Correlating Languages and Sentiment Analysis on the basis of Text-based Reviews

Abstract

Customer experiences, in the shape of online reviews, influence other customers and in general, contribute to build a perception of a destination. This work presents the conclusions of a survey to gather user text-based reviews about several categories of destination-related information (accommodation, restaurants, attractions and Points of Interest) from three well-known social media sources (Facebook, FourSquare and GooglePlaces) about eight worldwide destinations with a high overnight rate. Several hypotheses about the correlation between the language and sentiment features of the reviews have been validated over a large dataset of reviews. For example, the analysis detected that the highest number of reviews in a destination is written in the same official language spoken in that place. Furthermore, Dutch speaking people are more positive when writing a review. Finally, English, Italian and Spanish speakers seem to prefer FourSquare while German and French people are quite evenly distributed among FourSquare and GooglePlaces.

Keywords: social media; tourist reviews; destinations; sentiment analysis

1 Introduction

Social media are "Internet-based applications that build on the ideological and technological foundations of the Web 2.0 and that allow the creation and exchange of user-generated content" (Kaplan and Haenlen, 2010). Social media and user-generated content are completely reshaping the way tourism-related information is distributed and the way people make plans to travel (Suanpang, 2013). They continue growing and impacting the tourism and hospitality industry (Browning *et al*., 2013; Xiang and Gretzel, 2010).

The huge growth of these sources has forced the emergence of new research approaches in a wide range of disciplines (Schuckert *et al*., 2015), as the content generated must be systematically gathered, analysed and aggregated in order to leverage it (Hvass and Munar, 2012). Recent examples have demonstrated that the analysis of the text-based reviews and comments from customers can unveil interesting patterns, information and behaviours. Social media interactions generate valuable information that, if properly handled, enables new business strategies like targeted marketing campaigns (Hudson and Khal, 2013) or customised services (Ng and Lien, 2014). While some of them may include curiosities, others may be used to improve the global vision of the market and to perform a better profiling of the customers.

When talking about text-based reviews, it is often necessary to use some automatic approach to obtain the desired information from the text. Currently, there are many free and Open Source tools to analyse text and provide different degrees of automatic natural language understanding. This work presents the conclusions of a survey to gather user text-based reviews about several categories of destination-related information (accommodation, restaurants, attractions and Points of Interest) from three well-known social media sources (FaceBook[1], FourSquare[2] and GooglePlaces[3]). Natural Language Processing tools have been used to perform a basic sentiment polarity analysis, assigning a polarity value from 0 to 10 to all the reviews. The analysis has been based on several hypotheses about the relationship among several variables (destinations, languages, polarities) extracted from the reviews.

Some of the most relevant results of our analysis consist of the discovery of a correlation between the language of a review and the destination the review is about. The highest number of reviews about a touristic location is written in the same language spoken in that location. Secondly, Dutch people leave the most positive reviews. Finally, Foursquare seems the most used social network for tourism, especially for people speaking English.

The rest of this paper is structured as follows. Section 2 describes the research background related to social media and electronic Word-of-Mount as well as text mining and sentiment analysis in the tourism sector. Section 3 defines the methodology employed to gather and analyse the data from the chosen social networks, as well as the research hypothesis. Section 4 analyses some of the results obtained during the validation campaign, and proposes some discussion about it. Finally, Section 5 proposes several conclusions and future work.

2 Research background

2.1 Social media and electronic Word-of-Mouth (eWOM)

To reduce uncertainty and perceived risks, consumers often search for Word-of-Mouth (WOM) when making purchase decisions. Much previous research has presented extensive evidence showing the importance of WOM in purchase decision and choice behaviour. In the Internet era, the effect of WOM has been further enhanced in the form of electronic Word of Mouth (eWOM) (Litvin, Goldsmith and Pan, 2008; Hudson *et al.*, 2015). Advances in Information and Communication Technologies (ICT) have brought unprecedented opportunities and challenges to information-intensive industries (Munar *et al.,* 2013). Consumers can make their opinions easily accessible to other Internet users via message boards, Twitter, product review Websites or online communities.

The impact of eWOM is particularly important when it refers to experience goods, as the quality of those goods, such as accommodation or food&beverage services, is often unknown before consumption. Thus, consumers have to rely on eWOM to make inferences about the quality of such goods (Wirtz and Chew, 2002). Given the critical influence of eWOM on the hospitality industry, especially the hotel segment (Cantallops and Salvi, 2014), online text-based reviews have become a key component of hospitality management (Leung *et al.*, 2013). Hoteliers are increasingly aware of the need to develop strategies to address consumer reviews (Levy, Duan and Boo, 2013).

[1] www.facebook.com
[2] www.foursquare.com
[3] plus.google.com/u/0/local

Consumers write online reviews to indicate their level of satisfaction with the hotel (Liu *et al*., 2013) and inform other consumers on the Internet about their hotel stay experience (Park and Allen, 2013). Online text-based reviews have become one of the most important information sources in consumers' lodging decision making (Ye *et al*., 2011) and are used considerably to inform consumers of accommodation quality (Filieri and McLeay, 2014). Consumers tend not to book a hotel without seeking online reviews (Kim, Mattila and Baloglu, 2011).

A number of Websites specialized in tourism and hospitality has flourished on the Web (e.g., Trip-advisor, Hotels.com, Expedia, Yelp.com, Citysearch, Orbitz, Booking.com) and several social networks. Many of them enable users to exchange information, opinions or recommendations concerning certain destinations, hotels, and other tourist services (O'Connor, 2008; Ye *et al*., 2011; Liu and Park, 2015). These online platforms provide excellent tools for tourists to document and relive their travel experience such as expressing their satisfaction level with the hotel stay experience (Filieri and McLeay, 2014). Besides of the overall ratings, attribute ratings on hotel specific attributes such as service, location, price room and cleanliness are available to customers on social media platforms, and are commonly taken into account when customers evaluate a hotel (Ramanathan and Ramanathan, 2011; Zhang *et al*., 2013).

2.2 Text mining and sentiment analysis

The tremendous growth of these data-generating sources has inspired the development of new approaches to understand these phenomena in a variety of disciplines. Sentiment Analysis and Opinion Mining are closely related fields which refer to the application of Natural Language Processing (NLP) techniques to extract subjective information about how a person expresses a sentiment (negative, positive or neutral) about something (Pang and Lee, 2008; Liu, 2010; Liu, 2012). These tasks are increasingly important to determine the opinion about products and services, and brand reputation on the Internet (Cambria *et al*. 2013; He *et al*., 2013; Gräbner *et al*., 2012).

During the last years, many different approaches have appeared in the literature to mine information from customer-generated textual comments. Ghose, Ipeirotis and Li (2009) used a 4-grams Dynamic Language Model classifier to acquire a subjectivity confidence score for each sentence in a hotel review, and derived the mean and standard deviation of this score. The analysis of the content focused on polarity classification, sentiment classification of customer reviews, or the automated extraction of product attributes. Ye, Zhang and Law (2009) presented a study to analyse the existing approaches to perform automatic classifications based on the sentiment analysis of online reviews related to travel destinations. Furthermore, the study analyses different supervised machine learning algorithms and their effect on the different amount of training corpus to various performance measurements in terms of accuracy, precision, and recall in the sentiment classification of online reviews about tourist destinations.

Moreover, Lee, Singh and Chan (2011) used text mining techniques to extract keywords from descriptive comments of hotel customers in order to identify areas of service failures and recovery actions, and identify main topics based on the frequency of key terms. Finally, Kasper and Vela (2011) have implemented a service for hotel managers that collects customer reviews from various sites on the Web; analyzes and classifies the textural content of the review; and presents the results in a precise way.

The current research trend is focusing on micro-blogging messages, which include dealing with short texts with a very particular format and jargon (Martínez-Cámara *et al*., 2014), and using the so-called topic models (Blei, 2003) to detect the topics and sentiment expressed in customer reviews (Rossetti *et al*., 2015).

3 Methodology

This study focuses on customer text-based reviews about several types of destination-related information in eight different cities and regions worldwide (Amsterdam, Barcelona, Berlin, Dubai, London, Paris, Rome and the Tuscany region in Italy). Most of them are capital cities of relevant countries of Europe and all of them are destinations with high overnights rates. The case of Rome and Tuscany, both from Italy, has been selected to observe if there are similar patterns in the same country.

3.1 Research hypothesis

The Natural Language Processing and sentiment analysis tools have been used to validate the following hypothesis.

H1: There is a correlation between the language used to write the reviews and the official language at the destination

This hypothesis will investigate if a destination receives more reviews from people of the same country or from abroad (i.e. national tourism vs. international tourism). This fact can be inferred from the language of the reviews. It is assumed that the official language of the destination is the mother language of the tourists.

H2: The sentiment towards destination-related information depends on the nationality of the tourist

Several assumptions have been made to answer this question using the gathered data. The perception with the detected polarity for the analysed customer reviews is detected. Furthermore, it is assumed that the nationality of the reviewer is the same as the language used for the review. For example, we are assuming that a review in French belongs to a French tourist, which could be a very rough generalisation.

H3: Different languages are evenly distributed across studied social networks

The answer to this question will analyse if the three monitored social networks have an evenly distributed penetration with regard to the languages/countries. This hypothesis also assumes that there is a high correlation between languages and countries.

3.2 Data collection

As mentioned previously, the data used in this study has been retrieved from three popular social networks: Facebook, Foursquare and Google Places. Facebook is a general social network where registered users can build their own profile, establish relations with other users, create new content or pages about a topic (recipes, health, activities etc.) and review or express likes about users and their pages. Destinations can create their Webpage (a virtual wall) and publish multimedia content in order to retrieve feedback. In the case of FourSquare and GooglePlaces, they are more location-based social networks, allowing users to interact and comment about places they have visited (including any type of accommodation). Registered users in FourSquare can perform one or more check-ins in a place according to the places visited which may lead into a large availability of places. Furthermore, hoteliers can register and advertise their properties in Google Places, receiving reviews and feedback from customers.

For each social network, a tailored crawler was used to extract reviews about four types of categories or places within the destination-related information in the previously mentioned eight cities: accommodation, restaurants, Points of Interest (POI) and attractions. For this survey, accommodation refers to a place where people can sleep; a restaurant is a place where people can eat and drink; POI is a local service for tourists, such as an ATM or a library; and an attraction is a place which attracts tourists to a given location, such as a monument or a museum.

This process was done in two steps. First, each crawler searches for places within the geographical areas of the analysed destinations. A geographical area has been identified by a circle defined by a geographical coordinate (i.e. latitude and longitude), which represents its centre, and a radius. Secondly, reviews about the four types of places were retrieved using algorithms tailored for each social network.

The data gathering process extracted 553.347 places and 634.564 text-based reviews in total. Table 1 illustrates the number of extracted places and reviews for each destination, divided per social network. Due to some crawling problems, places in Tuscany were not extracted for Facebook.

Table 1. Tourpedia data content summary

Location	**Facebook**		**Foursquare**		**Google Places**	
	places	***reviews***	***places***	***reviews***	***places***	***reviews***
Amsterdam	583	7.725	7.735	27.537	13.635	7.623
Barcelona	455	3.732	6339	46.445	18.499	14.092
Berlin	2.084	3.078	21.875	43.816	39.765	22.630
Dubai	1.052	3.124	14.469	38.347	7.301	3.844
London	4.893	3.263	47.148	137.749	121.723	75.973
Paris	832	4.227	6.545	46.431	51.665	36.572
Rome	4.465	384	16.913	35.503	31.455	15.094

Tuscany	n.a.	n.a.	43.844	40.389	70.072	16.986

Table 2 shows some examples of text-based customer reviews gathered from the monitored social networks. Although all the examples are in English for better understanding of the paper, reviews in French, Italian, Spanish, German and Dutch have also been collected. The sample included both positive and negative reviews, some of them being more verbose while others are shorter and more direct.

Table 2. Some examples of gathered customer reviews (the original content is respected, including misspellings and typos)

Source	Examples of customer reviews (misspellings included)
Facebook	*One of my favorite hotels! very kind staff and great location!!! In the evening, a warm fireplace in the lobby and a wonderful mood!*
	Bar open until you decide to go to bed and largest towels ever !!!
	10 euros per day for wifi!,... Not acceptable in Europe !
Foursquare	*Perfect location.*
	Nice hosting service. Decent breakfast and honest staff, excepts some night shift recepcionists...
	Worst hostel ever! I could stay in better rooms with this money. Rooms are cold, there are bugs everywhere, sheets are not clean.
Google Places	*VERY POOR - Back-packers hostel, not a hotel. over 200 euros for the "family room"- very noisy, dirty very dusty, stains on the carpet.*
	Very noisy rooms. Cleaning staff continuosly enter in the room despite of "do not disturb" cartel. Reception staff not so helpful.
	Awesome hotel. Nice views (get an upper room) and it has all the top shelf frills you'd expect from a hotel like this.

The data gathering campaign was able to retrieve customer reviews ranging from December 2009 to May 2014. Table 3 illustrates the gathering periods for each social network (i.e. the earliest and the latest retrieved customer review).

Table 3. Date ranges of the gathered content

Social Media	Start Date	End Date

Facebook	Thu Dec 13 2012 09:44:29 GMT+0100 (CET)	Sat Feb 01 2014 21:07:46 GMT+0100 (CET)
Foursquare	Wed May 27 2009 11:10:11 GMT+0200 (CEST)	Wed Nov 13 2013 00:26:58 GMT+0100 (CET)
Google Places	Mon Dec 31 2009 01:00:00 GMT+0100 (CET)	Sun May 11 2014 22:52:21 GMT+0200 (CEST)

3.3 Data analysis

The customer review analysis process includes two main tasks: the identification of the language of the review and the analysis of its sentiment. OpeNER[4] tools have been used to fulfil this analysis. OpeNER is a Natural Language Processing framework that allows performing several types of text processing tasks in multiple languages (Agerri *et al.,* 2013; García-Pablos *et al.,* 2015).

The aggregated polarity for a place has been calculated in the following way. Let $M = (m_1, m_2, \ldots, m_k)$ be a set of social media networks, let j be a place available on all elements of the set M and let be $R_{ij}^{(m)} = (r_{1j}^{(m)}, r_{2j}^{(m)}, \ldots, r_{nj}^{(m)})$ the set of reviews associated to the place j on social media network m. Let o_{ji} be the sentiment polarity extracted through the polarity-tagger module from the review $r_{ji}^{(m)}$. The overall sentiment $s_j^{(m)}$ about the place j on social media m is calculated as the arithmetic mean of all opinions o_{ij} :

$$s_j^{(m)} = \sum_{i=1}^{n} \frac{o_{ij}}{n}$$

The overall sentiment s_j about the place j is calculated as a weighted arithmetic mean of all $s_j^{(m)}$.

$$s_j = \sum_{i=1}^{k} \frac{w_i s_j^{(i)}}{k}$$

where $\sum w_k = 1$. The sentiment s_j is then normalized on a range 1-10, where 1 indicates the lowest value and 10 the highest one.

4 Research hypotheses and discussion

The data gathered and processed has been analysed in order to validate the research hypotheses. The first one focuses on the correlation among the languages used to write the reviews and the official language at the destination. Figure 1 shows the distribution of reviews gathered for each of the targeted locations.

As it was foreseen, English has a strong presence in all destinations, probably due to its use as *lingua franca* (Truchot, 2002). Apart from the pervasive presence of reviews in English, it can be clearly observed that the official language (the major language for the destination) dominates over the rest, except in the case of Amsterdam, in which there are more reviews in English than in Dutch. According to several reports,

[4] www.opener-project.eu

about 94% of Dutch people can maintain a fluid conversation in English[5], a much higher percentage than people from other countries. This fact could explain the domination of English above Dutch in the reviews about Amsterdam.

In the case of Dubai, as Arabic has not been taken into account, English was expected to be the dominant language, as it is used for business, and around 75% of the population in Dubai are expatriates, most of whom speak English apart from their mother language[6].

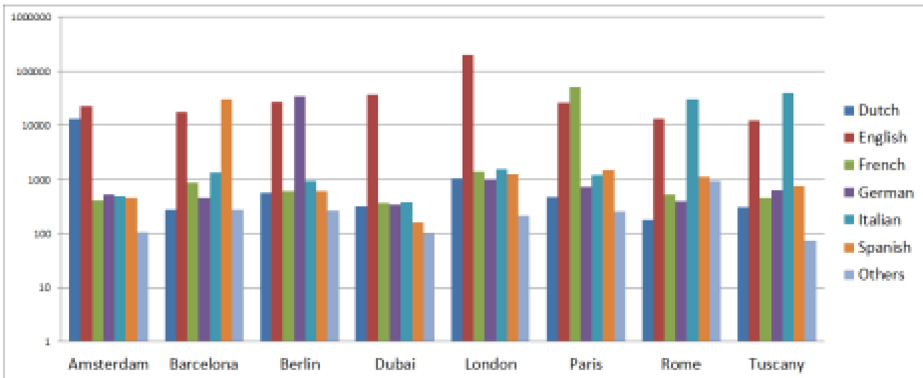

Fig. 1 Distribution of the language of the reviews per destination (logarithmic scale)

It is worth observing that both Rome and Tuscany share very similar distribution of languages, something that is not a surprise since both are placed in Italy. However, Rome seems to have more other languages marked as “others”, probably due to its capital role, its historic relevance or being the placement for the Vatican.

Figure 2 shows the complementary distribution, destination reviews (from the all four categories of the destination-related information) across each of the analysed languages. As it was expected, English is more evenly distributed among all the destinations.

---

[5] Special Eurobarometer 386: Europeans and their languages, June 2012 http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf Europeans and languages: http://ec.europa.eu/public_opinion/archives/ebs/ebs_237.en.pdf

[6] https://www.justlanded.com/english/Dubai/Dubai-Guide/Language/Languages

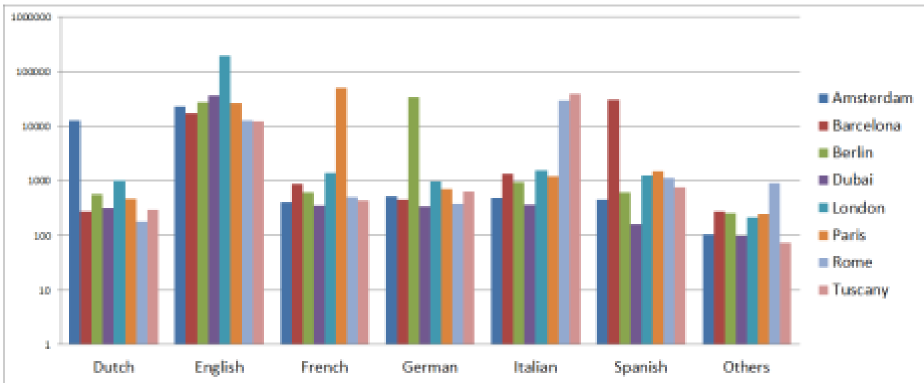

Fig. 2 Distribution of the destination reviews per language (logarithmic scale)

The relationship between the sentiment towards places and the nationality of the tourist was also analysed. Figure 3 displays the distribution of the detected polarity for each language. Unfortunately, it was not possible to detect sentiment for German due to a problem with the analysis tool. As it can be observed, there is a noticeable amount of Dutch reviews in the most positive side of the chart. Dutch people are known to be especially tolerant and open minded (Zick *et al.*, 2011), a fact that may affect the severity with which they evaluate place when they are traveling.

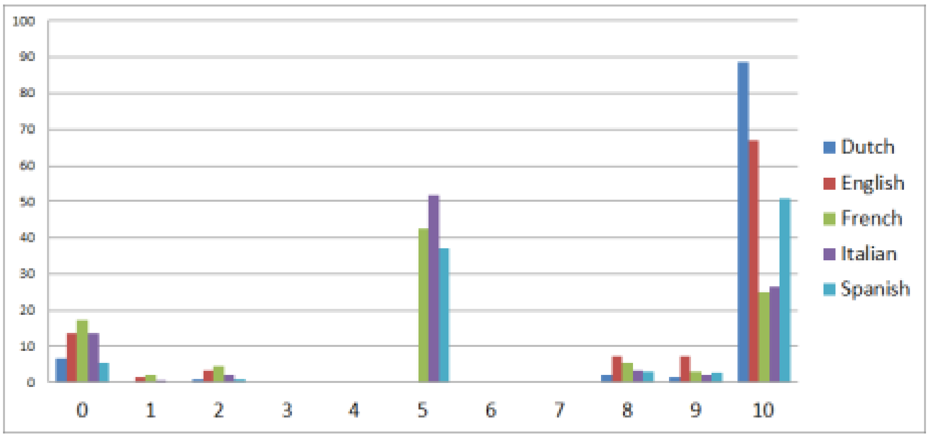

Fig. 3 Distribution of the language (percentage) of the reviews per detected polarity (polarity 0 means "very negative", polarity 10 means "very positive")

On the other hand, the languages with more neutral reviews[7] are French, Italian and Spanish. Three of them are Romance languages (Italian, Spanish and French) in opposition to the other three non-Romance languages (English, German and Dutch). This may indicate a bias on how the language families (or the cultures behind those

[7] Always assuming the polarity given by our system and our metric

language families) explain their experiences when visiting a place or reviewing an accommodation. Germanic languages are historically known to use a more direct communication, while Romance languages usually use smoother expressions and euphemisms or more elaborated discourses to point out the same facts[8]. In any case, the bias on the sentiment distribution for the different languages may resemble the different ways of expressing emotions for different cultures[9].

Furthermore, it can be observed that there are more reviews labelled as positives that negatives. This could be the actual tendency, although it is also known that negative opinions are more difficult to detect because of irony, sarcasm and other subtle ways to express dissatisfaction (Pang and Lee, 2008).

Finally, the distribution of the different languages across studied social networks was analysed (Fig. 4). It can be observed that FourSquare is very popular among English, Italian and Spanish speaking people. German and French people are quite evenly distributed among FourSquare and GooglePlaces. Facebook contains less content in general, probably because the interaction model is different and the content of the publication walls is more oriented to sharing multimedia content and special offers rather than to building a customer review ecosystem.

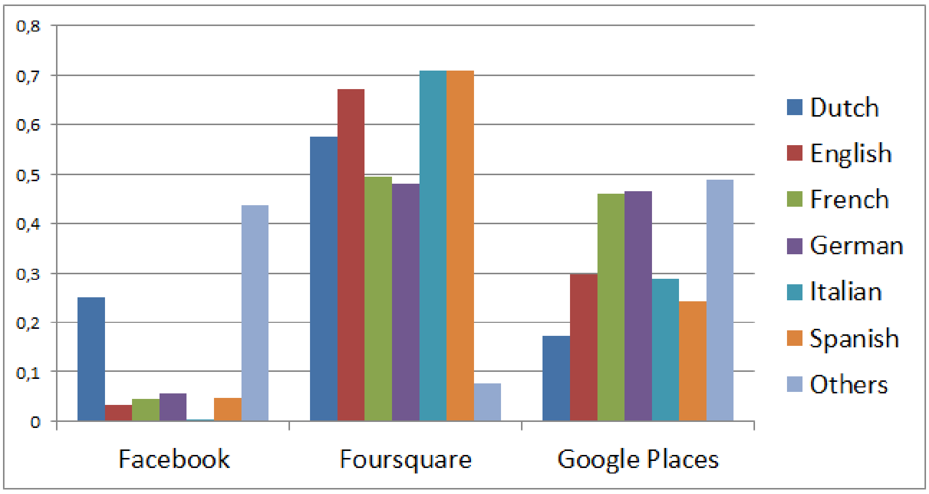

Fig. 4 Distribution of the language of the reviews per social network

It can also be observed the distribution of other languages apart from the ones selected for this study. For a more detailed analysis, it would be interesting to include more languages to be detected and aggregated, to find out whether these languages are Asian (Chinese, Japanese), Russian, Arabic, or local languages.

[8] http://www.cryptograph.com/englang.htm
[9] http://news.stanford.edu/news/2015/march/cultural-differences-sympathy-032525.html

5 Conclusions and future work

Social media and customer generated comments are become increasingly important to obtain relevant information about customers. This also applies to the several agents within the tourism value chain (destinations, accommodation, food&beverage), in which the customer generated opinions impact the perception of those agents and their associated products and services.

This paper deals an evaluation campaign in order to analyse with the correlation of the language and the sentiment analysis of multilingual text-based reviews from social media content. Three different social networks frequently used by destinations and tourists have been crawled in order to extract text-based reviews for eight relevant worldwide destinations. The language of the reviews has been automatically detected before performing a deep sentiment analysis to classify them into a sentiment scale.

Several research hypotheses have been defined and validated in order to motivate a discussion about the patterns detected and the generalization of the results to other destinations and languages. For example, the analysis detected that the highest number of reviews in a destination is written in the same official language spoken in that place. Furthermore, Dutch speaking people are more positive when writing a review. Finally, regarding the preference choices about the social media networks, English, Italian and Spanish speakers seem to prefer FourSquare while German and French people are quite evenly distributed among FourSquare and GooglePlaces.

Customer reviews can be further analysed to unveil more specific information, like the topics that are more frequently addressed, or to study the attributes of the categories that concern to different kind of tourists based on their language, behaviour and opinions.

References

Agerri Gascón, R., Cuadros Oller, M., Gaines, S., & Rigau Claramunt, G. (2013). Opener: Open polarity enhanced named entity recognition.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, (2), 15-21.

Cantallops, A.S., Salvi, F., 2014. New consumer behavior: a review of research one WOM and hotels. Int. J. Hosp. Manage. 36, 41–51.

Filieri, R., McLeay, F., 2014. E-WOM and accommodation: an analysis of the factors that influence travelers' adoption of information from online reviews. J. TravelRes. 53 (1), 44–57.

Garcıa-Pablos, A., Cuadros, M., & Linaza, M. T. (2015, January). OpeNER: Open Tools to Perform Natural Language Processing on Accommodation. In Information and Communication Technologies in Tourism 2015: Proceedings of the International Conference in Lugano, Switzerland, February 3-6, 2015 (p. 125). Springer.

Gräbner, D., Zanker, M., Fliedl, G., & Fuchs, M. (2012). Classification of customer reviews based on sentiment analysis.

Hales, T. C. (2007). Jordan's proof of the Jordan curve theorem. Studies in logic, grammar and rhetoric, 10(23), 45-60.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. International Journal of Information Management, (JUNE 2013).

Hudson, S., Roth, M. S., Madden, T. J., & Hudson, R. (2015). The effects of social media on emotions, brand relationship quality, and word of mouth: An empirical study of music festival attendees. Tourism Management, 47, 68-76.

Hvass, K. A., & Munar, A. M. (2012). The takeoff of social media in tourism. Journal of Vacation Marketing.

Kim, E.E.K., Mattila, A.S., Baloglu, S., 2011. Effects of gender and expertise on consumers' motivation to read online hotel reviews. Cornell Hosp. Q. 52 (4),399–406.

Levy, S.E., Duan, W., Boo, S., 2013. An analysis of one-star online reviews and responses in the Washington D.C., lodging market. Cornell Hosp. Q. 54 (1), 49–63.

Litvin, S.W., Goldsmith, R.E., Pan, B., 2008. Electronic word-of-mouth in hospitality and tourism management. Tour. Manage. 29 (3), 458–468.

Liu, B. (2010). Sentiment analysis and subjectivity. Handbook of natural language processing, 2, 627-666.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

Liu, S., Law, R., Rong, J., Li, G., Hall, J., 2013. Analyzing changes in hotel customers' expectations by trip mode. Int. J. Hosp. Manage. 34, 359–371.

Liu, Z., Park, S., 2015. What makes a useful online review? Implication for travel product websites. Tour. Manag. 47, 140–151.

Martínez-Cámara, E., Martín-Valdivia, M. T., Urena-López, L. A., & Montejo-Ráez, A. R. (2014). Sentiment analysis in twitter. Natural Language Engineering, 20(01), 1-28.

Milano, R., Baggio, R., & Piattelli, R. (2011). The effects of online social media on tourism websites. Information and Communication Technologies in Tourism 2011, 471–483.

Munar, A., Gyimóthy, S., & Cai, L. (2013). Tourism social media: a new research agenda. Tourism Social Media (Tourism Social Science Series, Volume 18). Bingley: Emerald Group Publishing Limited, 1-15.

Ng, E., & Lien, C. Y. (2014). Impact of social media in service innovations: an empirical study on the Australian hotel industry. Quality Innovation: Knowledge, Theory, and Practices, 390-405.

O'Connor, P., 2008. User-generated content and travel: a case study on TripAd-visor.com. In: O'Connor, P., Höpken, W., Gretzel, U. (Eds.), Information and Communication Technologies in Tourism, vol. 2008. Springer Verlag, Vienna,Austria, pp. 47–58.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1–135.

Park, S.-Y., Allen, J.P., 2013. Responding to online reviews: problem solving and engagement in hotels. Cornell Hosp. Q. 54 (1), 64–73.

Ramanathan, U., Ramanathan, R., 2011. Guests' perceptions on factors influencing customer loyalty: an analysis for UK hotels. Int. J. Contemp. Hosp. Manage. 23(1), 7–25.

Rossetti, M., Stella, F., Cao, L., & Zanker, M. (2015). Analysing User Reviews in Tourism with Topic Models. In Information and Communication Technologies in Tourism 2015 (pp. 47-58). Springer International Publishing.

Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. Journal of Travel & Tourism Marketing, 32(5), 608-621.

Suangpang, Pannee (2013). Travelers' behaviours and attitudes towards using social media for tourism. International Journal of Business Tourism Applied Sciences.

Truchot, C. (2002). Key Aspects of the Use of English in Europe. Council of Europe.

Wirtz, J., Chew, P., 2002. The effects of incentives, deal proneness, satisfaction and tie-strength on word-of-mouth behavior. Int. J. Serv. Ind. Manage. 13 (2), 141–162.

Ye, Q., Law, R., Gu, B., Chen, W., 2011. The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings. Comput. Hum. Behav. 27 (2), 634–639.

Zhang, Z., Zhang, Z., Wang, F., Law, R., Li, D., 2013. Factors influencing the effective-ness of online group buying in the restaurant industry. Int. J. Hosp. Manage. 35,237–245.

Zick, A., Küpper, B., & Hövermann, A. (2011). Intolerance, Prejudice and Discrimination-A European Report.