average of the standard deviations obtained for each hotel was compared with the ones obtained within each country. Results show how countries of origin were coherent on their evaluation, with no differences for UK and USA (Standard deviation is 1.45).

Countries of Origin	Switzer	France	Italy	UK	German	USA	Netherlar	Spain	Russia	Belgium
STDEV(s) Average	1.30	1.24	1.32	1.45	1.28	1.45	1.05	1.31	0.92	1.00

Table 3: Hotels evaluation coherence within each country of origin.

To investigate if being part of a country is predictive of an higher or lower hotel evaluation, we took into account the proportional difference of each country evaluation, based on the number of reviewers per each hotel.

Countries of Origin	Switzer	France	Italy	UK	German	USA	Netherlar	Spain	Russia	Belgium
% proportional difference	0.62%	1.15%	-1.73%	0.70%	-1.98%	3.00%	-1.17%	-1.91%	1.46%	-0.09%

Table 4: Proportional difference according to the country of origin.

Findings show that the proportional difference was lower mainly for Germany, Spain, Italy, and Netherland; these results allow to consider these countries as the ones that least appreciate Swiss hotels. On the contrary, users from USA seem to appreciate Swiss hotels more (with a percentage of 3%), followed by France, Russia and UK.

Swiss users showed coherence with the overall hotels evaluations; it is important to note that Swiss reviewers represent the 24% of the audience in range.

Frequencies analysis showed that the analyzed evaluations were mainly around the mean (Figure 3), and only few of them differ significantly. These results allowed to consider further research on the hotels group sample in order to investigate which factors can cause this effect.
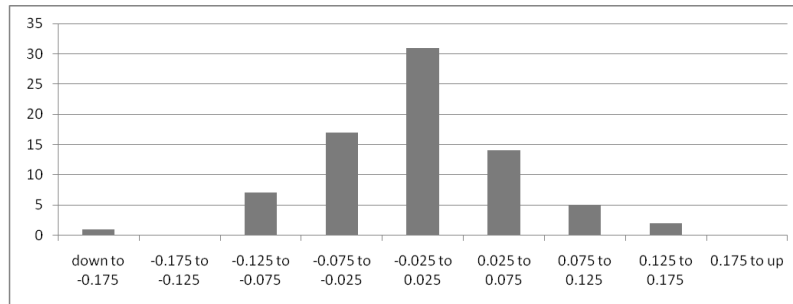

Figure 3. Frequencies down to -0.175 and up to 0.175

One of the possible factors could be the number of hotels' stars. Thus a further analysis was conducted in order to assess possible statistical significance using the number of the hotels' stars to segment the sample. Findings show that the hotel evaluation coherence within each country of origin remains constant without any significant change with respect to the segmentation per stars hotels (Table 5).

Countries of Origin	Switzer	France	Italy	UK	Germany	USA	Netherlan	Spain	Russia	Belgium
STDEV(s) Average - All	1.30	1.24	1.32	1.45	1.28	1.45	1.05	1.31	0.92	1.00
STDEV(s) Average 5*hotels	1.51	1.27	0.89	1.77	1.39	1.94	0.95	1.31	0.85	1.25
STDEV(s) Average 4*hotels	1.36	1.31	1.40	1.49	1.31	1.46	1.07	1.26	0.92	1.06
STDEV(s) Average 3*hotels	1.23	1.19	1.32	1.36	1.27	1.40	1.06	1.32	0.86	0.93
STDEV(s) Average 2*hotels	1.20	1.04	1.47	1.67	1.25	1.31	0.91	1.57	0.98	0.55

Table 5: Hotels evaluation coherence within each country of origin per stars hotel

Moreover, the analysis of the proportional difference shows that 5 and 2 star hotels are the ones which present different tendencies. This could be partially explained by the fact that among 77 hotels only 4 belong to the 5 star category and 3 to the 2 star. As an example it is possible to consider the Swiss HRW users (Table 6 column 1) which showed coherence within the overall hotel evaluations mainly for 3 and 4 star hotels. Less appreciation is shown for 5 and 2 star hotels.

Countries of Origin	Switzer	France	Italy	UK	Germany	USA	Netherlan	Spain	Russia	Belgium
% proportional difference ALL	0.62%	1.15%	-1.73%	0.70%	-1.98%	3.00%	-1.17%	-1.91%	1.46%	-0.09%
% proportional difference 5*hotels	-2.32%	9.28%	7.63%	-3.91%	-14.12%	-1.75%	2.60%	2.62%	10.56%	-7.30%
% proportional difference 4*hotels	0.11%	-0.97%	-2.19%	0.76%	-0.05%	3.23%	-0.04%	-1.79%	-0.60%	1.49%
% proportional difference 3*hotels	0.98%	1.99%	-2.13%	1.31%	-2.48%	3.18%	-2.46%	-1.76%	3.05%	-1.38%
% proportional difference 2*hotels	3.83%	4.31%	-3.36%	-2.93%	-2.57%	5.11%	-3.69%	-7.34%	0.86%	8.68%

Table 6: Proportional difference according to the country of origin per stars hotel.

5. Discussions and Conclusions

Technical instruments and statistical comparisons were applied to the analysis of the level of evaluation agreement about Swiss hotels (n=77) within HRWs. Tripadvisor and Booking seem to be similar on their hotel evaluations, and this result confirms the trend identified by Schegg and Fux (2010) between Tripadvisor and HolidayCheck.

Alternatively, it has been showed that users belonging to the same country of origin tend to be coherent with their hotel evaluation. Moreover, the analysis of the proportional difference highlighted which are the nationality groups (i.e. from the ten analyzed countries of origin) that are more likely to appreciate Swiss hotels. European countries (such as Germany, Spain, Italy, and Netherland) seemed to less appreciate Swiss hotels; in contrast with users from USA seem to appreciate more (with a percentage of 3%). 3 and 4 star hotels, which are the majority of the sample (n=70) indicate that nationality groupings are coherent in the evaluation also with the introduction of different clustering rules. These results allow researchers to design further works in this direction.

One of the main limitations of this study is that it is related to a group of hotels from only one specific country (Switzerland); and the results allow us to hypothesize some future works: (i) to test our approach using hotels from other countries, and extend the number of HRWs, in order to evaluate the coherence among systems and the influence of countries of origin; (ii) to group the hotels according to their actual star quality in order to investigate further type of coherence; (iii) to compare the results with the sentiment expressed on the reviews in order to understand which kind of factors are more likely to predict the single evaluation within a nationality. Therefore, the implication for future research could be on the prediction of evaluation from specific countries of origin. A complementary integration with the analysis of the online reputation (stakeholders opinions expressed online particularly on user-generated contents), can increase the awareness of the hotels presence online, and suggest actions to better communicate with prospective consumers.

Nevertheless, in line with the related work of Schegg and Fux (2010), a proactive use of customer evaluation presented on HRWs is an opportunity, particularly in the area of customer relationship management (CRM). The technical solutions presented in this study about data collection and analysis can finally contribute to methods for Internet-based data acquisition, and for the creation of value from online unstructured data.

References

Baumgartner, R., Gatterbauer, W., and Gottlob, G. (2009). Web Data Extraction System. In Ling Liu and M. Tamer Özsu, editor(s), Encyclopedia of Database Systems, 3465-3471, Springer US

Buhalis, D. (2000) Marketing the competitive destination in the future. Tourism Management 21 (1), 97–116.

Buhalis, D. (2003). eTourism: Information technology for strategic tourism management. Prentice Hall, Harlow.

Buhalis, D. (2003). eTourism: Information technology for strategic tourism management.

Callan, R. (1998). Attributional Analysis of Customers' Hotel Selection Criteria by U.K. Grading Scheme Categories. Journal of Travel Research. 36 (3): 20-34.

Cantoni, L., Tardini, S. (2010), The Internet and the Web, in D. Albertazzi & P. Cobley (eds), The media. An introduction (3rd ed.), New York et al.: Longman, 220-232

Chatterjee, P. (2001). Online reviews. Do consumers use them? In M. C. Gilly, & J. Myers-Levy, (Eds.), ACR2001 Proceedings (pp. 129-134). Provo, UT: Association for Consumer Research.

Dickinger, A. and Mazanec, J. (2008). Consumers' Preferred Criteria for Hotel Online Booking. in Peter O'Connor, Wolfram Hopken, Ulrike Gretzel (eds.), Information and Communication Technologies in Tourism 2008 (Proceedings of the International Conference in Innsbruck, Austria, Springer, Wien – New York 2008, pp. 244-253

Dubé, L. and Renaghan, L.M. (2000). Creating Visible Customer Value: How Customers View Best-Practice Champions. Cornell Hotel and Restaurant Administration Quarterly 41(1): 62-72.

Fesenmaier, D. R., Xiang, Z., Pan, B., & Law, R. (2010). An analysis of search engine use for travel planning. Paper presented at the Information and Communication Technologies in Tourism ENTER 2010, Lugano, Switzerland, pp. 381-392

Gretzel,U., Fesenmaier D.R., O'Leary, J.T. (2006). The transformation of consumer behaviour. In D. Buhalis, C.Costa (ed.), Tourism Business Frontiers, Elsevier, 9-18.

Inversini A., Marchiori E., Dedekind C., Cantoni L. (2010). Applying a Conceptual Framework to Analyze nline Reputation of Tourism Destinations. in Ulrike Gretzel, Rob Law & Matthias Fuchs (eds.), Information and Communication Technologies in Tourism 2010 (Proceedings of the International Conference in Lugano, Switzerland, February 10-12, 2010), Springer, Wien – New York 2010, pp. 321-332.

Inversini, A., Cantoni, L. and Buhalis, D. (2009) *Destinations' Information Competition and Web Reputation*, Journal of Information Technology & Tourism, 11 (2009), pp. 221-234

Inversini, A., Buhalis, D. (2009).Information Convergence in the Long Tail. The Case of Tourism Destination InformationIn. In W. Hopken, U. Gretzel & R. Law (Eds.), Information and Communication Technologies in Tourism 2009 – Proceedings of the International Conference in Amsterdam, Netherland (pp. 381-392). Wien: Springer.

Inversini, A., Marchiori, E., Dedekind, C., Cantoni, L.(2010) Applying a Cenceptual Framework to Analyze Web Reputation of Tourism Destinations. In U. Gretzel, R. Law, & M. Fuchs (Eds.), Information and Communication Technologies in Tourism 2010 – Proceedings of the International Conference in Lugano, Switzerland (pp. 321-332). Wien: Springer.

Kittur, A., Kraut, R. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. Proceedings of the 2008 ACM conference on Computer supported cooperative work.

Marchiori, E., Inversini, A., Cantoni, L. (2010) Measuring the Online Reputation of Sustainable Tourism Destinations. Presented at 14th International Conference on Corporate Reputation, Brand, Identity and Competitiveness, Rio de Janeiro – Brazil

Marcussen, C. (2009). Trends in European Internet Distribution - of Travel and Tourism Services. Updated 23 March 2009. http://www.crt.dk/uk/staff/chm/trends.htm

Nielsen Global Online Consumer Survey (April 2009). Global Advertising: Consumers Trust Real Friends and Virtual Strangers the Most. http://blog.nielsen.com/nielsenwire/consumer/global-advertising-consumers-trust-real-friends-and-virtual-strangers-the-most/

O'Connor, P. (2008). User-Generated Content and Travel: A Case Study on Tripadvisor.Com. in Peter O'Connor, Wolfram Hopken, Ulrike Gretzel (eds.), Information and Communication Technologies in Tourism 2008 (Proceedings of the International Conference in Innsbruck, Austria, Springer, Wien – New York 2008, pp. 47-58.

O'Reilly, T. (2005). What Is Web 2.0. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/ what-is-web-20.html

Poon, A. (1993). Tourism, Technology and Competitive Strategies, Wallingford, CT: CAB International, Oxford.

Schegg, R. and Fux, M. (2010). A Comparative Analysis of Content in Traditional Survey versus Hotel Review Websites. in Ulrike Gretzel, Rob Law & Matthias Fuchs (eds.), Information and Communication Technologies in Tourism 2010 (Proceedings of the International Conference in Lugano, Switzerland, February 10-12, 2010), Springer, Wien – New York 2010, pp. 429-440.

Segaran, T. (2007). Programming Collective Intelligence: Building Smart Web 2.0 Applications. O'Reilly, Sebastopol, CA, USA

Shao, J. and Gretzel, U. (2010). Looking Does Not Automatically Lead to Booking: An Analysis of Clickstreams on a Chinese Travel Agency Website. in Ulrike Gretzel, Rob Law & Matthias Fuchs (eds.), Information and Communication Technologies in Tourism 2010 (Proceedings of the International Conference in Lugano, Switzerland, February 10-12, 2010), Springer, Wien – New York 2010, pp. 197-208.

Sheldon, P. J. (1997). Tourism information technology. Cab International New York.

Smith, D., Menon, S. and Sivakumar, K. (2007). Online peer and editorial recommendation, trust and choice in virtual markets. Journal of Interactive Marketing, 19 (3) pp. 15-37.

Stuckenschmidt, H. and van Harmelen, F. (2005) Information Sharing on the Semantic Web. Berlin. Springer

Surowiecki, J. (2005). The Wisdom of Crowds. 306, Anchor Books, New York.

Werthner, H., & Klein, S. (1999). Information technology and tourism: A challenging relationship SpringerWien.

Xiang, Z. and Gretzel, U. (2009). Role of Social Media in Online Travel Information Search. Journal of Tourism Management, 31: 2, 179-188

Xiang, Z., Wöber, K., & Fesenmaier, D. R. (2008). The representation of the tourism domain in search engines. Journal of Travel Research. 47: 137-150