Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 24 (2017) 1–8

Transportation Research Procedia

www.elsevier.com/locate/procedia

3rd Conference on Sustainable Urban Mobility, 3rd CSUM 2016, 26 – 27 May 2016, Volos, Greece

Sustainability analysis on Urban Mobility based on Social Media content

Ainhoa Serna[a*], Jon Kepa Gerrikagoitia[b], Unai Bernabé[a], Tomás Ruiz[c]

[a]*Faculty of Engineering, Information Systems Department, Mondragon Unibertsitatea, Loramendi, 4, Arrasate, 20500, Spain*
[b]*IK4-IDEKO, Intelligent Software department, Arriaga Kalea, 2, 20870 Elgoibar, Spain*
[c]*Transport Department, School of Civil Engineering, Universitat Politècnica de Valencia, Camino de Vera s/n, 46022, Valencia, Spain.*

Abstract

Urban transport became an important element in the promotion of strategies towards sustainability, in fact one of the challenges posed by booming urban populations is the question of mobility. Traditional travel survey methods used to study urban mobility are very expensive, and the data collected are of poor quality. This is mainly explained because of the difficulty of getting a representative sample of the population, and the lack of motivated participants. Therefore, travel surveys are carried out less and less frequently, and the result is that good travel data is not available to develop mobility and travel behaviour studies. Information and Communication Technologies (ICT) offer the opportunity to improve traditional travel survey methods, decreasing bias in the data, reducing respondent burden, and increasing data quality. On the other hand, nowadays the User Generated Content (UGC) is growing very fast in Internet. Social media have become a valuable source for knowledge but there is a big gap in the automatic Sentiment Analysis with Semantic taxonomy annotation of online textual content. The aim of this research is to identify sustainability issues related to urban mobility based in the perceptions and experiences that underlie in the UGC. The methodology follows a quantitative and qualitative content analysis using Sentiment Analysis techniques. This paper demonstrates empirically the feasibility of the automatic identification of the Sustainable Urban Mobility problems in the discourses generated by the UGC, through a powerful ad-hoc software combining Natural Language Processing and Sentiment Analysis field tools. The main contribution of this work is the development of a tool and methodology on sustainability analysis on urban environment. Our approach enriches the data of the traditional surveys, extends traditional analysis with Big-Data methods, using data mining algorithms and Natural Language Processing techniques to extract urban mobility information from Social Media data. These data include important information about activities and travels, and can help to improve our understanding of urban mobility.

Peer-review under responsibility of the organizing committee of the 3rd CSUM 2016.

Keywords: Urban mobility; Transportation; Social Media; Sustainable Urban Transport; Sentiment Analysis; Natural Language Processing;

*Corresponding author. Tel.: +(34) 647504056. E-mail address: aserna@mondragon.edu

1. Introduction

Cities are drivers of economic development and growth, in fact about 85% of EU's GDP is generated in urban areas. Moreover, 40% of all CO2 emissions and 70% of emissions of other pollutants are caused by urban traffic. As stated by the European Commission (EC) Action Plan on urban mobility, cities are a critical part of the transport system as

2352-1465
Peer-review under responsibility of the organizing committee of the 3rd CSUM 2016.
10.1016/j.trpro.2017.05.059

more than 72% of Europeans live in an urban area (UN, 2007) and this percentage has a tendency to increase. This is acknowledged in the Sustainable Urban Transport Plans report[1] (2010).

According to J. B. S. O de Andrade Guerra et al. (2015), one estimate suggests that by 2055, 75% of the world population will be living in urban areas. Currently, 50% of the world population lives in large cities (de Oliveira et al., 2013), and cities generate 70% of global GDP (Global City Indicators, 2012). Viewing climate change forecast, cities will have to take steps to avoid economic and environmental degradation.

Primary concern of EU citizens is Urban Mobility. In a survey conducted by Eurobarometer[2] in July 2007, the 90% of Europeans said that the traffic situation in their area should be improved (Eurobarometer, 2007).

On the other hand, the opinions and experiences are central to almost all human activities and are key factors in influencing our behaviour. Our beliefs and perceptions of reality, and the decisions we make, are largely conditioned by how others see and evaluate the world. For this reason, when we have to make a decision often seek the opinions of others (Liu, 2012). The user-generated content UGC (User Generated Content), and in particular, the online comments have allowed substantial changes in the dynamic of entire sectors. Furthermore, the number of topics related to mobility is relevant on Traveller social networks like Minube, TripAdvisor, since social media is a great tool for communication and meeting point. Therefore, the analysis of that type of information can be very important for travel behaviour analysis.

Information and Communication Technologies (ICT) offer the opportunity to improve traditional survey methods to collect travel behaviour data, decreasing bias in the data, reducing respondent burden, and increasing data quality. Nowadays the User Generated Content (UGC) is growing very fast in Internet. Social media have become a valuable source for knowledge but there is a big gap in the automatic Sentiment Analysis with Semantic taxonomy annotation of online textual content.

The aim of this research is to identify sustainability issues related to urban mobility based in the perceptions and experiences that underlie in the UGC. The methodology follows a quantitative and qualitative content analysis using Sentiment Analysis techniques. This paper demonstrates empirically the feasibility of the automatic identification of the Sustainable Urban Mobility problems in the discourses generated by the UGC, through a powerful ad-hoc software combining Natural Language Processing and Sentiment Analysis field tools.

The main contribution of this work is the development of a tool and methodology on sustainability analysis on urban environment. Our approach enriches the data of the traditional surveys, extends traditional analysis with Big-Data methods, using data mining algorithms and Natural Language Processing techniques to extract urban mobility information from Social Media data.

The structure of the paper is as follows: section 2 presents an overview of research in the field of sustainable urban mobility, as well as the contributions of this article. After that, section 3 explains the methodology followed in this study. Section 4 outlines the case study, section 5 shows the results, and in the last section, conclusions and future lines are explained.

2. Related Work

It is noteworthy that are fairly recent investigations of Social Media analysis in the field of urban transport to explore from another perspective urban mobility. Lately, substantial increase in research in this area is appreciated.

Focusing on sustainable urban transport, Silvia Gabrielli et al, (2014) identified several challenges related to motivating change towards sustainable urban mobility. They reviewed the literature to support a larger adoption of sustainable mobility choices (e.g., use of public transport services) by urban travellers, several forms of persuasive solutions have been proposed (Dourish, 2010; Fogg, 1998, 2003). Also, some critical analyses of the key assumptions and limitations of the persuasive sustainability systems developed so far are explained (Brynjarsdottir et al., 2012; DiSalvo, Sengers, & Brynjarsdóttir, 2010; Dourish, 2010; Brynjarsdottir et al., 2012).

[1] http://www.europarl.europa.eu/studies

[2] http://ec.europa.eu/public_opinion/flash/fl_206b_en.pdf

Besides, recent investigations include Social media analysis in the study of urban transport. The use of on-line social networks for conducting transport surveys is presented in the paper "Use of social media for transport data collection" by Dimitrius Efthymioua and Constantinos Antoniou (2012).

Grant-Muller (2015) identifies the main requirements for a social media information harvesting methodology in the transport context and highlights the challenges involved. The work demonstrates that information harvested from social media can complement, enrich (or even replace) traditional data collection. Whilst further research is needed to develop automatic or semi-automatic methodologies for harvesting and analysing transport-related social media information, new skills are also needed in the sector to maximise the benefits of this new information source.

Gal-Tzur et al, (2014) explores two sides to engagement with social media – firstly the potential uses of social media by transport service suppliers and secondly the potential value to policy development of shared transport related information by the public. Samples of stakeholders of different sizes were selected for the study, reflecting alternative modes and providing a variety of services and functions in the transport sector. The findings give insights into the practices of organizations of different size, function and longevity of social media use. The early results of a study to harvest freely available transport information from the public and transport system users are also presented, demonstrating that transport policy relevant information can be harvested from online social media sources.

Furthermore, Susan Bregman in the book "Uses of Social media in Public Transportation" (2012) explores the use of social media among transit agencies and documents successful practices in the United States and Canada. A review of relevant literature was combined with findings from a survey of selected transit agencies. Based on survey results, several case examples were developed to describe innovative and successful practices in more detail.

Grant-Muller's contribution (2014) describes technical challenges related to mining social media data within the transport context, laying the foundation for further research in this field. Following an overview of the text mining process to extract relevant information from the corpus, a review of the challenges this approach holds for the transport sector is given. These include ontologies, sentiment analysis, location names and measuring accuracy. Finally, institutional issues associated to the increasing use of social media are highlighted, concluding that social media information has not yet been fully explored.

Serna et al (2014, 2015) and Serna, Gerrikagoitia & Bernabé (2016a, 2016b) with their investigation contribute to the auto-coding analysis of different Social Media sources through ontologies, and several lexical databases resources. Moreover, they created a model that enables the categorization among other things of the infrastructure that included transport mode. Although the focus of his research was the domain of tourism.

3. Methodology

The methodology of this research is a quantitative and qualitative content analysis following the Walle (1997) approach. As the purpose of this research is to have a general overview of perceived sentiments, large quantities of data are analysed. For the specific issue of transport mode classifications, apart from quantitative analyses, a qualitative approach is necessary to understand their deeper meanings and likely interpretations by individuals (Mariné Roig, 2013).

The methodology used consists of four main phases: data acquisition; data preparation for analysis; data curation and data storage (Serna et al, 2016).

3.1. Data Acquisition

It is the process of gathering, filtering and cleaning the unstructured data before making them persistent in a storage solution on which data analysis can be carried out. In this step the information is extracted either through API's ("Application Programming Interface") from on-line travel social network Minube. Minube provides an API allowing searches by territory and city. Minube is a travel experience community with over 1,684,399 users. The comments refer to experiences that include information about mobility as main category. Furthermore, concepts like buildings (museums, cathedrals), locations accommodation, entertainment, sports, food (restaurants), shops, museums…are also taken into account.

3.2. Data Preparation for Analysis

This process is concerned with making the raw data amenable to use in decision-making. The data analysis module has several steps: first, the comments are loaded one by one and after that, the texts are corrected using Aspell, a spell checker that is customized with localism and abbreviation. Once the text is corrected, each word is morphosyntactically noted, using Freeling (Padró & Stanilovsky, 2012). After that, common nouns and adjectives are considered and ordered by number of occurrences.

3.3. Data Curation

This phase is the active management of data over its life-cycle to ensure that it meets the necessary data quality requirements for its effective usage. In addition, categories are identified and a sentiment analysis process is carried out. For this purpose, WordNet aligned with SUMO ontology, and ad-hoc software is used. WordNet is a lexical database that relates hyponyms/ hypernyms with sets of synonyms called synsets, which can be interpreted as specialization relations between conceptual categories. Among all identified categories, only those related to the transport mode corresponding to the category *Transportation*, *Walking* and *TransportationDevice* are selected. Sentiment analysis is done at different levels: word level, sentence level, comment level and capital city level, thereby a general overview and a detail level with greater granularity.

3.4. Data Storage

This process is the responsible of storing and managing data in a scalable way satisfying the needs of the applications that require access to the data. The downloaded comments after the data curation phase are saved in XML format in a native XML database called Apache Solr. One of the most important features of this open source database is that it is optimized for indexing large volumes of data in real time.

4. Case Study

In the case study, data from Minube Traveller Social Network has been analysed. In this way, data from Minube referring to Capital cities of all provinces in Spain have been selected and downloaded. In total 43,251 comments written in Spanish language corresponding to 52 capital cities.

For each capital city two different type of analysis have been done. The first type of analysis extracts nouns from the comments, obtains the number of mentions and the category for each noun, adding the polarity analysis, as we can observe in Table 1 and Table 2.

For the second analysis, a particular treatment is made for the nouns identified as “Transportation”, “Walking” or “TransportationDevice”. Their related attributes are obtained in order to get detailed information about them. Some examples of comments for Bilbao, Valencia and Madrid cities are shown (positive attributes in green and negative ones in red).

Bilbao city (810 comments)

For me something more than just a means of transport. A new underground, care, different. And with a very modern design. Full of light and symmetries.

Valencia city (1,854 comments)

A spectacular festival ..... To live them day and night on the street. Light, colourful, firecrackers and gunpowder in every corner of the city. The Fallas monuments are amazing ..., at least the ***underground*** *works night and day.*

Madrid city (8,597 comments)

For me begging seems a problem. In the underground there is little security and too many freeloaders.

5. Results

After analysing the 52 capital cities' comments mentioned above, Tables 1 and 2 show the results of Bilbao and Valencia cities respectively. The first column reflects the most relevant words. Note that the two first concepts are not about Transport, but they had been maintained to show the whole process and how the ontology categorizes every word contained in the comments (*Building, GeopoliticalArea, Region*…). The second column shows the associated sentiment, the third one the number of occurrences of the concept and the last one represents the category.

Table 1. Bilbao city Transportation Modes.

Concept	Positive	Mentions	Ontology Category
restaurant	75%	260	[Building]
city	72%	251	[GeopoliticalArea]
underground	90%	72	[TransportationDevice]
car	90%	34	[TransportationDevice]
bus	100%	33	[TransportationDevice]
train	100%	19	[TransportationDevice]
boat	72%	13	[TransportationDevice]
elevator	71%	13	[TransportationDevice]
walking	100%	11	[Walking]
funicular	75%	10	[TransportationDevice]
bicycle	100%	8	[TransportationDevice]
vessel	52%	5	[TransportationDevice]
rowing boat	60%	3	[TransportationDevice]
taxi	50%	3	[TransportationDevice]

Table 2. Valencia city Transportation Modes.

Concept	Positive	Mentions	Ontology Category
restaurant	72%	683	[GeopoliticalArea]
city	74%	638	[Building]
underground	50%	145	[TransportationDevice]
bus	100%	46	[TransportationDevice]
rowing boat	71%	34	[TransportationDevice]
walking	95%	33	[Walking]
car	40%	25	[TransportationDevice]
train	90%	24	[TransportationDevice]
plane	55%	19	[TransportationDevice]
boat	69%	18	[TransportationDevice]
bicycle	100%	17	[TransportationDevice]
cruise	58%	8	[Transportation]
railway	64%	6	[TransportationDevice]
taxi	62%	6	[TransportationDevice]
catamaran	83%	4	[TransportationDevice]
carriage	44%	3	[TransportationDevice]
sailboat	83%	1	[TransportationDevice]
canoe	56%	1	[TransportationDevice]

As can be observed in the results, the most commented mode of transport in Madrid, Bilbao and Valencia is the *underground* with 460, 72 and 145 mentions, respectively. In Valencia the second most commented mode of transport is *bus* with 46 mentions, then *rowing boat* with 34 mentions, followed by *walking* with 33 mentions.

For both Madrid and Bilbao, the first 4 modes of transport are the same: *underground*, *car*, *bus* and *train*.

Respecting to type of transport, as Bilbao and Valencia are coastal cities, sea transportation modes have presence. In Bilbao *boat, vessel* and *rowing boat* nouns are mentioned and in Valencia *rowing boat, boat, cruise, catamaran, sailboat* and *canoe*.

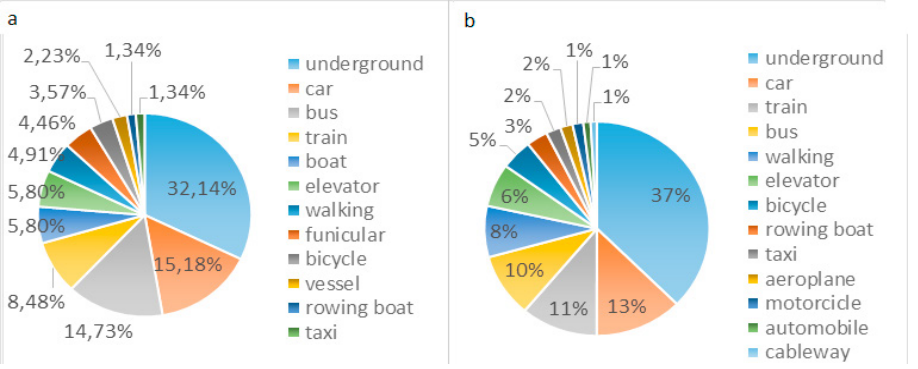

Figure 1. (a) Bilbao Transportation Modes; (b) Madrid Transportation Modes

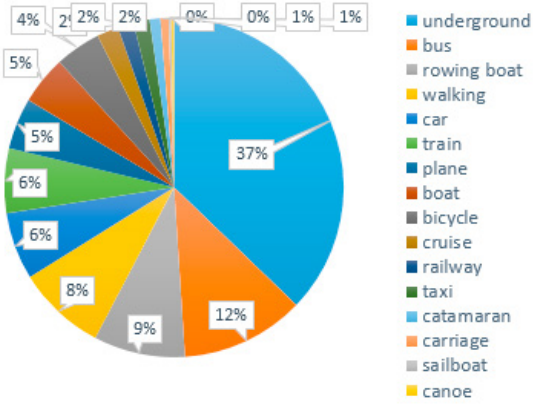

Figure 2. Valencia Transportation Modes

Figures 1 and 2 show the percentage of mentions of the different transport modes in Bilbao, Madrid and Valencia respectively. In Bilbao the 70% of the comments are about 4 transport modes: underground 32%, car 15%, bus 15% and train 8%. 5% of the comments have to do with walking mobility.

In Madrid 71% of the comments represent 4 transport modes and in Valencia 66%, but noteworthy that walking appears in fifth place and has a representative value with 8%. In Valencia there is a bigger variety of transport modes mentioned than in the other two cities, 16 in total; 6 maritime transport, 1 air transport and 9 land transport.

In Bilbao there are eight land and three maritime transport modes. Air transport is not mention at all. However, in Madrid there are 56 references to bike and 22 to plane.

The best valued transport modes in Bilbao are bus and train with 100% of positivity followed by underground and car with 90 % of positivity.

In Valencia is *bus* with 100% positivity followed by *train* 90%. The worst valued transport modes are *underground* and *car* with 50% and 40% of positivity, respectively. In Madrid, the transport mode with more positive feedback is *bus* with 95% of positivity, followed by *underground* with 60% of positivity. The worst rated are *train* and *car* with 50% of positivity.

It is noteworthy that the positive attributes that describe *bus* and *train* in Bilbao are: *great coverage, new, ease of mobility, magnificent, well connected, thrilling*. The positive feedback describing car are: *recommended, easily accessible*. The attribute related to negative perceptions is: *too many cars*. Positive attributes associated to *underground* are: *easy to use, fast, attractive*. Negative attribute related to underground is *price*.

In Valencia the positive attributes associated to bus are: well connected, many bus lines. The positive attributes about *train* are *modern*, *practical*. However, the negative aspect is *price*. The attributes that are positive about

underground are: *works day and night, good access, easy to move, good*. The negative attribute is: *away*. The positive aspects about car are: *practical, visible, wide*. By contrast the negatives are: *not recommended, attention, shitty*.

In Madrid the positive attributes associated to *bus* are: *numerous*, *cheap*, *well connected*. The negative aspects are: *strange people, endless queues*. The positive aspects about *underground* are: *improved a lot, good quality, restored*. The negative aspects: *poor security, dangerous, beggar*. The positive aspects about *train* and *car* are: *frequency, strategic position, accessible, comfortable* and the negatives: *inconvenient*, *not recommendable, delicate, expensive, uncomfortable, abandoned,* and *dirty*.

6. Conclusions

The majority of the literature about Social Media use in Urban Transport appears from 2012 onwards. This explosion reflects the increasing interest and maybe relevance of the topic in urban transport research.

Underground, along with other rail public transport systems (train and tram) are the most efficient transport modes in terms of energy efficiency and environmental impact. The main advantage of the underground with regard to public transport modes that share their path with traffic (such as bus) is that travel time and regularity does not depend on traffic, nor they are influenced by congestion. To guarantee the right to universal mobility of citizens is necessary to have a public transport network that allows access to different spaces and public facilities, as well as industrial areas, centres of economic activity and entertainment.

Besides, the bicycle is considered an important transport mode to bring sustainability within unsustainable daily routines of short urban journeys. But in the research results described in this paper, the presence of this transport mode is almost inexistent. This fact may encourage policy makers to outline future mobility and infrastructure agendas.

Making urban mobility more inclusive, accessible, efficient and environmentally friendly is challenging. The development of urban transport will evolve based on many factors, such us technological progress, socioeconomic and demographic changes, and environmental developments. In this context, once the method and results have been assessed, the presented approach can be used by public bodies as a valid strategic innovation tool for decision making because it is the responsibility of government to respond to this collective need, as the demand for mobility of the population is closely related to urban policies and territory planning.

Furthermore, as a future work, additional data sources directly related to transport may extend the scope of the tourism experiences on-line social network used in the current research.

Acknowledgements

This research has been supported by:

Cost Action TU1305 Social Networks and Travel Behaviour. COST is supported by the EU Framework Programme Horizon 2020.

Spain R+D programs: Ministry of Economy and competitiveness. State Program of Research, Development and Innovation Facing the Challenges of the Company under the State Plan of Scientific and Technical Research and Innovation. Spanish government MINERVA- Innovative Travel Data Collection Methods for Transport Planning coordinated by Universitat Politècnica de Valencia and Mondragon Unibertsitatea.

References

Bregman, S. (2012). *Uses of social media in public transportation* (Vol. 99). Transportation Research Board.

Brynjarsdottir, H., Håkansson, M., Pierce, J., Baumer, E., DiSalvo, C., & Sengers, P. (2012). Sustainably unpersuaded: How persuasion narrows our vision of sustainability. In *Proceedings of the 2012 ACM annual conference on human factors in computing systems* (pp. 947–956). ACM.

de Andrade Guerra, J. B. S. O., Pereira Ribeiro, J. M., Fernandez, F., Bailey, C., Barbosa, S. B., & da Silva Neiva, S. (2015). The adoption of strategies for sustainable cities: A comparative study between Newcastle and Florianópolis focused on urban mobility. *Journal of Cleaner Production*. doi:10.1016/j.jclepro.2015.07.135

de Oliveira, J. A. P., Doll, C. N., Balaban, O., Jiang, P., Dreyfus, M., Suwa, A., ... & Dirgahayani, P. (2013). Green economy and governance in cities: assessing good governance in key urban economic processes. *Journal of Cleaner Production*, 58, 138-152.

DiSalvo, C., Sengers, P., & Brynjarsdóttir, H. (2010). Mapping the landscape of sustainable HCI. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1975–1984). ACM.

Dourish, P. (2010). HCI and environmental sustainability: The politics of design and the design of politics. In *Proceedings of the 8th ACM conference on designing interactive systems* (pp. 1–10). ACM.

Efthymiou, D., & Antoniou, C. (2012). Use of social media for transport data collection. *Procedia-Social and Behavioral Sciences*, 48, 775-785.

Fistung, F. D., Popescu, T., & Sima, C. (2015). Interferences between Sustainable Mobility and Economic Development in Romania. *Procedia Economics and Finance*, 22, 36–44. doi:10.1016/S2212-5671(15)00224-5

Fogg, B. J. (1998). Persuasive computers: Perspectives and research directions. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 225–232). ACM Press/Addison-Wesley Publishing Co.

Fogg, B. J. (2003). Persuasive technology: Using computers to change what we think and do. San Francisco, CA, USA: Morgan Kaufmann Publishers.

Gabrielli, S., Forbes, P., Jylhä, A., Wells, S., Sirén, M., Hemminki, S., .. & Jacucci, G. (2014). Design challenges in motivating change for sustainable urban mobility. *Computers in Human Behavior*, 41, 416-423.

Gal-Tzur, A., Grant-Muller, S. M., Minkov, E., & Nocera, S. (2014). The impact of social media usage on transport policy: issues, challenges and recommendations. *Procedia-Social and Behavioral Sciences*, 111, 937-946.

Grant-Muller, S. M., Gal-Tzur, A., Minkov, E., Kuflik, T., Nocera, S., & Shoor, I. (2015). Transport Policy: Social Media and User-Generated Content in a Changing Information Paradigm. In *Social Media for Government Services* (pp. 325-366). Springer International Publishing.

Grant-Muller, S. M., Gal-Tzur, A., Minkov, E., Nocera, S., Kuflik, T., & Shoor, I. (2014). Enhancing transport data collection through social media sources: methods, challenges and opportunities for textual data. *IET Intelligent Transport Systems*, 9(4), 407-417.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.

Mariné Roig, E. (2013). From the projected to the transmitted image: The 2.0 construction of tourist destination image and identity in Catalonia.

Serna, A., Gerrikagoitia, J. K., & Alzua, A. (2014). Towards a better understanding of the cognitive destination image of Euskadi-Basque Country based on the analysis of UGC. In *Information and Communication Technologies in Tourism 2014* (pp. 395-407). Springer International Publishing.

Serna, A., Gerrikagoitia, J. K., & Bernabé, U. (2016a). Discovery and Classification of the Underlying Emotions in the User Generated Content (UGC). In *Information and Communication Technologies in Tourism 2016* (pp. 225-237). Springer International Publishing.

Serna, A., Gerrikagoitia, J. K., & Bernabé, U. (2016b). Identificación de las fortalezas y flaquezas en los puntos de interés religiosos a través del Análisis de Sentimiento. *The 10th Professional Seminar for Church Communications*. Rome, Italy.

Serna, A., Marchiori, E., Gerrikagoitia, J. K., Alzua-Sorzabal, A., & Cantoni, L. (2015). An auto-coding process for testing the cognitive-affective and conative model of destination image. In *Information and Communication Technologies in Tourism 2015* (pp. 111-123). Springer International Publishing.

Walle, A. H. (1997). Quantitative versus qualitative tourism research. *Annals of Tourism Research*, *24*(3), 524-536.