

Filtering opinions in Spanish with Topics of Tourist Interest for the Sentiment Analysis Task

José de Jesús Ceballos-Mejía^{1,*}, Martha Angélica Parra-Urías¹ and Miguel Ángel Álvarez-Carmona^{2,3}

¹*Instituto Tecnológico de Tepic (ITT), Nayarit, Mexico*

²*Centro de Investigación en Matemáticas (CIMAT), Sede Monterrey, Mexico*

³*Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT), CDMX, Mexico*

Abstract

This paper presents a proposal by the ITT Team to participate in the Rest-Mex 2023 forum, focusing on filtering opinions in Spanish with topics of tourist interest for the Sentiment Analysis task. Our objective is to conduct a sentiment analysis that is tailored to the key areas that significantly impact tourists' experiences. We have identified 11 specific topics: Activities, Friendliness, Climate, Food, Insecurity, Cleanliness, Nature, Pandemic, Prices, Transportation, and Location, which encompass a wide range of factors critical for tourists when evaluating and making decisions about their travel experiences. By leveraging topic modeling techniques, we aim to enhance the accuracy and granularity of sentiment analysis by considering these specific areas of interest. Our proposed approach combines machine learning methods and feature engineering to classify opinions as positive, negative, or neutral based on their sentiment towards these topics. The results demonstrate the effectiveness of our approach in filtering Spanish opinions related to tourist topics and provide valuable insights for tourism-related decision-making processes.

Keywords

Rest-Mex, Sentiment Analysis, Tourism, Topics, NLP

1. Introduction

The advent of social media and online platforms has transformed the way people communicate and share their opinions, experiences, and feedback [1, 2]. This shift has had a profound impact on various industries, including tourism, where customer opinions play a vital role in shaping services and improving customer satisfaction [3]. In the digital era, sentiment analysis has emerged as a powerful tool for extracting insights from the vast amount of user-generated content available [4]. However, in the context of the tourism domain, a more focused and topic-specific approach to sentiment analysis is necessary to better understand and address the needs and preferences of tourists.

In this paper, we present a proposal to participate in the Rest-Mex 2023 [5] forum as the ITT Team. We propose to filter opinions in Spanish using specific topics of interest related to tourism. Our aim is to conduct sentiment analysis that is tailored to the key areas that significantly

IberLEF 2023, September 2023, Jaén, Spain

*Corresponding author.

✉ miguel.alvarez@cimat.mx (M. Á. Álvarez-Carmona)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

impact tourists' experiences. We focus on 11 specific topics: Activities, Friendliness, Climate, Food, Insecurity, Cleanliness, Nature, Pandemic, Prices, Transportation, and Location. These topics encompass a wide range of factors that are critical for tourists when evaluating and making decisions about their travel experiences.

The study of sentiment analysis within the tourism domain is not novel, as previous research has explored general sentiment classification in this context[6, 7, 8, 9, 10]. However, limited attention has been given to topic-based sentiment analysis that specifically targets areas of interest relevant to tourists. By focusing on these specific topics, we aim to provide a more nuanced understanding of sentiment distribution and customer preferences within the Spanish tourism industry.

To achieve these objectives, we outline a methodology that encompasses data collection, preprocessing, topic classification, sentiment classification, and model evaluation.

To accomplish this, the Rest-Mex dataset is utilized [5]. In the 2023 edition, the organizers have introduced an extension to the sentiment analysis task that has been in development since 2021 [9]. The objective is to predict three aspects based on a tourist's opinion about a place:

1. The polarity of the opinion, represented as an integer value ranging from 1 to 5.
2. The type of tourist place, which can be categorized as an attraction, a hotel, or a restaurant.
3. The country where the tourist place is located, which can be Mexico, Cuba, or Colombia.

The training data for the polarity task this year comprises over 250,000 opinions, while the test data consists of more than 100,000 opinions. Typically, three separate models would be constructed using this data, one for each aspect to be predicted (polarity, type, and country). However, given the substantial size of the data collection, training and testing three distinct models could be time-consuming [6, 11, 12].

By filtering opinions in Spanish with a focus on topics of tourist interest, we aim to contribute to the advancement of sentiment analysis in the tourism domain and assist industry stakeholders in better understanding and catering to the needs and preferences of their customers.

2. Sentiment Analysis in Tourism

Sentiment analysis, also known as opinion mining, is a subfield of natural language processing (NLP) that focuses on extracting and understanding sentiments, opinions, and attitudes expressed in textual data. In the context of the tourism industry, sentiment analysis plays a crucial role in analyzing customer opinions and feedback to gain insights into their experiences, preferences, and satisfaction levels[1]. By automatically classifying sentiments as positive, negative, or neutral, sentiment analysis provides valuable information for businesses to improve their services and make data-driven decisions.

In recent years, sentiment analysis has gained significant attention in the tourism domain, as online platforms and social media have become popular platforms for travelers to share their opinions and experiences. The analysis of sentiment in tourism texts provides valuable insights into various aspects, such as destination preferences, hotel experiences, restaurant reviews, and attraction satisfaction. However, performing sentiment analysis in the tourism domain poses specific challenges that need to be addressed for accurate and meaningful results [4].

2.1. Challenges in Sentiment Analysis in Tourism

Subjectivity and Context: Tourism-related opinions are inherently subjective and context-dependent. The sentiment expressed in a review can vary based on individual preferences, cultural differences, and personal experiences. Interpreting sentiment accurately requires considering the context in which the opinions are expressed, including the specific tourist place, the reviewer's background, and the overall experience [3].

Domain-specific Language and Slang: Tourism texts often include domain-specific language, jargon, and slang. For example, travelers may use specific terms to describe activities, attractions, or local customs. Understanding and appropriately handling these linguistic nuances is crucial for accurate sentiment classification[13].

Aspect-based Analysis: Sentiment analysis in tourism should go beyond general sentiment classification and delve into aspect-based analysis. Aspects or topics of interest, such as activities, friendliness, cleanliness, and prices, have a significant impact on tourists' opinions. Analyzing sentiments at the aspect level provides deeper insights into specific areas of interest and helps identify strengths and weaknesses in tourism offerings[14].

Multilingual Analysis: The tourism industry attracts a diverse range of travelers from various countries and cultures, resulting in the need for sentiment analysis in multiple languages. Different languages may require specific language models, preprocessing techniques, and sentiment lexicons for accurate sentiment classification [15].

2.2. Methodologies for Sentiment Analysis in Tourism

Various methodologies have been employed for sentiment analysis in the tourism domain, ranging from traditional machine learning approaches to more advanced deep learning techniques. Common steps in sentiment analysis include data collection, preprocessing, feature extraction, sentiment classification, and evaluation. However, considering the challenges specific to tourism sentiment analysis, certain adaptations and techniques can be employed:

Aspect-based Sentiment Analysis: To capture sentiments related to specific topics of interest in tourism, such as activities, friendliness, climate, and prices, aspect-based sentiment analysis techniques can be employed. This involves identifying and classifying sentiments for each aspect separately, providing a more granular understanding of customer opinions.

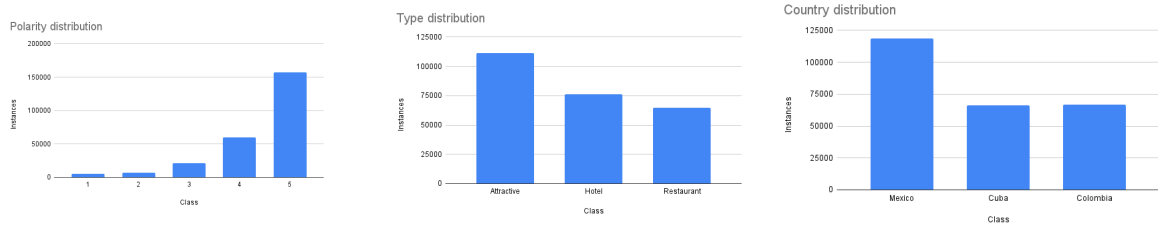
Domain-specific Lexicons: Building or utilizing domain-specific sentiment lexicons can enhance sentiment analysis accuracy in tourism. These lexicons contain domain-specific terms and their associated sentiment polarities, enabling better recognition of sentiment-bearing words and phrases in tourism texts.

Machine Learning and Deep Learning Approaches: Traditional machine learning algorithms, such as Support Vector Machines (SVM), Naïve Bayes, and Random Forests, have been widely used for sentiment classification. Additionally, deep learning techniques, including recurrent neural networks (RNNs) and transformer-based models, such as BERT[16].

3. Rest-Mex Corpus

The Rest-Mex 2023 corpus, curated by the organizers, comprises a total of 251,702 opinions collected from TripAdvisor. This dataset serves as the foundation for training and evaluating sentiment analysis models in the context of tourism.

For the task of Polarity classification, the opinions are categorized into five classes, ranging from the worst polarity (class 1) to the best polarity (class 5). Figure 1a illustrates the distribution of these classes, revealing a clear imbalance within the dataset.



(a) Distribution of Polarity

(b) Distribution of Country

(c) Distribution of Country

Figure 1: Traits distributions for the Rest-Mex 2023

To determine the Type of place, the opinions are classified into three categories: Attractive, Hotel, and Restaurant. Figure 1b presents the distribution of these classes. While there is not as pronounced an imbalance as in the Polarity classification, some variations can still be observed.

Lastly, the opinions are labeled with the Country of origin of the place the tourist visited, resulting in three classes: Mexico, Cuba, and Colombia. The distribution of these classes is shown in Figure 1c.

From the distributions presented, it is evident that the Polarity classification exhibits the greatest imbalance among the three traits. Conversely, the Type classification demonstrates a relatively more balanced distribution, while the Country classification falls somewhere in between. These class distribution variations must be considered when developing and evaluating sentiment analysis models using the Rest-Mex corpus.

4. Proposal

The hypothesis of this work is that by using only opinions where words from certain themes appear, it is possible to classify the features of polarity, type, and country with competitive results.

To test this hypothesis we propose 3 phases. First, define the topics of interest, second, extract representative words from each topic, and third, propose an algorithm that filters opinions from the corpus.

Each of the 3 phases is described below.

4.1. Important topics for tourism

Tourism is a thriving industry that attracts millions of visitors each year. When choosing a travel destination, tourists consider various factors to ensure an enjoyable and memorable experience. The following topics play a crucial role in determining the appeal and success of a tourist destination.

4.1.1. Activities

Engaging activities and attractions are vital for tourism. Tourists seek destinations that offer a diverse range of recreational options, such as adventure sports, cultural events, historical landmarks, museums, and natural wonders. The availability of exciting activities enhances the overall experience and creates lasting memories for travelers.

4.1.2. Friendliness

Friendliness and hospitality of the local population significantly impact the tourism industry. Travelers often prefer destinations where they feel welcomed and can interact with friendly locals. Warm and welcoming communities create a positive environment, fostering cultural exchange and leaving tourists with a sense of belonging.

4.1.3. Climate

Climate plays a pivotal role in choosing a travel destination. Different individuals have varied preferences, but favorable weather conditions are generally preferred. Whether it's a tropical paradise, a winter wonderland, or a moderate climate, the suitability of the weather for outdoor activities greatly influences tourists' decisions.

4.1.4. Food

Food is an integral part of the tourism experience. Culinary delights and local cuisine are major attractions for travelers. The availability of diverse food options, ranging from street food to fine dining experiences, allows tourists to explore the unique flavors and culinary traditions of a particular region.

4.1.5. Insecurity

Tourist destinations must prioritize safety and security. Travelers are more likely to choose locations with low crime rates and effective security measures. Providing a sense of safety and peace of mind allows tourists to relax and enjoy their visit without concerns about personal well-being.

4.1.6. Cleanliness

Cleanliness and hygiene are critical aspects that influence tourists' perceptions of a destination. Maintaining cleanliness in public spaces, accommodations, and tourist attractions creates a

positive image and promotes visitor satisfaction. Well-maintained environments contribute to a pleasant and enjoyable stay.

4.1.7. Nature

The natural beauty and conservation efforts of a destination attract eco-tourists and nature enthusiasts. Scenic landscapes, national parks, wildlife reserves, and opportunities for outdoor activities like hiking, birdwatching, and nature photography contribute to the overall appeal of a tourist destination.

4.1.8. Pandemic

In light of the ongoing global pandemic, tourists prioritize destinations that prioritize health and safety measures. Transparent communication about vaccination requirements, testing protocols, and adherence to public health guidelines instills confidence in travelers and ensures their well-being during their visit.

4.1.9. Prices

Affordability is a significant consideration for tourists. Reasonable prices for accommodations, transportation, attractions, and dining options make a destination more accessible and appealing to a broader range of travelers. Competitive pricing ensures a competitive edge in attracting visitors.

4.1.10. Transportation

Efficient transportation systems are crucial for the success of tourism. Easy accessibility and connectivity within a destination, as well as convenient modes of transportation, including airports, public transit, and reliable road networks, enhance the overall travel experience and make it more convenient for tourists to explore different attractions.

4.1.11. Location

The geographical location of a tourist destination plays a vital role. Proximity to other attractions, accessibility from major cities or transportation hubs, and unique geographical features all contribute to the attractiveness of a location. A desirable location can make a destination more appealing and increase its competitiveness in the tourism market.

4.2. Found important words

With the important topics definition, the question is how we can find descriptive words for each topic. For this, we propose to use similar words to the name of each topic. We use the online system named RelatedWords. This system receives as input a word and returns a set of words related to the input. For this work we use the top 20 words related to each topic.

4.3. Filtering opinions

To filter the opinions, we will only use those with a number greater than k words related to the topics.

For that we propose the algorithm ???. In this algorithm, it can be seen that we pass a different parameter k for each polarity class (between 1 and 5). The function c returns the polarity class of an opinion i .

Algorithm 1 Filtering opinions

```
Input Input Output Output Filtering opinions Filtering opinions
opinions, topics, k opinions Filtered
words = []
for t in topics: words += topics[t]
opinionesFiltered = []
for i in range(len(opinions)):
    text = opinions[i]
    if len(set(text.lower().split()) ∩ words) > k[c(opinions[i])]:
        opinionesFiltered.append(text)
opinionesFiltered
```

4.4. Text classification based on Beto

The Beto classifier is a state-of-the-art natural language processing (NLP) model developed by the research team at the University of Chile. It is specifically designed for processing text in the Spanish language. Beto is built upon the Transformer architecture, which has been widely successful in various NLP tasks.

The Transformer architecture, originally introduced in the "Attention is All You Need" paper by Vaswani et al., revolutionized NLP by replacing recurrent neural networks (RNNs) with attention mechanisms. Transformers excel in capturing long-range dependencies in sequences, making them highly effective in language modeling and other text-related tasks.

The Beto model, like its English counterpart BERT (Bidirectional Encoder Representations from Transformers), is a variant of the Transformer model. It utilizes a bidirectional architecture that allows the model to capture both the left and right context of each word or token in a sentence. This enables the model to understand the meaning and context of words based on their surrounding words, greatly improving its ability to grasp the semantics of the text.

Beto is pre-trained on a massive amount of Spanish text data, typically using a masked language modeling objective. During pre-training, the model learns to predict missing words in a sentence based on the surrounding context. This process helps the model develop a deep understanding of the Spanish language.

After pre-training, Beto can be fine-tuned for specific downstream tasks such as text classification, named entity recognition, sentiment analysis, and more. Fine-tuning involves training the model on a smaller, task-specific dataset to adapt it to a particular NLP task. By fine-tuning Beto on a specific task, it can leverage its pre-trained knowledge to achieve high performance and accuracy in that task.

The Beto classifier has gained popularity and achieved excellent results in various Spanish NLP benchmarks and competitions. Its ability to capture context, semantics, and syntactic information makes it a powerful tool for analyzing and understanding Spanish text, opening up numerous possibilities for applications in areas such as sentiment analysis, document classification, and information retrieval.

Table 1 shows the features used in the implementation used for the classifier

Table 1

Parameters of the Beto classifier

Parameter	Value
Model	Beto-cased
Max length	32
Optimizer	Adam
Learning rate	$5e - 5$
Steps	$1e - 08$
Epochs	4

5. Results

To evaluate the models in the Train partition of the Rest-Mex corpus, we propose a separation of the data of 70 % for training and the rest for testing.

Table 2

Train results

k	$Polarity_{F-measure}$	$Type_{F-measure}$	$Country_{F-measure}$
[10, 10, 10, 10, 10]	0.35	0.90	0.87
[1, 1, 1, 1, 1]	0.51	0.97	0.89
[5, 5, 5, 5, 5]	0.42	0.95	0.84
[1, 2, 3, 4, 5]	0.50	0.95	0.83
[1, 1, 1, 5, 5]	0.50	0.94	0.81
[0, 0, 5, 6, 6]	0.50	0.94	0.80
[0, 0, 2, 3, 4]	0.56	0.96	0.84

Table 2 presents the results of the training process, evaluating the performance of the Beto classifier with different parameter settings, represented by the k values. The three evaluation metrics used are $Polarity_{F-measure}$, $Type_{F-measure}$, and $Country_{F-measure}$.

Observing the results, we can make the following observations:

The $Polarity_{F-measure}$ ranges from 0.35 to 0.56, indicating the classifier’s ability to correctly identify the polarity (positive, negative, or neutral) of the text. The highest value of 0.56 is achieved when using the k value [0, 0, 2, 3, 4]. That is, the more instances of a class are required to filter the opinions with more words.

The $Type_{F-measure}$ ranges from 0.90 to 0.97, indicating the classifier’s performance in correctly identifying the type of text. The highest value of 0.97 is achieved with the k value [1, 1, 1, 1, 1].

The $Country_{F-measure}$ ranges from 0.80 to 0.89, representing the classifier's ability to identify the country associated with the text. The highest value of 0.89 is achieved with the k value [1, 1, 1, 1, 1].

Overall, we can observe that different k values have an impact on the performance of the classifier for each evaluation metric. It is important to note that the choice of k values depends on the specific requirements and objectives of the task at hand. A higher $F-measure$ value indicates better performance, so selecting the appropriate k values based on the desired outcome is crucial.

5.1. Test partition results

For this edition, the organizers of Rest-Mex propose some evaluation metrics that give greater weight to correctly classify the negative classes of polarity.

To assess the effectiveness of the polarity classifier, the organizers propose the equation 1. This metric gives the additive inverse of importance according to the percentage of instances of a class in the test collection.

$$Res_p(k) = \frac{\sum_{i=1}^{|C|} ((1 - \frac{T_{C_i}}{T_C}) * F_i(k))}{\sum_{i=1}^{|C|} 1 - \frac{T_{C_i}}{T_C}} \quad (1)$$

Finally, to obtain a unique value per participant, they propose a combination of the results as indicated by the equation 2. It is important to mention that in the same way, greater weight is given to the result of polarity than to the other two traits.

$$Sentiment(k) = \frac{2 * Res_p(k) + F_T(k) + F_C(k)}{4} \quad (2)$$

Table 3 shows the results obtained in the Rest-Mex 2023.

Table 3
Rest-Mex 2023 results

k	Sentiment(k)	$Polarity_{F-measure}$	$Type_{F-measure}$	$Country_{F-measure}$
[10, 10, 10, 10, 10]	0,40	0,27	0,66	0,55
[1, 1, 1, 1, 1]	0,68	0,48	0,96	0,87
[5, 5, 5, 5, 5]	0,60	0,43	0,90	0,72
[1, 2, 3, 4, 5]	0,56	0,34	0,90	0,76
[1, 1, 1, 5, 5]	0,58	0,32	0,94	0,80
[0, 0, 5, 6, 6]	0,51	0,21	0,95	0,77
[0, 0, 2, 3, 4]	0,56	0,28	0,95	0,82

Observing the results, we can make the following observations:

The $Sentiment(k)$ column shows the sentiment score associated with each k value. The scores range from 0.40 to 0.68, indicating the overall sentiment captured by the classifier for each parameter setting. Higher scores generally indicate more positive sentiment.

The $Polarity_{F-measure}$ ranges from 0.27 to 0.48, indicating the classifier's ability to correctly identify the polarity (positive, negative, or neutral) of the text. The highest value of 0.48 is achieved with the k value [1, 1, 1, 1, 1].

The $Type_{F-measure}$ ranges from 0.66 to 0.96, representing the classifier's performance in correctly identifying the type of text. The highest value of 0.96 is achieved with the k value [1, 1, 1, 1, 1].

The $Country_{F-measure}$ ranges from 0.55 to 0.87, indicating the classifier's ability to identify the country associated with the text. The highest value of 0.87 is achieved with the k value [1, 1, 1, 1, 1].

6. Conclusions

In this paper, we have presented a proposal to participate in the Rest-Mex 2023 forum as the ITT Team, with a focus on filtering opinions in Spanish using specific topics of interest related to tourism. Our objective is to conduct sentiment analysis that is tailored to the key areas that significantly impact tourists' experiences. We have identified 11 specific topics, including Activities, Friendliness, Climate, Food, Insecurity, Cleanliness, Nature, Pandemic, Prices, Transportation, and Location, which encompass a wide range of factors critical for tourists when evaluating and making decisions about their travel experiences.

To identify the relevant topics within the opinions, we employed a simple approach. This allowed us to uncover the underlying themes and subjects discussed in the opinions, enabling us to focus on specific aspects of interest to tourists. By associating each opinion with the most relevant topic, we enhanced the sentiment analysis process and captured the nuanced sentiment expressed in the context of these key areas.

The results of our experiments demonstrated the effectiveness of our proposed approach in filtering opinions related to tourist topics in Spanish. By considering the specific aspects of interest to tourists, our method provided more targeted and relevant sentiment analysis results compared to traditional approaches that do not incorporate topic information. The inclusion of specific topics allowed for a comprehensive understanding of tourists' experiences and facilitated more informed decision-making for potential travelers.

In conclusion, our proposed approach offers a valuable contribution to the field of sentiment analysis by presenting a tailored method for filtering opinions in Spanish with specific topics of tourist interest. By focusing on the key areas that significantly impact tourists' experiences, we enhance the accuracy and granularity of sentiment analysis, providing more targeted insights for tourism-related decision-making. We anticipate that our approach can be further extended and applied to other languages and domains, promoting improved sentiment analysis in various contexts.

Acknowledgments

The authors thank the Mexican Academy of Tourism Research (AMIT) for their support of the project "Creation of a labeled database related to tourist destinations for training artificial intelligence models for classifying relevant topics" through the call "I Research Projects 2022"

References

- [1] E. Olmos-Martínez, M. Á. Álvarez-Carmona, R. Aranda, A. Díaz-Pacheco, What does the media tell us about a destination? the cancu case, seen from the usa, canada, and mexico, *International Journal of Tourism Cities* (2023).
- [2] R. Guerrero-Rodríguez, M. Á. Álvarez-Carmona, R. Aranda, A. P. López-Monroy, Studying online travel reviews related to tourist attractions using nlp methods: the case of guanajuato, mexico, *Current issues in tourism* 26 (2023) 289–304.
- [3] M. A. Alvarez-Carmona, R. Aranda, A. Rodriguez-Gonzalez, D. Fajardo-Delgado, M. G. A. Sanchez, H. Perez-Espinosa, J. Martinez-Miranda, R. Guerrero-Rodríguez, L. Bustio-Martínez, A. D. Pacheco, Natural language processing applied to tourism research: A systematic review and future research directions, *Journal of King Saud University-Computer and Information Sciences* (2022).
- [4] A. Diaz-Pacheco, M. Á. Álvarez-Carmona, R. Guerrero-Rodríguez, L. A. C. Chávez, A. Y. Rodríguez-González, J. P. Ramírez-Silva, R. Aranda, Artificial intelligence methods to support the research of destination image in tourism. a systematic review, *Journal of Experimental & Theoretical Artificial Intelligence* (2022) 1–31.
- [5] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Muñis-Sánchez, A. P. Pastor-López, F. Sánchez-Vega, Overview of rest-mex at iberlef 2023: Research on sentiment analysis task for mexican tourist texts, *Procesamiento del Lenguaje Natural* 71 (2023).
- [6] M. A. Alvarez-Carmona, A. P. López-Monroy, M. Montes-y Gómez, L. Villasenor-Pineda, H. Jair-Escalante, Inaoe's participation at pan'15: Author profiling task, *Working Notes Papers of the CLEF 103* (2015).
- [7] M. Á. Álvarez-Carmona, E. Guzmán-Falcón, M. Montes-y Gómez, H. J. Escalante, L. Villasenor-Pineda, V. Reyes-Meza, A. Rico-Sulayes, Overview of mex-a3t at ibereval 2018: Authorship and aggressiveness analysis in mexican spanish tweets, in: *Notebook papers of 3rd sepln workshop on evaluation of human language technologies for iberian languages (ibereval)*, seville, spain, volume 6, 2018.
- [8] M. E. Aragón, M. A. A. Carmona, M. Montes-y Gómez, H. J. Escalante, L. V. Pineda, D. Moctezuma, Overview of mex-a3t at iberlef 2019: Authorship and aggressiveness analysis in mexican spanish tweets., in: *IberLEF@ SEPLN, 2019*, pp. 478–494.
- [9] M. Á. Álvarez-Carmona, R. Aranda, S. Arce-Cardenas, D. Fajardo-Delgado, R. Guerrero-Rodríguez, A. P. López-Monroy, J. Martínez-Miranda, H. Pérez-Espinosa, A. Y. Rodríguez-González, Overview of rest-mex at iberlef 2021: recommendation system for text mexican tourism 67 (2021). doi:<https://doi.org/10.26342/2021-67-14>.
- [10] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, D. Fajardo-Delgado, R. Guerrero-Rodríguez, L. Bustio-Martínez, Overview of rest-mex at iberlef 2022: Recommendation system, sentiment analysis and covid semaphore prediction for mexican tourist texts, *Procesamiento del Lenguaje Natural* 69 (2022) 289–299.
- [11] S. Arce-Cardenas, D. Fajardo-Delgado, M. Á. Álvarez-Carmona, J. P. Ramírez-Silva, A tourist recommendation system: a study case in mexico, in: *Advances in Soft Computing: 20th Mexican International Conference on Artificial Intelligence, MICAI 2021, Mexico City, Mexico, October 25–30, 2021, Proceedings, Part II* 20, Springer, 2021, pp. 184–195.

- [12] L. Bustio-Martínez, M. A. Álvarez-Carmona, V. Herrera-Semenets, C. Feregrino-Uribe, R. Cumplido, A lightweight data representation for phishing urls detection in iot environments, *Information Sciences* 603 (2022) 42–59.
- [13] M. A. Álvarez-Carmona, M. Franco-Salvador, E. Villatoro-Tello, M. Montes-y Gómez, P. Rosso, L. Villaseñor-Pineda, Semantically-informed distance and similarity measures for paraphrase plagiarism identification, *Journal of Intelligent & Fuzzy Systems* 34 (2018) 2983–2990.
- [14] M. A. Álvarez-Carmona, R. Aranda, R. Guerrero-Rodríguez, A. Y. Rodríguez-González, A. P. López-Monroy, A combination of sentiment analysis systems for the study of online travel reviews: Many heads are better than one, *Computación y Sistemas* 26 (2022). doi:<https://doi.org/10.13053/CyS-26-2-4055>.
- [15] M. Á. Álvarez-Carmona, E. Villatoro-Tello, L. Villaseñor-Pineda, M. Montes-y Gómez, Classifying the social media author profile through a multimodal representation, in: *Intelligent Technologies: Concepts, Applications, and Future Directions*, Springer, 2022, pp. 57–81.
- [16] M. Á. Álvarez-Carmona, R. Aranda, A. Y. Rodríguez-González, L. Pellegrin, H. Carlos, Classifying the mexican epidemiological semaphore colour from the covid-19 text spanish news, *Journal of Information Science* (2022). doi:<https://doi.org/10.1177/01655515221100952>.