

# Varanasi Tourism in Question Answer System Track:IIIT SURAT @ FIRE'25 Shared Task<sup>\*</sup>

Ritesh Kumar<sup>1,\*†</sup>, Sumit Chand Jaiswal<sup>2†</sup> and Dhiraj Bhatia<sup>3†</sup>

<sup>1</sup>Department of Computer Science and Engineering, Indian Institute of Information Technology Surat

<sup>2</sup>Department of Computer Science and Engineering, Indian Institute of Information Technology Surat

<sup>3</sup>Department of Biological Sciences and Engineering, Indian Institute of Technology Gandhinagar

## Abstract

This paper presents our approach to the VATIKA: Varanasi Tourism in Question Answering System Track at FIRE 2025, conducted by the Indian Institute of Information Technology Surat. The task focuses on developing a domain-specific Question Answering (QA) system for tourism-related queries in Hindi, particularly centered on the culturally significant city of Varanasi. To address this challenge, we propose a hybrid architecture that integrates semantic retrieval with extractive question answering. Our system leverages Facebook AI Similarity Search (FAISS) for efficient similarity search in high-dimensional vector spaces. Contextual embeddings are generated using IndicBERT, a multilingual ALBERT-based transformer model pretrained on major Indic languages. These embeddings are indexed within FAISS to enable fast and accurate retrieval of semantically relevant contexts for a given user query. The retrieved context is subsequently processed by a fine-tuned IndicBERT-based extractive QA model, which predicts the start and end token positions of the answer span within the passage. This two-stage retrieval and comprehension framework improves computational efficiency while maintaining contextual relevance. We submitted three system runs for the shared task. Although IndicBERT proved effective for both embedding generation and question answering, the overall performance was constrained by challenges in capturing nuanced linguistic characteristics of pure Hindi text, particularly domain-specific expressions and culturally grounded references. Our findings highlight the importance of domain adaptation and language-specific fine-tuning for Hindi QA systems. Future improvements may include enhanced Hindi-specific pretraining, incorporation of linguistic features, and improved retrieval strategies to better address semantic variability in tourism-related queries.

## Keywords

VATIKA, FASIS, ALBERT, IndicBERT

## 1. Introduction

Tourism plays a vital role in India's socio-economic development by generating income, creating employment opportunities, and supporting local businesses [1]. Beyond its economic contribution, tourism also facilitates cultural exchange, promotes the preservation of heritage, and accelerates infrastructural growth [2]. By attracting global visitors, tourism not only enhances international visibility but also instills regional pride, positioning itself as a key driver of sustainable development and global cooperation in the travel and tourism sector [3]. Among India's most prominent destinations, Varanasi (Kashi) holds a unique position as one of the world's oldest living cities. It is revered as a cultural and spiritual hub, attracting millions of domestic and international tourists seeking spiritual awakening, cultural enrichment, and experiential travel [4]. Known for its Bhakti-Bhaav (devotional ethos), Varanasi reflects the living traditions of India and continues to be a vibrant center for pilgrimage, cultural festivities, and heritage tourism [5].

Despite its global significance, the tourism experience in Varanasi can often be hindered by limited access to authentic and structured information. Tourists frequently seek reliable guidance regarding

*Forum for Information Retrieval Evaluation, December 17-20, 2025, Varanasi, India*

<sup>\*</sup> You can use this document as the template for preparing your publication. We recommend using the latest version of the ceurart style.

<sup>\*</sup> Corresponding author.

<sup>†</sup> These authors contributed equally.

✉ [riteshkumar@iiitsurat.ac.in](mailto:riteshkumar@iiitsurat.ac.in) (R. Kumar)

🆔 0000-0003-1932-8615 (R. Kumar)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

religious rituals such as the Ganga Aarti, local services including cruise rides, food courts, public facilities, travel agencies, ashrams, temples, kunds, museums, and general cultural events [6]. Traditional modes of information dissemination, such as guidebooks or physical helpdesks, are often insufficient in meeting the diverse and immediate queries posed by modern tourists, particularly in Indian languages [7].

In this context, Natural Language Processing (NLP) offers promising solutions through Question Answering (QA) systems, which are designed to automatically respond to user queries in natural language using structured databases or unstructured text resources [8]. By combining domain specificity with multilingual capabilities, QA systems can enhance tourist experiences by offering precise, user-friendly, and context-aware information [9].

However, Hindi presents challenges including morphological richness, free word order, and limited annotated datasets. To address these issues, we participated in the VATIKA shared task at FIRE 2025, proposing a retrieval-augmented QA framework integrating FAISS with IndicBERT.

Our contributions are:

- A domain-specific Hindi QA pipeline for tourism.
- Integration of FAISS for semantic retrieval.
- Fine-tuning IndicBERT for extractive QA.
- Empirical evaluation on VATIKA Test Data-II.

Organization of rest of the paper is as follows. Section 2 describes dataset, Section 3 describes about Methodology we have used and Section 4 discuss about Results and Analysis. Finally, we conclude in Section 5 with directions for future work.

## 2. Related Work

QA systems evolved significantly with the introduction of datasets such as SQuAD [9]. Transformer-based architectures like BERT [10] improved contextual understanding, while ALBERT [11] reduced parameter redundancy.

Multilingual BERT extended support to multiple languages but showed limitations for low-resource languages. AI4Bharat introduced IndicBERT [12], a multilingual ALBERT-based model trained on Indian languages.

Retrieval-Augmented QA approaches combine semantic retrieval with answer extraction. FAISS [13] enables efficient similarity search in high-dimensional vector spaces and is widely used in open-domain QA frameworks. Tourism-based QA systems remain underexplored for Indian languages. Our work contributes by combining IndicBERT with FAISS for Hindi tourism QA.

For multilingual settings, mBERT and XLM-R [14] extended Transformer architectures to support multiple languages. However, studies have shown that multilingual models often underperform in low-resource languages due to limited language-specific supervision. To address this gap, AI4Bharat introduced IndicBERT [12], a multilingual ALBERT-based model trained specifically on Indian language corpora. IndicBERT demonstrated promising results in tasks such as classification, NER, and QA across several Indic languages. In addition to extractive QA, open-domain QA systems have gained attention. Dense Passage Retrieval (DPR) [15] introduced dual-encoder retrieval models that learn dense embeddings for efficient document retrieval. Retrieval-Augmented Generation (RAG) [16] combined retrieval with generative models, improving answer quality in open-domain settings. These systems highlight the importance of retrieval mechanisms for improving context relevance. For Indian language QA, research remains comparatively limited. Several studies have explored Hindi QA using mBERT and multilingual Transformer models, but domain-specific tourism QA datasets have been scarce. The FIRE evaluation campaigns have played a crucial role in promoting Indian language IR and QA research. The VATIKA shared task focuses specifically on tourism queries in Hindi, providing a structured benchmark for evaluating domain-adapted QA systems.

In the tourism domain, conversational agents and chatbots have been proposed to assist travelers with itinerary planning and local information access. However, many of these systems are primarily

English-centric and rely on generative approaches without robust domain grounding. Our work differs by focusing on extractive QA with domain-specific retrieval in Hindi, leveraging IndicBERT embeddings and FAISS-based semantic indexing.

Overall, prior research highlights three important directions:

- leveraging Transformer-based contextual encoders,
- integrating retrieval mechanisms for improved relevance, and
- adapting models to low-resource languages. Our approach builds upon these principles to design a retrieval-augmented Hindi tourism QA system tailored to the VATIKA benchmark.

### 3. Data

Test collection provided by FIRE 2025 VATIKA organizers for shared task [17]. The test data is divided into 2 parts: Test Data-I and Test Data-II. Test Data-I was provided for the initial stage. Test Data-II was provided for the final submission of system. Focused on the culturally significant city of Varanasi, the dataset captures realistic queries that travelers and pilgrims commonly raise regarding locations, services, logistics, and spiritual landmarks.

VATIKA is distinctive in its coverage of ten tourism-relevant domains: Ganga Aarti, Cruise, Food Court, Public Toilet, Kund, Museum, General Queries, Ashram, Temple, and Travel. Each domain contains carefully curated Hindi passages (in Devanagari script), paired with multiple question–answer sets. The dataset is designed to simulate authentic information-seeking behavior by including questions that span factual, navigational, and experiential types, thereby ensuring comprehensive coverage of diverse tourist concerns. Entirely developed in Hindi, VATIKA provides paragraph-level contexts with associated QA pairs, making it a valuable linguistic resource for the Indian tourism sector. It supports both open-domain QA and contextual MRC-style QA, offering researchers and developers a benchmark for building and evaluating robust, user-centric systems tailored to Indian language contexts. From the provided dataset from split section we have Train and Validation that consists of 5,538 contexts, 13,408 QA pairs and 1,158 contexts, 2,963 QA pairs respectively.

### 4. Methodology and Experimental Setup

The system operates in two main phases: training and inference. Data preprocessing was required for the task. We present a Hindi extractive question-answering (QA) framework that integrates text normalization, data structuring, tokenization, fine-tuning, and context retrieval into a unified pipeline. Hindi text, including both questions and answers, is first standardized using the *indic-nlp-library* to ensure consistency and improve downstream performance. Question–answer pairs and their contexts are then parsed from JSON data, with the script augmenting each context to guarantee inclusion of the correct answer span—an essential step for extractive QA tasks. The text is subsequently tokenized with the IndicBERT tokenizer, which converts input into numerical *input<sub>i</sub>ds* and *attention<sub>m</sub>ask* tensors while mapping character-level answer boundaries to token-level indices for supervised training. A pre-trained IndicBERT model is fine-tuned for QA using the *AutoModelForQuestionAnswering* class, and the Hugging Face Trainer API manages optimization under specified hyperparameters such as epoch count, batch size, and logging frequency. For inference, the system employs an embedding model with FAISS-based semantic search to retrieve the most relevant context for a given user query. The fine-tuned model then predicts the start and end tokens of the answer span, which are decoded back into fluent Hindi text, with fallback responses provided when no confident answer is available.

### 5. Results and Analysis

For this work, we employ IndicBERT (ai4bharat/indic-bert), a multilingual ALBERT-based model pre-trained on 12 major Indic languages, including Hindi. The model is utilized for two core tasks within our system:

- **Embedding Generation:** IndicBERT is used to encode text into dense vector representations that capture semantic meaning. These embeddings form the foundation of the system’s similarity search component.
- **Question Answering:** The model is further fine-tuned for extractive QA, enabling it to predict the start and end positions of answers within a given passage, thereby supporting precise information retrieval.

To enable efficient retrieval, we integrate FAISS for similarity search in high-dimensional vector spaces. All contextual embeddings produced by IndicBERT are indexed using FAISS. Upon receiving a query, the system retrieves the most semantically relevant context from the index, which is subsequently passed to the QA model. This significantly enhances both the accuracy and efficiency of the system. Our implementation relies heavily on the Hugging Face ecosystem, particularly the transformers and datasets libraries. The transformers library provides access to the pre-trained IndicBERT model and the Trainer API for fine-tuning, while the datasets library supports preprocessing, formatting, and efficient handling of training and validation data.

Our model’s performance is hindered by its struggle with the nuances of pure Hindi datasets, particularly when utilizing the AI4Bharat IndicBERT model. Key contributing factors include potential mismatches between the model’s training data and our specific dataset, as well as limitations in capturing linguistic intricacies unique to Hindi. These challenges suggest avenues for improvement, such as fine-tuning the model with Hindi-specific datasets or incorporating additional linguistic features tailored to the language. The scores obtained by our three runs are given in Table 1. Here, we have used Test Data-II for testing our system. The official evaluation measure by VATIKA’25 are F1 score, BLEU score, and ROUGE-L score. Our best performance is by *IIIT Surat-03-05072025* where we use IndicBERT.

**Table 1**

Results- The official evaluation measure by VATIKA’25

Team	submission ID	BLEU-1	ROUGE-1	QA-F1
IIIT Surat	IIIT Surat-01-05072025	6.8	0.00613155	0.0.061806397
IIIT Surat	IIIT Surat-02-05072025	6.8	0.00613155	0.0.061806397
IIIT Surat	IIIT Surat-03-05072025	6.8	0.00613155	0.0.061806397

## 6. Conclusion and Future Work

This year we participated in the Shared Task of VATIKA: Varanasi Tourism in Question Answer System. We tried for the integration of FAISS that enables efficient similarity search in high-dimensional vector spaces. By indexing the contextual embeddings generated by IndicBERT, the system can quickly identify the most semantically relevant context for a given query. Passing this retrieved context to the QA model ensures more accurate and context-aware responses, thereby enhancing both the effectiveness and efficiency of the overall system. While there can be no denial of the fact that our overall performance is dismal, initial results are suggestive as to what should be done next. Future efforts may also explore conversational agents capable of handling multi-turn dialogues, allowing tourists to refine and contextualize their queries in real time. Integration with real-time services, such as transport schedules, weather updates, and ticketing platforms, could further enhance the system’s practical utility. Additionally, personalized recommendation systems based on user preferences (e.g., spiritual, cultural, or culinary tourism) represent an interesting research direction. We shall be exploring some of these tasks in the coming days.

## 7. Acknowledgment

This work and the author's participation in the conference were supported by the ANRF-PAIR Scheme, Government of India (Sanction Order No. ANRF/PAIR/2025/000008/PAIR).

## 8. Declaration on Generative AI

During the preparation of this work, the authors used OpenAI-GPT-4: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

## References

- [1] W. Travel, T. E. Impact, World travel & tourism council (wttc), Travel & Tourism Economic Impact (2024).
- [2] K. S. Rana, R. V. Manshin, Contribution of tourism in india's gdp in pre-and post-pandemic scenarios, in: Sustainable Development of Transport: Economy, Transformation, Logistics and ESG Agenda. Volume 2, Springer, 2025, pp. 223–231.
- [3] D. R. Hall, G. Richards, Tourism and sustainable community development, volume 1, Routledge London, 2000.
- [4] R. P. Singh, Professor rb singh (1955~ 2021), an icon of indian geography: A passage on the path of lineage, legacy and liminality, Space and Culture 9 (2021) 06–49.
- [5] D. P. Mitra, Religious tourism and ascetic integrity: A sociological study of economic dependency and sacred authenticity in varanasi (2017).
- [6] T. N. Feizabadi, A critical review of the sustainability of tourism in varanasi, Department of Geography, Banaras Hindu University (2015).
- [7] H. Sharma, S. Pande, Religious tourism in uttar pradesh: A case study of varanasi, CASEPEDIA (2022).
- [8] J. L. Crawford, Linguistic changes in spontaneous speech for detecting parkinson's disease using large language models, PLOS Digital Health 4 (2025) e0000757.
- [9] P. Rajpurkar, J. Zhang, K. Lopyrev, P. Liang, Squad: 100,000+ questions for machine comprehension of text, arXiv preprint arXiv:1606.05250 (2016).
- [10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), 2019, pp. 4171–4186.
- [11] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, Albert: A lite bert for self-supervised learning of language representations, arXiv preprint arXiv:1909.11942 (2019).
- [12] D. Kakwani, A. Kunchukuttan, S. Golla, G. NC, A. Bhattacharyya, M. M. Khapra, P. Kumar, Indicnlp suite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for indian languages, in: Findings of the association for computational linguistics: EMNLP 2020, 2020, pp. 4948–4961.
- [13] J. Johnson, M. Douze, H. Jégou, Billion-scale similarity search with gpus, IEEE Transactions on Big Data 7 (2019) 535–547.
- [14] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Proceedings of the 58th annual meeting of the association for computational linguistics, 2020, pp. 8440–8451.
- [15] V. Karpukhin, B. Oguz, S. Min, P. S. Lewis, L. Wu, S. Edunov, D. Chen, W.-t. Yih, Dense passage retrieval for open-domain question answering., in: EMNLP (1), 2020, pp. 6769–6781.

- [16] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al., Retrieval-augmented generation for knowledge-intensive nlp tasks, *Advances in neural information processing systems* 33 (2020) 9459–9474.
- [17] P. Gatla, Anushka, N. Kanwar, G. Sahoo, R. K. Mundotiya, Tourism question answer system in indian language using domain-adapted foundation models, *arXiv preprint* (2025).