Available online at
https://mijuitm.com.my

e-ISSN: 2735-0703

Mathematical Sciences and Informatics Journal 6(2) 2025, 193 – 202.

Mathematical Sciences and Informatics Journal

Ontology Driven Platform for Tourism Insights in Malaysia

Ummi Maisarah Izani[1*], Nur Syadhila Che Lah[1]

[1]*Department of Computing, Universiti Teknologi PETRONAS, Malaysia*

ARTICLE INFO

Article history:
Received 28 August 2025
Revised 23 September 2025
Accepted 29 September 2025
Online first
Published 31 October 2025

Keywords:
Ontology-Driven Platform
Smart Tourism
Semantic Integration
Personalized Recommendation
Tourism Dashboard

DOI:
10.24191/mij.v6i2.9044

ABSTRACT

Tourism in Malaysia is characterized by its cultural diversity, rich heritage, and a wide range of experiences spanning from local attractions to unique culinary offerings. However, fragmented data sources and lack of semantic integration often hinder tourists and stakeholders from gaining meaningful insights. This research presents a digital ontology-driven platform designed to unify and contextualize tourism-related information through a structured knowledge representation framework. The proposed ontology model captures and interlinks key tourism domains including local attractions and food options, enabling intelligent query processing, personalized recommendation, and contextual awareness. This research follows the CRISP-DM methodology, ensuring a structured approach from data collection and preprocessing to ontology modelling and visualization. By leveraging ontology based reasoning and semantic relationships, the platform supports enhanced decision making for both tourists and tourism authorities. A prototype dashboard is developed to visually demonstrate tourism insights, offering users an intuitive interface to explore categorized information based on location, preference, and service availability. This integrated approach aims to support Malaysia's smart tourism initiatives and elevate the digital experience for both domestic and international travellers.

1. INTRODUCTION

Tourism is a vital sector in Malaysia's economy, drawing millions of visitors each year and generating large volumes of data from various sources such as travel websites, reviews, and booking systems. Despite this abundance, the data remain fragmented and unstructured, which limits its usefulness for both tourists and stakeholders. Traditional keyword-based search engines and databases struggle to provide personalised or context-aware results, often failing to capture traveller preferences, activity relationships, or destination-specific nuances (Sastriani et al., 2020). This creates a gap in delivering meaningful insights, affecting both user experience and strategic decision-making in the tourism industry. As tourism becomes more data-driven, there is a pressing need for an integrated system that can categorize and connect tourism information effectively. Ontology provides a solution through structured knowledge representation, enabling semantic

[1*] Corresponding author. *E-mail address*: ummi_21000213@utp.edu.my
https://doi.org/10.24191/mij.v6i2.9044

search, contextual recommendations, and intelligent data inference (Camacho & Cruz, 2022). Unlike static relational databases, ontology models define entities and their relationships, allowing more dynamic and personalised information retrieval. Major technology companies already employ ontology-based knowledge graphs to enhance search experiences, proving its relevance. This research proposes a novel ontology framework for Malaysia's tourism system, developed using Protégé and integrated with Power BI for interactive visualisations. Through SPARQL queries and structured categorization of data, the system supports improved decision-making and smarter tourism insights, aiming to enhance Malaysia's digital tourism capabilities.

Tourism-related data are highly fragmented and unstructured, scattered across multiple platforms such as review websites, social media, government reports, and travel portals, making effective retrieval and analysis challenging (Ordóñez Martínez, 2024). Tourists may leave reviews on TripAdvisor, share experiences on Instagram, and search for recommendations on Google, yet these data points remain disconnected. This lack of integration hinders data-driven decision-making. Moreover, limited collaboration among key tourism stakeholders including government agencies, travel businesses, and policymakers further exacerbates the issue, as each entity operates independently with its own datasets. This siloed approach leads to redundant efforts and disjointed planning, ultimately impeding comprehensive tourism development (Maimaitiaili, 2024). Most tourism platforms still rely on basic keyword-based search mechanisms, offering generic recommendations that often overlook user preferences, travel history, or dynamic conditions like weather. As a result, tourists receive non-personalised results, which diminish the overall experience; for example, a search for "best places in Penang" may yield a list of popular attractions without distinguishing between cultural sites, beaches, or hidden gems tailored to the traveller's interests (Dung et al., 2021). A major challenge also lies in extracting meaningful insights from unstructured content such as blogs, reviews, and social media posts. These free-text formats are difficult for traditional systems to process. For instance, if many reviews indicate a destination is overcrowded, existing systems may fail to detect and respond to this trend. Scholar Camacho and Cruz (2022) highlighted that ontology-based frameworks can organize such data semantically, enabling predictive analytics and supporting more adaptive and intelligent tourism services.

This research aims to develop a tourism ontology framework that structures tourism related knowledge into meaningful entities and relationships, thereby enhancing search and recommendation capabilities. By integrating diverse data sources including travel platforms, online reviews, and internet platform into a unified and structured knowledge representation, the framework addresses the current fragmentation of tourism information. To improve data accessibility and retrieval, SPARQL queries are implemented to enable semantic search and reasoning, ensuring more relevant and context-aware search results. Furthermore, the research utilizes Power BI to visualize tourism insights, allowing stakeholders to interact with the structured data through interactive and user-friendly dashboards.

2. LITERATURE REVIEW

In recent years, the tourism industry has increasingly embraced digital transformation, aiming to provide personalised, context-aware, and intelligent services to travellers (Hrankai & Mak, 2025). Amid this evolution, ontology has emerged as a pivotal tool for structuring, managing, and interpreting complex and heterogeneous tourism data. Ontology is commonly defined as "a formal and explicit specification of a shared conceptualization" (Gruber, 1993), and it plays a crucial role in knowledge representation, enabling semantic interoperability and reasoning across diverse tourism platforms.

Dung et al. (2021) proposed the Tourism Ontology for Machine Learning (TOML), a framework designed to enhance recommendation systems by integrating domain knowledge with machine learning models. The TOML framework enables tourism-related entities to be encoded into vector spaces, allowing for semantic reasoning and personalised recommendations. Unlike traditional content-based or collaborative filtering approaches, ontology-based systems like TOML mitigate cold-start problems and

https://doi.org/10.24191/mij.v6i2.9044

improve contextual awareness by leveraging structured domain knowledge. This is particularly beneficial in complex domains like tourism where user preferences are influenced by multiple contextual factors (Dung et al., 2021). Zhang et al. (2023) proposed a Tourism-Type Ontology Framework that organizes tourism knowledge into six major categories: themes, attractions, tourist characteristics, spacetime, experiences, and transportation. Their ontology was developed through grounded theory analysis of over 230 tourism types and 150 naming elements. By offering a hierarchical and semantically consistent model, the framework supports tourism analytics and domain-specific search, especially for destinations categorized by activity type and environmental features such as heritage, beach, or mountain tourism (Zhang et al., 2023).

She et al. (2018) introduced a property graph ontology-based recommender system that models tourism entities and user profiles using Neo4j graph database technology. The system incorporates Point of Interest (POI) categorization, star ratings, and user preferences to offer personalised recommendations. By emphasizing relationships between entities, such as “Located In” and “Preferred By,” the property graph structure enables efficient data retrieval and contextual matching. This approach is especially useful in the food tourism domain, where traveller decisions often depend on nuanced preferences such as cuisine type, ambience, and proximity. Although these studies demonstrate the potential of ontology to improve personalisation and contextual awareness in tourism services, they remain largely domain-specific, focusing on either recommendations or activity classification. Few frameworks attempt to unify multiple tourism domains into a single ontology or operationalize semantic reasoning through interactive visualisation dashboards. This study addresses these gaps by developing an ontology framework that integrates destinations and restaurants into one structure, validated through SPARQL queries and visualized via Power BI dashboards, thereby advancing Malaysia’s smart tourism initiatives

3. METHODOLOGY

This paper adopted the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to guide the development of a digital tourism ontology and dashboard tailored for Malaysia’s tourism sector (Azeroual et al., 2025). The approach was selected to ensure a structured and iterative workflow from problem understanding through data collection, modelling, and deployment. The methodology was driven by two core objectives: (1) to address the fragmentation of tourism data across destination and culinary, and (2) to enable semantic search, contextual recommendations, and visual analytics through ontology modelling and Power BI integration. The CRISP-DM methodology was implemented in six iterative phases. These phases, adapted to this study, are further detailed in Table 1 and illustrated in Fig. 1.

Table 1. CRISP-DM application stages

Phase	Description
Business Understanding	Defined the core problem of fragmented and unstructured tourism data and established system goals
Data Understanding	Identified and collected raw data from various open platforms such as TripAdvisor, Kaggle.
Data Preparation	Cleaned and reclassified data into ontology-relevant formats, resolving duplication and ambiguity
Ontology Modelling	Constructed classes and properties in Protégé using OWL, focusing on Destination and Restaurant domains
Evaluation	Designed SPARQL queries to validate the ontology’s semantic relationships and answer tourism-related questions
Deployment	Exported outputs into Power BI, enabling interactive dashboards for real-time exploration and filtering

https://doi.org/10.24191/mij.v6i2.9044

Fig. 1. CRISP-DM process adapted for tourism ontology development

3.1 Ontology Modelling

Ontology modelling was carried out using Protégé software and the Web Ontology Language (OWL). Core classes included Destination and Restaurant, each with relevant subclasses such as Heritage and Cuisine. These two domains were selected as they represent Malaysia's most data-rich tourism sectors and align with national smart tourism priorities. The ontology was populated with publicly available data from platforms such as TripAdvisor and Kaggle. Object and data properties (e.g., hasLocation, hasRating, hasCategory) were defined to formalize relationships between entities. This structure enabled semantic reasoning, ensuring that queries could traverse multiple relationships to return context-aware results.

3.2 Evaluation Through SPARQL Queries

The evaluation phase focused on verifying the ontology's ability to support meaningful and domain-specific queries. The queries tested hierarchical retrieval (class–subclass), attribute-based filtering (e.g., location, rating), and cross-domain integration (e.g., linking restaurants with locations). Their successful execution confirmed the logical consistency and completeness of the ontology. For example, one query retrieved heritage destinations located in Penang, and another extracted restaurant in Kuala Lumpur offering Japanese cuisine. These evaluations validated the correctness of class hierarchies and relationships and demonstrated the ontology's capacity to achieve semantic integration across domains.

3.3 Deployment via Power BI

The deployment phase involved exporting the ontology outputs into Microsoft Power BI. This enabled the construction of interactive dashboards that visualize categorized tourism insights, allowing users to filter results by state, tourism type, or service availability. The integration of Power BI ensured that the ontology was not only theoretically validated but also practically operationalized, providing an intuitive decision-support tool for stakeholders.

4. DISCUSSION AND FINDINGS

The developed ontology framework was evaluated through a series of SPARQL queries and Power BI dashboards to validate its effectiveness in retrieving structured tourism insights. The findings from this evaluation are categorized into three main domains: Destinations and Restaurants.

https://doi.org/10.24191/mij.v6i2.9044

4.1 Ontology Structuring and Class Population

The ontology was designed around two core domains under the superclass Thing: Destination and Restaurant. Each domain was semantically enriched with logical subclass hierarchies and aligned with actual tourism datasets obtained from official portals and public sources. The subclasses and data coverage are summarized in Table 2.

Table 2. Class hierarchy

Domain	Subclasses / Types
Destination	Geographical_Environment: (Mountain, Forest, Island, Beach, Lake, Cool_Highland_Climate) Activities: (Eco_Tourism, Heritage, Theme_Park)
Restaurant	Asian, Fusion, Western, Middle_Eastern, Buffet, Pastry, Street_Food

As illustrated in Fig. 2, the ontology framework is divided into two main categories: Restaurant and Destination. The Restaurant domain includes subclasses such as Street_Food, Fusion, Western, Buffet, Pastry, Asian, and Middle_Eastern, with Asian further refined into Malay, Chinese, Indian, Japanese, Korean, Thai, and Vietnamese cuisines. The Destination domain is organized into Activities (Eco_Tourism, Urban_Attraction, Heritage, Theme_Park) and Geographical_Environment (Lake, Island, Forest, Beach, Mountain, Cool_Highland_Climate). This hierarchical organisation ensures that diverse culinary and destination features are semantically represented, enabling precise and context-aware queries such as "Japanese restaurants in Kuala Lumpur" or "eco-tourism sites in forest regions."

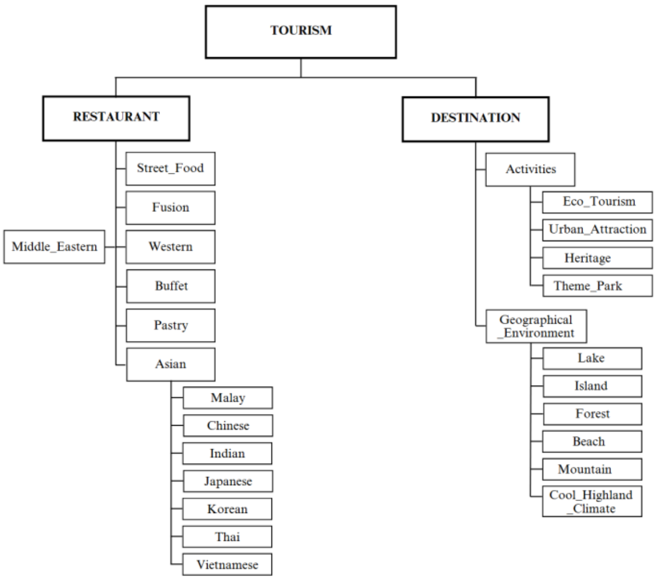

Fig. 2. Ontology structure for the tourism domain, showing the classification of restaurants and destinations into hierarchical subclasses

4.2 Semantic Validation via SPARQL Queries

The functionality of the developed ontology was evaluated through a series of structured SPARQL queries across three key domains: Destinations and Restaurants. These queries validated the ontology's ability to support precise data retrieval based on subclass relationships, location constraints, and attribute filters such as rating or offered services. This section discusses the query performance, supported by outputs and interpretation, to demonstrate how semantic modelling significantly improves tourism insight extraction compared to traditional flat data models.

https://doi.org/10.24191/mij.v6i2.9044

4.2.1 Destination Domain

The Destination class comprises two main subclass groups: Activities and Geographical_Environment. SPARQL queries shown in Fig. 3 validated that each instance was retrievable based on its subclass and location. The query successfully retrieved thematic destinations like Taman Negara and Sunway Lagoon, demonstrating structured classification. This reflects the ontology structure proposed by Zhang et al. (2023), who used hierarchical categories such as space-time and tourist experience to improve tourism data retrieval. Therefore, these results confirm that the class hierarchy and hasLocation relationships are functioning as intended.

```
SPARQL query:

PREFIX tourism: <http://www.semanticweb.org/tourism#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?destination ?category ?location
WHERE {
   ?destination rdf:type ?category .
   ?category rdfs:subClassOf tourism:Activities .  # Ensure it checks the correct subclass
   ?destination tourism:hasLocation ?location .
}
```

Fig. 3. Retrieve destinations under activities

4.2.2 Restaurant Domain

Restaurants in the ontology were categorized under seven distinct cuisine types: Asian, Fusion, Western, Middle_Eastern, Street_Food, Pastry and Buffet. A key objective was to test multi-criteria filtering by category, rating, and location. This confirms that ontology can support preference-based filtering, enabling a more tailored restaurant search experience. The semantic tagging of cuisine and rating attributes ensures that tourists receive results matching specific culinary preferences, helping differentiate among hundreds of choices. This aligns with She et al. (2018), who demonstrated the effectiveness of POI property-graph ontology to personalize culinary recommendations based on user preference and rating. Fig. 4 shows the restaurants with ratings retrieval process.

```
SPARQL query:

PREFIX tourism: <http://www.semanticweb.org/tourism#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?restaurant ?rating
WHERE {
   ?restaurant rdf:type ?cuisine .  # Step 1: Get all restaurants and their cuisine type
   ?cuisine rdfs:subClassOf tourism:Restaurant .  # Step 2: Ensure the cuisine is a subclass of Restaurant
   ?restaurant tourism:hasRating ?rating .  # Step 3: Retrieve the restaurant's rating

}
```

Fig. 4. Retrieve restaurants with ratings

4.3 Dashboard Implementation in Power Bi

The semantic outputs generated from the ontology were transformed from RDF format into CSV files to facilitate integration with Microsoft Power BI. This step ensured that the ontology was not only theoretically validated through SPARQL queries but also practically operationalized for stakeholder use. The Tourism Insight Dashboard was designed with two interactive modules, as shown in Fig. 5 - 7. The Overview module (Fig. 5) presents high-level tourism statistics such as total arrivals, monthly growth rates, and distributions by country of origin. This module confirmed that the ontology could integrate macro-level

https://doi.org/10.24191/mij.v6i2.9044

datasets with semantic structures, enabling policymakers to observe seasonal patterns and track key performance trends more effectively than static reporting.

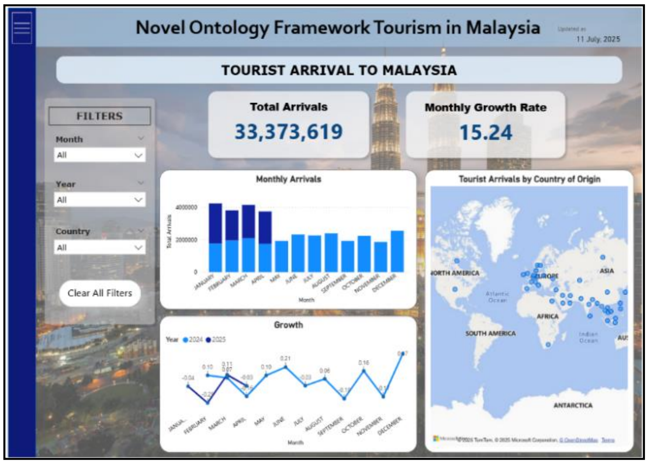

Fig. 5. Overview module of the tourism insight dashboard

The Destination Insights (Fig. 6) module enables filtering by geographical categories such as heritage sites, eco-tourism, mountains, and lakes. The outputs are mapped by state, allowing users to visualize the spatial distribution of destinations. The results highlighted that eco-tourism sites were heavily concentrated in East Malaysia, demonstrating the ontology's ability to expose regional imbalances and support targeted tourism planning.

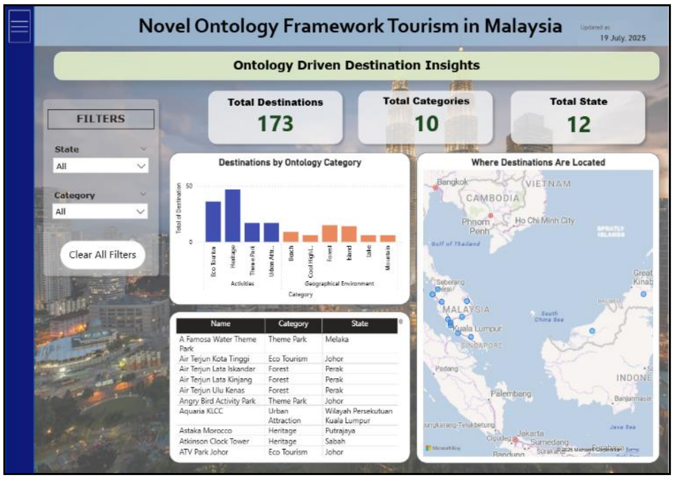

Fig. 6. Destination insights module

The Restaurant Insights (Fig. 7) module organizes results by cuisine type, rating, and location. This functionality supports personalised exploration by allowing users to compare restaurants based on semantic categories and user preferences. The analysis revealed that Asian cuisines dominated the dataset, validating the ontology's ability to reflect Malaysia's culinary diversity while also highlighting underrepresentation of Western and Fusion cuisine categories.

https://doi.org/10.24191/mij.v6i2.9044

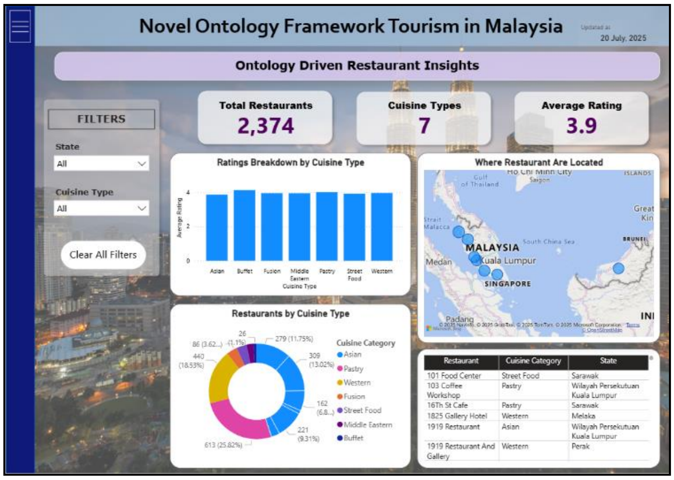

Fig. 7. Restaurant insights module.

5. CONCLUSION

This study addressed the fragmented and unstructured nature of tourism information systems in Malaysia by developing an ontology-driven digital platform. The framework semantically integrates two key domains: Destinations and Restaurants, which are connected through structured subclass hierarchies and logical relationships. Developed using Protégé and visualized with Microsoft Power BI, the platform supports accurate classification, structured querying, and personalised recommendations via SPARQL. Unlike prior ontology studies that focused on single domains or lacked visualisation support, this research contributes by unifying multiple tourism domains into a single knowledge base and operationalizing it through interactive dashboards.

The evaluation demonstrated that SPARQL queries accurately retrieved domain-specific insights such as heritage destinations and cuisine-based restaurants. These outputs were effectively transformed into Power BI dashboards, enabling users to explore tourism data through interactive filters, maps, and visual summaries. This dual validation confirmed both the technical soundness of the ontology and its practical value as a decision-support tool. Despite its contributions, the study faced challenges in cleaning and harmonizing heterogeneous data sources, which required extensive manual preprocessing. This highlights the urgent need for standardized and machine-readable tourism datasets in Malaysia.

Future work will focus on extending the ontology to additional domains such as accommodation and cultural events, integrating real-time data streams (e.g., transport, weather, and event listings), and incorporating user feedback mechanisms for dynamic personalisation. These enhancements would further strengthen the platform's ability to support adaptive, intelligent, and data-driven tourism services. Overall, this research demonstrates that ontology-based knowledge representation combined with visual analytics can improve the usability of tourism data, support informed decision-making by tourists, businesses, and policymakers, and contribute to Malaysia's broader smart tourism initiatives.

6. ACKNOWLEDGEMENTS/ FUNDING

The authors would like to acknowledge the support of Universiti Teknologi PETRONAS (UTP) and the Department of Computing, Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia, for providing the research facilities, technical guidance, and academic environment necessary for the successful completion of this study.

https://doi.org/10.24191/mij.v6i2.9044

7. CONFLICT OF INTEREST STATEMENT

The authors affirm that this research was conducted without any commercial, financial, or personal relationships that could be construed as a potential conflict of interest. No competing interests were identified regarding the funding, data sources, or institutional support for this study.

8. AUTHORS' CONTRIBUTIONS

Ummi Maisarah Binti Izani conceptualized the research framework, conducted data collection and preprocessing, developed the ontology using Protégé, implemented SPARQL queries, and designed the interactive dashboard in Power BI. She also wrote, revised, and finalized the manuscript. Nur Syadhila Che Lah provided supervisory support, guided the research direction and methodology, reviewed the ontology model and visualisations, and approved the final version of the manuscript for submission.

REFERENCES

Azeroual, O., Nacheva, R., Nikiforova, A., & Störl, U. (2025). A CRISP-DM and predictive analytics framework for enhanced decision-making in research information management systems. *Informatica, 49*(18), 67–86. https://doi.org/10.31449/inf.v49i18.5613

Camacho, V. T., & Cruz, J. (2022). Ontology-based context-aware recommender system application for tourism. *arXiv*. https://doi.org/10.48550/arXiv.2301.00768

Dung, H. Q., Le, L. T. Q., Nguyen, H. H. T., Truong, T. Q., & Nguyen-Dinh, C. H. (2021). A novel ontology framework supporting model-based tourism recommender. *IAES International Journal of Artificial Intelligence, 10*(4), 1060–1068. https://doi.org/10.11591/ijai.v10.i4.pp1060-1068

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199–220. https://doi.org/10.1006/knac.1993.1008

Hrankai, R., & Mak, B. (2025). Bridging the affordance–actualization gap in user preferences for AI-assisted trip planning. *Journal of Travel Research*. Advance online publication. https://doi.org/10.1177/00472875251322518

Maimaitiaili, Y. (2024). Integrated tourism: A holistic approach to resolving fragmentation challenges in tourism governance. *SHS Web of Conferences, 187*, 03033. https://doi.org/10.1051/shsconf/202418703033

Ordóñez Martínez, M. D. (2024). *Tourism data spaces in smart tourism destinations: Conceptual and methodological approach* [Doctoral thesis, Universitat de les Illes Balears]. UIB Research Portal. https://portalinvestigacio.uib.cat/documentos/66f6f0d1ae35f26423e035a4

Sastriani, R., Baizal, Z. K. A., & Kusumo, D. S. (2020). Ontology-based semantic search on tourism information search system. *Indonesian Journal on Computing (Indo-JC), 5*(1), 103–114. https://doi.org/10.34818/INDOJC.2020.5.1.397

She, K.-S., Haw, S.-C., Loh, Y.-G., & Chua, F.-F. (2018). AK Tourism: A property graph ontology-based tourism recommender system. In *Proceedings of the Knowledge Management International Conference (KMICe)*. Universiti Utara Malaysia Repository. https://repo.uum.edu.my/id/eprint/25195/ UUM Repository

https://doi.org/10.24191/mij.v6i2.9044

Zhang, P., Wang, J., & Li, R. (2023). Tourism-type ontology framework for tourism-type classification, naming, and knowledge organization. *Heliyon, 9*(4), e15192. https://doi.org/10.1016/j.heliyon.2023.e15192

Zhang, X., Zhao, C., & Wang, X. (2015). A survey on knowledge representation in materials science and engineering: An ontological perspective. *Computers & Industry, 73*, 8–22. https://doi.org/10.1016/j.compind.2015.07.005