

Tourist Information Evaluation Using a Social Network

Paola Cortez, Serguei Levachkine, and Carlos de la Cruz

Abstract—Currently the time to start planning a trip is common that the user makes use of the Internet in order to obtain through other people's comments certain recommendations for a hotel, what to see, how transported, etc. But when seeking information about a hotel, there are specialized websites through the feedback of users recommend to hotels, leaving out the user comments. This paper proposes a methodology to recommend hotels classifying the comments made by users. As a case study using the comments found on the TripAdvisor site belonging to a Mexican port hotels. The objective is to analyze ontologies through the collective intelligence source that is in the comments.

Index Terms—Recommender, e-tourism, semantic web, social networks, collective intelligence.

I. INTRODUCTION

Currently when planning a trip is common for people to perform an Internet search, since it is the main source of information on tourist destinations for travelers. Simply perform a search on a destination to display millions of results of websites that contain comments, this causes the user to see oversaturated with information, and therefore spend reading only some few comments or otherwise resort to a travel agency.

Internet is considered as the largest travel agency, to be available around the clock, and having too many Web sites dedicated to tourism. People often want to do a search to know the recommendations or opinions that other users have made through specialized tourism sites, tourist spot on the next visit. It is for this reason that the tourism sector is a suitable candidate for the Semantic Web, as mentioned in [1].

In a study by [2] found that the user when making queries related to tourism, access the Web to answer the following questions: Where to go? What means of transportation used to get to the place to visit? What hotels are recommended to stay?

What are the places to visit on the next trip? But although it is known that the common user questions, there has been much emphasis on trying to answer precisely what the user wants to know. For instance, [3] proposed a system that recommends sights Web-based tourist profiling where uses ontologies to make recommendations that satisfy this. Therefore the use of semantics and ontologies could support the user and help him/her to find answers to commonly

asked questions.

This paper proposes comment classifier hotels of the port of Mazatlan Sinaloa, Mexico. This port has been chosen to be in one of the countries that capture currency each year from tourism, in addition Mexico is considered as a favorite tourist destination among foreign and domestic tourism [4]. The goal is to support the potential tourist in search for a hotel, to classify automatically comments, made to a particular hotel. In the classification process ontologies are used in order to capture the semantics of the comment. Then the user is seen the result of the recommendation and hotel location on a map. We have used the Web site TripAdvisor as a supplier of comments.

This work is organized as follows: Section II presents the state of the art related to this proposal; Section III describes the methodology used to carry out the process of gathering information and classification of reviews; Section IV presents the results of the designed system and in Section V we discuss the conclusions and future work.

II. THE STATE-OF-THE ART

Tourism has become the world's largest industry and its growth shows a steady increase year by year. Given this scenario, it highlights the fact that Mexico is among the top countries in the world that receives more visitors each year. There are currently only specialized Web sites provide information on Mexico (<http://www.visitmexico.com/>), however they do not support the tourist in decision-making, it is the responsibility of the tourist to analyze all that information. However if it wants to continue as a preferred tourist destination among tourists, they should propose me

The information currently in the sights is wasted due to the interoperability of the same. However in [5] they mentioned that the problem of interoperability in tourism data can be solved with the emphasis on the combination of social consensus supported by the application of new technologies. This is supported by the source of information in the comments given by the various tourist Web sites. Note that travel blogs have a source of information based on the experiences of tourists [6], leading to a collective intelligence, and it is notorious to observe the increased participation of users in recent years in these blogs.

Day-by-day increased communication via computer, allowed greater exchange of information. In [7] there are some benefits of using social (information) networks involving people who have common interests, but different experiences. One advantage of having information networks is that they allow informal communication, and this increases the chances of exchanging information, fostering a source of confidence among users. In [8] it is commonly referred to sites with information exchange to analyze the collective behavior, because it is difficult if not possible to

Manuscript received November 9, 2013; revised March 25, 2014.

Paola Cortez is with Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas, IPN/México (e-mail: pcortez@ipn.mx).

Serguei Levachkine is with Centro de Investigación en Computación, IPN/México (e-mail: serguei@ipn.mx).

Carlos de la Cruz is with the Professional Interdisciplinary Unit in Engineering and Advanced Technologies, México (e-mail: cdelacruz@ipn.mx).

analyze a user individually.

There are currently OntoGuate systems like [9] which aim to create a knowledge management system for various sectors including tourism highlights. This system is based on an ontology that can capture information from all stakeholders (users, businesses) through an online system, and specific searches.

The Project on Tour [10] works as an assistant focused on semantic search suggests the tourist hotels. This project is based on the use of ontologies.

There are projects like Harmonise [11], which aims to enable interoperability between different ontologies developed to strengthen standards in the tourism sector, this project has been developed for the European Union countries.

There are Web sites like TripAdvisor tourist sector [12] or Expedia [13], however these Web systems are dedicated to give recommendations based on the feedback from tourists and neglecting the information given by the user comments. In [14] they mentioned that when analyzing the collective intelligence expressed in comments we should consider some factors such as: The users are from different countries, have different culture, share certain common affinities, etc., which may cause the resulting comments a bit confusing when you read.

Analyzing a comment you must have a goal to know what is going to get from this, as [15] indicated. Then identifying keywords through comments made in an information network can establish a connection between where the user has made a comment. In [16], they also agree that a few keywords in the review help to identify where the user has made that comment.

Other factors may be considered in a review, as described in [17], where they only used the coordinates and time where the comment was made to make a geographical characterization of an area.

Other advances found in the analysis of social networks are in [18], where it is mentioned that it is possible to obtain a measure of semantic similarity in data using weight in concepts. They propose the use of a pre-defined ontology for data classification. They mention that the data analysis has a main goal the interoperability data.

In [19] tourism information analysis is done using tweets. They classify and identify concepts in different categories. The concepts analysis is done using fuzzification techniques and a count is made to identify classification corresponds to each concept.

Among the tourist sites there is a source of information or knowledge that needs to be explored in order to obtain information synthesized and especially useful for users. In conclusion, we must harness the collective knowledge to meet collective needs. This paper proposes the use of ontologies to perform automatic classification of the comments made at hotels, in order to support the tourist in decision making when choosing a hotel.

III. METHODOLOGY

Mexico through the Ministry of Tourism (SECTUR) makes available to users only Web pages that provide tourist information, however it does not help the tourist in a more

thorough search. Then the user wants to have a review of a hotel, he should spend a lot of time to read some comments to form their own opinion. This is a waste of time for the potential tourist. This proposal proceeded to analyze the comments and found certain patterns of keywords, which were repeated when reading the comments. These keywords have common sense, and considering that the Spanish language has a rich vocabulary; these words were classified through ontologies. These concepts are within each ontologies have some weight, which is linked to the frequency of use of the word in the analyzed comments. Fig. 1 shows a fragment of our methodology.

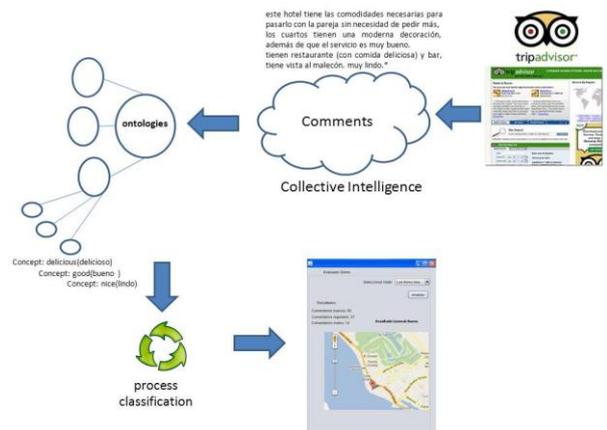


Fig. 1. Methodology used to classify the comments. ¹

Fig. 1 shows that we get the reviews of the hotels (in this case we used the hotels located in the port of Mazatlan, Sinaloa), from the TripAdvisor website, then we proceed to analyze all the comments of each hotel using ontologies. This is intended to prevent the user from having to read all comments. So that once the user selects a hotel is automatically given a recommendation expressed in qualitative terms: For example the comments on hotel "x" say that it is regular.

We have considered the following classifications: "bad", "fair" and "good" to recommend a hotel and we are defined them in the following table:

TABLE I: RATINGS SYSTEM MAKES ENTRIES

Classification	Definition
Bad	The bad concept states that the hotel is considered bad in terms of amenities and services offered. Therefore there is not a convenient option to stay.
Fair	The concept fair means that the hotel is a good place, but not the best option. The fair term suggests that the hotel lacks some amenities or services.
Good	The good concept indicates that the hotel meets the expectations of being a great place to stay. In a matter of amenities is quite acceptable.

The qualitative result is provided by the system once the comments are analyzed using ontologies; this is achieved with a strong recommendation of the hotel. The usefulness of analyzing the comments by ontology is that it avoids the subjectivity of the user to read reviews, and makes use of

¹Comment translation of Fig. 1 (from Spanish). This hotel has the amenities you need to spend with the family without having to ask for more, the rooms have a modern décor and the service is very good. There is restaurant (with delicious food) and bar, overlooks the boardwalk. Very nice. *

collective intelligence embedded in these reviews.

Here are the steps that comprise the methodology used in this research.

A. Gathering Information

The TripAdvisor website is considered one of the most visited travel websites according to [20]. Through this site the user can search for hotels in certain cities, and see a list of detailed comments. On this site there is an active involvement of users, because after commenting on a hotel they provide its rating. But reading the comments, sometimes there are some discrepancies between the written text and the rating given by the user.

For this research we were used selected comments on this website for two weeks of the Easter, resulting in an amount of 686 comments analyzed. It is noteworthy that the comments collected were those found in Spanish.

TABLE II: INFORMATION ABOUT EXPERIMENTS

Activities	Results
Analyzed geographic zone	Mazatlán, Sinaloa
Period of analysis	March 29 - April 19
Number of analyzed comments	686
Number of hotels in the geographic zone	74

B. Ontologies Design

After considering the comments, they were classified into three groups: feedback “bad”, “fair” and “good” (see Table 1). It should be emphasized that the site TripAdvisor has five classifications of hotels which are: *bad*, *poor*, *fair*, *very good*, and *excellent*. We reduce the number of these classifications for the sake of simplicity. The goal of simplification is to avoid ambiguities, because a user in most cases confuses the difference between a *very bad* hotel and a *bad* hotel.

Within the classifications proposed we proceeded to find the recurring words in the comments and once they were identified we were searched respective synonyms to further expand the number of words. It highlights the fact that some words are expressions composed of words frequently discussed in the comments. Later with the words classified, three ontologies were created to thereby perform the analysis of the comments. In total 46 concepts are handled for an ontology called *bad*, 28 concepts for the ontology called *fair* and 68 concepts for the ontology called *good*.

For example, a fragment of the ontology *fair* is shown in Fig. 2.

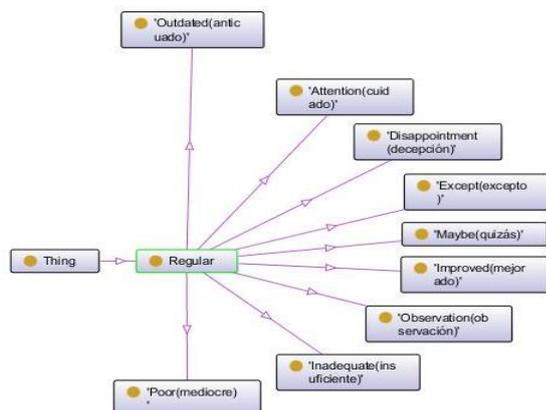


Fig. 2. Methodology used to classify the comments.

In the following, there are some of the most representative items (with their respective weight) of each ontology (Fig. 3):

Good	Fair	Bad
Excellent (5)	Improve (5)	Worse (5)
Great (5)	Insufficient (5)	Awful (5)
Good (4)	Improved (4)	Dirt (4)
Cute (3)	Unfortunately (3)	Dirty (3)
Friendly (2)	Observations (2)	Disgust (2)
Servicable (1)	Except (1)	Lack (1)

Fig. 3. Some ontological concepts according to the proposed classification.

Once we have all three ontologies, we proceeded to give a weight to each of the concepts within them. The weight is a function of the number of times that a word appears in the total feedback analyzed. For example the concept *excellent* weighs 5 unlike *cute* concept has a weight 3.

C. Comments Classification

After considering the comments of each hotel, we proceeded to analyze each of them, so that we do three summations corresponding to the sum of the concepts weight of the *good* ontology, *fair* ontology and *bad* ontology. A weight is added for each ontology concepts found within the comment. The following formula (1), explains the classification performed:

$$\sum_1^n \text{concept}(\text{weight}) \quad (1)$$

Finally we review the values of the summations and one that has a higher value, which is predominant in rating the comment. This first approximation allowed comments evaluation that resulted a bit ambiguous. But after this satisfactory results were obtained.

Once we determine the amount of good, fair and bad comments for every hotel, we took the greatest value given by (1) to say whether the hotel is *good*, *fair* or *bad*.

IV. RESULTS

This methodology was tested with individual comments in order to verify its validity. And these ones were obtained from TripAdvisor. In the following, there are some comments that were reviewed and evaluated.

I do not recommend at least to stay in this hotel, transforming your holiday in the most terrible experience thanks to the tantrums that can one go to see the depressing rooms, horrible mattresses full of ants, or not being able to sit in the chairs as it hits you painting with the thousand of dandruff, smelly sheets as well as their pillows, it is the worst hotel I have ever been, I do not recommend it at all.

This comment highlights to the naked eye the fact of being dissatisfied with the service provided by the hotel. Evaluating this comment, we obtained that the number of items classified as bad is higher with respect to other remaining concepts classified in ontologies. So this comment is classified as *bad*.

More precisely, according to the formula (1) we obtain:

$$\sum_1^{46} \text{concept}(\text{weight}) = 34$$

$$\sum_1^{28} \text{concept}(\text{weight}) = 7$$

$$\sum_1^{28} \text{concept}(\text{weight}) = 7$$

Considering the value of the summations yields the highest value, which is the sum of ontology bad. So the comment is classified as *bad*.

Another comment that has been analyzed and was rated good is:

*This hotel has the amenities you need to spend with the family without having to ask for more, rooms have **modern** decor, besides that the service is **very good** they have a restaurant with **delicious** food and bar, it overlooks the boardwalk **very cute**.*

Once again in this comment, more concepts classified as good are repeated, and to read the text carefully note that this comment does not disqualify the hotel at any time.

$$\sum_1^{46} \text{concept}(\text{weight}) = 0$$

$$\sum_1^{28} \text{concept}(\text{weight}) = 4$$

$$\sum_1^{68} \text{concept}(\text{weight}) = 17$$

Considering the value of the summations, the value of the sum of the ontology good is the predominant to classify the comment.

To determine the recommendation of the hotel, we considered the amount of *good*, *fair* and *bad* comments. The greater is the amount of comments bigger is its influence when classifying the hotel.

The system functionality is shown in Fig. 4.

On the screen you select the hotel that you want to see the analysis of the comments. In order to be more understandable, the number of comments classified as good, fair and bad is shown. Finally, the map showing the location of the hotel and the result of the process of the comments evaluation is displayed.

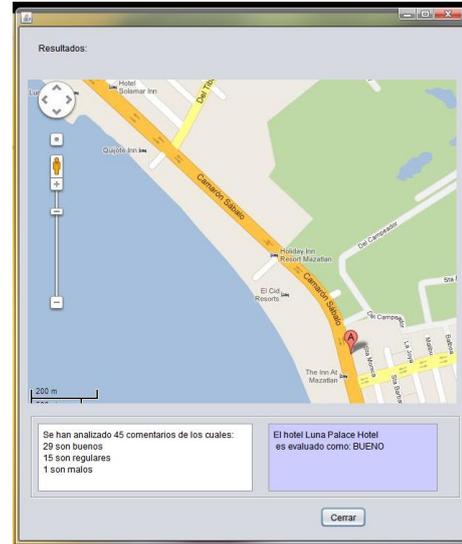
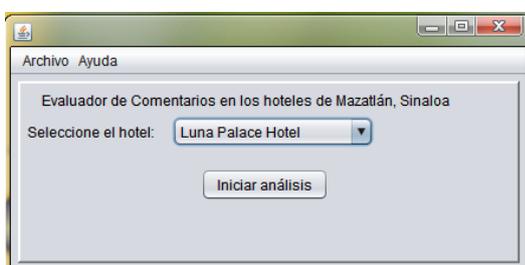


Fig. 4. Simplified system interfaces.

V. CONCLUSIONS AND FUTURE WORK

The Web has become a source of information which people uses searching for travel information. However most of the time that information is expressed in the form of comments, which results in a waste of time the user spent to read the information and analyze it. This paper presents an approach to computer analysis of the comments found on a website like TripAdvisor in order to obtain automatic classification of a hotel. We propose the use of ontologies to assist the analysis of these comments. The purpose is to prevent the vote as a parameter of recommendation and use collective intelligence embedded into these comments.

The major challenge encountered is the accuracy of the results, because sometimes the comments are confusing. An important factor is that the Spanish language allows ambiguity, by having lots of synonyms. However with data statistical analysis, using semantic similarity technique is possible to have more relied on results. And this represents future work.

ACKNOWLEDGMENT

The authors wish to thank to National Institute Polytechnic (IPN), for the facilities provided.

REFERENCES

- [1] P. Cech and V. Bures, "Advanced technologies in e-tourism," in *Proc. the 9th WSEAS International Conference on Applied Computer Science*, 2009, pp. 85-92.
- [2] V. Waralak, "Learning semantic web from e-tourism," in *Proc. the 2nd KES International Conference on Agent and Multi-agent Systems: Technologies and Applications*, 2008, pp. 516-525.
- [3] C. Paola, L. Serguei, and M. Miguel, "La web social y la inteligencia colectiva: un enfoque general," in *Proc. Congreso de Computación y Electronica*, Zacatecas, México, 2011.
- [4] Secretaria de Turismo. (2010). Sección de publicaciones. [Online]. Available: http://www.sectur.gob.mx/es/sectur/sect_704_publicaciones.
- [5] C. Paola, L. Serguei, and V. Karina, "Usando el perfil del turista para apoyar el e-tourism," in *Proc. Congreso Internacional Tendencias Tecnológicas en Computación*, 2012, México.
- [6] C. Bosangit, S. McCabe, S. Hibbert, "What is told in travel blogs? Exploring travel blogs for consumer narrative analysis," *Information and Communication Technologies in Tourism*, 2009, pp. 61-71.
- [7] D. Zhao and M. Rosson, "How and why people twitter: the role that micro-blogging plays in informal communication at work," in *Proc.*

the ACM 2009 International Conference on Supporting Group Work, 2009, pp. 243-252.

- [8] F. Abel, O. Gao, G. Houben, and K. Tao, "Semantic enrichment of twitter posts for user profile construction on the social web," in *Proc. The 8th Extended Semantic Web Conference on the Semantic Web: Research and Applications*, 2011, vol. part II, pp. 375-389.
- [9] J. Gramajo. (2008). Sistemas de gestión de conocimiento basados en ontologías aplicados a sectores de desarrollo social y económico. [Online]. Available: <http://ontoguate.wordpress.com/>
- [10] K. Siorpaes. (2004). On tour system design. [Online]. Available: <http://e-tourism.derl.at/ont/docu2004/OnTour%20-%20System%20Design.pdf>
- [11] R. Tolksdorf and R. Eckstein, *Tourism Standards*, XML Clearinghouse Report 12.
- [12] TripAdvisor. El sitio de viajes más grande del mundo. [Online]. Available: <http://www.tripadvisor.com.mx/>
- [13] Expedia. Travel site web. [Online]. Available: <http://www.expedia.mx/>
- [14] C. Paola and S. Levachkine. "Hacia el Turismo personalizado en México: retos y oportunidades," in *Proc. XIII Congreso Nacional y VII Internacional de Investigación Turística*, 2011, México.
- [15] S. Abrol and L. Ikhan, "TWinner: understanding news queries with geo-content using Twitter," presented at the 6th Workshop on Geographic Information Retrieval.
- [16] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: A content-based approach to geo-locating twitter users," in *Proc. the 19th ACM International Conference on Information and Knowledge Management*, 2010, pp. 759-768.
- [17] S. Wakamiya, R. Lee, and K. Sumiya, "Urban area characterization based on semantics of crowd activities in Twitter," in *Proc. the 4th International Conference on GeoSpatial Semantics*, 2011, pp. 108-123.
- [18] M. Opuszko, "Using semantic web technologies for classification analysis in social networks," *Forschungsberichte des Instituts für Wirtschaftsinformatik der Universität Leipzig, Interuniversitäres Doktorandenseminar Wirtschaftsinformatik der Universitäten Chemnitz, Dresden, Freiberg, Halle-Wittenberg, Jena und Leipzig*, 2011, pp. 11-21.
- [19] S. Kazutaka, I. Shunsuke, and E. Tsutomu, "On-site Likelihood Identification of Tweets for Tourism Information Analysis," in *Proc. the 2012 International Conference on Advanced Applied Informatics*, 2012, pp. 117-122.

- [20] Alexa the web information company. [Online]. Available: http://www.alexa.com/search?q=tripadvisor&r=topsites_index&p=big_top Accedido el 26 de Septiembre de 2012



Paola Cortez has received the master degree in computer science from Computer Research Center of the National Polytechnic Institute (IPN). Since 2007, she is a research professor of UPIITA-IPN. Her research area is the social network analysis, expert systems, user profiling analysis and collective intelligence.



Serguei Levachkine has received the Ph.D. in applied and computational mathematics. Actually he is a research professor in Computer Research Center, IPN. His current research efforts are mainly in artificial intelligence, cognitive science, human-computer interaction, information retrieval and integration, intelligent geographic information systems (GIS), image processing, pattern recognition, semantic web, text processing. He is a SNI member.



Carlos de la Cruz has a master degree in computer science with specialization in Software Engineering from the National Center of Research and Technological Development located in the city of Cuernavaca, Morelos in Mexico. Since 1998, he has been worked in Professional Interdisciplinary Unit in Engineering and Advanced Technologies (UPIITA) of National Polytechnic Institute (IPN). His principal activity is research professor at the Academy of

Telematic, who teaches and directs research and technological development projects. These investigations derived titling projects so that students can get the engineering degree