RESEARCH-ARTICLE

UGuideRAG: Intent-Enhanced Retrieval-Augmented Generation with User-Generated Content for Personalized Urban Tourism

Published: 03 November 2025

Citation in BibTeX format

SIGSPATIAL '25: 33rd ACM International Conference on Advances in Geographic Information Systems
November 3 - 6, 2025
MN, Minneapolis, USA

Conference Sponsors:
SIGSPATIAL

SIGSPATIAL '25: Proceedings of the 33rd ACM International Conference on Advances in Geographic Information Systems (November 2025)
https://doi.org/10.1145/3748636.3762712
ISBN: 9798400720864

UGuideRAG: Intent-Enhanced Retrieval-Augmented Generation with User-Generated Content for Personalized Urban Tourism

Jing Tang*
Department of Geography
University of Zurich
Zürich, Switzerland
jing.tang@uzh.ch

Inhye Kong
Department of Geography
University of Zurich
Zürich, Switzerland
inhye.kong@geo.uzh.ch

Zhaonan Wang†
Division of Arts and Sciences
NYU Shanghai
Shanghai, China
zhaonan.wang@nyu.edu

Abstract

Citywalk, as an increasingly popular form of urban tourism, emphasizes immersive, diverse, and personalized exploration over conventional sightseeing. These features evolving tourist expectations pose new challenges for intelligent itinerary planning, particularly in capturing the rich experiential attributes of visitor attractions and aligning them with ambiguous and underspecified natural language queries. We propose UGuideRAG (User-Generated Content-Guided RAG), a modular framework that leverages user-generated content to construct a comprehensive attraction database, employs large language models for intent-enhanced retrieval and recommendation, and incorporates spatial optimization to ensure coherent itinerary planning. By bridging the gap between partially expressed user goals and the multi-dimensional nature of urban experiences, UGuideRAG enables more insightful and personalized trip recommendations. Experiments on real-world datasets demonstrate that our framework consistently surpasses existing methods in producing contextually relevant, user-centered, and spatially optimized urban tourism itineraries. Source codes are available at https://github.com/tangjsysu/UGuideRAG

CCS Concepts

• **Information systems** → **Web mining**; Recommender systems;
• **Computing methodologies** → *Natural language processing*.

Keywords

Urban Tourism, Recommender Systems, Large Language Models, User-Generated Content, Travel Personalization

ACM Reference Format:
Jing Tang, Inhye Kong, and Zhaonan Wang. 2025. UGuideRAG: Intent-Enhanced Retrieval-Augmented Generation with User-Generated Content for Personalized Urban Tourism. In *Proceedings of 33rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2025) (SIGSPATIAL '25)*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3748636.3762712

1 Introduction

Citywalk has become an increasingly popular form of urban tourism, defined as "a recreational activity including strolling across metropolitan regions to acquire certain experiences while engaging in behaviors that seek diversity" [1]. Unlike traditional sightseeing, citywalk prioritizes immersive, self-guided exploration and spontaneous engagement with local culture. Tourists in this context seek experiences that are emotionally resonant, perception-rich, and contextually embedded, rather than merely visiting top-rated attractions [2].

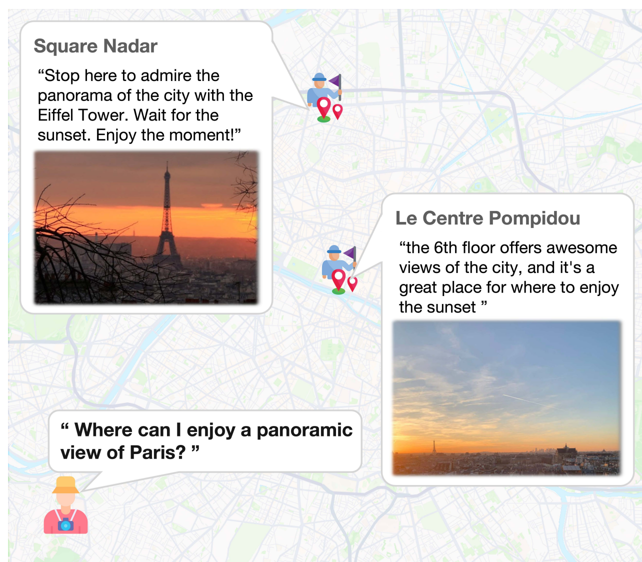


Figure 1: *"Where can I enjoy a panoramic view of Paris?"* This figure illustrates how user-generated content can reveal hidden scenic viewpoints that are beyond guidebooks and 2D maps.[1]

In response to the growing demand for more dynamic, personalized, and diverse tourism experiences, user-generated content (UGC) has emerged as a vital source of travel information. Unlike official sources, UGC offers authenticity, emotional depth, and localized insights that help tourists discover hidden or underrepresented

*This work was conducted during a research assistantship at NYU Shanghai, where the author is also affiliated with Shanghai Key Laboratory of Urban Design and Urban Science.
†Corresponding author, also affiliated with NYU Shanghai Center for Data Science and Shanghai Key Laboratory of Urban Design and Urban Science.

[1]Photo sources: https://maps.app.goo.gl/fMibardwj3SBE1uM8, https://maps.app.goo.gl/4i48B3rnDuAXVXQf6.

SIGSPATIAL '25, Minneapolis, MN, USA

ACM ISBN 979-8-4007-2086-4/2025/11
https://doi.org/10.1145/3748636.3762712

places [3–5]. As illustrated in Figure 1, UGC can reveal detailed and hidden aspects of visitor attractions (VAs) that are often missing from official descriptions or curated travel guides. For instance, while Le Centre Pompidou is widely known for its modern art exhibitions, UGC highlights an alternative facet — its rooftop being appreciated as a scenic viewpoint. At the same time, user reviews surface lesser-known places such as Square Nadar, which offers panoramic views but rarely appears in conventional itineraries. These examples demonstrate how UGC helps uncover both the subtle characteristics of well-known attractions and lesser-known spots in the city.

While UGC reveals rich and experiential knowledge about VAs, effectively incorporating this information into personalized travel planning remains a significant challenge. In the context of city-walk scenarios, travelers often seek curated one-day itineraries that balance iconic landmarks with local discoveries. UGC holds great potential to support such route recommendations by surfacing nuanced, experience-driven insights that are typically absent from official descriptions. However, existing UGC-based recommendation methods often fall short in leveraging this rich information. Earlier approaches primarily relied on shallow text-mining techniques such as Latent Dirichlet Allocation (LDA) and other statistical approaches, which extract only salient keywords while ignoring contextual and perceptual depth [6, 7]. As a result, these methods struggle to represent attraction features comprehensively and fail to align with the multifaceted and detailed preferences of travelers.

Recent advances in large language models (LLMs) have shown new potential to improve intent understanding and semantic matching in tourism recommendations. Systems such as *ITINERA* [8] leverage LLMs to parse natural language queries into structured sub-requirements and retrieve relevant attractions through semantic matching. While these systems represent a major step forward, they still face challenges in handling user inputs that are often ambiguous, incomplete, and highly faceted [9–11]. As a result, the alignment between partially expressed user intent and the complex, multi-dimensional features of VAs remains limited.

To address these challenges, we propose *UGuideRAG* (User-Generated Content-Guided Retrieval-Augmented Generation), a modular recommendation framework designed for personalized and fine-grained urban tourism. UGuideRAG consists of four components: (1) *UGC-based Attraction Database Construction (UADC)*, which aggregates and structures UGC to enrich VAs with descriptive, experiential, and contextual information that goes beyond official categorizations; (2) *Intent-Enhanced Retriever (IER)*, which decomposes user queries into structured intents across experiential dimensions using LLMs and retrieves semantically relevant content; (3) *LLM-based Reranker (LRR)*, which scores retrieved candidates based on their semantic relevance to the user query; and (4) *Cluster-aware Spatial Optimization (CSO)*, which constructs personalized and spatially coherent itineraries for urban travel.

Our overall contributions are as follows:

(1) Grounded in tourism research, we define a set of perception-aligned attraction features comprising *landscape and content*, *activities*, and *atmosphere*, and employ LLMs to extract these structured features from unstructured user-generated content, providing the data foundation for personalized recommendations.
(2) We propose an intent-enhanced RAG architecture, in which the retrieval module is guided by LLM-based decomposition of user queries into structured intents across multiple experiential dimensions. Retrieved candidates are then re-ranked using an LLM based on their alignment with the user query, enhancing semantic precision while supporting more personalized and diverse itinerary generation.
(3) We conduct extensive experiments across multiple cities, demonstrating that UGuideRAG generates personalized and spatially coherent itineraries that outperform existing baselines in urban travel recommendations.

2 Related Work

2.1 User-Generated Content in Tourism

User-generated content (UGC), such as reviews, photos, and blogs, serves as useful information in tourism planning, significantly influencing travel decision-making processes [12, 13]. UGC provides first-person descriptions, opinions, and multimedia information related to destinations, services, and points of interest (POIs) [14]. Compared to official promotional materials or market-generated content (MGC), UGC captures a richer spectrum of the tourist experience, offering experiential detail, emotional depth, as well as local and hidden insights [15]. Analyzing UGC enables the understanding of tourists' perceptions, preferences, satisfaction levels, and emerging trends [16, 17].

Traditional UGC analysis methods include sentiment analysis, keyword extraction, and topic modeling techniques such as LDA and statistical language models [18, 19]. These models help uncover important terms and latent themes within tourism-related UGC [20, 21]. For example, Liang et al. employed LDA to extract thematic tags from UGC and construct a feature tag library, which was then used to match attractions with user preferences based on fuzzy label similarity [6]. Missaoui et al. developed a mobile tourism recommender system that builds multi-layer user profiles using statistical language models derived from UGC and matches them with attraction profiles through content-based filtering and contextual pre-filtering [7].

However, these methods still face limitations in capturing deep perceptual understanding. They tend to generate coarse-grained topics (e.g., "beach," "service") and struggle to distinguish subtle perceptual qualities such as "tranquil, secluded beach" versus "busy, crowded beach" [22–24]. Rooted in the bag-of-words assumption, such models often overlook contextual subtleties and deeper semantic meanings that are crucial for interpreting user perception [25]. Progress in tourism recommendation systems that aim for deep experience matching depends largely on advances in extracting and representing fine-grained, context-aware features of VAs from UGC [26]. The emergence of LLMs has made it increasingly feasible to capture more comprehensive and nuanced representations of attractions, enabling more accurate and personalized tourism recommendations [27].

2.2 RAG in Recommendation Systems

In recent years, LLMs have shown advancement in understanding and processing natural language. However, challenges such as hallucinations [28] and inefficiencies in fine-tuning [29] continue to affect their reliability in real-world applications. One promising solution is Retrieval-Augmented Generation (RAG), which combines external information retrieval with generative modeling to enrich input representations and improve the quality of generated content [30].

RAG has been widely adopted for its strong ability to interpret user needs expressed in natural language [31], and it has demonstrated great potential in modeling user preferences and delivering personalized recommendations [32–35]. For example, Di Palma proposed a simple RAG-based recommendation model that leverages structured knowledge from movie and book datasets to enhance recommendation relevance [32]. Yu et al. introduced Spatial-RAG, an extension of the RAG framework that integrates both semantic and spatial retrieval to support spatial reasoning tasks, enabling LLMs to generate geographically grounded and contextually relevant responses based on user preferences and real-world spatial constraints [35].

The RAG framework typically adopts a dual-module architecture consisting of a retrieval module and a reader module, which jointly improve the relevance and informativeness of generated outputs. However, the effectiveness of the retrieval component is often hindered by ambiguous or underspecified user queries, leading to suboptimal retrieval results and degraded overall performance. Recent research has shown that rewriting and expanding user intent representations within input prompts can significantly enhance RAG's performance by improving retrieval quality and alignment with user needs [36, 37].

This issue is particularly pronounced in tourism recommendation scenarios, where user demands extend beyond simple keywords to include nuanced expectations for experiences, emotional responses, and environmental contexts [38]. While previous efforts, such as Tang's method of extracting positive and negative query components and computing embedding similarities for POI recommendation, have shown initial success, they often fall short in capturing the full breadth of user expectations [8]. This results in imprecise POI retrieval and limited recommendation diversity. These challenges highlight the importance of developing methods that better capture and represent implicit user intent to support personalized recommendations in complex, experience-driven domains such as tourism.

3 Problem Formulation

We define the personalized urban itinerary recommendation task as follows.

Let $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ denote the set of all available VAs in a given city. Each attraction $v_i \in \mathcal{V}$ is associated with experiential features primarily derived from UGC.

Given a natural language user query q, the goal is to generate a personalized and spatially coherent one-day itinerary:

$$\mathcal{V}_{\text{order}} = [v_{o_1}, v_{o_2}, \ldots, v_{o_M}]$$

where: $v_{o_i} \in \mathcal{V}$, $M \geq n_{\min}$ denotes the minimum number of attractions required for a feasible one-day itinerary, and the sequence is optimized for both semantic alignment with q and spatial efficiency.

This task presents three major challenges: 1) How to extract structured, perception-aligned features from noisy, unstructured UGC for each $v_i \in \mathcal{V}$. 2) Given a free-form user query q, how to retrieve and rank a candidate subset $\mathcal{V}_{\text{top-}k} \subset \mathcal{V}$ that is semantically aligned with user preferences. 3) How to select and order a final subset $\mathcal{V}_{\text{order}} \subset \mathcal{V}$ that forms an itinerary aligned with user intent while remaining spatially coherent.

4 Methodology

4.1 UGC-based Attraction Database Construction

4.1.1 Modeling the Core Experience of Visitor Attractions.

According to Pearce's definition, an attraction is a "named site with a specific human or natural feature which is the focus of visitor and management attention" [39]. When visiting a destination, various attributes or features, often referred to as pull factors shape tourists' travel experiences [40, 41]. These factors can be categorized into three main types: Physical Environment, Service Quality, and Core Experience [42].

Here, we focus on VAs, where the Core Experience plays a crucial role in attracting tourists and shaping their travel experiences. It includes elements such as content and presentation [42], entertainment, fun, emotions, atmosphere [43], novelty [44], and authenticity [45], all of which define the fundamental visitor experience. This study synthesizes these aspects into three key dimensions of the VA core experience: landscape and content, activities, and atmosphere.

4.1.2 LLM-based Feature Extraction and Attraction Representation.

According to Leask [46], VAs can be categorized into types such as theme parks and amusement venues, museums and galleries, natural sites, animal-related locations, visitor centers, religious sites, and heritage sites. Following this taxonomy, a comprehensive list of VAs was compiled using data retrieved from Google Maps[2] and supplemented with additional entries from TripAdvisor[3] and OpenStreetMap[4] to ensure coverage of sites that may not be listed on a single platform.

User-generated reviews for each VA are collected from Google Maps using Selenium[5], sorted by relevance to prioritize informative and detailed content. These reviews contain information such as ratings, narratives, and emotional expressions, offering valuable insights into tourist perceptions, satisfaction levels, and site-specific features.

For each VA v_i, we prompt LLMs to analyze its collected reviews R_i, extracting structured textual features across the three experiential dimensions defined in Section 3.1.1: landscape and content f_i^{lan}, activities f_i^{act}, and atmosphere f_i^{atm}, along with a general summary d_i describing the overall character of the site.

To illustrate this process, we present an example for the *Pont Neuf* in Paris. Table 1 shows selected user reviews, which reflect visitor attention to scenic views, nearby landmarks, and atmospheric qualities. Based on these reviews, LLMs extract structured experiential features summarized in Table 2, revealing landscape attributes (e.g., river vistas and historical architecture), common activities (e.g.,

[2]https://www.google.com/maps
[3]https://www.tripadvisor.com/
[4]https://www.openstreetmap.org/
[5]https://www.selenium.dev/

Table 1: Selected Google Map Reviews for Pont Neuf with Ratings

Rating	Review
5.0	*"Pont Neuf is a beautiful destination to visit in the evening, offering stunning views of the city and the Seine River. As the sun begins to set, the lights of the city come to life, casting a romantic and picturesque ambiance on the bridge. At night, Pont Neuf is illuminated, providing a beautiful backdrop for a romantic stroll or a relaxing evening walk."*
5.0	*"Walking around Paris is one of the best activities one can do when there. This is an amazing sunset spot by the Seine river. Very close to both Notre Dame and Louvre museum. Highly recommend walking around the area and soaking in Paris. Also a great picnic spot near the river."*
4.0	*"Built in 1607 and still look great and solid and probably the most picturesque of all the Parisian bridges. It is made of two spans due to small island in between. This is also where you can go for a boat cruise near the very top of the island. Nice to get views on both sides of the Seine."*

Table 2: LLM-Extracted Experiential Features for Pont Neuf

Dimension	Extracted Feature Description
Landscape & Content	Oldest stone bridge in Paris with iconic Seine River views, nearby parks, and historic features like the Henri IV statue. Features scenic vistas of landmarks like Notre Dame and Eiffel Tower.
Activities	Walking, river cruises, photography, sunset viewing, sightseeing landmarks, and boarding Vedettes tour boats.
Atmosphere	Historic yet vibrant, blending romantic charm with lively crowds. Offers peaceful spots for relaxation amid bustling artistic and cultural energy.

walking, sunset watching, river cruises), and perceptual atmosphere (e.g., romantic).

All textual features are encoded using an embedding model $\psi(\cdot)$, producing the following embeddings:

$$\begin{aligned} e_i^{\text{lan}} &= \psi(f_i^{\text{lan}}), \quad e_i^{\text{act}} = \psi(f_i^{\text{act}}), \\ e_i^{\text{atm}} &= \psi(f_i^{\text{atm}}), \quad e_i^{\text{des}} = \psi(d_i) \end{aligned} \tag{1}$$

These dimension-specific representations are stored as part of the attraction embedding database:

$$\mathcal{V} = \left\{ \left(e_i^{\text{lan}}, e_i^{\text{act}}, e_i^{\text{atm}}, e_i^{\text{des}} \right) \right\}_{i=1}^{N} \tag{2}$$

This structured database captures the nuanced characteristics of each VA across multiple experiential facets and serves as the foundation for semantic retrieval, re-ranking, and itinerary construction in downstream modules.

4.2 Intent-Enhanced Retriever

As previously discussed, to address the challenges posed by ambiguous, incomplete, or semantically faceted user queries, we propose a retrieval module that performs intent decomposition and structured semantic alignment. Leveraging the reasoning capabilities of LLMs, each user query q is parsed into three intent components corresponding to core dimensions of attraction experience: expected *landscape and content* (r^{lan}), *activities* (r^{act}), and *atmosphere* (r^{atm}).

Each intent component $r^d \in \{r^{\text{lan}}, r^{\text{act}}, r^{\text{atm}}\}$ is then projected into the embedding space using the same embedding model $\psi(\cdot)$ employed for VA feature representation, yielding separate query embeddings for each experiential dimension $\varphi^d = \psi(r^d)$. Correspondingly, each VA $v_i \in \mathcal{V}$ is represented by a tuple of embeddings $\{e_i^{\text{lan}}, e_i^{\text{act}}, e_i^{\text{atm}}\}$, which encode its semantic profile across the three experiential dimensions.

We compute the cosine similarity for each dimension d as:

$$\cos(\varphi^d, e_i^d) = \frac{\varphi^d \cdot e_i^d}{\|\varphi^d\| \cdot \|e_i^d\|} \tag{3}$$

Using this, the composite relevance score for each candidate VA is defined as:

$$\begin{aligned} \text{Score}_i = & \; w_{\text{lan}} \cdot \cos(\varphi^{\text{lan}}, e_i^{\text{lan}}) \\ & + w_{\text{act}} \cdot \cos(\varphi^{\text{act}}, e_i^{\text{act}}) \\ & + w_{\text{atm}} \cdot \cos(\varphi^{\text{atm}}, e_i^{\text{atm}}) \end{aligned} \tag{4}$$

Here, $w_{\text{lan}}, w_{\text{act}}, w_{\text{atm}} \in [0, 1]$ control the relative contribution of each experiential dimension.

By performing this dimension-aware matching, the retriever is able to more robustly align user intent with semantically rich and structurally organized attraction profiles—thereby improving recall for nuanced or under-specified queries. The final ranked list is obtained by computing Score_i for all $i \in \{1, \ldots, N\}$, and selecting the top-k candidates:

$$\mathcal{V}_{\text{top-}k} = \text{Top-}k \left(\{\text{Score}_i\}_{i=1}^{N} \right) \tag{5}$$

The resulting set $\mathcal{V}_{\text{top-}k} \subset \mathcal{V}$ serves as the input to the subsequent re-ranking stage, where contextual reasoning is applied via LLMs to refine semantic alignment and preference fit.

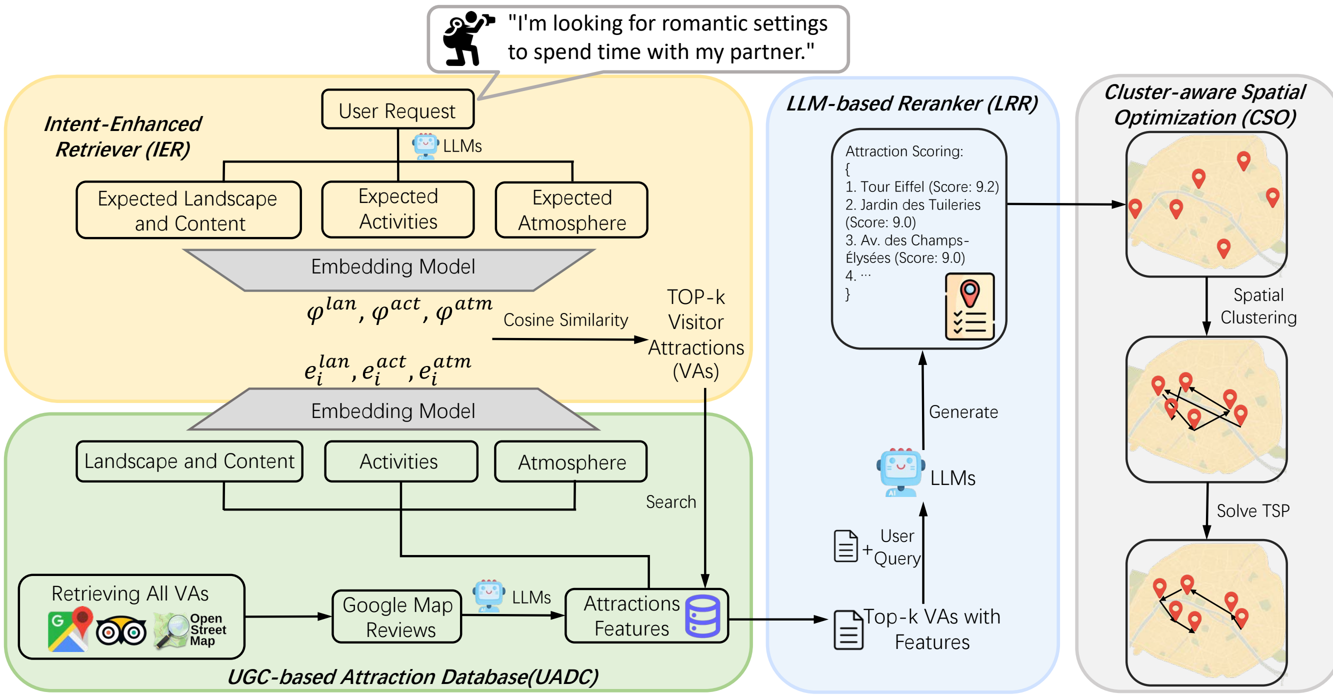


Figure 2: Illustration of the proposed *UGuideRAG* framework

4.3 LLM-based Re-ranking of Retrieved VAs

While embedding-based retrieval provides a coarse-grained semantic alignment between structured user intent and candidate attractions, it lacks the capacity to perform fine-grained contextual reasoning. To address this, we introduce a second-stage re-ranking module that leverages the inference capabilities of LLMs to evaluate each retrieved candidate in the full context of the original query.

Given the Top-k retrieved VAs $\mathcal{V}_{\text{top-}k}$, we construct a natural language prompt for each candidate $v_i \in \mathcal{V}_{\text{top-}k}$ that integrates: (1) the user's original query q; and (2) the structured attribute descriptions of v_i, including its landscape and content features f_i^{lan}, activities f_i^{act}, and atmosphere f_i^{atm}. These prompts are passed to the LLM, which performs context-aware semantic matching between the user query and each candidate's experiential attributes.

Formally, the LLM outputs a contextual alignment score $s_i^{\text{LLM}} \in [0, 10]$, representing the degree to which the candidate satisfies the user's latent preferences as expressed in natural language. This rescoring process enables reasoning over implicit user goals, complex lexical variations, and nuanced feature combinations that are often poorly represented in fixed vector spaces.

The re-ranked list $\mathcal{V}_{\text{rerank}}$ is obtained by sorting the candidates in descending order of s_i^{LLM}. This re-ranking stage enhances the semantic fidelity and personalization of the final recommendation results, bridging the gap between discrete feature embeddings and holistic user intent understanding.

4.4 Cluster-Aware Spatial Optimization

To ensure that the recommended VAs form a spatially coherent and walkable itinerary, we introduce a two-step cluster-aware optimization process. The first step selects geographically compact VA groups via spatial clustering, while the second step optimizes the visiting order of selected VAs using a genetic algorithm to minimize travel distance.

Step 1: Spatial Clustering for VA Selection. Given the Top-k candidate attractions ranked by semantic relevance, we apply an incremental clustering algorithm that evaluates each VA based on its proximity to existing clusters. A new VA is assigned to a cluster if it lies within a specified distance threshold τ of any member in that cluster; otherwise, a new cluster is created. Clusters with fewer than n_{cmin} members are discarded. The process continues until the number of VAs in valid clusters exceeds a minimum threshold n_{min}. This approach ensures that only sufficiently dense and spatially compact clusters are retained, such that each valid cluster contains at least n_{cmin} VAs within a walking distance threshold τ, allowing users to conveniently visit them on foot. Travel between clusters, in contrast, can then be planned using alternative transportation modes, thereby reducing the overall travel burden. The clustering procedure is outlined in Algorithm 1.

Step 2: Genetic Algorithm for VA Ordering. To determine an optimal visiting order among the selected candidate VAs, we employ a genetic algorithm that minimizes the path length between locations. The population is initialized with random permutations of the VA list. In each generation, individuals are evaluated using a fitness

Algorithm 1 Spatial Clustering for VA Selection

Require: Sorted list of VAs $\mathcal{V}_{\text{rerank}} = \{v_1, v_2, \ldots, v_k\}$ by LLM score, distance threshold τ, minimum total VAs n_{min}, minimum cluster size n_{cmin}
Ensure: Candidate VAs list $\mathcal{V}_c$

$C \leftarrow [\ ]$ {Initialize empty list of clusters}
$\mathcal{V}_{\text{rerank}} \leftarrow \emptyset$
for $i = 1$ to k **do**
  $v \leftarrow v_i$
  assigned $\leftarrow$ **false**
  for all $C_j \in C$ **do**
    if $\exists v' \in C_j$ such that $\text{dist}(v, v') < \tau$ **then**
      $C_j \leftarrow C_j \cup \{v\}$; assigned $\leftarrow$ **true**
      break
    end if
  end for
  if not assigned **then**
    $C \leftarrow C \cup \{\{v\}\}$ {Create new cluster}
  end if
  $\mathcal{V}_c \leftarrow \bigcup\{C \in C : |C| \geq n_{cmin}\}$
  if $|\mathcal{V}_c| \geq n_{min}$ **then**
    break
  end if
end for
return $\mathcal{V}_c$

Algorithm 2 Genetic Algorithm for VA Ordering

Require: Candidate VAs $\mathcal{V}_c$, distance matrix $\mathcal{D}$
Ensure: Ordered list of candidate VAs $\mathcal{V}_{\text{order}}$

$\mathcal{P} \leftarrow \{P_1, P_2, \ldots, P_g\}$ {Initialize population}
$t \leftarrow 0$ {Initialize the generation count}
while $t < t_{\max}$ **do**
  for $i = 1$ to g **do**
    fitness(P) {Calculate the fitness score for each P_i in $\mathcal{P}$}
  end for
  for $j = 1$ to $\frac{g}{2}$ **do**
    $P_a, P_b \leftarrow$ selection() {Select two parent routes P_a, P_b based on their fitness}
    $C_a, C_b \leftarrow$ crossover(P_a, P_b) {Crossover between parents to generate children}
    $C_a, C_b \leftarrow$ mutation(C_a, C_b) {Apply mutation to children to introduce variability}
    $P_{\text{new}} \leftarrow$ add(C_a, C_b) {Add C_a, C_b to a new population P_{new}}
  end for
  $t \leftarrow t + 1$ {Increment generation count}
end while
$\mathcal{V}_{\text{order}} \leftarrow P_{\text{best}}$ {Return the best route P_{best} based on the highest fitness score}

function based on the total distance traveled. Selection, crossover, and mutation operations are applied iteratively to evolve better route candidates. The algorithm terminates after a fixed number of generations, and the best individual is returned as the optimized sequence. The complete procedure is shown in Algorithm 2.

This two-step spatial optimization not only ensures spatial coherence but also enhances user experience by promoting smooth and walkable exploration.

5 Experiments

5.1 Experiments Setting

5.1.1 *Experimental Cities.*

To evaluate the proposed recommendation framework in realistic urban tourism settings, we conducted experiments on two culturally rich European destinations: Paris and the Rome–Vatican region. A total of 981 VAs were collected in Paris, while 867 VAs were collected for the Rome–Vatican area. For each attraction, user reviews from Google Maps were also collected to extract descriptive and perceptual features, which were then used to construct a semantically enriched attraction database.

5.1.2 *User Queries Generation.*

To simulate diverse user intentions grounded in psychological, cultural, and social motivations, we adopt a multidimensional framework for generating natural language queries. This framework supports two complementary generation strategies, each rooted in established tourism typologies and contextualized through travel companionship and urban cultural profiles.

First, we draw upon Elands and Lengkeek's [47] refinement of Cohen's [48] tourist experience theory. Their work provides a detailed typology of *modes of experience* in tourism, outlining a spectrum of motivations—*Amusement*, *Change*, *Interest*, *Rapture*, and *Dedication*—that represent distinct experiential orientations. These motivational profiles were systematically paired with travel companion contexts (e.g., *alone*, *with a partner*, *with young or older children*, or *with friends without children*) and grounded in the cultural and experiential characteristics of real-world cities to simulate general leisure and meaning-seeking user queries in urban environments.

Second, recognizing that cultural tourism constitutes a significant subset of urban tourism, we incorporate McKercher 's [49] typology of cultural tourists. This model distinguishes five types of cultural tourists—*Purposeful*, *Serendipitous*, *Sightseeing*, *Casual*, and *incidental*—based on the centrality of cultural motivations and the depth of cultural engagement. These types were similarly paired with travel companion profiles and enriched with city-specific cultural assets to guide LLMs in simulating user queries reflecting diverse forms of culturally oriented intent.

The user query generation process is illustrated in Figure 3, which summarizes how motivational typologies, cultural intent categories, travel context, and cultural and experiential characteristics of cities were combined to construct a semantically diverse and realistic set of user queries. In our generated queries, 35.3% of the Paris queries and 36.1% of the Rome-Vantican queries were derived from the cultural tourist typology, with the remainder grounded in the mode of experience framework.

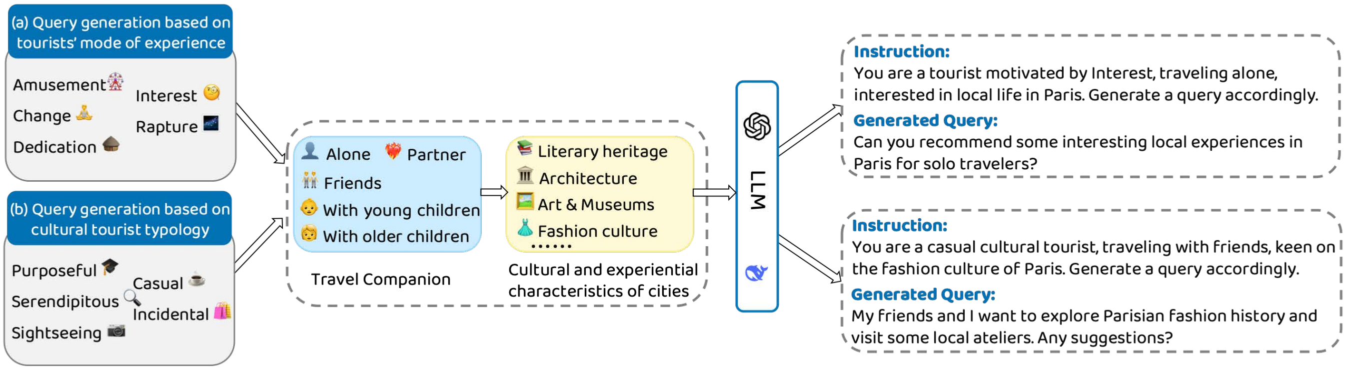


Figure 3: An overview of query generation based on tourists' motivation types, travel companions, and urban characteristics. Two generation pathways are illustrated: (a) experience-based typology and (b) cultural tourist typology. These factors are composed into LLM prompts to simulate diverse queries.

5.2 Evaluation Metrics

We adopt a combination of semantic and spatial metrics to evaluate the relevance, efficiency, and spatial coherence of each generated itinerary. Let $\mathcal{V}_{\text{order}} = \{v_1, v_2, \ldots, v_N\}$ denote the ordered set of selected attractions in a given itinerary, and let $\mathbf{q}$ denote the user query. The Euclidean distance between two attractions v_i and v_j is denoted by $d(v_i, v_j)$.

Hit Rate (HR). Hit Rate measures the proportion of attractions in the itinerary that are semantically relevant to the user query $\mathbf{q}$:

$$\text{HR} = \frac{1}{|\mathcal{V}_{\text{order}}|} \sum_{v \in \mathcal{V}_{\text{order}}} \mathbb{I}\big[\text{Relevant}(v, \mathbf{q}) = 1\big], \tag{6}$$

where $\mathbb{I}[\cdot]$ is the indicator function. Relevance is assessed via LLM judgement and verified through human annotation.

Average Margin (AM). Average Margin measures the difference in total Euclidean distance between the generated itinerary and the optimal Traveling Salesman Problem (TSP) solution over the same set of attractions:

$$\text{AM} = D(\mathcal{V}_{\text{order}}) - D^*(\mathcal{V}_{\text{order}}), \tag{7}$$

where $D(\cdot)$ denotes the total distance of the visiting order, and $D^*(\cdot)$ is the optimal TSP distance over the same set.

Travel Distance (TD). Travel Distance is the total Euclidean distance incurred when visiting attractions in the recommended order:

$$\text{TD} = \sum_{i=1}^{N-1} d(v_i, v_{i+1}). \tag{8}$$

Spatial Tightness (ST). Spatial Tightness measures how spatially clustered the selected attractions are, regardless of their visiting order:

$$\text{ST} = \frac{1}{N} \sum_{i=1}^{N} \min_{j \neq i} d(v_i, v_j). \tag{9}$$

5.3 Results

We evaluate the full UGuideRAG framework against ITINERA[6] [8], a recent LLM-based itinerary recommendation baseline. Table 3 presents a comparison of the two systems on the Paris and Rome-Vatican datasets.

In terms of semantic alignment, UGuideRAG achieves significantly higher HR in both cities—78.5% in Paris and 72.7% in Rome-Vatican—compared to ITINERA (42.3% and 33.8%, respectively). These results indicate a stronger match between the user query and the recommended VAs.

For spatial metrics, both systems achieve nearly identical AM values, suggesting comparable efficiency in visiting order relative to the optimal TSP baseline. Despite variations across cities, UGuideRAG maintains TD values around 6000 meters, which translates to a feasible walking distance for a day itinerary, ensuring practical usability for urban tourists. Additionally, UGuideRAG consistently achieves low ST values across both cities, indicating that the recommended attractions are geographically well-clustered and exhibit strong walkable connectivity.

Taken together, these findings demonstrate that UGuideRAG delivers substantially improved semantic relevance while maintaining competitive spatial performance. This highlights its potential to improve user satisfaction through context-aware itinerary recommendations without imposing additional travel burden.

5.4 Ablation Study

To assess the individual contributions of each module in UGuideRAG framework, we conduct an ablation study on both the Paris and Rome-Vatican datasets (Table 4). We examine four ablated variants: (1) without intent decomposition and UGC-derived VA features (w/o Intent Decomposition & UGC), (2) without intent decomposition (w/o Intent Decomposition), (3) without the LLM-based reranker (w/o LRR), and (4) without cluster-aware spatial optimization (w/o CSO), keeping all other components intact. In addition, we also compare against two modified variants of the ITINERA. Since ITINERA's original density-based clustering is not well-suited for high-density

[6]Results are obtained using the authors' original implementation.

Table 3: Comparison between *UGuideRAG* and *ITINERA* across the Paris and Rome-Vatican datasets.

Methods	Paris				Rome-Vatican			
	HR↑ (%)	AM↓ (m)	TD↓ (m)	ST↓ (m)	HR↑ (%)	AM↓ (m)	TD↓ (m)	ST↓ (m)
ITINERA	42.3	**455.8**	**5816.3**	441.4	33.8	**446.0**	5367.3	413.5
UGuideRAG	**78.5**	651.2	6781.4	**301.8**	**72.7**	582.4	**5227.1**	**231.8**

Table 4: Performance comparison across different ablation settings on Paris and Rome-Vatican datasets.

City	Method	HR↑ (%)	AM↓ (m)	TD↓ (m)	ST↓ (m)
Paris	**UGuideRAG (Full)**	78.5	651.2	6781.4	**301.8**
	w/o Intent Decomposition & UGC	52.0	805.7	6752.1	358.5
	w/o Intent Decomposition	66.9	**539.2**	6370.7	329.3
	w/o LRR	66.4	658.9	6249.9	310.7
	w/o CSO	**80.1**	16424.4	36592.7	1535.7
	ITINERA (w/ UGuideRAG's CSO, w/o LRR)	59.0	651.4	**6001.1**	310.2
	ITINERA (w/ UGuideRAG's CSO and LRR)	72.6	627.0	6423.6	312.0
Rome-Vatican	**UGuideRAG (Full)**	**72.7**	582.4	5227.1	231.8
	w/o Intent Decomposition & UGC	52.7	**369.6**	**3990.1**	**231.0**
	w/o Intent Decomposition	64.0	668.3	5476.3	255.5
	w/o LRR	63.4	563.3	4736.5	242.9
	w/o CSO	72.1	12035.4	21998.2	815.6
	ITINERA (w/ UGuideRAG's CSO, w/o LRR)	56.7	512.6	5702.8	267.5
	ITINERA (w/ UGuideRAG's CSO and LRR)	65.5	669.8	5064.2	259.2

VA regions such as Paris and Rome-Vatican, we re-implement ITINERA using UGuideRAG's clustering strategy: (5) ITINERA with UGuideRAG's CSO only, and (6) ITINERA with UGuideRAG's CSO and LRR.

The w/o Intent Decomposition & UGC variant relies on each VA's Wikipedia [7] summary for attraction matching, without leveraging structured user intent and UGC-derived VA features. It operates on a reduced attraction pool due to the limited availability of Wikipedia descriptions (454 VAs for Paris and 467 for Rome). This setting yields the lowest HR across both cities, with a notable performance drop compared to the w/o Intent Decomposition variant. These results further highlight the foundational importance of extracting rich experiential VA features from UGC for personalized, fine-grained recommendations.

When the LLM-based reranker module is removed, the system experiences a noticeable drop in recommendation performance, highlighting the importance of fine-grained contextual ranking. While intent decomposition ensures that retrieval broadly aligns with user intent, removing the LLM-based reranker limits the system's ability to distinguish fine-grained semantics beyond what embeddings can represent.

Removing the intent decomposition module leads to a significant drop in HR, as the system fails to infer user intent from multi-faceted, ambiguous, or implicit queries. This degrades retrieval quality and limits the effectiveness of downstream LLM-based reranking.

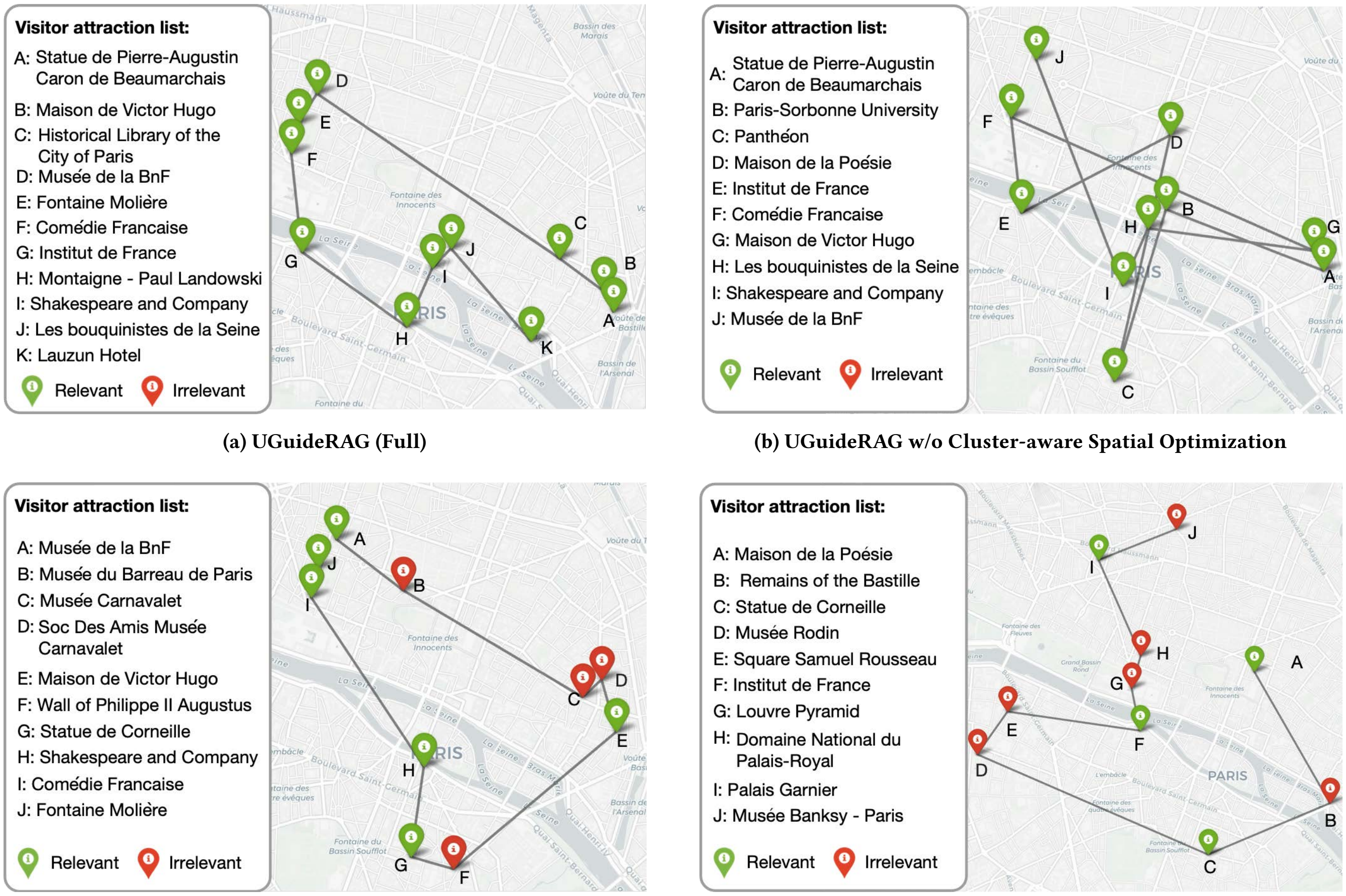
[7] https://en.wikipedia.org/

Although HR remains high, removing the CSO module results in severe degradation of spatial metrics. In particular, AM increases by over 25×, while TD and ST increase by approximately 5×. This indicates that although the selected attractions are semantically relevant, they are spatially scattered and inefficiently ordered. Therefore, CSO is essential for producing spatially coherent and walkable itineraries.

By replacing ITINERA's original CSO method with that of UGuideRAG, ITINERA achieves notable improvements in the semantic relevance of recommended VAs while simultaneously yielding lower spatial tightness values, indicating more compact and walkable clusters. This demonstrates that UGuideRAG's spatial optimization strategy is particularly effective in attraction-dense cities such as Paris and Rome-Vatican, as it preserves walking feasibility while ensuring closer semantic alignment with user queries.

Integrating the LRR module onto this enhanced baseline dramatically boosts ITINERA's Hit Rate by 13.6pp in Paris and 8.8pp in Rome-Vatican. These gains are comparable to those seen in UGuideRAG (12.1pp and 9.3pp, respectively), demonstrating the LRR's versatility in refining semantic relevance. Crucially, the persistent performance gap between the systems confirms that UGuideRAG's primary competitive advantage stems from its intent-aware retrieval, which provides a stronger set of initial candidates for the downstream modules to optimize.

Together, these results demonstrate that the effectiveness of UGuideRAG arises from the complementary contributions of all its



(a) UGuideRAG (Full)

(b) UGuideRAG w/o Cluster-aware Spatial Optimization

(c) UGuideRAG w/o LLM-based Reranker

(d) ITINERA

Figure 4: Case study comparison of recommended attractions across methods for the query *"I'm interested in French literature"*.

modules. UGC-derived VA features provide rich semantic signals, intent decomposition ensures accurate retrieval, and the LLM-based reranker refines results with fine-grained contextual reasoning. Meanwhile, the CSO module achieves a balance between semantic relevance and travel burden, producing coherent and walkable itineraries. Each component is indispensable, and only their integration delivers recommendations that are both semantically aligned and practically feasible.

5.5 Case Study

To further demonstrate the effectiveness of our framework, we present a case study based on the user query: *"I'm interested in French literature. What places do you recommend?"* We compare the outputs of four systems previously introduced in the ablation study: the full **UGuideRAG**, its variants **w/o Cluster-aware Spatial Optimization** and **w/o LLM Reranker**, and the baseline **ITINERA**.

Figure 4 shows the recommended itineraries generated by each method. The selected VAs are labeled alphabetically (A–K), with names listed in each subfigure's legend. Detailed descriptions of all VAs will be included in the supplementary materials.

The full UGuideRAG framework produces the most semantically aligned and diverse itinerary. It identifies a rich mix of attractions closely related to the theme of French literature, including iconic author residences such as *Maison de Victor Hugo* and *Lauzun Hotel* (associated with Baudelaire), as well as sculptures and monuments dedicated to French playwrights, including *Fontaine Molière* and the *Statue de Beaumarchais*. The itinerary also features cultural landmarks like the *Musée de la BnF*, the *Institut de France*, and the historic literary theater *Comédie Française*, along with experiential VAs such as the riverside secondhand book market *Les bouquinistes de la Seine* and the renowned bookstore *Shakespeare and Company*. These results highlight UGuideRAG's strength in identifying attractions related to French literary culture, including historic author residences, public monuments, national literary institutions, and reader-focused VAs such as secondhand book markets and independent bookstores. The resulting itinerary combines well-established landmarks with immersive experiences, offering a coherent and multifaceted exploration of the literary landscape of the city.

Removing the CSO module does not significantly alter the set of selected attractions but results in a disorganized and spatially scattered itinerary. The absence of spatial coherence highlights

CSO's essential role in optimizing the visit order and improving overall travel feasibility without sacrificing semantic alignment.

The variant without the LRR still benefits from intent decomposition and successfully retrieves many relevant sites, including *Victor Hugo's house*, *Shakespeare and Company*, *Fontaine Molière*, *Statue de Corneille*, *Comédie Française*, and the *Musée de la BnF*. However, it also includes more marginally relevant or thematically ambiguous places such as the *Musée du Barreau de Paris* and the *Musée Carnavalet*, reflecting a lack of contextual nuance. Despite this, its output is notably more on-topic and diverse than *ItiNera*, suggesting that even in the absence of reranking, structured intent modeling significantly improves semantic relevance in retrieval.

ITINERA, which extracts sub-requirements directly from the user query without performing user intent reasoning, yields the least thematically aligned list. It does include clearly literary venues—such as *Maison de la Poésie*, *Statue de Corneille*, *Institut de France*, and *Palais Garnier*—but many recommendations are only weakly related to literature or off-theme, including *Remains of the Batille*, *Musée Rodin*, *Square Samuel Rousseau*, *Louvre Pyramid*, *Domaine National du Palais-Royal*, *Musée Banksy – Paris*. This outcome highlights a limitation of direct query parsing: although the user intent is clearly stated, the system often returns VAs that are superficially relevant but misaligned with the intended literary theme.

This case illustrates the importance of UGuideRAG's intent decomposition strategy as a key enabling component. By structuring user queries into experiential dimensions—landscape and content, activities, and atmosphere—the system establishes a meaningful foundation for subsequent semantic alignment. However, this potential is fully realized through the addition of the LLM-based reranker module, which enables deep contextual understanding and nuanced evaluation of candidate attractions based on the user's full intent. Together, these components allow UGuideRAG to generate personalized itineraries that are semantically aligned, experientially coherent, and spatially optimized. This results in a more interpretable, engaging and meaningful travel experience.

6 Conclusion

In this paper, we present UGuideRAG, a modular framework for fine-grained and personalized urban tourism recommendation. By leveraging user-generated content and the semantic reasoning capabilities of LLMs, UGuideRAG bridges the gap between ambiguous, perception-driven user queries and the rich experiential features of urban visitor attractions. Through structured intent modeling, semantic reranking, and spatial optimization, our system delivers itineraries that are both thematically aligned and spatially coherent. Extensive experiments across multiple cities demonstrate that UGuideRAG outperforms existing baselines in aligning with user intent and supporting meaningful urban exploration. Potential future directions include extending the framework to multi-day itinerary planning and integrating visual content from UGC to enhance personalization and contextual relevance of recommendations.

Acknowledgments

This work was partially supported by Shanghai Pujiang Program (Grant No. 24PJA094) and Shanghai Key Laboratory of Urban Design and Urban Science, NYU Shanghai Open Topic Grant (No. 2024ZNWang_LOUD).

References

[1] Zhenbin Wang, Hui Zhang, Sridar Ramachandran, and Suiying Cheng. Why are individuals tracing travel trends? a case study of city walk in malaysia. *PloS one*, 20(2):e0309493, 2025.

[2] Dongdong Wu. City walk in a gap day: potential and opportunities for tourism and leisure. *Tourism Review*, 79(9):1576–1581, 2024.

[3] Athena HN Mak. Online destination image: Comparing national tourism organisation's and tourists' perspectives. *Tourism management*, 60:280–297, 2017.

[4] Estela Marine-Roig and Salvador Anton Clavé. Tourism analytics with massive user-generated content: A case study of barcelona. *Journal of destination marketing & management*, 4(3):162–172, 2015.

[5] Nuno Antonio, Marisol B Correia, and Filipa Perdigão Ribeiro. Exploring user-generated content for improving destination knowledge: The case of two world heritage cities. *Sustainability*, 12(22):9654, 2020.

[6] Kaibo Liang, Huwei Liu, Man Shan, Junhui Zhao, Xiaolan Li, and Li Zhou. Enhancing scenic recommendation and tour route personalization in tourism using ugc text mining. *Applied Intelligence*, 54(1):1063–1098, 2024.

[7] Sondess Missaoui, Faten Kassem, Marco Viviani, Alessandra Agostini, Rim Faiz, and Gabriella Pasi. Looker: a mobile, personalized recommender system in the tourism domain based on social media user-generated content. *Personal and Ubiquitous Computing*, 23:181–197, 2019.

[8] Yihong Tang, Zhaokai Wang, Ao Qu, Yihao Yan, Zhaofeng Wu, Dingyi Zhuang, Jushi Kai, Kebing Hou, Xiaotong Guo, Han Zheng, et al. Itinera: Integrating spatial optimization with large language models for open-domain urban itinerary planning. *arXiv preprint arXiv:2402.07204*, 2024.

[9] I. Kostric, K. Balog, and F. Radlinski. Generating usage-related questions for preference elicitation in conversational recommender systems. *ACM Transactions on Recommender Systems*, 2024. Forthcoming.

[10] Kimiya Keyvan and Jimmy Xiangji Huang. How to approach ambiguous queries in conversational search: A survey of techniques, approaches, tools, and challenges. *ACM Computing Surveys*, 55(6):1–40, 2022.

[11] Jiani Huang, Shijie Wang, Liang-bo Ning, Wenqi Fan, Shuaiqiang Wang, Dawei Yin, and Qing Li. Towards next-generation recommender systems: A benchmark for personalized recommendation assistant with llms. *arXiv preprint arXiv:2503.09382*, 2025.

[12] Ulrike Gretzel and Kye-Hyoung Yoo. Use and impact of online travel reviews. In *Information and communication technologies in tourism 2008*, pages 35–46. Springer Vienna, 2008.

[13] Zheng Xiang and Ulrike Gretzel. Role of social media in online travel information search. *Tourism management*, 31(2):179–188, 2010.

[14] Stephen W Litvin, Ronald E Goldsmith, and Bob Pan. Electronic word-of-mouth in hospitality and tourism management. *Tourism management*, 29(3), 2008.

[15] Ahmed Abdelkader, Sabry Elshennawy, Mahmoud Fekry, and Lixian Li. The impact of user-generated content on tourist visit intentions: The mediating role of destination imagery. *Sustainability*, 15(4):3117, 2023.

[16] Qingbin Ye, Rob Law, Bin Gu, and Weimin Chen. The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel bookings. *Computers in human behavior*, 27(2):709–716, 2011.

[17] Kurt Philander and Ya Zhong. Twitter sentiment analysis: Capturing sentiment from integrated resort reviews. *International Journal of Hospitality Management*, 55:16–24, 2016.

[18] Mohamed Mostafa Mostafa. More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, 40(10):4241–4251, 2013.

[19] Seshadri Tirunillai and Gerard J Tellis. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, 51(4):463–479, 2014.

[20] Jay M Ponte and W Bruce Croft. A language modeling approach to information retrieval. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 275–281. ACM, 1998.

[21] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[22] Jonathan Chang, Sean Gerrish, Chong Wang, Jevan L Boyd-Graber, and David M Blei. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*, volume 22, 2009.

[23] Rajan Arun, Vasantha Suresh, C V Madhavan, and M Narasimha Murthy. On finding the natural number of topics with latent dirichlet allocation: Some new metrics. In *Pacific-Asia conference on knowledge discovery and data mining*, pages

391–402. Springer, Berlin, Heidelberg, 2010.

[24] Samuel Brody and Noémie Elhadad. An unsupervised aspect-sentiment model for online reviews. In *Proceedings of the 2010 annual conference of the North American chapter of the association for computational linguistics*, pages 804–812, 2010.

[25] Daniel Jurafsky and James H Martin. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Pearson Prentice Hall, 2009.

[26] Marko Tkalčič and Berardina De Carolis. *Personality and emotions in user modeling and recommender systems*, pages 1–2. Springer, Berlin, Heidelberg, 2013.

[27] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. Technical Report 8, OpenAI blog, 2019.

[28] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *International Conference on Learning Representations (ICLR)*, January 2023.

[29] Zong Du, Tianyu Zhang, and Zihan Zhang. MoDS: The Mirage of Data Sophistication for Instruction Tuning. *arXiv preprint arXiv:2310.06037*, 2023.

[30] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Ott, Wen-tau Chen, Alexis Conneau, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

[31] Yifan Zhang, Shiyu Qiao, Jiazhen Zhang, Tzu-Heng Lin, Chen Gao, and Yihong Li. A survey of large language model empowered agents for recommendation and search: Towards next-generation information retrieval. *arXiv preprint arXiv:2503.05659*, 2025.

[32] Daniele Di Palma. Retrieval-augmented recommender system: Enhancing recommender systems with large language models. *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 1369–1373, September 2023.

[33] Wenqi Fan. Recommender systems in the era of large language models (llms). *IEEE Transactions on Knowledge and Data Engineering*, pages 1–20, 2024.

[34] Yuxia Lu, Juntao Bao, Yimu Song, Zhaohua Ma, Shanshan Cui, Yiming Wu, and Xiangnan He. Revcore: Review-augmented conversational recommendation. *arXiv preprint arXiv:2106.00957*, 2021.

[35] Dongjie Yu, Ruocheng Bao, Gengchen Mai, and Ling Zhao. Spatial-rag: Spatial retrieval augmented generation for real-world spatial reasoning questions. *arXiv preprint arXiv:2502.18470*, 2025.

[36] Yuxiang Shi, Xiao Zi, Zhaoyu Shi, Hongbo Zhang, Qijiong Wu, and Meng Xu. Eragent: Enhancing retrieval-augmented language models with improved accuracy, efficiency, and personalization. *arXiv preprint arXiv:2405.06683*, 2024.

[37] Xinbei Ma, Yeyun Gong, Pengcheng He, Hai Zhao, and Nan Duan. Query rewriting in retrieval-augmented large language models. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5303–5315, December 2023.

[38] Theano S. Terkenli. Research advances in tourism-landscape interrelations: An editorial. *Land*, 10(9):944, 2021.

[39] Philip L Pearce and Peter F Stringer. Psychology and tourism. *Annals of tourism Research*, 18(1):136–154, 1991.

[40] Erdogan H Ekiz and Catheryn Khoo-Lattimore. Destination india: Investigating the impact of goa's attributes on families' leisure travel experience. *Tourism: An International Interdisciplinary Journal*, 62(2):165–180, 2014.

[41] David B Klenosky. The "pull" of tourism destinations: A means-end investigation. *Journal of travel research*, 40(4):396–403, 2002.

[42] Lea Susanne Faerber, Julian Hofmann, Dennis Ahrholdt, and Oliver Schnittka. When are visitors actually satisfied at visitor attractions? what we know from more than 30 years of research. *Tourism Management*, 84:104284, 2021.

[43] Gary L Geissler and Conway T Rucks. The overall theme park experience: A visitor satisfaction tracking study. *Journal of Vacation Marketing*, 17(2):127–138, 2011.

[44] Susanne HG Poulsson and Sudhir H Kale. The experience economy and commercial experiences. *The marketing review*, 4(3):267–277, 2004.

[45] Gianna M Moscardo and Philip L Pearce. Historic theme parks: An australian experience in authenticity. *Annals of tourism research*, 13(3):467–479, 1986.

[46] Anna Leask. Progress in visitor attraction research: Towards more effective management. *Tourism management*, 31(2):155–166, 2010.

[47] Birgit Elands and Jaap Lengkeek. Typical tourists. *Research into the theoretical and methodological foundations of a typology of tourism and recreation experiences. Mansholt Studies*, 21, 2000.

[48] Erik Cohen. A phenomenology of tourist experiences. *Sociology*, 13(2):179–201, 1979.

[49] Bob McKercher and Hilary Du Cros. Testing a cultural tourism typology. *International journal of tourism research*, 5(1):45–58, 2003.

A Prompt Design for Retrieval and Reranking Modules

A.1 Prompt for Intent-Enhanced Retriever (IER)

Prompt for Intent-Enhanced Retriever

```
Hello, you are now a travel analysis expert specializing in {city}. Your task is to decompose the user's travel query into multiple independent experiential requirements based on the following three dimensions:

1. **Landscape and Content**: Includes tangible and intangible visual, physical, natural, man-made, and informational elements that define the attraction's environment. This can include natural scenery (e.g., mountains, rivers, beaches), architectural features, cultural or historical elements, artworks, and designed spaces.
2. **Activities**: Refers to specific actions or engagements the user intends to undertake, such as walking, sightseeing, dining, learning, photography, or attending events.
3. **Atmosphere**: Refers to the mood, tone, or emotional/sensory experience the user is seeking, such as romantic, peaceful, lively, historic, or adventurous.
---
### Output Format:

You should return a list where each item is a dictionary representing an **independent requirement**, with the following key-value pairs:
- **expected landscape and content**: Describe what kind of natural or built environments, scenery, or informational features the user wants to experience. If relevant, include people or cultural references.
- **expected activities**: Describe what specific actions or experiences the user wants to engage in. Include any associated people or contexts if mentioned.
- **expected atmosphere**: Describe the mood, tone, or emotional quality the user is looking for.
**Do not include any explanations or code. Only return the list.**
The format should be exactly like this:
[
  {{
    "expected landscape and content": "...",
    "expected activities": "...",
    "expected atmosphere": "..."
  }},
  ...
]
---
### User Input
{user_input}
---
### Task Overview
Your goal is to analyze and break down the **user input** into multiple independent experiential requirements along the three dimensions defined above. Be precise, grounded, and consistent with the definitions.
Now return your output in the required format.
```

The IER module transforms a natural language user query into a set of structured experiential requirements along three core dimensions grounded in tourism theory:

- **Landscape and Content**: Refers to tangible and intangible elements of the physical or cultural environment, including natural scenery, built features, historical elements, or designed spaces.
- **Activities**: Captures the user's intended actions, such as sightseeing, walking, dining, attending events, or photography.
- **Atmosphere**: Represents the emotional tone, mood, or ambiance the user is seeking, such as peaceful, romantic, lively, or adventurous.

The LLM is instructed to return a list of dictionaries, each representing one independent experiential requirement across these three dimensions. This structured representation is subsequently used to guide the retrieval process.

A.2 Prompt for LLM-based Reranker (LRR)

Prompt for LLM-based Reranker

```
You are an AI travel planning assistant specializing in {city}.
Your task is to assign a **suitability score** (from 0 to 10) to each of the candidate attractions based on the user's query.
### Scoring Guidelines
For each attraction, evaluate and assign a **total score** between 0 and 10, considering:
1. **Content Relevance (0–10)**: How well the attraction matches the user's desired themes, activities, and atmosphere.
2. **Negative Filtering**: Strongly penalize attractions containing user-prohibited or mismatched elements.
3. **Do NOT consider coordinates or spatial information.**
---
### Input Data
【User Query】
{user_input}
【Candidate Attractions】
Each attraction is represented as a dictionary with the following fields:
{
 "id": "1",
 "name": "Attraction Name",
 "landscape and content": "Description of physical landscape and cultural/historical content",
 "activities": "Available or typical activities for visitors",
 "atmosphere": "General vibe, ambiance, or emotional tone of the place"
}
All candidate attractions are included in the following list:
{attractions_list}
---
### Output Format
Return a **JSON object** where each key is an attraction ID and the value is a float score between 0 and 10 (inclusive). For example:
{{
 "1": 8.5,
 "2": 3.0,
 "3": 0.0
}}
---
### Output Requirements
- Only return the JSON object.
- Do NOT explain your reasoning.
- Do NOT include rankings, coordinates, or any formatting other than valid JSON.
- Scores must be float numbers between 0 and 10 (inclusive).

Begin scoring now.
```

The LRR module evaluates the semantic relevance of retrieved attractions by assigning a scalar *suitability score* (0–10) to each candidate based on its match to the user query's themes, activities, and atmosphere. Each candidate attraction is structured as a dictionary with the following fields: `id`, `name`, `landscape and content`, `activities`, and `atmosphere`. The LLM is asked to score relevance without using any spatial information, and to output a clean JSON object suitable for downstream itinerary generation.