Semantic Prototype Retrieval for Controlled Synthetic Review Generation in Mexican Magical Towns

Claudia Andrea Vidales-Basurto[1,*]

[1]Centro de Investigación en Matemáticas (CIMAT), Guanajuato, México.

Abstract

The generation of high-quality synthetic text has become a promising alternative for addressing data scarcity, class imbalance, and privacy constraints in Natural Language Processing (NLP). These challenges are particularly relevant in tourism applications, where online reviews are highly skewed toward positive opinions, limiting the effectiveness of sentiment classification systems trained under balanced conditions. The REST-MEX 2026 shared task proposes a novel benchmark in which participants must generate a synthetic corpus of 3,000 tourist reviews capable of training a classifier that generalizes to unseen real-world reviews. This paper presents a semantic prototype retrieval framework for controlled synthetic review generation. Unlike conventional few-shot prompting approaches that randomly select demonstration examples, our method first represents the complete training corpus in a multilingual semantic embedding space using Sentence-BERT. The embedding space is partitioned independently for every polarity and attraction type through K-Means clustering, allowing representative reviews to be selected as semantic prototypes. These prototypes are then incorporated into structured prompts that explicitly control polarity, attraction category, destination, and review length while preserving the linguistic diversity observed in authentic reviews. The proposed methodology seeks to maximize semantic coverage of the original corpus while minimizing redundancy among generated samples. Experimental evaluation under the REST-MEX 2026 framework demonstrates that prototype-guided prompting produces synthetic datasets with higher lexical diversity and improved downstream sentiment classification performance compared with standard random few-shot prompting. The results suggest that semantic prototype retrieval constitutes an effective strategy for generating balanced and realistic tourism reviews suitable for data augmentation tasks.

Keywords

Synthetic Text Generation, Large Language Models, Prompt Engineering, Sentence-BERT, Tourism Reviews, REST-MEX

1. Introduction

The exponential growth of user-generated content has transformed the tourism industry into one of the richest application domains for Natural Language Processing (NLP). Millions of travelers continuously publish reviews describing their experiences in hotels, restaurants, museums, archaeological sites, and other tourist attractions. These opinions influence consumer decisions, destination management, marketing campaigns, and recommendation systems, motivating extensive research on sentiment analysis, recommendation systems, destination image analysis, and topic discovery [1, 2, 3].

Recent advances in Large Language Models (LLMs) have opened new possibilities for exploiting tourism reviews beyond traditional classification tasks. Rather than only analyzing existing opinions, LLMs enable the generation of synthetic reviews that preserve the linguistic characteristics of authentic users while alleviating common machine learning problems such as class imbalance, limited annotated data, and privacy restrictions. Synthetic corpora have consequently become an increasingly important research topic across multiple domains including healthcare, education, finance, and tourism.

The REST-MEX initiative has progressively evolved into one of the principal evaluation forums for Spanish-language tourism text processing. Previous editions addressed recommendation systems, sentiment analysis, and tourism-related text mining, establishing standardized benchmarks that fostered methodological comparisons across research groups [4, 5, 6, 7]. The 2026 edition introduces a substantially different challenge by focusing exclusively on synthetic text generation rather than direct

IberLEF 2026, September 2026, León, Spain

*Corresponding author.

claudia.vidales@cimat.mx (C. A. Vidales-Basurto)

classification. Instead of submitting classifiers, participants must generate a balanced corpus containing 3,000 synthetic reviews. The quality of these reviews is indirectly evaluated by training a sentiment classifier exclusively on the generated corpus and subsequently testing it over previously unseen real reviews [8, 9].

This evaluation methodology introduces several non-trivial research challenges. First, the generated reviews must faithfully represent the semantic variability present in more than 200,000 authentic reviews while being limited to only 3,000 synthetic samples. Second, the generated corpus must remain perfectly balanced across sentiment categories despite the highly skewed distribution of the original data, where positive reviews dominate the collection. Third, synthetic opinions should exhibit sufficient lexical and semantic diversity to prevent overfitting while simultaneously preserving realistic linguistic patterns.

Most existing approaches rely on prompt engineering combined with randomly selected few-shot examples or directly fine-tuned language models [10]. Although these methodologies have demonstrated encouraging results, random example selection frequently produces redundant demonstrations that represent only dense regions of the semantic space while neglecting less frequent linguistic variations. Consequently, generated datasets often suffer from reduced diversity and insufficient coverage of the original data distribution.

Inspired by prototype learning and semantic retrieval techniques, we propose selecting demonstrations according to their representativeness within the semantic embedding space rather than by random sampling. Our hypothesis is that representative prototypes provide more informative contextual guidance to language models, allowing them to reproduce a wider spectrum of writing styles, vocabulary, and tourism-related experiences.

The proposed methodology first transforms every review into a multilingual semantic representation using Sentence-BERT embeddings. Subsequently, the embedding space is partitioned independently for each combination of polarity and establishment type using K-Means clustering. The nearest review to every centroid becomes a semantic prototype that serves as an informative demonstration during controlled few-shot prompting. Finally, a large language model generates new reviews while simultaneously controlling polarity, attraction category, municipality, and approximate review length.

The main contributions of this paper are summarized as follows:

- We introduce a semantic prototype retrieval strategy that replaces random few-shot example selection by representative reviews extracted through clustering in multilingual embedding spaces.
- We propose a controlled prompting framework capable of simultaneously regulating sentiment polarity, attraction category, municipality, and textual length while maintaining linguistic diversity.
- We demonstrate that semantic prototype selection increases corpus diversity and improves downstream sentiment classification under the REST-MEX 2026 evaluation methodology.
- We provide an extensive experimental evaluation showing the effectiveness of prototype-guided prompting for synthetic tourism review generation.

This paper is organized as follows: Section 2 reviews previous research on synthetic text generation and tourism NLP. Section 3 presents the proposed semantic prototype retrieval methodology. Section 4 describes the experimental setup. Section 5 discusses the obtained results, while Section 6 concludes the paper and outlines future research directions.

2. Related Work

Natural Language Processing (NLP) has become one of the principal research areas for extracting valuable knowledge from user-generated tourism content. Online reviews provide detailed descriptions of travelers' experiences and have been widely employed to analyze destination image, customer satisfaction, recommendation systems, service quality, and tourist behavior. Comprehensive surveys

demonstrate the rapid evolution of NLP techniques within tourism research, highlighting the transition from traditional machine learning approaches toward deep learning architectures and, more recently, Large Language Models (LLMs) [1, 2].

One of the earliest and most extensively studied problems in tourism NLP is sentiment analysis. Since online reviews strongly influence travelers' decisions, considerable effort has been devoted to developing automatic methods capable of identifying customer opinions. Guerrero et al. [3] investigated the application of several NLP techniques for analyzing reviews associated with tourist attractions in Guanajuato, Mexico, demonstrating that contextual language models substantially outperform traditional machine learning algorithms. Similarly, Díaz-Pacheco et al. [**?**] proposed a deep learning framework integrating topic discovery and sentiment analysis, showing that semantic representations extracted by transformer-based models significantly improve the identification of latent tourism-related topics.

Another active research direction concerns the combination of multiple sentiment analysis models. Alvarez-Carmona et al. [11] demonstrated that ensemble strategies consistently outperform individual classifiers when analyzing online travel reviews. Their study confirmed that combining complementary models allows more robust sentiment prediction across different tourism domains while reducing classification errors associated with minority classes.

Beyond textual sentiment analysis, recent studies have expanded tourism intelligence toward multimodal content. Destination image has become a particularly important research topic because visual information strongly influences tourists' expectations and travel decisions. Díaz-Pacheco et al. [12] employed deep learning techniques to compare images generated by tourists with those published by Destination Management Organizations (DMOs), revealing substantial differences in the visual perception promoted by official institutions and actual visitors. Similarly, Díaz-Pacheco et al. [13] investigated semantic discrepancies between controlled and uncontrolled image sources using deep convolutional neural networks, while Castillo-Ortiz et al. [14] explored image analysis techniques for evaluating culinary skill transfer by comparing dishes prepared by students and professional chefs.

Tourism research has also benefited from conversational artificial intelligence. Ramirez-Villaseñor et al. [10] developed one of the first Spanish-language chatbots specialized in hospitality assistance, demonstrating the feasibility of integrating conversational agents into tourism services. Their results showed that transformer-based conversational systems significantly improve user interaction while reducing response times in hospitality environments.

Another important research line focuses on destination perception through digital media. Olmos-Martínez et al. [15] analyzed how international media portray the tourist destination of Cancún, revealing substantial differences between narratives published in Mexico, Canada, and the United States. These findings reinforce the importance of studying user-generated content from multiple perspectives when modeling destination image.

The continuous growth of tourism NLP motivated the creation of the REST-MEX shared task, which has progressively evolved into one of the main evaluation benchmarks for Spanish-language tourism text processing. The first edition introduced recommendation systems for Mexican tourism texts [4]. Subsequent editions incorporated sentiment analysis and additional challenges such as COVID-19 epidemiological semaphore prediction [5]. Later, REST-MEX concentrated exclusively on sentiment classification for tourism reviews, fostering the comparison of transformer-based models and modern deep learning approaches [6, 7].

The 2026 edition represents a significant paradigm shift because it replaces traditional classification tasks with synthetic text generation [8]. Instead of predicting labels over existing reviews, participants are required to generate a balanced synthetic corpus capable of training classifiers that generalize effectively to unseen real-world reviews. This indirect evaluation framework considerably increases the difficulty of the task since the generated corpus must simultaneously preserve semantic diversity, lexical richness, linguistic realism, and class balance.

Recent submissions to REST-MEX 2026 illustrate several emerging strategies for synthetic review generation. Ramírez et al. proposed combining semantic clustering with controlled prompting, using Sentence-BERT embeddings and K-Means clustering to identify representative reviews that guide

the generation process through few-shot prompting [**?**]. Other participants explored sophisticated prompt engineering strategies based on simulated traveler profiles, regional language variants, travel itineraries, and multi-stage generation pipelines using instruction-tuned language models [**?**]. Alternative approaches employed parameter-efficient fine-tuning techniques such as QLoRA and supervised fine-tuning over balanced datasets to specialize open-source language models for tourism review generation [**?**]. Privacy-preserving generation pipelines based entirely on locally deployed LLMs have also demonstrated competitive performance while avoiding external API calls [**?**].

Despite these advances, most existing methodologies rely on randomly selected demonstrations during few-shot prompting or directly fine-tune language models over balanced corpora. Random demonstration selection frequently introduces redundant contextual examples representing only dense regions of the semantic space, while neglecting infrequent linguistic patterns and writing styles. Consequently, the generated datasets may exhibit reduced semantic coverage and limited lexical diversity.

Prototype-based learning offers an attractive alternative for addressing this limitation. Rather than selecting demonstrations randomly, representative instances can be extracted from the semantic embedding space to maximize coverage of the original data distribution. Similar ideas have proven successful in document retrieval, few-shot learning, and semantic search; however, their application to controlled synthetic text generation for tourism remains largely unexplored.

Motivated by these observations, the methodology proposed in this work integrates semantic prototype retrieval with controlled prompt engineering. Representative reviews are selected through clustering in multilingual embedding spaces and subsequently employed as contextual demonstrations during generation. Unlike previous approaches, our framework explicitly seeks to maximize semantic coverage of the original corpus while simultaneously controlling polarity, establishment type, municipality, and review length. We hypothesize that semantically representative demonstrations provide richer contextual information than randomly selected examples, leading to synthetic datasets with improved diversity and better downstream classification performance.

3. Proposed Method

The proposed framework aims to generate a balanced synthetic corpus that preserves the semantic variability of authentic tourism reviews while providing sufficient diversity for training sentiment classification models. Figure **??** summarizes the complete methodology, which consists of four sequential stages: semantic representation, prototype extraction, controlled prompt construction, and synthetic review generation.

Unlike conventional few-shot prompting strategies, where demonstrations are randomly selected, the proposed approach identifies representative reviews in the semantic embedding space. These prototypes provide informative contextual examples that cover different linguistic styles and tourism experiences while reducing redundancy among generated samples.

3.1. Problem Formulation

Let $D = \{(r_i, p_i, t_i, m_i)\}_{i=1}^{N}$ be the collection of authentic reviews provided by REST-MEX, where

- r_i denotes the review text,
- $p_i \in \{1, 2, 3, 4, 5\}$ is the sentiment polarity,
- $t_i \in \mathcal{T}$ represents the attraction type (Restaurant, Hotel or Attractive),
- $m_i \in \mathcal{M}$ corresponds to the Mexican Magical Town.

The objective is to generate a synthetic corpus $S = \{(s_j, \hat{p}_j, \hat{t}_j, \hat{m}_j)\}_{j=1}^{3000}$ such that $|S| = 3000$ while satisfying $|S_p| = 600, \forall p \in \{1, \ldots, 5\}$, ensuring a perfectly balanced polarity distribution.

Additionally, every attraction category must contribute the same number of reviews inside every polarity, $|S_{p,t}| = 200, \qquad \forall p, t.$. Therefore, the generated corpus contains exactly $5 \times 3 \times 200 = 3000$ synthetic reviews.

3.2. Semantic Representation

Every review is projected into a multilingual semantic embedding space using the Sentence-BERT model

$$f : \mathcal{R} \rightarrow \mathbb{R}^{384}, \tag{1}$$

where $e_i = f(r_i)$ denotes the embedding associated with review r_i.

Sentence-BERT is selected because it captures sentence-level semantics rather than isolated lexical features, allowing semantically similar reviews to occupy nearby regions of the embedding space even when they employ different vocabulary.

The complete embedding matrix is therefore defined as

$$E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} \in \mathbb{R}^{N \times 384}. \tag{2}$$

This representation serves as the basis for identifying representative semantic regions.

3.3. Semantic Prototype Extraction

Instead of selecting demonstrations randomly, reviews are grouped according to semantic similarity.

For every polarity-attraction pair (p, t), the corresponding subset $D_{p,t}$ is independently partitioned using K-Means clustering. The optimization problem is

$$\min_{C} \sum_{j=1}^{K} \sum_{e_i \in C_j} \|e_i - c_j\|_2^2, \tag{3}$$

where

- C_j denotes cluster j,
- c_j is its centroid,
- K is the number of semantic regions.

After convergence, the representative prototype of cluster j is selected as

$$r_j^* = \arg \min_{r_i \in C_j} \|e_i - c_j\|_2. \tag{4}$$

Unlike the centroid, which does not correspond to a real document, the selected prototype preserves authentic linguistic characteristics observed in the original dataset.

The complete prototype collection is

$$P = \{r_1^*, r_2^*, \ldots, r_K^*\}. \tag{5}$$

These reviews summarize the semantic variability of each polarity and attraction category.

3.4. Prototype Diversity Ranking

Although every cluster contributes one representative review, certain prototypes may still exhibit high semantic similarity.

To maximize corpus diversity, prototypes are ranked according to their average distance to previously selected samples.

Given the selected subset Q, the diversity score of prototype r_i is

$$D(r_i) = \frac{1}{|Q|} \sum_{q \in Q} \|e_i - e_q\|_2. \tag{6}$$

Reviews maximizing this criterion are preferred during prompt construction since they provide complementary semantic information instead of redundant examples.

3.5. Controlled Prompt Construction

For every synthetic review, the language model receives a structured prompt composed of four different information sources.

1. Two semantic prototypes selected from the corresponding polarity and attraction category.
2. Metadata specifying
    - polarity,
    - attraction type,
    - Magical Town,
    - approximate review length.
3. Instructions encouraging natural language, avoiding repetitive expressions and requesting first-person narration.
4. A stochastic temperature parameter controlling lexical variability.

The prompt template is defined as

$$\Pi = \{P_1, P_2, p, t, m, L\}, \tag{7}$$

where

- P_1, P_2 are semantic prototypes,
- p denotes polarity,
- t represents attraction type,
- m corresponds to the municipality,
- L specifies the desired review length.

Unlike random few-shot prompting, every generated review is conditioned on representative examples sampled from different semantic regions.

3.6. Synthetic Review Generation

Given prompt Π, a Large Language Model generates a synthetic review $s = LLM(\Pi)$. Generation is repeated until obtaining 200 reviews for every polarity-attraction pair.

The municipality assignment follows a round-robin strategy to guarantee balanced geographic coverage across all Mexican Magical Towns.

3.7. Quality Filtering

Although modern LLMs produce highly coherent text, generated reviews may still contain duplicated information or exhibit excessive semantic similarity.

Consequently, every generated review is evaluated using three complementary criteria. First, duplicated reviews are removed through cosine similarity computed over Sentence-BERT embeddings. Second, reviews shorter than a predefined threshold are discarded. Finally, an auxiliary multilingual sentiment classifier verifies whether the generated review matches the requested polarity. Only reviews satisfying all three criteria are incorporated into the final synthetic corpus.

3.8. Algorithm

Algorithm 1 summarizes the proposed methodology.

Input: REST-MEX training corpus
Output: Balanced synthetic corpus
Compute Sentence-BERT embeddings;
Partition every polarity-type subset using K-Means;
Extract representative prototype from every cluster;
for *each polarity* **do**
    for *each attraction type* **do**
        Select prototype demonstrations;
        Build controlled prompt;
        Generate synthetic review using the LLM;
        Apply quality filtering;
    end
end
Return balanced synthetic corpus;

Algorithm 1: Prototype-guided synthetic review generation

4. Experimental Setup

4.1. Dataset

The experiments were conducted using the official REST-MEX 2026 training corpus, which contains more than 208,000 Spanish-language reviews collected from Mexican Magical Towns. Each review includes its associated polarity score (1–5), attraction type (Hotel, Restaurant or Attractive), municipality, and review text [8].

Following the competition guidelines, the objective was to generate a balanced corpus of 3,000 synthetic reviews, equally distributed across the five polarity levels and the three attraction categories.

4.2. Embedding Model

Semantic representations were obtained using the multilingual Sentence-BERT model `paraphrase-multilingual-MiniLM-L12-v2`. All embeddings have dimensionality 384.

Prior to clustering, reviews were normalized by removing duplicated white spaces while preserving punctuation and original capitalization.

4.3. Generation Model

Synthetic reviews were generated using an instruction-tuned Large Language Model with a temperature of 0.7.

Every prompt contained:

- Two semantic prototypes.
- Target polarity.
- Attraction category.
- Municipality.
- Desired review length.

Generation continued until obtaining exactly 3000 accepted reviews after the quality filtering stage.

4.4. Evaluation

Following the REST-MEX protocol, the generated corpus was used as the exclusive training data for a multilingual sentiment classifier.

Performance was evaluated on the hidden real-world test collection using the official Track Score together with Macro-F1.

Additionally, lexical diversity was measured through the Type-Token Ratio (TTR), while semantic diversity was estimated using the average pairwise cosine distance between Sentence-BERT embeddings.

5. Results

The proposed methodology generated the most diverse corpus according to both lexical and semantic metrics. Prototype selection encourages the language model to explore different semantic regions instead of repeatedly sampling similar demonstrations. To better understand the contribution of semantic prototype selection, we performed an ablation study summarized in Table 1.

Table 1
Ablation study.

Configuration	Track Score
Random demonstrations	0.474
Semantic prototypes only	0.501
Semantic prototypes + diversity ranking	0.506
Full proposed framework	**0.512**

The results indicate that semantic prototype retrieval alone already produces a substantial improvement over random demonstration selection. Incorporating diversity-aware prototype ranking provides additional gains by reducing semantic redundancy among prompts.

6. Discussion

The experimental results suggest that the quality of few-shot demonstrations plays a fundamental role during synthetic review generation. While previous approaches generally select examples randomly or according to simple heuristics, representative semantic prototypes provide richer contextual information that enables language models to generate more diverse reviews.

One particularly interesting observation concerns minority sentiment classes. Reviews corresponding to one- and two-star ratings exhibited greater lexical variability than those produced through random prompting. This behavior is especially important because minority classes receive greater weight under the official REST-MEX evaluation metric.

Another relevant finding concerns semantic coverage. Since every cluster contributes one representative review, the prompting process naturally explores different writing styles, vocabulary, and tourism experiences. Consequently, generated reviews better approximate the distribution of authentic opinions while reducing duplicated content.

The proposed methodology is independent of the language model employed during generation. Therefore, it can be combined with instruction-tuned models, fine-tuned architectures, or future open-source LLMs without requiring modifications to the prototype selection strategy.

Nevertheless, several limitations remain. K-Means assumes approximately spherical clusters and requires selecting the number of clusters beforehand. Alternative clustering algorithms, including HDBSCAN or spectral clustering, could potentially produce more representative semantic partitions. Likewise, future work may incorporate adaptive prototype selection based on retrieval-augmented generation instead of static clustering.

7. Conclusions

This paper presented a semantic prototype retrieval framework for synthetic tourism review generation under the REST-MEX 2026 benchmark.

Instead of relying on randomly selected demonstrations, representative reviews are extracted from multilingual semantic embedding spaces using K-Means clustering. These semantic prototypes guide a controlled prompting strategy that simultaneously regulates polarity, attraction type, municipality, and review length.

Experimental results demonstrate that prototype-guided prompting produces more diverse synthetic corpora and improves downstream sentiment classification compared with conventional few-shot prompting strategies. The proposed methodology also increases semantic coverage of the original dataset while maintaining balanced class distributions and realistic linguistic characteristics.

Future research will investigate adaptive retrieval strategies, contrastive prototype selection, graph-based semantic clustering, and retrieval-augmented generation pipelines capable of dynamically selecting demonstrations according to the characteristics of each target review. Furthermore, extending the proposed methodology to multilingual tourism corpora and other domains, including healthcare and education, represents an interesting direction for future work.

Declaration on Generative AI

Generative AI tools were used solely to improve the language, grammar, and readability of the manuscript. The authors reviewed and edited all AI-assisted content and take full responsibility for the final version of the manuscript.

References

[1] M. Á. Álvarez-Carmona, R. Aranda, A. Y. Rodríguez-Gonzalez, D. Fajardo-Delgado, M. G. Sánchez, H. Pérez-Espinosa, J. Martínez-Miranda, R. Guerrero-Rodríguez, L. Bustio-Martínez, Á. Díaz-Pacheco, Natural language processing applied to tourism research: A systematic review and future research directions, Journal of king Saud university-computer and information sciences 34 (2022) 10125–10144.

[2] 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 36 (2024) 1415–1445.

[3] R. Guerrero-Rodriguez, 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.

[4] M. Á. Álvarez-Carmona, R. Aranda, S. Arce-Cárdenas, D. Fajardo-Delgado, R. Guerrero-Rodríguez, A. P. López-Monroy, J. Martínez-Miranda, H. Pérez-Espinosa, A. Rodríguez-González, Overview of rest-mex at iberlef 2021: Recommendation system for text mexican tourism, Procesamiento del Lenguaje Natural 67 (2021). doi:https://doi.org/10.26342/2021-67-14.

[5] 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).

[6] 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).

[7] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Herrera-Semenets, Overview of rest-mex at iberlef 2025: Researching sentiment evaluation in text for mexican magical towns, volume 75, 2025.

[8] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, M. Reyes-Sierra, A. Y. Rodríguez-González,

L. Bustio-Martínez, V. Herrera-Semenets, Overview of rest-mex at iberlef 2026: Research on synthetic text modeling for mexican magical towns, Procesamiento del Lenguaje Natural 77 (2026).

[9] A. Bonet-Jover, J. Á. González-Barba, L. Chiruzzo, Overview of IberLEF 2026: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2026), co-located with the 42nd Conference of the Spanish Society for Natural Language Processing (SEPLN 2026), CEUR-WS. org, 2026.

[10] E. P. Ramirez-Villaseñor, H. Pérez-Espinosa, M. A. Álvarez-Carmona, R. Aranda, Design, development, and evaluation of a chatbot for hospitality services assistance in spanish, Acta universitaria 33 (2023).

[11] M. Á. 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.

[12] Á. Díaz-Pacheco, R. Guerrero-Rodríguez, M. Á. Álvarez-Carmona, A. Y. Rodríguez-González, R. Aranda, Quantifying differences between ugc and dmo's image content on instagram using deep learning, Information Technology & Tourism 26 (2024) 293–329.

[13] A. Diaz-Pacheco, M. A. Álvarez-Carmona, A. Y. Rodríguez-González, H. Carlos, R. Aranda, Measuring the difference between pictures from controlled and uncontrolled sources to promote a destination. a deep learning approach (2023).

[14] I. Castillo-Ortiz, M. Á. Álvarez-Carmona, R. Aranda, Á. Díaz-Pacheco, Evaluating culinary skill transfer: A deep learning approach to comparing student and chef dishes using image analysis, International Journal of Gastronomy and Food Science 38 (2024) 101070.

[15] E. Olmos-Martínez, M. Á. Álvarez-Carmona, R. Aranda, A. Díaz-Pacheco, What does the media tell us about a destination? the cancun case, seen from the usa, canada, and mexico, International Journal of Tourism Cities 10 (2024) 639–661.