

Natural Language Processing for Multilingual Intelligent Travel Assistants: A Framework for Context-Aware and Inclusive Smart Tourism

Dr. Aamir Junaid Ahmad

S S Systems Pvt. Ltd., Patna, India

Email: aamirjunaid@yahoo.com

Abstract—Travel services increasingly depend on digital interfaces through which visitors search for destinations, transport, accommodation, attractions, local information, and assistance. Natural Language Processing (NLP) provides a direct mechanism for converting human language into searchable, actionable information and can therefore become a key intelligence layer for smart tourism systems. This paper proposes a conceptual framework for multilingual intelligent travel assistants that combines natural-language query understanding, semantic search, conversational interaction, sentiment analysis, recommendation, and context-aware response generation. The framework is motivated by earlier work on natural-language searching of social networking sites, semantic-web based retrieval, query classification, and social-network traceability, and extends these ideas to travel information environments. The paper also considers Hindi, Urdu, and code-switched language settings, drawing on research by Shagufta Fatema on multilingual chatbots, generative AI for Hindi and Urdu language teaching, and sentiment analysis of Hindi and Urdu e-learning feedback. The proposed framework identifies a layered architecture from language input to intent detection, entity extraction, semantic retrieval, personalization, response generation, and feedback learning. The study is conceptual and does not claim experimental performance results. It concludes that multilingual NLP can improve accessibility and reduce the friction of travel information discovery, while privacy, hallucination, bias, multilingual resource

scarcity, and evaluation remain important research challenges.

Keywords—Natural language processing; smart tourism; multilingual chatbot; semantic search; travel assistant; sentiment analysis; conversational AI; Hindi; Urdu; personalization

I. INTRODUCTION

Travel is an information-intensive activity. Before and during a journey, users ask questions such as where to go, how to reach a destination, what is available nearby, whether a place is suitable for a family, how crowded a location may be, and which services match their preferences. Traditional search interfaces generally expect keywords, menus, filters, or predefined categories. NLP enables a different interaction model in which a traveler can express an information need in ordinary language.

The relevance of NLP to information discovery has been demonstrated in earlier research by Ahmad and Priyadarshini. Their work on searching social networking sites using NLP proposed representing and processing profile information so that users could search through English-language queries rather than depending only on a known person's name [1]. Their subsequent work combined semantic-web concepts and NLP for social-network searching [2], while their 2020 study described a user interface for social-network traceability using NLP, query optimization, and classification [3]. These contributions provide a useful conceptual foundation for travel systems because travel search also requires transformation

of natural-language requests into structured retrieval conditions.

Smart tourism extends conventional digital travel services by connecting data, users, devices, and intelligent services. In this setting, a multilingual travel assistant can act as a natural-language gateway to heterogeneous sources such as maps, hotel databases, public transport information, attraction descriptions, reviews, weather feeds, and Internet of Things (IoT) devices. The assistant does not need to replace existing information systems; instead, it can provide a common conversational layer over them.

The multilingual dimension is particularly important in India and other linguistically diverse travel markets. Research by Fatema has examined multilingual chatbots for Hindi and Urdu conversational practice [6], generative AI for Hindi and Urdu language learning [7], AI-assisted language teaching [8], and sentiment analysis of Hindi and Urdu e-learning feedback [9]. These studies highlight practical issues including code-switching, script diversity, morphological complexity, limited labeled resources, and the need for culturally appropriate language technologies. Such issues are equally relevant when a traveler communicates with a digital assistant using Hindi-English or Urdu-English expressions.

II. OBJECTIVES OF THE STUDY

- To examine how NLP can transform natural-language travel questions into structured and context-aware information requests.
- To propose a conceptual architecture for a multilingual intelligent travel assistant.
- To connect semantic search and query classification with recommendation and conversational interaction.
- To identify the role of Hindi, Urdu, and code-switched language processing in inclusive smart tourism.

- To identify opportunities and limitations for future empirical evaluation.

III. METHODOLOGY

This paper follows a conceptual and literature-informed methodology. First, relevant prior work by the author on NLP, semantic search, query classification, and social-network traceability was examined. Second, published work by Fatema on multilingual conversational AI, Hindi-Urdu language learning, AI-assisted language teaching, and sentiment analysis was reviewed. Third, these concepts were mapped to travel-information tasks to construct an application-oriented framework. The resulting framework is not presented as a tested system; rather, it identifies components, information flows, research questions, and evaluation dimensions for future implementation.

IV. NLP IN SMART TRAVEL INFORMATION SYSTEMS

A. Natural-Language Query Understanding

A travel assistant should interpret a request beyond isolated keywords. For example, the query “I need a quiet family place near Patna for a one-day trip with vegetarian food” contains a destination context, temporal scope, user preference, group characteristic, activity requirement, and food preference. An NLP pipeline can identify intent, named entities, attributes, constraints, and implicit relationships. The result can be represented as a structured query for downstream retrieval.

This idea parallels the approach in Ahmad and Priyadarshini [1] and [3], where natural-language input is processed into conditions that can be used to retrieve relevant records. In a travel environment, the records may represent destinations, hotels, restaurants, transport options, attractions, or reviews rather than social-network profiles.

B. Semantic Search and Query Expansion

Keyword matching can fail when travelers use synonyms, abbreviations, spelling variations, or

colloquial expressions. Semantic search can map different expressions to a shared meaning representation. A request for “cheap places to stay close to the railway station” should be able to retrieve accommodation records containing terms such as budget, affordable, near station, or walking distance. The semantic-web and NLP perspective discussed in [2] is therefore relevant to travel retrieval, particularly when heterogeneous datasets use different terminology.

C. Intent and Entity Recognition

The assistant can classify requests into intents such as destination discovery, route planning, accommodation search, attraction discovery, complaint, review request, booking assistance, or emergency information. Named-entity recognition can identify locations, dates, transport modes, hotels, landmarks, currencies, and other travel entities. Query classification can then determine which external service or database should receive the structured request.

D. Sentiment and Opinion Understanding

Travel decisions are strongly influenced by user-generated reviews. Sentiment analysis can summarize whether travelers perceive an attraction, hotel, restaurant, or transport service positively or negatively. Fatema's work on sentiment analysis for Hindi and Urdu e-learning feedback [9] is not a tourism study, but its language-processing concerns are transferable to multilingual opinion mining. The same methodological issues—code-switching, script variation, limited labeled data, and language-specific sentiment expressions—can affect multilingual travel-review analysis.

V. MULTILINGUAL CONVERSATIONAL TRAVEL ASSISTANT

A multilingual assistant should support both language selection and mixed-language interaction. A traveler may begin in Hindi, insert English hotel names, switch to Urdu for a follow-up question, and use local place names that do not translate directly. A practical architecture therefore requires language identification, normalization, multilingual intent classification,

entity extraction, semantic retrieval, and response generation.

Fatema's work on multilingual Hindi-Urdu chatbots [6] provides a directly relevant conceptual reference for conversational practice, while her work on generative AI for Hindi and Urdu [7] illustrates how large language models can support language-specific content generation. Her later discussion of AI for Hindi-Urdu instruction [8] reinforces the importance of adapting AI systems to morphologically rich and comparatively low-resource languages.

The proposed travel assistant should use a retrieval-grounded approach rather than generating travel facts solely from a language model. Verified travel databases and current service APIs should provide factual grounding, while the NLP layer converts the user's language into a structured information request. This separation can reduce unsupported responses and makes it easier to audit the origin of recommendations.

VI. PROPOSED SYSTEM ARCHITECTURE

Layer	Primary Function	Travel Example
1. User Input	Text or voice capture	“Find a hotel near the airport”
2. Language Layer	Language ID, normalization, transliteration	Hindi-English mixed query
3. NLP Layer	Intent, entities, constraints	hotel + airport + proximity
4. Semantic Layer	Ontology, embeddings, query expansion	nearby $\approx$ close to
5. Data Layer	Search verified travel sources	Hotels, maps, reviews
6. Personalization	Preferences and context	budget, family, accessibility
7. Response Layer	Grounded conversational	Shortlist with reasons

	answer	
8. Feedback Layer	Corrections and sentiment	User accepts/rejects result

The architecture extends the query-processing logic described in the author's earlier NLP and social-network traceability research [1]–[3] into a travel context. It also incorporates multilingual and sentiment dimensions reflected in Fatema's studies [6]–[9].

VII. APPLICATION SCENARIOS

A. Conversational Destination Discovery

A traveler can describe desired experiences instead of navigating category menus. The system can extract constraints such as distance, budget, duration, interests, accessibility, and group type. Semantic retrieval can then identify candidate destinations and rank them according to explicit user constraints.

B. Multilingual Travel Support

Tourists who are more comfortable communicating in regional languages can interact with the assistant in Hindi or Urdu, while the underlying databases may remain multilingual or English-oriented. Translation and transliteration can act as intermediate representations. However, place names and culturally specific expressions should be preserved where translation would create ambiguity.

C. Review and Sentiment Analysis

The system can aggregate multilingual reviews and classify opinions about cleanliness, hospitality, food, safety, transport, crowding, and value. Rather than producing a single unexplained score, the assistant can present topic-level sentiment with supporting review evidence. This approach is consistent with the broader concern in Fatema's sentiment-analysis research that regional-language feedback requires language-sensitive processing [9].

D. Conversational Itinerary Refinement

An itinerary can be progressively refined through dialogue. For example, after receiving a first itinerary, a traveler may ask to reduce walking, add a museum, remove an expensive restaurant, or shorten the trip. NLP can maintain conversational context and update only the relevant constraints.

E. Accessibility and Inclusive Travel

Natural-language interfaces may reduce interaction barriers for users who find complex travel websites difficult to navigate. Voice input, regional-language responses, simplified explanations, and accessibility-aware entity extraction can make smart tourism systems more inclusive. Future work should evaluate these benefits empirically across age groups, language backgrounds, and accessibility needs.

VIII. BENEFITS AND RESEARCH CHALLENGES

Dimension	Potential Benefit	Research Challenge
Search	Natural-language discovery	Ambiguous user intent
Personalization	Preference-aware results	Privacy and consent
Multilingual use	Regional-language access	Low-resource datasets
Reviews	Topic and sentiment summaries	Sarcasm and code-switching
Conversation	Iterative itinerary refinement	Context retention
Grounding	More reliable travel facts	Data freshness
Accessibility	Voice and simplified interaction	Accent and dialect variation

IX. DISCUSSION

The proposed framework demonstrates a clear conceptual connection between the author's earlier NLP research and emerging smart-tourism applications. The social-network studies treated natural language as an interface for searching structured information [1]–[3]. Travel systems contain similarly heterogeneous records and can benefit from the same principle: users should be able to express information needs naturally while the system performs classification, semantic interpretation, filtering, and retrieval behind the interface.

The multilingual research of Fatema adds another dimension. Her studies demonstrate that conversational and generative AI for Hindi and Urdu must address language-specific issues rather than simply applying English-centric NLP assumptions [6]–[8]. For tourism, these concerns become more significant because users may combine regional languages with English names of hotels, transport providers, applications, or landmarks. Sentiment analysis adds a further requirement because travel reviews frequently contain informal language, mixed scripts, abbreviations, and culturally specific expressions.

A major design principle is therefore separation of language intelligence from factual data. NLP should interpret the request; verified databases should supply current facts; a recommendation component should apply user constraints; and the response generator should clearly distinguish retrieved facts from generated explanations. This architecture also provides an opportunity for explainability because the system can expose which constraints and sources influenced an answer.

The framework should not be interpreted as evidence that NLP automatically improves travel outcomes. Such a conclusion requires controlled experiments involving accuracy, task completion time, user satisfaction, multilingual intent recognition, recommendation relevance, and factuality. The present study instead provides a

research model that can guide such evaluation.

X. EVALUATION FRAMEWORK FOR FUTURE STUDIES

A future prototype can be evaluated using a multilingual travel-query dataset containing English, Hindi, Urdu, Hindi-English, and Urdu-English requests. Each query can be annotated for intent, entities, constraints, sentiment where applicable, and expected information sources. Evaluation can then be conducted at both NLP and user-experience levels.

- Intent classification accuracy and macro-F1 across languages and code-switched queries.
- Named-entity extraction precision, recall, and F1 for destinations, hotels, attractions, dates, and transport entities.
- Semantic retrieval relevance using precision@k, recall@k, and normalized discounted cumulative gain.
- Response factuality measured against verified travel records and source attribution.
- Task completion time and number of conversational turns required to satisfy a travel request.
- User satisfaction, perceived usefulness, trust, and accessibility across language groups.

XI. ETHICAL, PRIVACY, AND IMPLEMENTATION CONSIDERATIONS

Travel assistants may process sensitive contextual information such as location, movement patterns, family composition, spending preferences, and travel history. Data minimization, consent, retention controls, and secure processing should therefore be built into the system. Personalization should be optional rather than silently inferred whenever possible.

Multilingual systems also require bias evaluation. A model that performs well in English but poorly in Hindi, Urdu, or mixed-language queries may produce unequal service quality. Dialectal and script variation should be included in evaluation

datasets. In addition, generated responses should be grounded in current sources because opening hours, fares, availability, routes, and travel restrictions can change.

The research direction should also distinguish between language generation and decision authority. The assistant can explain alternatives and retrieve information, while the traveler remains responsible for selecting and booking services. This distinction is particularly important when the system presents safety, financial, or time-sensitive travel information.

XII. FUTURE RESEARCH DIRECTIONS

Future research can develop a benchmark dataset for multilingual smart-tourism queries, especially Hindi-English and Urdu-English code-switching. Another direction is integration with IoT and real-time travel infrastructure so that natural-language requests can retrieve live transport, baggage, crowd, or environmental information. Multimodal assistants could additionally process photographs, maps, voice messages, and text together.

A further research opportunity is personalized travel dialogue based on long-term preferences while preserving privacy. Explainable retrieval could show which query constraints produced each result. Finally, domain-specific language models for tourism should be evaluated against general-purpose large language models on factuality, multilingual intent understanding, cultural appropriateness, and robustness to ambiguous travel queries.

XIII. CONCLUSION

Natural Language Processing can serve as an important interaction and intelligence layer for smart tourism by allowing travelers to communicate information needs in natural language rather than navigating rigid search interfaces. This paper proposed a conceptual multilingual intelligent travel assistant combining intent detection, entity extraction, semantic search, sentiment analysis, personalization, retrieval grounding, and conversational response generation.

The framework builds directly on the author's earlier research into NLP-based social-network searching, semantic-web search, and social-network traceability [1]–[3]. It also connects with Fatema's research on Hindi-Urdu multilingual chatbots, generative AI, AI-assisted language teaching, and sentiment analysis [6]–[9]. The combined perspective suggests a research pathway from natural-language query processing toward inclusive, context-aware travel information systems.

The proposed framework remains conceptual; empirical validation is required before claims about performance or user benefits can be established. Future prototypes should therefore focus on multilingual datasets, measurable retrieval and classification performance, factual grounding, user studies, privacy, and fairness. With these safeguards, NLP can contribute to making digital travel services more conversational, accessible, and responsive to diverse travelers.

ACKNOWLEDGMENT

The author acknowledges the researchers whose prior work on NLP, semantic search, multilingual conversational AI, and language technologies informed the conceptual development of this paper.

REFERENCES

- [1] A. J. Ahmad and S. Priyadarshini, "Searching a social networking site using natural language processing technique," *International Journal of Current Advanced Research*, vol. 6, no. 3, pp. 2473–2476, 2017, doi: 10.24327/ijcar.2017.2476.0031.
- [2] A. J. Ahmad and S. Priyadarshini, "Searching social networking sites using semantic web and natural language processing," *International Journal of Engineering, Applied and Management Sciences Paradigms*, vol. 54, no. 3, pp. 34–40, 2019.
- [3] A. J. Ahmad and S. Priyadarshini, "Design and development of a friendly user interface for building social network traceability system," *Social Network Analysis and Mining*, vol. 10, art. 59, 2020, doi: 10.1007/s13278-020-00675-2.

- [4] A. J. Ahmad, Y. Misra, and A. K. Srivastava, "Secure information collection and energy efficient in heterogeneous sensor networks using machine learning with the Internet of Things," *Wireless Communications and Mobile Computing*, 2022.
- [5] A. J. Ahmad, V. Sharma, E. Ramirez-Asis, and D. Krah, "On the Internet of Things, blockchain technology for supply chain management," *Wireless Communications and Mobile Computing*, vol. 2022, art. 9185731, 2022.
- [6] S. Fatema, "Developing multilingual chatbots for conversational practice in Hindi and Urdu: An AI-driven solution," *Webology*, vol. 18, no. 6, 2021.
- [7] S. Fatema, "Generative AI for creative content generation in Hindi and Urdu language teaching: A pilot study," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 12, pp. 582–595, 2022.
- [8] S. Fatema, "Harnessing AI for effective language teaching: Innovations in Hindi Urdu instructions," *Journal of Computational Analysis and Applications*, vol. 33, no. 8, pp. 6570–6583, 2024.
- [9] S. Fatema, "Sentiment analysis of learner feedback in Hindi and Urdu e-learning platforms: An ML approach to course improvement," *International Journal of Applied Engineering & Technology*, vol. 5, no. 4, 2023.
- [10] D. Buhalis and A. Amaranggana, "Smart tourism destinations enhancing tourism experience through personalisation of services," *Information and Communication Technologies in Tourism*, pp. 377–389, 2015.