Guest Reputation Indexes to Analyze the Hotel's Online Reputation Using Data Extracted from OTAs

RUI CHOUPINA, MARISOL B. CORREIA, CÉLIA M.Q. RAMOS
Escola Superior de Gestão, Hotelaria e Turismo
University of Algarve
Campus da Penha, 8005-139 Faro, PORTUGAL
{rmchoupina, mcorreia, cmramos}@ualg.pt

DANIEL MARTINS
Instituto Superior de Engenharia
University of the Algarve Campus da Penha, 8005-139 Faro, PORTUGAL
djmartins@ualg.pt

FRANCISCO SERRA
Escola Superior de Gestão, Hotelaria e Turismo
University of Algarve
Campus da Penha, 8005-139 Faro, PORTUGAL
fserra@ualg.pt

Abstract: Nowadays many travelers use online travel agency (OTAs) to book flights, hotel rooms, rent-a-cars, cruises or entire vacation packages. Usually OTAs allow their users to give scores and to write reviews about what was used. Each OTA defines the terms and conditions for guest rating or review score and hoteliers are giving increasing importance to the scores and reviews their guests do in OTAs. This paper proposes two guest reputation index to help hoteliers to monitorize their presence in OTAs. The Aggregated Guest Reputation Index (AGRI), which shows the positioning of a hotel in different OTAs and it is calculated from the scores obtained by the hotels in those OTAs. Another one, the Semantic Guest Reputation Index (SGRI), which incorporates the social reputation of a hotel and that can be visualized through the development of word clouds or tag clouds. Examples of usage of these indexes are given with data extracted from 5-stars hotels in the Algarve, south region of Portugal, that are available on Booking and Expedia.

Key-Words: Guest reputation index, Hospitality, Online Travel Agency, Positioning of hotels, Online reputation, Sentiment analysis, Word Cloud.

1 Introduction

The management of rates on hotel management is becoming increasingly complex and it is very difficult to understand the value that hotels present, in a geographical area or in a class of services with similar features.

With the quantity of information that daily circulates through the web and a number of users estimated at 3 billion in 2015 [1], there is a lot of information about competitors, the hospitality industry and about consumers trends. This information is increasingly more accessible to organizations at lower costs, presenting a new challenge on creating platforms that are able to deal with this huge amount of information that organizations have at their disposal. This is the "big data challenge" [2], that allows that organizations have turned their focus to collect information not only from internal sources, but from external ones [3].

The web 2.0 with its strong interactive component allows its users to consult the static contents and to share and exchange information within the virtual community, which is extremely dynamic and influential in the consumers decision-making. Virtual relationships currently established between the hospitality industry and its guests provide valuable information, allowing a continuous assessment by the hotel managers in its management decisions, guests' feedback and the behavior of their competitors [4].

ISBN: 978-1-61804-313-9

The sale of hotel rooms by online channel, particularly through the various OTAs (Online Travel Agencies) that exist in the market, assumes an increasingly importance [5]. Many travelers consult different websites before booking online or to contact a hotel booking service, which reinforces the idea of the increasingly important role that the OTAs have in choosing a particular hotel.

OTAs are the fastest growing segment of the travel industry. Booking, Expedia [6], Travelocity, Priceline. Orbitz and Kayak are some examples of OTAs. Travelers can use these OTAs to search for flights, hotel rooms, rent-a-cars, and so on. For example, Expedia collect and aggregate data from thousands of travel service providers, allowing to book flights, hotel rooms, rental cars, cruises or entire vacation packages.

More and more booking traffic is to be carried over the traditional channels (travel agencies) to the individual customers and to corporate travel planners, which use the online intermediaries (OTAs) for information queries and to obtain pricing information and online reputation of the hotel [7].

This paper proposes two different guest reputation indexes. The first one, the Aggregated Guest Reputation Index (AGRI), which shows the positioning of a hotel in different OTAs and it is calculated from the scores obtained by the hotels in those OTAs. The second one, the Semantic Guest Reputation Index (SGRI), which incorporates the social reputation of a hotel and that can be visualized through the development of word clouds (also known as tag clouds), which enables a facilitated and a graphically attractive visualization of the characteristics most mentioned by the guests of a hotel in their reviews in the OTAs.

The AGRI and SGRI can be considered as a new Key Performance Indicators (KPI), to be included in techniques for an efficient optimization of occupancy and rates of hotel accommodations, known as Smart Revenue Management (SRM) [8], [9].

Prices and types of rooms, capacity, facilities, amenities, and reviews from the hotel guests, among others are some of the functionalities extracted by webcrawlers [8], [9], which run periodically through the webpages of the different OTAs, over different periods of time, in order to get suitable data. For this work, two different OTAs were analyzed: Booking and Expedia.

After the extraction of the information available in OTAs, it is necessary to perform its analysis to make available to hoteliers of valid and easily legible information about their hotels and about their competitive set, in order to enable valuable and quick decision-making.

This paper is structured as follows: Section 2 presents two scenarios for the calculation of an aggregated guest reputation index, while section 3 explains a semantic guest reputation index, which can be developed using the sentiment analysis or opinion mining approach or in a simpler manner, using word clouds. Finally, section 4 presents the conclusion and some guidelines for future work.

2 Aggregated Guest Reputation Index (AGRI)

As presented in [8] and [9], the webcrawler performs the extraction of several items with information about the hotel in a given OTA, for instance: the available rooms, prices, features, amenities, policies, guest reviews and so on. On Booking and Expedia, only the person who booked and completed a stay at that hotel can write reviews and/or gives scores to that hotel. On Expedia, this rate is called the Guest Rating; on Booking is considered the Review Score. Any of these two designations are used in this paper. On TripAdvisor [10] any person can leave a review about a hotel, a restaurant, and so on. They do not need to book and to complete a stay in that hotel. This is one of the reasons why TripAdvisor is not considered in the calculation of the proposed indexes in this paper [11].

The webcrawler also extracts the review score that is based on a given number of reviews and the score breakdown that rate several information dimensions. On Booking these dimensions are: Cleanliness, Comfort, Location, Facilities, Staff, Value for Money and Free Wi-Fi; on Expedia they are: Room Cleanliness, Service & Staff, Room Comfort and Hotel Condition.

Another important information related to guest reviews is the customer segment that the guest belongs. For example, Booking displays the following segments: All reviewers, Families, Couples, Group of friends, Solo travelers, Business travelers, while Expedia shows the following ones: Everyone, Couples, Families, Getaway with friends, Business travelers, Overnight stay before destination, Personal event, Spa, Golf and other.

In the process of organizing the information extracted by the webcrawler, it was found that the Review Score on Booking does not correspond to the average of the ratings of each dimension. Figure 1 shows an example for a hotel on Booking, where

ISBN: 978-1-61804-313-9

the Review Score does not correspond to the average of the Score Breakdown.

On the contrary, the Guest Rating on Expedia corresponds to the average of the scores of the different dimensions analyzed. Figure 2 displays an example of the calculation of the Guest Rating of a hotel on Expedia.

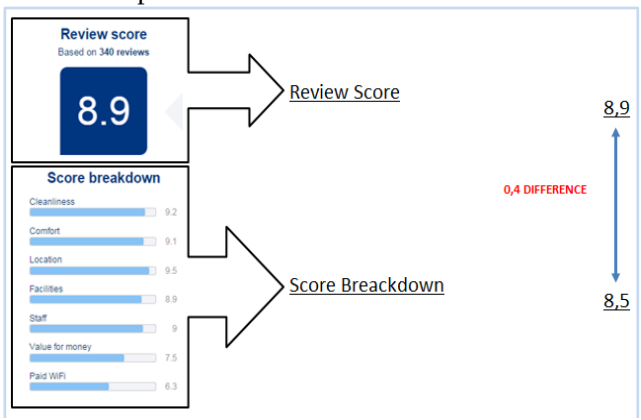


Fig. 1 – Differences between Review Score and Score Breakdown.

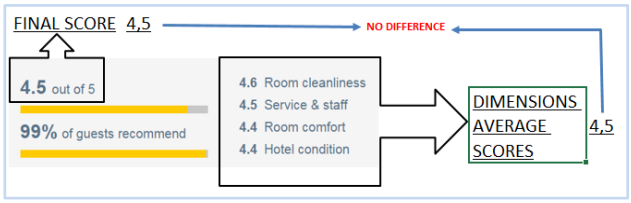


Fig. 2 – Guest Rating corresponds to the average of the dimensions scores.

To overcome these differences that exists with Booking, a first approach to the calculation of an Aggregated Guest Reputation Index (AGRI) was done using the following items:

1) Number of reviews by OTA;
2) Weight of each OTA in the total number of reviews;
3) Review Score by OTA;
4) Review Score by Segment;
5) Number of reviews per Segment;
6) Weight of each Segment in the total number of reviews.

Some of these items are extracted by the webcrawler, others are calculated, as items 2) and 6). As the number of reviews in one OTA can be different from that number of reviews in another OTA, the weight of each OTA in the total number of reviews and the weight of each segment in the total number of reviews are considered and calculated.

Another important aspect is that OTAs can use different rating scales. While on Booking scores are assigned on a scale of 1-10, on Expedia the scale chosen is 1-5. So, the normalization of the scale has to be done. In this paper, the normalization or standardisation to the 1-10 scale was chosen for two reasons: the first one because it is considered easier to have 1-10 scale than 1-5; the other one is because Booking continues to be considered the number one OTA in the world.

Table 1 displays an example for a hotel for the two OTAs considered, for the dimensions of each OTA and showing only one segment, the Families one. The other segments are not displayed only for lack of space. Furthermore, it was considered that the Solo travelers of Booking match the Personal Event of Expedia. Finally, the numbers from 1 to 6 showed in the columns of table 1 correspond to the items presented for the calculation of the AGRI.

HOTEL		Reviews			Families		
	3	1	2		5	4	6
	Score	N°	Weight Number Total Reviews		Score	N° Reviews	%
GRI	4,5	865	100%		4,2	130	15,0%
Booking	4,3	540	62,4%		4,2	79	14,6%
Cleanliness	4,5	540	62,39%		4,3	79	14,6%
Comfort	4,3				4,2		
Location	4,9				4,8		
Facilities	4,2				4,1		
Staff	4,5				4,2		
Value for Money	3,7				3,6		
Expedia	4,6	325	37,6%		4,2	51	15,7%
Room Cleanliness	4,7	325	37,55%		4,4	51	15,7%
Service & Staff	4,6				3,9		
Room Comfort	4,5				4,1		
Hotel Condition	4,6				4,4		

Table 1 – Review Score and one dimension of Score Breakdown.

Having in mind this information, several scenarios can be drawn. Table 2 shows an example of one of them: the calculation of an AGRI as the weighted average of the scores obtained in two OTAs using the weight that each OTA has in the total number of analysed reviews.

Scenario 1					
OTA	Nº Reviews	% Total Nº Reviews	Score	Normalized Score	Weighted Average 1
Booking	220	65,1%	9,2	9,2	9,3
Expedia	118	34,9%	4,8	9,6	
Total	*338*	*100%*			

Table 2 – Calculation of the weighted average using a weighted factor in function of total number of reviews

Another scenario can be using a weighting factor that can be defined by the user. In this case, the weighted factor "number of bookings received Year-To-Date by each channel" (tables 3 and 4) was used.

ISBN: 978-1-61804-313-9

OTA	Nº bookings received by each channel YTD	Weighted Factor
Booking	100	91%
Expedia	10	9%
Total	*110*	*100%*

Table 3 – Calculation of a weighted factor in function of the number of bookings received Year-To-Date (YTD) by each channel.

Scenario 2				
OTA	Weighted Factor	Score	Normalized Score	Weighted Average 2
Booking	91%	9,2	9,2	9,2
Expedia	9%	4,8	9,6	
Total	*100%*			

Table 4 – Calculation of weighted average using weighted factor defined by the user in function of number of bookings received Year-To-Date by each channel.

Taking into account the need to present reliable results and to allow a scalability and a rapid information integration in any hotel revenue management system, several scenarios for the calculation of the AGRI can be proposed. The hoteliers have to choose the scenario that for them provided the most consistent information with the reality of their hotel units.

It is important to refer that the information displayed by the OTAs changes very rapidly. OTAs are constantly improving the interface and changing the type of information displayed. For this reason, the dimensions and segments presented in this paper for each OTA can change from one day to another.

2.1 Application of the AGRI to Algarve 5-stars hotels

Next, the calculation of the AGRI, using the first scenario is demonstrated with the information extracted by the webcrawler for forty 5-stars hotels of the Algarve region, in the south of Portugal, that are available on Booking and on Expedia.

The hotel designation, score and number of reviews were extracted by the webcrawler for each OTA and are displayed in table 5. The calculation of the AGRI was performed using the total number of reviews of the analyzed hotels in the two OTAs considered.

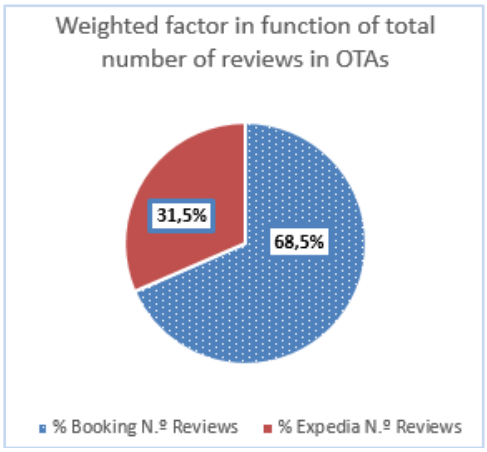
The AGRI can be used to develop KPIs (Key Performance Indicators), which can provide valuable information about the hotel positioning compared with the other 5-stars hotels segment or compared with the competitive set.

For each hotel, the hotelier can analyse the weight that each OTA has on the number of reviews posted about the hotel. For example, for all the forty 5-stars hotels of the Algarve that are available on Booking and Expedia, graph 1 shows that Expedia generated 31,5% of the total number of reviews, while Booking generated 68,5%. As expected, this reinforces the perception that Booking takes an increasingly important role in the number of bookings generated by OTAs. It is important to stress the importance of choosing a common time horizon to the analysis to be done.

	Booking			Expedia					
HOTEL	Nº reviews	Weighted Factor	Score	Nº review	Weighted Factor	Score	Norm. Score	Reviews Total	AGRI
Hotel1	116	65,2%	9,5	62	34,8%	4,8	9,6	178	9,5
Hotel2	226	65,5%	9,2	119	34,5%	4,8	9,6	345	9,3
Hotel3	484	56,9%	9	367	43,1%	4,8	9,6	851	9,3
Hotel4	210	53,8%	8,9	180	46,2%	4,8	9,6	390	9,2
Hotel5	371	91,6%	9,2	34	8,4%	4,6	9,2	405	9,2
Hotel6	501	81,7%	9,2	112	18,3%	4,6	9,2	613	9,2
Hotel7	174	70,2%	9	74	29,8%	4,8	9,6	248	9,2
Hotel8	35	66,0%	9	18	34,0%	4,7	9,4	53	9,1
Hotel9	78	91,8%	9,1	7	8,2%	4,6	9,2	85	9,1
Hotel10	690	66,5%	8,9	347	33,5%	4,7	9,4	1037	9,1
Hotel11	30	66,7%	8,9	15	33,3%	4,7	9,4	45	9,1
Hotel12	188	67,4%	9	91	32,6%	4,6	9,2	279	9,1
Hotel13	140	67,3%	8,9	68	32,7%	4,7	9,3	208	9,0
Hotel14	237	76,5%	8,9	73	23,5%	4,7	9,4	310	9,0
Hotel15	477	61,8%	8,9	295	38,2%	4,6	9,2	772	9,0
Hotel16	148	55,8%	8,7	117	44,2%	4,7	9,4	265	9,0
Hotel17	200	39,5%	8,6	306	60,5%	4,6	9,2	506	9,0
Hotel18	150	48,7%	8,6	158	51,3%	4,6	9,2	308	8,9
Hotel19	147	82,1%	8,8	32	17,9%	4,5	9,0	179	8,8
Hotel20	91	84,3%	8,8	17	15,7%	4,5	9,0	108	8,8
Hotel21	399	76,4%	8,7	123	23,6%	4,6	9,2	522	8,8
Hotel22	644	71,6%	8,8	256	28,4%	4,4	8,8	900	8,8
Hotel23	404	80,5%	8,7	98	19,5%	4,6	9,2	502	8,8
Hotel24	88	57,5%	8,5	65	42,5%	4,6	9,2	153	8,8
Hotel25	667	92,0%	8,7	58	8,0%	4,5	9,0	725	8,7
Hotel26	508	95,5%	8,7	24	4,5%	4,4	8,8	532	8,7
Hotel27	365	60,9%	8,4	234	39,1%	4,5	9,0	599	8,6
Hotel28	593	76,5%	8,5	182	23,5%	4,5	9,0	775	8,6
Hotel29	165	62,0%	8,5	101	38,0%	4,4	8,8	266	8,6
Hotel30	430	48,7%	8,5	453	51,3%	4,3	8,6	883	8,6
Hotel31	71	28,2%	8,4	181	71,8%	4,3	8,6	252	8,5
Hotel32	145	97,3%	8,5	4	2,7%	4,8	9,6	149	8,5
Hotel33	169	38,5%	8	270	61,5%	4,4	8,8	439	8,5
Hotel34	411	89,5%	8,5	48	10,5%	4,1	8,2	459	8,5
Hotel35	116	61,1%	8,4	74	38,9%	4,2	8,4	190	8,4
Hotel36	54	56,8%	8,4	41	43,2%	4,2	8,4	95	8,4
Hotel37	318	86,9%	8,3	48	13,1%	4	8,0	366	8,3
Hotel38	171	77,4%	8,2	50	22,6%	4,2	8,4	221	8,2
Hotel39	215	68,7%	8,1	98	31,3%	4,2	8,4	313	8,2
Hotel40	467	69,7%	8	203	30,3%	4,2	8,4	670	8,1

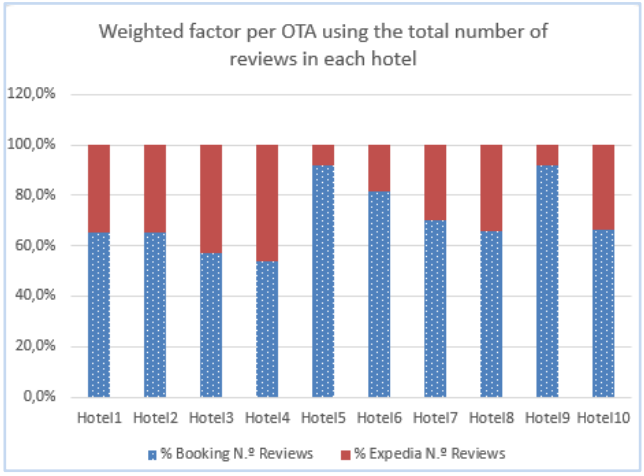
Table 5 – AGRI calculation for 5-stars hotels of the Algarve that are available on Booking and on Expedia

ISBN: 978-1-61804-313-9



Graph 1 – Weighted factor calculated using the total number of reviews of the Algarve 5-stars hotels on Booking and on Expedia.

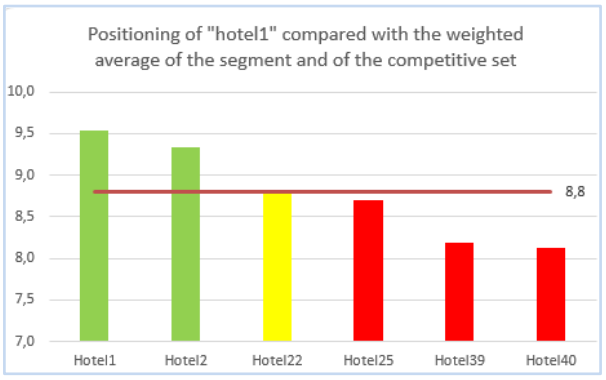
Graph 2 displays the first ten hotels displayed in table 5. Using the information provided by the hotel revenue management system or by the Property Management System (PMS), it is possible and recommended an analysis of the number of bookings raised by each OTA and the number of reviews that were generated by this same channel, allowing to analyse what is the type of guests that is more interactive and participative.



Graph 2 – Weighted factor per OTA using the total number of reviews of each hotel.

Graph 3 shows the positioning of one hotel, denoted as “hotel1” in relation to the average of the forty 5-stars hotels segment (score 8,8) and also in relation to its most direct competitors, its competitive set, which are highlighted with yellow colour in table 5. This is an example of a KPI that can be performed and that is easy to read and to understand.

Finally, saving these KPIs along time, allows that the hotelier can check the evolution of the AGRIs over time and compared it to the competitive set.



Graph 3 – Positioning of “Hotel1” compared with the weighted average of the segment of 5-stars hotels of the Algarve and of the competitive set.

3 Semantic Guest Reputation Index (SGRI)

In addition to the analysis that was done using the AGRI proposed above, the text of the reviews extracted by the webcrawler in each OTA can be used in a different manner and need a different analysis.

The reviews reflect the opinions of the guests and can highlight aspects and items that are more or less valued for them.

Each OTA has its specifications on how it is possible to write reviews on the website. For example, on Booking the guest can give a positive review that is displayed with a green plus sign or a negative one as grey minus sign. For Expedia, the designation is different, the positive reviews are designated as “Pros”, the negative as “Cons” and there is a different designation “Location”, where reviews about the localization can be given.

Another important aspect of the reviews is the language used to write them. On Booking there is the possibility to have reviews in 17 languages as can be seen in figure 3.

ISBN: 978-1-61804-313-9

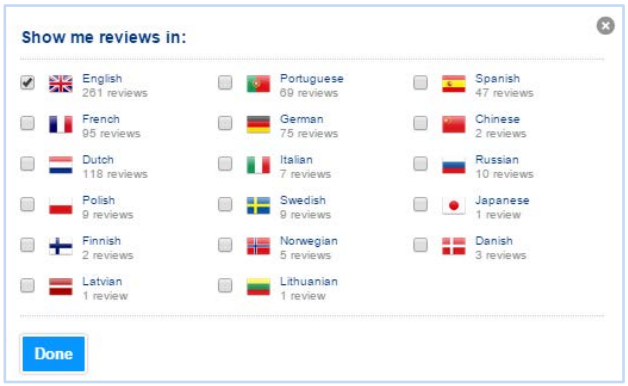


Fig. 3 – Languages of the reviews on Booking.

On Expedia the languages used are 13. Czech, Russian, Polish and Slovenian are not considered on Expedia but used on Booking.

Another important difference between Booking and Expedia is concerning to the reviews showed, Booking shows the reviews posted by guests during the 14 past months whether Expedia never delete reviews apart if the hotel ask it (for example following a refurbishment or a property change of ownership).

A relevant aspect is that reviews give valuable information about the hotel, what is going well or bad with the hotel, related to the various dimensions presented before. Recently, Booking changed the way as reviews are showed. They began to display a summary of the reviews, given information about the number of positive and negative reviews about Location, Staff, Price, Bathroom, and so on. Figure 4 shows an example for a hotel.

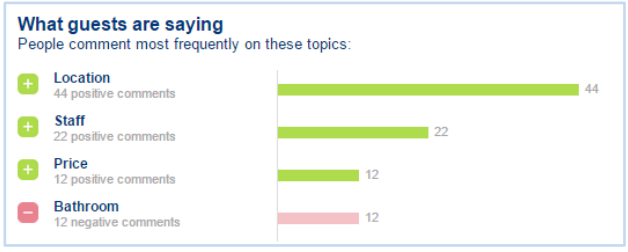


Fig. 4 – Total number of positive and negative reviews for a hotel on Booking.

It is also important to note that the review can be considered positive or negative by the guest, but the text of the review itself can give a slightly different information, which can be important for the hotelier. For instance, the positive review displayed in figure 5 (extracted from Booking) says that staff is friendly but minimal and it was considered by the guest as a positive review.

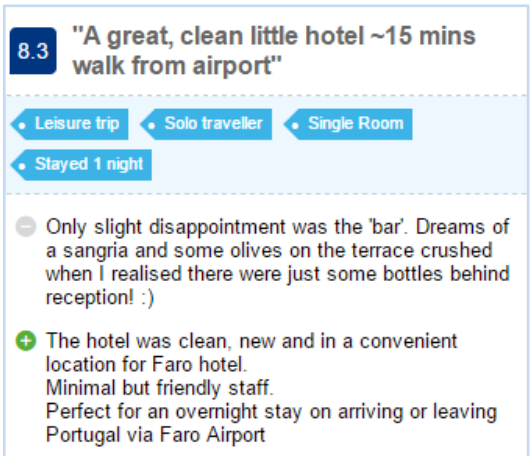


Fig. 5 – Example of a positive and a negative review for a hotel on Booking.

Finally, there are techniques that allow to extract and to evaluate the sentiment expressed in textual data. Sentiment Analysis (also known as Opinion Mining) allows this evaluation.

3.1 Sentiment Analysis or Opinion Mining

Sentiment analysis or Opinion Mining refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in different source materials, generally from text. These fields of knowledge are at the crossroads of information retrieval and computational linguistics and have a rich set of applications [12], since ranging from tracking users' opinions about products or services, to customer relationship management until analysis of hotel guest reviews on OTAs.

To get good results with Sentiment Analysis (or Opinion Mining), it is necessary to implement previously several techniques to sentiment analysis itself, which can help to get the polarity of the text.

As the text in reviews is normally written in an informal way, it is necessary the preprocessing of the text to correct grammatical and orthographical errors, which can difficult the search of relevant information [13].

Generally, Sentiment Analysis involves several phases: extraction and preprocessing of text; natural processing language and sentiment analysis itself.

In the extraction and text preprocessing the abbreviations and linguistics contractions are corrected in order to obtain words that exists in a given language. In informal text, it can also occur the repetition of letters in words to give emphasis (for example, "baaaaaad" instead of "bad").

The natural language processing also involve several steps [13], since the division of the text in

ISBN: 978-1-61804-313-9

simpler terms (tokenization) until complex ones as parsing (phrase chunking). Another step is POS Tagging (part-of-speech tagging), which determines the grammatical class of each component of the analyzed sentences. The Apache OpenNLP library [14] is the most used software for the processing of natural language. It is a machine learning based toolkit, which supports the most common NLP tasks and also includes maximum entropy and perceptron based machine learning.

Finally, in the sentiment analysis phase the subjects of the text are identified. The names that expose the subjects of the text and the adjectives that characterize those subjects as positive or negatives are analyzed in this phase. The terms are analyzed according to their grammatical class. These operations use databases and lexical resources. SentiWordNet [15] [16] is an example of a tool that can be used to perform the Sentiment Analysis (or Opinion Mining). SentiWordNet extends WordNet's usability by another dimension. WordNet as explained in [9] and in [17] is a "dictionary of meanings", which integrates the functions of a dictionary and a thesaurus. In WordNet, nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms, called *synsets*, each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations.

Figure 6 displays a flowchart for the Sentiment Analysis process applied on reviews using SentiWordNet. According to [16], after preprocessing the text, it is reduced to its contents words in a normalized form. For each of the words, SentiWordNet retrieves the synsets that contain each word. If SentiWordNet does not find any synset for that word, the sentiment score is defined to zero. On the contrary, if more than one synset are returned, the word sense disambiguation is necessary. According to those authors, there are several ways to perform word sense disambiguation using WordNet, one of them is using the Lesk algorithms, which disambiguate calculating overlaps of the context words and the synsets' glosses. Finally scores are then given. Generally the scale [-1.0; 1.0] is used, the -1.0 corresponding to the most negative sentiment, 0 to a neutral sentiment and 1.0 the most positive.

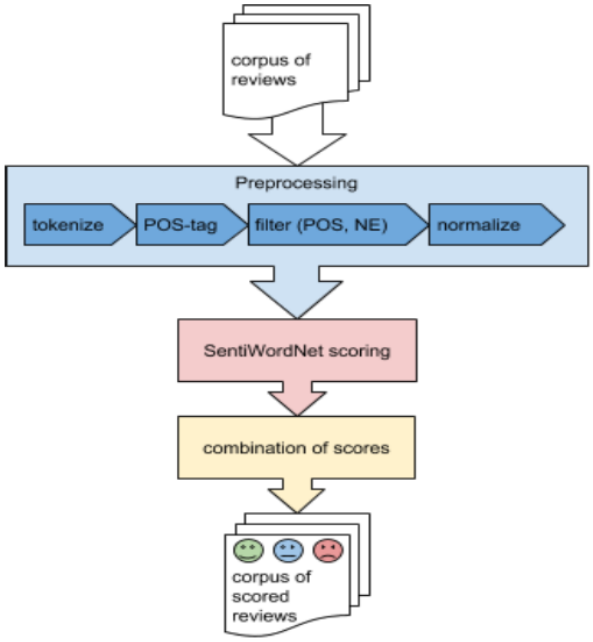


Fig. 6 – Flowchart for Sentiment Analysis (or Opinion Mining) process applied on reviews using SentiWordNet. Source: [16]

Other solutions are acquiring existing software in the market. These solutions allow an easiest implementation of the Sentiment Analysis process, however implies an extra cost with the acquisition of this software. Table 6 listed three of several software available in the market. These solutions are currently used by several leading companies of different areas such as hospitality, consulting and social networking among others.

A different and easier approach to the semantic guest reputation index is to use word clouds to graphically see the most mentioned words present in the reviews.

SOFTWARE	LANGUAGES	FEATURES
Repustate	English, French, Spanish, German, Italian, Russian, Chinese and Arabic	Sentimental Analysis; Semantic Analysis; Repustate Server; Repustate's Excel plugin
ReviewPro	Source language of comments	Global Review Index™ (GRI); Revenue Optimizer; Guest Survey; Sentiment Analysis
TrustYou	Semantic analysis in more than 20 languages with an accuracy of 90-95% for most languages	Reputation Marketing, Reputation Surveys, Reputation Monitoring, Mata Review API

Table 6 – Software for sentiment analysis or opinion mining of guest reviews. Source: [18], [19], [20].

3.2 Word Clouds

Word clouds (also known as tag clouds or text clouds) is a visual representation for text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text. The frequency of each word/tag can be shown with a

ISBN: 978-1-61804-313-9

different font size or color. Wordle, Tagxedo, Tagcrowd and Wordaizer are examples of word clouds software available in the market.

In this paper, the Wordaizer [17] software was used to analyse the extracted guest reviews. This software allows a word counting, revealing the items most cited by guests in their reviews. Obviously, the software does not identify sentiment, feeling or opinion associated to reviews. As was explained in section 3.1, Booking and Expedia have the reviews classified in positive (pros) and negatives (cons), so two different reviews groups (Positives and Negatives) were created and analysed with Wordaizer. Twenty five reviews on Booking and on Expedia of one of the forty 5-stars hotels of the Algarve are presented in figures 7 and 8.

As can be seen in figure 7, room, staff, hotel, breakfast, excellent, great, restaurant, facilities, services are some of the nouns and adjectives most used in the positive reviews.

Fig. 7 – Word cloud of positive reviews.

Fig. 8 – Word cloud of negative reviews.

On the other hand, swimming pool, little, hotel, euros, warm, breakfast, nothing, reception, short are of the nouns and adjectives most used in the negative reviews.

In conclusion, word clouds can give to hoteliers an easy and fast way to visualize the positive and negative reviews.

4 Conclusion

This paper presents recent developments to hotel's online reputation management, which aims the development of smart automatic techniques for an efficient optimization of occupancy and rates of hotel accommodations [8] [9].

In this paper two different guest reputation indexes were proposed. The Aggregated Guest Reputation Index (AGRI), which shows the positioning of a hotel in different OTAs and that is calculated from the scores obtained by the hotels in those OTAs. The Semantic Guest Reputation Index (SGRI), which incorporates the reviews given by the hotel guests who booked the hotel room using an OTA.

The AGRI proposed can use two different scenarios: the first one that calculated the AGRI as the weighted average of the scores obtained in various OTAs using the weight that each OTA has in the total number of reviews analysed. The second one, using a weighting factor that can be defined by the hotelier.

The SGRI can also be developed using two approaches. One, using Sentiment Analysis (or opinion mining) that identifies the sentiment, feeling or opinion expressed in reviews; other, the analysis and visualization of word clouds (or tag clouds) that graphically shows the words most cited by guests in the reviews. Each one of the approaches can give valuable information to hoteliers to monitorize the social reputation and positioning of hotels in OTA. Furthermore, hoteliers can anticipate and influence consumer behavior in order to maximize revenue.

The results achieved in this paper open multiple paths for future work. One of them is to study and to compare the software presented in table 6, the other one is to implement the Sentiment Analysis using SentiWordNet or other similar software, and create a new indicator that integrates the concepts of AGRI, SGRI and social networks into a single indicator.

Acknowledgements: This work was supported by project SRM QREN I&DT, n.º 38962, CEFAGE (PEst- C/EGE/UI4007/2013) and CEG-IST – Universidade de Lisboa. We also thanks to project

leader VisualForma - Tecnologias de Informação S.A.

References

[1] M. Kende, "Internet Society Global Internet Report 2014," [Online]. Available: https://www.internetsociety.org/sites/default/files/Global_Internet_Report_2014_0.pdf. [Accessed in 25 04 2015].

[2] S. Chaudhuri, U. Dayal and V. Narasayya, "An overview of business intelligence technology," *Communications of the ACM*, vol. 54(8), pp. 88-98, 2011.

[3] M. Castellanos, F. Daniel, I. Garrigós and J. N. Mazón, "Business Intelligence and the Web," *Information Systems Frontiers*, vol. 15(3), pp. 307-309, 2013.

[4] D. Ruzic, B. Andrlic and I. Ruzic, "Web 2.0 Promotion Techniques in Hospitality Industry," *International Journal of Management Cases. Special Issue: CIRCLE Conference*, pp. 310-319, 2011.

[5] J. Hao, Y. Yu, R. Law, D. Ka and C. Fong, "A genetic algorithm-based learning approach to understand customer satisfaction with OTA websites," *Tourism Management*, vol. 48, pp. 231-241, 2015.

[6] R. Law and F. Chen, "Internet and Tourism–Part II: Expedia," *Journal of Travel & Tourism Marketing*, pp. 83-87, 2008.

[7] B. Carroll and J. Siguaw, "The evolution of electronic distribution: Effects on hotels and intermediaries," *Cornell Hotel and Restaurant Administration Quartely*, vol. 44(4), pp. 38-50, 2003.

[8] D. Martins, R. Lam, J. Rodrigues, P. Cardoso and F. Serra, "A Web Crawler Framework for Revenue Management," em *14th Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED '15), in Advances in Electrical and Computer Engineering*, Tenerife, España, 2015.

[9] C. M. Q. Ramos, M. B. Correia, J. M. F. Rodrigues, D. Martins and F. Serra, "Big Data Warehouse Framework for Smart Revenue Management," em *3rd NAUN Int. Conf. on Management, Marketing, Tourism, Retail, Finance and Computer Applications MATREFC '15)*, Tenerife, Canary Islands, Spain, 2015.

[10] R. Law, "Internet and Tourism–Part XXI: TripAdvisor," *Journal of Travel & Tourism Marketing*, pp. 75-77, 2008.

[11] R. Ali and Skift, "The Top Online Travel Booking Sites for January 2014," [Online]. Available: http://skift.com/2014/02/24/the-top-online-travel-booking-sites-for-january-2014/. [Accessed in 31 05 2015].

[12] A. Esuli and F. Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining," em *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*, Italy, 2006.

[13] D. Teixeira and I. Azevedo, "Análise de opiniões expressas nas redes sociais," *Revista Ibérica de Sistemas e Tecnologias de Informação*, vol. 8, pp. 53-65, 2011.

[14] "Apache OpenNLP," [Online]. Available: https://opennlp.apache.org/. [Accessed in 01 05 2015].

[15] S. Baccianella, A. Esuli and F. Sebastiani, "SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," em *Proceedings of the Seventh International Conference on*

Language Resources and Evaluation (LREC'10, Malta, 2010.

[16] J. Kreutzer and N. Witte, “Opinion Mining using SentiWordNet,” Uppsala University, [Online]. Available: http://stp.lingfil.uu.se/~santinim/sais/Ass1_Essays/Neele_Julia_SentiWordNet_V01.pdf. [Accessed in 01 05 2015].

[17] “Wordnet Princeton,” [Online]. Available: http://wordnet.princeton.edu/. [Accessed in 01 05 2015].

[18] “Repustate - Enterprise scale text analytics,” [Online]. Available: https://www.repustate.com/. [Accessed in 01 05 2015].

[19] “Reviewpro - Guest Intelligence,” [Online]. Available: http://www.reviewpro.com/. [Accessed in 01 05 2015].

[20] “Trustyou,” [Online]. Available: http://www.trustyou.com/. [Accessed in 01 05 2015].

[21] E. Boiy, P. Hens, K. Deschacht and M. F. Moens, “Automatic sentiment analysis of online text,” 2007.

[22] “Manual of Wordaizer,” [Online]. Available: http://www.mosaizer.com/Wordaizer/WordaizerTutorial.pdf. [Accessed in 01 05 2015].