TRAVALL: Travel for All

Abhijay Singh
2021226

Amil Bhagat
2021309

Milind Jain
2021165

Kanishk Goel
2021325

1. Problem Statement

The central challenge addressed by this project is the conspicuous absence of a unified, accessible platform that consolidates comprehensive travel information, personalized planning tools, and inclusive destination recommendations catering to a broad spectrum of travelers, especially those with special needs. The current travel and tourism landscape is fragmented, with most existing platforms and services offering piecemeal solutions that fall short of addressing the full range of accessibility concerns. This gap significantly hinders the ability of individuals requiring special accommodations to engage confidently and fully in travel experiences.

2. Motivation

The concept of travel is inherently bound to the desire for exploration, understanding, and interaction with diverse cultures and environments. However, the accessibility of travel and tourism remains a significant challenge, particularly for individuals with special needs or disabilities. These travelers often encounter barriers that limit their full participation in travel experiences, from insufficient information on accessible facilities to a lack of personalized, adaptive services that cater to their specific requirements.

Recognizing these challenges, the motivation behind the development of Travall, our all-accessible travel guide, stems from a deep commitment to inclusivity and the democratization of travel. Our goal is to transform the landscape of travel planning and destination discovery by ensuring that all individuals, regardless of physical limitations or special needs, can experience the world with confidence and ease. This initiative is not only a response to a clear market need but also a moral imperative to promote equality and accessibility in one of the most universally valued human activities—travel.

3. Novelty of Travall

The innovative aspect of Travall is its comprehensive approach to creating a travel recommendation system that combines accessibility features, personalized planning, and inclusivity. This makes it a unique and comprehensive travel companion. Below are the key novel features that define Travall:

3.1. Integration of Accessibility Features

Travall stands out by embedding detailed accessibility information directly into the travel recommendation process. This includes data on wheelchair-friendly facilities, sensory-sensitive accommodations, and options suitable for various physical and cognitive conditions. This comprehensive integration ensures that all travelers, regardless of their special needs, can find travel options that are not only accessible but also enjoyable.

3.2. Personalized Travel Recommendations

Utilizing advanced algorithms and natural language processing, Travall processes reviews and user input to offer personalized travel suggestions. Whether a user is interested in culinary experiences, cultural activities, or specific accommodation needs, Travall tailors its recommendations to match individual preferences and accessibility requirements. This personalization extends beyond standard travel queries, offering a deeper level of customization based on nuanced user profiles.

3.3. Crowdsourced Reviews and Real-Time Data

Travall leverages the power of crowdsourcing to gather and update information about destinations, accommodations, and activities. This ensures that the information is not only up-to-date but also reflective of actual user experiences and accessibility standards. By incorporating real-time data from users, Travall can adapt and refine its recommendations to reflect the most current conditions and user feedback.

3.4. Benefits to Diverse User Groups

- **Travelers with Special Needs:** Individuals who require specific accommodations for mobility, sensory, or cognitive conditions gain access to tailored travel recommendations that consider their unique needs.

- **Individuals with Mobility Issues:** Including those who use wheelchairs, walkers, or have limited mobility, who are seeking travel destinations that accommodate and facilitate their travel experiences.

- **General Travel Enthusiasts:** Those looking for unique and personalized travel experiences based on their preferences for food, markets, activities, and more.
- **Tourism Industry Stakeholders:** Businesses and service providers in the tourism industry can better cater to a diverse clientele, tapping into an underserved market segment and enhancing their service offerings.

By addressing these points, Travall not only fills a crucial gap in the current travel industry but also pioneers new methodologies for inclusive and personalized travel experiences. This approach is set to redefine how travel planning is conducted, making it more accessible, enjoyable, and convenient for all types of travelers.

4. Methodology

The methodology of Travall involves a detailed multi-step process that combines advanced natural language processing (NLP) techniques and machine learning algorithms to provide personalized and accessible travel recommendations. This approach spans from initial data collection to sophisticated model training, followed by an integrated recommendation generation process. Each step of this methodology is outlined below, highlighting how we address the challenge of creating accessible travel experiences.

4.1. Data Collection and Preprocessing

The project started with compiling a review dataset that included city and place columns, with reviews for those places concatenated into single text blocks per place. This dataset serves as the foundation for understanding travelers' sentiments and experiences at various destinations.

4.2. Feature Extraction Using TF-IDF

To focus on the most important aspects of the reviews, we conducted a TF-IDF analysis on the concatenated texts to identify the top 100 most relevant words. This method helps highlight key themes and important descriptors within the reviews, which are critical for our feature analysis and subsequent steps.

4.3. Embedding Generation with Sentence-BERT

With the significant words extracted via TF-IDF, embeddings were generated using Sentence-BERT, a derivative of BERT optimized for sentence-level processing. These embeddings capture the semantic nuances of the identified words, facilitating a deeper analytical capability in later processes.

4.4. Fine-Tuning BERT for Accessibility Ratings

A separate dataset was prepared by fine-tuning a BERT model specifically designed for Question Answering tasks (BertQuestionAnswerLM) on descriptions of travel destinations. This fine-tuning process was targeted to derive accessibility ratings for each destination, focusing particularly on wheelchair access and visual accessibility. This tailored model allowed us to extract detailed accessibility information from descriptive texts, which is critical for our system's inclusivity.

4.5. Combining Datasets

The Sentence-BERT embeddings and the outputs from the fine-tuned BertQuestionAnswerLM were merged into a comprehensive dataset. This combination not only provides semantic understanding and relevance matching from the embeddings but also integrates essential accessibility ratings, enabling a more nuanced and tailored recommendation process.

4.6. Implementation of Model 2: Retrieval-Augmented Generation (RAG)

In addition to our primary models, we implemented a Retrieval-Augmented Generation (RAG) model on web-scraped data from Wikipedia to supplement our system with additional contextual information. This model dynamically retrieves and generates content relevant to the user's natural language queries, providing a rich layer of supplemental information.

4.7. User Query Processing and Recommendation Generation

User queries are preprocessed and transformed into embeddings using Sentence-BERT, ensuring semantic alignment with the data in our system. These embeddings are then used to perform similarity assessments with destination embeddings.

4.8. Cosine Similarity and Accessibility Filtering

The system calculates cosine similarity between user query embeddings and destination embeddings to identify potential matches. These matches are then filtered and ranked according to their accessibility scores based on user-specific requirements such as wheelchair access or visual aids.

4.9. Ranking and Final Recommendation

The final recommendations are ranked based on a combination of cosine similarity and the accessibility scores, ensuring that the most relevant and accessible options are presented to the user. This ensures that our recommendations are not only personalized to the user's preferences but are also practical and usable for individuals with specific accessibility needs.

Through this detailed and careful methodology, Travall aims to revolutionize the travel industry by providing a highly accessible, personalized travel planning tool that caters to the needs of all travelers, especially those with special requirements, ensuring an inclusive and enjoyable travel experience.

5. Dataset Description

Figure 1. Word Cloud

5.1. Reviews Dataset

The foundational dataset for Travall comprises user-generated reviews from various cities in India, including Delhi, Agra, Bengaluru, Chennai, Hyderabad, Jaipur, Kolkata, Mumbai, Pune, and Udaipur. From each city dataset, we identified the top 15 places with the highest volume of reviews, indicative of public interest and rich data for analysis.

5.2. Accessibility Ratings

Using a fine-tuned language model (LLM), we generated accessibility scores for each identified place. The model was specifically designed to parse reviews and descriptions to assess factors such as wheelchair accessibility and visual impairments, ensuring our travel recommendations support inclusivity.

5.3. Integration with Wikipedia Data

We enriched our dataset with Wikipedia pages corresponding to the most reviewed places. The content aids our Retrieval-Augmented Generation (RAG) model, which uses the Wikipedia information to generate and retrieve contextually relevant content in response to user queries.

5.4. Data Preparation and Preprocessing

To maintain consistency and accuracy, we performed thorough preprocessing on the review and Wikipedia data. This involved cleaning text, normalizing place names, and structuring the data for analysis and embedding generation.

5.5. Usage in Travall

The amalgamation of reviews, accessibility ratings, and encyclopedic content constitutes the core of Travall's recommendation engine. This integration allows Travall to provide tailored, accessible travel suggestions suitable for a broad spectrum of travelers, emphasizing those with special needs.

5.6. Example of Preprocessed Data

An example of the preprocessed data is provided in the table below:

City	Place	Embeddings	Accessibility Score

Table 1. Sample Data from the Travall Reviews Dataset

6. Code Implementation

6.1. Data Preprocessing and Analysis

The initial phase involves data preprocessing and analysis to prepare the dataset for further processing. This includes rearranging columns, cleaning text, and extracting relevant features such as TF-IDF vectors and BERT embeddings.

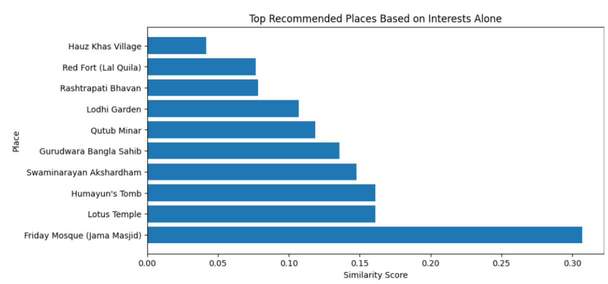

Figure 2. Results with No accessibility requirements

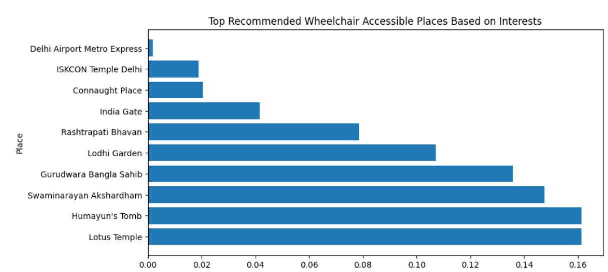

Figure 3. Results with wheelchair accessibility requirements

6.2. User Query Processing and Recommendation

User queries are preprocessed, and places are recommended based on similarity scores computed using TF-IDF vectors and BERT embeddings. Accessibility preferences are integrated into the recommendation process using combined scores.

6.3. Integration with Language Models

Language models, such as GPT-3.5, are utilized to provide conversational responses to user queries. Context retrieval and response generation are implemented to enhance the user experience.

6.4. Web Scraping for Wikipedia Content

Web scraping techniques are employed to extract content from Wikipedia pages related to various places. The scraped content is then cleaned and processed to extract relevant information.

6.5. BERT-based Question Answering for Accessibility Ratings

BERT-based question answering is implemented to extract accessibility information from reviews. Similarity scores are calculated between answers and review texts to assess accessibility.

7. Evaluation

In this section, we present the evaluation of our travel recommendation system, including baseline implementation, experimentation with different embeddings, model finetuning, and parameter tuning.

7.1. Baseline Implementation

We started our evaluation by implementing a simple baseline approach, which served as a reference point for comparing the performance of more sophisticated methods.

7.2. Experimentation with Embeddings

Next, we experimented with different embeddings for calculating cosine similarity between user queries and place reviews. We initially tried FastText and BERT embeddings and observed improvements in recommendation accuracy compared to the baseline.

To further enhance our system's performance, we fine-tuned a Sentence-BERT model on our dataset. This allowed us to capture more nuanced semantic similarities between user queries and reviews, resulting in better recommendation outcomes.

7.3. Cross-Checking with RAG

To validate the recommendations generated by our system, we employed a RAG (Retrieval-Augmented Generation) model to cross-check the results. This served as an additional measure to ensure the quality and relevance of the recommendations provided to users.

7.4. Model for Accessibility Ratings

Additionally, we trained a separate model to predict accessibility ratings for places in our dataset. This model was fine-tuned using Gemini API, providing valuable insights into the accessibility aspects of recommended places.

7.5. Tuning Parameters for Relevancy Reviews

Finally, we conducted experiments to determine the optimal number of words for relevancy reviews using TF-IDF. By experimenting with different values (50, 100, 200), we found that using 100 words yielded the best results in terms of recommendation accuracy.

7.6. Conclusion

Through systematic evaluation and experimentation, we have demonstrated the effectiveness of our travel recommendation system. By leveraging advanced techniques such as fine-tuned embeddings and model finetuning, we have significantly improved recommendation accuracy and user experience. Our approach also ensures the consideration of accessibility aspects, making our system more inclusive and informative for all users.