Human-centric Faceted Approach for Ontology Construction†

Shiv Shakti Ghosh*, Subhashis Das**, and Sunil Kumar Chatterjee***

Jadavpur University—Library & Information Science—Main Campus. 188, Raja S.C. Mallick Rd , Kolkata, West Bengal 700032, India,
*<shivu@drtc.isibang.ac.in>, ***<sunilkc61@gmail.com>
**University of Trento—ICT-Doctoral School—Via Sommarive, 9 I-38123 Povo, Trento, TRENTO 38123, Italy, <subhashis.das@unitn.it>

Shiv Shakti Ghosh is doing his PhD and is a senior research fellow in the Department of Library & Information Science, Jadavpur University, India. In 2013, he completed his bachelors in mathematics from Serampore College, India and obtained his master's degree in library and information science (MSLIS) from the Documentation Research and Training Centre (DRTC), Indian Statistical Institute, India in 2015. His research interests include information science and arts and humanities.

Subhashis Das currently works as a post-doctoral researcher at the Department of Information Engineering and Computer Science (DISI), Università degli Studi di Trento, Italy. He obtained his Ph.D from the ICT-Doctoral School, University of Trento, Trento, Italy. Subhashis does research in information science, computing in health science, arts and humanities and geoinformatics (GIS).

Sunil Kumar Chatterjee is a professor in the Department of Library & Information Science, Jadavpur University, India. In 1987, he obtained his MSc in chemistry. In 1990, he got the associateship in information science from INSDOC, India and completed his master's degree in library and information science from Burdwan University, India in 1991. He obtained his PhD in 2005. Having begun as a researcher in library and information science during the early nineties, he has contributed papers in several peer reviewed journals and conferences of national and international repute. His research areas include scientometrics, informetrics and information science.

Shiv Shakti Ghosh, Subhashis Das, and Sunil Kumar Chatterjee. 2020. "Human-centric Faceted Approach for Ontology Construction." *Knowledge Organization* 47(1): 31-44. 39 references. DOI:10.5771/0943-7444-2020-1-31.

Abstract: In this paper, we propose an ontology building method, called human-centric faceted approach for ontology construction (HCFOC). HCFOC uses the human-centric approach, improvised with the idea of selective dissemination of information (SDI), to deal with context. Further, this ontology construction process makes use of facet analysis and an analytico-synthetic classification approach. This novel fusion contributes to the originality of HCFOC and distinguishes it from other existing ontology construction methodologies. Based on HCFOC, an ontology of the tourism domain has been designed using the Protégé-5.5.0 ontology editor. The HCFOC methodology has provided the necessary flexibility, extensibility, robustness and has facilitated the capturing of background knowledge. It models the tourism ontology in such a way that it is able to deal with the context of a tourist's information need with precision. This is evident from the result that more than 90% of the user's queries were successfully met. The use of domain knowledge and techniques from both library and information science and computer science has helped in the realization of the desired purpose of this ontology construction process. It is envisaged that HCFOC will have implications for ontology developers. The demonstrated tourism ontology can support any tourism information retrieval system.

Received: 8 May 2019; Revised: 17 June 2019, 8 July 2019, 20 August 2019, 2 September 2019, 25 October 2019, 5 November 2019: Accepted: 15 November 2019

Keywords: ontology, tourism, information, domain, concepts, terms

† The link for accessing the tourismontology.owl file can be found at (https://webprotege.stanford.edu/#projects/00bef201-8f71-4f68-b507-b04d4bd9b99d/edit/Classes)

1.0 Introduction

Recent years have witnessed an increase in the use of ontology for knowledge representation, sharing and distribution. As defined in Studer et al. (1998), ontology is a formal, explicit specification of a shared conceptualization. Concepts belonging to a domain of discourse are described with the help of properties. The description present in the properties refers to the different features and attributes of the concepts. Thus, semantic relationships are established between the concepts. As an outcome, background knowledge or relevant semantic information pertaining to the domain of discourse gets encapsulated. In totality, an ontology attempts to model a domain of discourse. The modeling here refers to either an attempt to describe or categorize objects belonging to the domain of discourse.

As remarked by Smiraglia (2015, 19), knowledge organization studies are focusing towards a domain-analytical direction. Hjørland (2017) has termed domain analysis as the theorization and analytical approach to library and information science and knowledge organization. For domain analysis, many ontology construction methodologies have conceived the concept of facet analysis (see study by (Prieto-Diaz 2003), (Giunchiglia et al. 2009) and (Das and Roy 2016)) and analytico-synthetic classification propounded by Ranganathan (1967). The analytico-synthetic approach consists of two phases. In the first phase, known as the analysis phase, compound and complex ideas are fragmented into fundamental ideas. After analyzing their characteristics, these fundamental ideas are grouped or clustered according to similarity. This first phase is completed by following the first-link-downwards and last-link-upwards approach. The difference between these two approaches are characterized in the way they start approaching concepts, where a concept, as defined by Dahlberg (1978) is a knowledge unit, and the statements about its referent are the characteristics of the given concept. The first-link-downwards starts the analysis from the root concepts in the domain and then gradually narrows down to more specific concepts. Thus, it proceeds from abstract level to a concrete level. The last-link-upwards technique identifies and studies the characteristics of base concepts and assembles them depending upon their similarity of features. In this way, by continuing the process of clubbing together a large or universal concept is formed. By repeating this process, the root concept is reached. Thus, the bottom up approach proceeds from concrete level to an abstract level. In the second phase, mentioned as the synthesis phase, semantic relationships are established between concepts. This approach helps in identifying facets, where a facet, as has been described in Giunchiglia and Dutta (2011) as a hierarchy of homogenous group of terms (nodes), each term denoting a primitive atomic concept.

Plenty of research is being conducted to determine context. The possibility of using ontology as a tool for context management has also been endeavored. The methodologies for developing or constructing these ontologies depict the human-centric approach to deal with context, its analysis and development. Further, in the field of library and information science, the idea of selective dissemination of information (SDI) propounded by Luhn (1961) has been in use for quite a considerable amount of time. SDI has been routinely used to deal with "context." SDI got involved with the aim of catering information to those who found it most useful, or, in other words, it aimed to cater information according to the right context. Additionally, the aim was to prevent communication of misinformation or, information that is out of the context. The system is to be fed with a database of users' profiles containing areas of interest belonging to individual end users. SDI advocates that when a particular item or information is to be disseminated, it is to be done after comparing its information pattern or trend with profiles of the members present in the system. The system may select as many recipients as it finds suitable. There is a concept of weeding out imbibed within SDI as it has been recommended to delete those document patterns that have become obsolete. Furthermore, interests sustaining for longer time must be preserved. This sense of weeding out of obsolete information patterns and preservation of information sustaining for longer time periods has particularly motivated the evaluation step in HCFOC. New concepts and its representatives must be sustained for longer periods and must be inserted and embedded within the ontology with the objective to saturate it. The requirement of addition or deletion of outdated information patterns must be identified by repeating the evaluation step of HCFOC. The idea of SDI is quite similar to the human-centric approach followed while dealing with context. Lamsfus (2009) has familiarized the human-centric approach with relevance to contextual information. The proposed human-centric faceted approach for ontology construction (HCFOC), discussed in this paper, uses a synthesis of both these similar approaches to correctly deal with context. Further, the analytico-synthetic approach has been also utilized for designing this methodology. Tourism has been chosen as the domain for exemplifying this process. Since people moving from one place to another may need relevant information for enhancing their mobility at any time, it becomes crucial to accurately determine the context of a tourist's information requirement or information seeking pattern. Opting tourism as a domain for study and experimentation is readily providing a scope to test the context-modeling capability of the HCFOC methodology. Smiraglia (2015, 19) has listed the domains that have been taken up recently for knowledge organization studies, which clearly reveals that not even once the tourism domain was taken up for study.

Furthermore, many regions and countries rely on tourism as the main source of revenue generation, which contributes to the national GDP. This domain has become the focus of a lot of economic activities. According to the World Tourism Organization (UNWTO) (2017), "the business volume of tourism equals or even surpasses that of oil exports, food products or automobiles." Besides this, it has direct effects on educational, cultural and social sectors. Transportation, hospitality and entertainment services also harness benefits from this industry. Mobility of people has increased significantly over time. Quite naturally, the tourism domain is also experiencing a surge in information and knowledge handling like never before. And to deal with this, innovative approaches and applications are required. So, an effort has been made to build a model for the tourism domain using ontology.

The rest of the paper is organized as follows: Section 2.0 provides a brief literature review on this domain and discusses the related works. Section 3.0 explains the HCFOC methodology exemplifying its use for constructing the tourism ontology. Section 4.0 concludes the paper while discussing avenues for future work.

2.0 Ontologies related to tourism

Since the efficiency of ontologies in decision making has been proven, a lot of research is focusing on ontology construction methodologies. Qiu et al. (2018) have used a combination of rule-based (for concept and relationships extraction), statistics-based (for ranking the concepts) and cluster-based methods (for clustering and constructing taxonomy) for constructing ontologies. Nguyen and Lu (2016) have developed ontologies for web pages. The steps followed are requirement analysis, conceptualization and implementation. Yang et al. (2017) have proposed DOCM or domain ontology construction method. The methodology involves requirements and domain knowledge analysis followed by establishment, evaluation and modification effort assessment of the ontology. Further, a method has been proposed to evaluate the modification effort on the ontology. Suárez et al. (2015) have developed the NeOn Methodology framework based on glossary of processes and activities and ontology building scenarios, networks and life-cycle models.

The tourism domain is also experiencing a surge in the use of ontologies for information dissemination, decision making and fabrication of recommendation systems. Chu et al. (2016) have constructed a tourism recommender system. For this, users in the database have been categorized as related and unrelated. Different kinds of algorithms have been used to deal with the contexts of users belonging to different categories. Al-Hassan and Lu (2015) have discussed the use of ontology and the defined relationships and attributes within it to find semantic similarities between items for use in an e-Government tourism service recommendation system. The Harmonise ontology proposed by Fodor and Werthner (2005) focuses on tourism data exchange. The QALL-ME framework by Ferrandez et al. (2011) has been mapped with Princeton WordNet by Miller (1995) and the Suggested Upper Merged Ontology (SUMO) and thus it has a strong foundation knowledge base. GETESS by Staab et al. (1999) deploys natural language processing (NLP) and semantic web methods to answer user queries using web-based information exchange and distribution. Existing ontologies on the tourism domain have been listed in Mathur et al. (2015). The Mondeca ontology has been built using concepts from the thesaurus developed by the World Tourism Organization (UNWTO) (2001). The OnTour ontology by the eTourism Semantic Web Portal describes the concepts of location coordinates, accommodation, date and time of certain events, etc. for tourism information dissemination. The purpose of the HiTouch Ontology and the TAGA ontologies is to cater to travel agents. Gregor et al. (2016) have proposed a methodology using semantic clustering algorithms to create ontology for intelligent transportation systems. Frikha et al. (2016) and Lee et al. (2017) have shown the use of ontology for medical tourism and leisure tour recommender systems respectively. She et al. (2018) have deployed property graph ontology for a tourism recommender system. Special efforts for understanding and modeling the context in tourism can be seen in (Kashevnik 2017).

3.0 The HCFOC methodology and the tourism ontology

The human-centric faceted ontology construction (HCFOC) methodology consists of eight steps. The ontology construction process demonstrates the synthesis of the idea of SDI from the field of library and information science with the idea of human-centric approach to deal with the context of the primary information seeker. Further, the analytico-synthetic classification approach has been also used to capture the necessary and relevant background or inherent knowledge. This methodology evolved while attempts were being made to develop an ontology for the tourism domain, which has been also shown here. The tourism ontology has been developed using the Protégé-5.5.0 (https://protege.stanford.edu) ontology editor. It is a free, open-source ontology editor developed by the Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine. The Protégé OWL (web ontology language) ontologies consist of classes, properties (object properties and data properties) and individuals. Classes are sets of individuals that are objects of the domain. Object properties are relations between the

objects. Data properties are relations between the objects and data types.

3.1 Step 0: domain selection

A field or area of study is to be selected on which the ontology will be created. The field or area of study is commonly referred to as domain. The tourism domain is the centre of a lot of economic activities. Contextual information is highly sought after in this domain. The aim of the HCFOC methodology is to understand and deal with context in a comprehensive and precise manner. Based on previous studies and reviewing of existing literature the tourism domain has persistently appealed as a perfect domain for testing and implementing this methodology. Moreover, as this domain is multifaceted, the analytico-synthetic approach ingrained in the HCFOC methodology will also be tested.

3.2 Step 1: focus map creation

The contextual queries of the primary information seeker in the selected domain are usually considered to ascertain the focus. Ascertainment of the focus helps in aligning with or modifying the purpose and scope of the ontology. This step is instrumental in making the ontology capable of dealing with the context of the primary information seeker. Correct and comprehensive understanding of the context helps in increasing the precision of the answers, responses and results obtained in return. To achieve this, the HCFOC methodology uses a synthesis of two ideas, namely, the human-centric approach, to deal with context, and selective dissemination of information (SDI). Following this synthesis, a prospective map of depicting the contextual behavior of the primary information seeker has been included in this step. This map also represents the purpose and scope of the ontology under construction.

Competency questions were obtained from prospective tourism information users, which include professors, research scholars, students and heads of travel agencies from India and Italy. Some of the frequently asked questions are: a) Where can I stay during my visit to Kolkata?; b) Which rivers pass through India?; c) List all the mountains in India?; d) List of cuisines of India?; e) List of bridges in India?; f) Which deserts are located in India?; g) How to reach Mumbai from Kolkata?; h) What is the local language of Salem, Tamil Nadu?; i) How is the weather of Bangalore in June?; and, j) What are the drinks available in Goa?

Accordingly, to answer the aforementioned questions, some of the terms that need to be considered are: address, administrative division, drinks, alcohol, artifact, structure, bridge, bus, car, transport, city, contact, email, fax, country code, country, geo-coordinate, location, height, hotel, language, weather, latitude, longitude, mountain, landform, desert, postal code, river, basin, vehicle, website, etc. An analysis of these questions revealed the approach and information seeking behavior in this domain, and the primary information seeker was identified as a tourist. So, this ontology has been conceived keeping the tourist at the centre. Further, this analysis also helped us in forming a prospective map of the context of tourists' information needs. This map also represents the purpose and scope of the tourism ontology.

3.3 Step 2: information acquisition

A footprint of the ideas, concepts and their features, present in the existing information sources on the domain under consideration, is to be formed. This footprint is to be matched against the map created in the previous step, and the overlapping areas are to be identified. Information sources on all such overlapping areas are to be studied and the collected information must be consolidated.

Figure 1(a) depicts the general information acquisition process, and Figure 1(b) shows how tourism information is scattered over different domains like geography, cadastre, automotive and economy. Our present requirements are only a subset of information from each domain. For instance, monuments and heritage buildings are usually collected from the cadastre database, but we might not need property tax or building material information in the context of tourism application.

Information for developing the tourism ontology has been obtained from different governmental and non-governmental tourism websites, reference tools on tourism, travelers and users of tourism information. Several individual travel experiences that were shared over the internet were studied. Information provided by users in travel groups prevalent on social media platforms were also scrutinized. The results of the assessment done on these resources have been explained in Ghosh and Chatterjee (2019). A reliable information base was formed after compiling all such information.

3.4 Step 3: term identification

Terms representing the ideas, concepts and their features belonging to the overlapping areas found in the previous step are to be identified. Different tourism related terms were identified from the compiled information base formed in the "information acquisition" step. The World Tourism Organization (UNWTO) thesaurus (World Tourism Organization 2001) has been used as a guide to tourism terminology. WordNet was rigorously consulted in this process. Besides this, Schema.org (https://schema.org/Thing) and the INSPIRE (https://inspire.ec.europa.eu/)

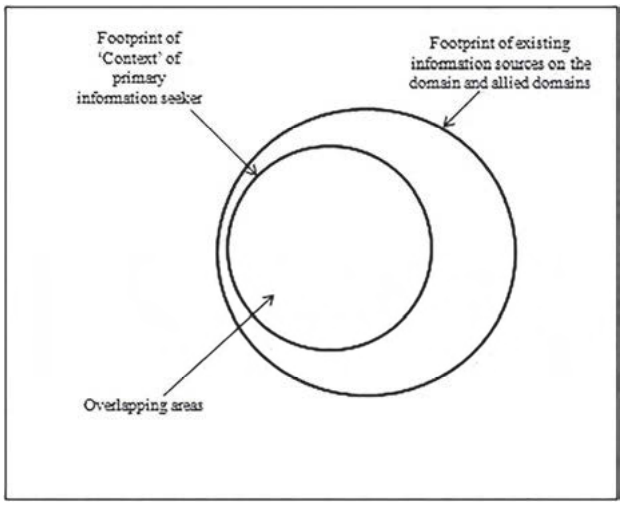

Figure 1(a). Information acquisition

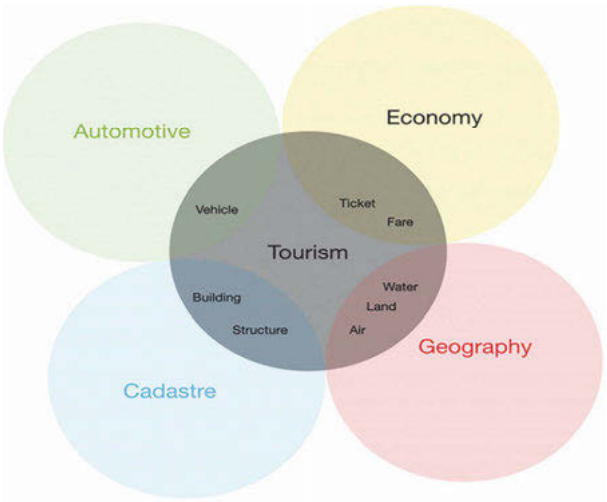

Figure 1(b). Domains intersecting with tourism

knowledge base were also consulted. This helped in resolving the ambiguity involved in understanding the concepts. Also, the perfect term to represent the concepts could be identified. Some of the terms are: artifact, structure, lodging, hostel, monument, telpherage, biome, cuisine, drink, event, mountain, hill, plateau, location and person.

3.5 Step 4: analysis.

The terms obtained in the previous step have to be analyzed for identification of differences and similarities. Terms denoting compound and complex ideas have to be disintegrated into terms denoting simple ideas. Keeping in mind the purpose, scope and context of the primary information seeker, it is to be analyzed whether a term will be used to denote a class / sub-class / instance / relation or attribute (of object or of data).

For the tourism ontology, the analysis was continued with reckon to the purpose, scope and context of a tourist. Besides other principles as proposed in Ranganathan (1967), the "principle of context" and "canon of relevance" was widely used in the procedure. Terms with similar features were grouped together. For example, terms like arena, camp, hospice and hostel were found to be similar. They have been grouped together.

Analysis has been done without user participation as it difficult to involve users. This is because not all users are domain experts. Many of them are leisure travelers. Some of them are also not aware of the exact piece of information they want (very similar to users in libraries, where often the librarians try to find out the exact needs of the users). Certain categories of users have been involved for defining the competency questions and their respective evaluation. Many user centric questions were formulated, which proved to be helpful in designing and evaluating the tourism ontology. Domain expertise and proficiency and friendliness with the vocabularies and tools used in the study were an essential requirement of this step. Due to lack of required ability, not many users were enthusiastic or eligible to participate in this step. Some were also wary of the amount of time this work would require. Thus, though we had initially thought of involving the users in this step, we dropped this idea later.

3.6 Step 5: knowledge synthesis

After the analysis carried out in the previous step, the ideas were clustered or grouped together based on similarity in characteristic, and the categories have to be labeled. In this step, for classifying the facets, the first link downwards and last link upwards approach was followed. The first link downwards approach proceeds from abstractness to concreteness. While the last link upwards approach proceeds from specific concepts towards generic concepts. Following these two approaches leads to an overall increase in the degree of robustness. Knowledge on the domain is synthesized in this step by establishing relationships between the concepts.

For the tourism ontology, facet discovery and inventory control were guided by the "principle of context" and "principle of helpful sequence" as proposed in Ranganathan (1967). For example, the group containing the terms arena, camp, hospice and hostel were listed under "lodging." The features of the concepts represented in one group were used to establish relationships between concepts represented in other groups. For example, "lodging" has features like price, location etc., which have been used to establish its relationship with other concepts. Facets contain sub-facets. For example, "lodging" has been listed under "structure," which has been again listed under "artifact."

3.7 Step 6: knowledge representation

This step consists of the following sub steps:

Term standardization: There may exist many terms to denote a single concept. However, such synonymous terms may differ among themselves based on usage. The use of an appropriate and proximate term has to be decided. Such proximate terminology should suit the context of the ontology model. Also, the terms chosen should be most frequently used or collected from a standard vocabulary on the domain on which the ontology is being constructed. Use of popular terms increases user friendliness while use of terms from standard vocabulary helps in interoperability. Use of a term by domain experts in their written and verbal communication influences its selection for use in the ontology. The terms that have been standardized for use in the tourism ontology have been enlisted by consulting different controlled vocabularies on tourism. For example, the word "artifact" has been chosen to represent the concept "any object made by human beings, especially with a view to subsequent use." The term "artifact" has been chosen from those terms representing the sense of product such as, merchandise, produce, creation; examples of terms representing the sense of relic such as, antique, monument (representing the sense relic); examples of terms representing the sense such as, handiwork, artisanship (representing the sense handicraft).

Ordering: There exist many criteria for ordering the standardized terms within the array. Some of the criteria as mentioned in Ranganathan (1967) are existent classification schemes on the domain, alphabetical order, decreasing or increasing complexity, increasing or decreasing extension, etc. Whichever criterion is chosen, it must be kept in mind that the ordering must help in reaching the desired aim of the ontology.

This step was completed keeping in mind the purpose and scope of the tourism ontology and the context of the tourist's information need. In Table 1, the outcome of ordering is visible.

Modeling: The HCFOC methodology uses the DERA (domain, entity, relation and attribute) framework mentioned in Giunchiglia and Dutta (2011) for structuring the facets of the domain under consideration. Here, the idea of entity has been derived from Bhattacharyya (1975). Ranganathan's faceted classification (1989) divides knowledge into five fundamental categories, namely, "personality" (P), "matter" (M), "energy" (E), "space" (S) and "time" (T), known by the acronym PMEST. Bhattacharyya (1981) further refined the division into four main categories, namely, "discipline" (or domain) (D), "entity" (E), "property" (P) and "action" (A), and an additional special category called "modifier" (m), known by the acronym DEPA. The DERA framework advocates organization of knowledge into domains. Further, each domain should be organized using facets. Giunchiglia et al., (2014) shows that DERA allows addition of domains, facets and terms into the ontology, whenever required, and its exploration for automatic rea soning via direct encoding into description logics (DL).

Class	Thing Event MentalObject Cuisine AfricanCuisine … PhysicalObject Artifact Handicraft Structure Building HealthcareFacility Hospital … Brothel Library … Fountain … TransportationSystem AirTransportationSystem … Location AdministrativeDivision … Person TravelAgent … Substance Drink …
Object Properties	hasCreator isBasedFrom hasArrivalPoint …
Data Properties	hasPrice hasIdentifier hasGeocoordinate … hasCheckInTime hasRating …

Table 1. Partial list of classes, object properties and data properties from the tourism ontology.

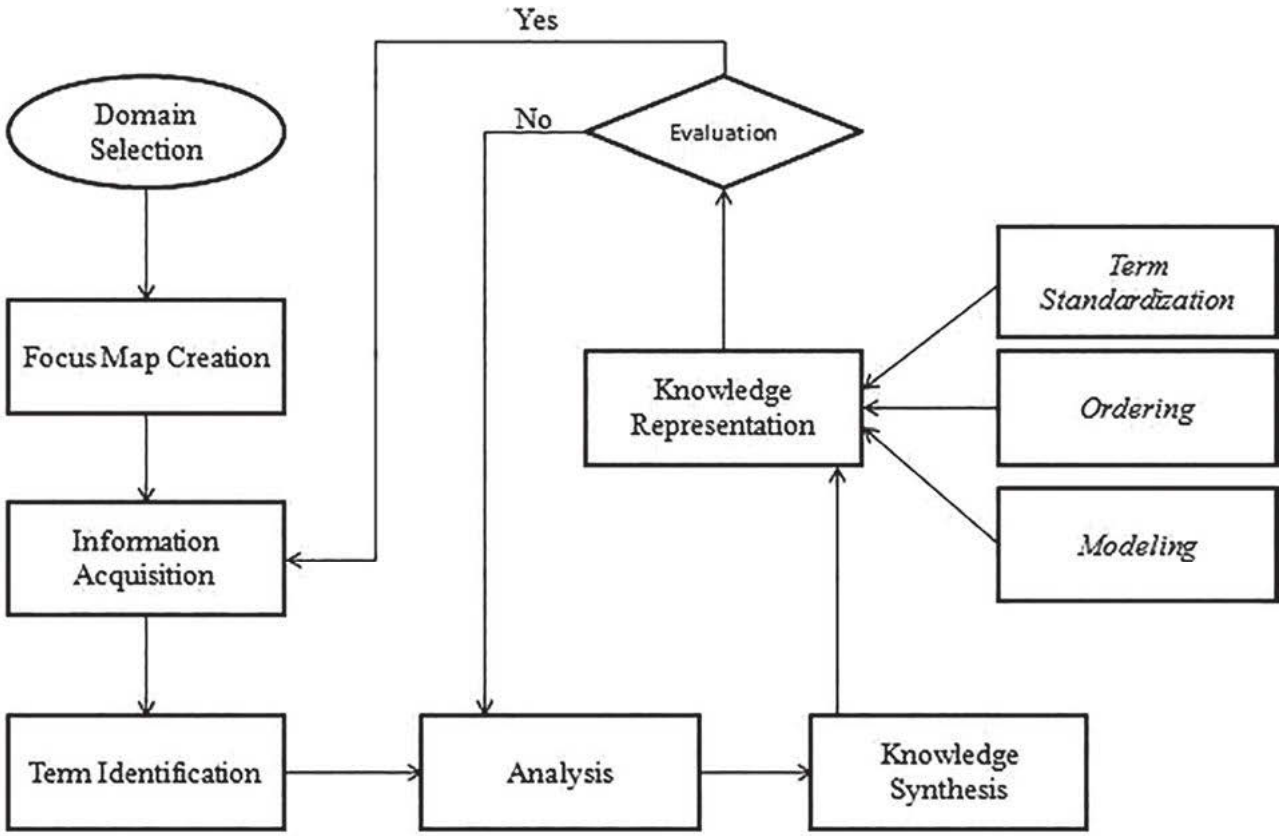

Figure 2. Flowchart of HCFOC methodology.

Due to these features of DERA, the characteristic features of DL, like, soundness, decidability and decision procedures will be inherited into the system. Furthermore, since addition of facets can be done any time, the use of DERA makes the system expandable.

All that has been developed in the previous steps are to be put into action in this penultimate step using DERA. The domain knowledge that has been synthesized in the previous steps is to be expressed in this step by clearly establishing the relationships between the concepts. For this ontology on the tourism domain D, the set of facets Event, MentalObject, PhysicalObject, Substance, etc. belong to the element E, the set of facets hasCreator, hasIdentifier, isBasedFrom, isLocatedAt, etc. belong to the element R and the set of facets Cost, Name, Currency etc. belong to the element A. An example of an established relationship is, Taj Mahal (Subject) isLocatedAt (Predicate) Agra (Object). Similarly, Feni (Subject) isBasedFrom (Predicate) Goa (Object). Here, Taj_Mahal, Agra, Feni, Goa are instances of the classes Monument, AdministrativeDivision, AlcoholicDrink and AdministrativeDivision respectively.

3.8 Step 7: evaluation

In this step, the ontology is put to test. It is to be tested whether the ontology is fulfilling its purpose according to the specifications. It is to be found out whether the ontology is able to deal with the context intended for, thereby addressing other aims and objectives. Necessary corrections are to be made, if required, by going back and repeating the steps in order.

The syntactic correctness and consistency of the tourism ontology were checked in Protégé, using the HermiT OWL reasoner. The HCFOC methodology itself ensures the completeness and conciseness of the tourism ontology. The usability of the tourism ontology from a tourist's point of view has been gauged. It has been found out whether the ontology is capable of understanding the tourist's context with the help of competency queries. Use of competency queries as an evaluation method is one of the best available methods to evaluate an ontology, as has been suggested by Abacha et al. (2013) and Bezerra et al. (2013). Competency queries provided the way to check the entity (E) facet, relation (R) facet and attribute (A) facet together, which are embedded in the form of natural lan-

guage in a given question. For example, queries like "What are the amenities (X) provided by the hotels (Y) nearby to the place?" From this natural language question, we can derive "identification" using a general query pattern:

> Give me all X in Y AND WHERE.property.True. Identification: "concepts" and "properties." Entity: hotel, place relation (R) nearby, place and attribute (A) amenities. Boolean.

The evaluation of the tourism ontology was carried out by research scholars and students belonging to Jadavpur University, Kolkata who are quite enthusiastic when it comes to tourism. Queries collected from them helped in checking the elements entity, relation and attribute embedded in natural language together. They have analyzed whether the ontology is able to meet their criteria of needs. The group of evaluators and the group from whom the competency questions had been obtained were disjoint. The evaluators were asked to imagine that they are going to visit a place as a tourist. Now, based on the above situation, they were asked to enlist the questions whose answers they would like to know before the visit.

The purpose of the ontology constructors behind setting up this situation and asking the evaluators to complete the task was to list as many user queries as possible. This helped in understanding the appropriateness of the ontology framework. This also helped in finding out whether the ontology could meet the purpose for which it was built. The comments posted by the evaluators served as a potential feedback on the usefulness of the tourism ontology and how it could be improved further. The listed questions provided an in-depth insight on the information seeking behavior of the tourists.

Key terms were extracted from the questions listed by the users. Then, it was found out manually for each extracted key term whether, the term, or a synonymous term, or the concept denoted by the term is present in the ontology or not. The questions with all key terms having representation in the ontology were marked as fully answerable. The questions with some key terms having representation in the ontology were marked as partially answerable. The questions with all key terms having no representation in the ontology were marked as not answerable. The fully answerable, partially answerable and not answerable questions were pointed out to the evaluators after the marking. 90.19% of the queries posed by the evaluators were fully answerable. 2.94% of the queries were partially answerable. 6.86% were unanswerable. The concepts denoted by the key terms present in the questions that were partially

Query	Extracted key terms or concepts
What is the currency of the place?	<currency, location>
What are the festivals that will be held in Paris during the time of visit?	<festival, location, time>
Which mode of transport is to be availed to visit the place?	<transport, location>
Where to stay during the visit to the place?	<accommodation, location>
What are the local foods available at that place?	<cuisine, location>

Table 2. List of some of the fully answerable queries and the extracted key terms or concepts.

Query	Extracted key terms or concepts
Which are the tourist spots adjacent to the place?	<tourist spot, adjacent, location>
Is the place secure for tourists?	<location, security, tourists>
Are there any nearby markets to the place?	<markets, nearby, location>

Table 3. List of some of the partially answerable queries and the extracted key terms or concepts.

Query	Extracted key terms or concepts
Will water be available on the highways connecting two places?	<water, highways, location>
What are the amenities provided by the hotels nearby to the place?	<amenities, hotels, nearby, location>
What is the socio-political scenario of the place?	<socio-political, location>
What are the fields that an educational institution located at a place specializes in?	<education, institute, specialization>
How is the mobile network connectivity at the place?	<mobile, network, connectivity, location>

Table 4. List of some of the unanswerable queries and the extracted key terms or concepts.

Evaluators	No. of queries	No. of Fully answerable queries	No. of Partially answerable queries	No. of Not answerable queries
Evaluator 1	9	8	1	0
Evaluator 2	6	6	0	0
Evaluator 3	7	6	0	1
Evaluator 4	7	6	0	1
Evaluator 5	6	5	1	0
Evaluator 6	6	5	0	1
Evaluator 7	7	7	0	0
Evaluator 8	11	10	0	1
Evaluator 9	8	8	0	0
Evaluator 10	4	4	0	0
Evaluator 11	7	4	1	2
Evaluator 12	8	7	0	1
Evaluator 13	5	5	0	0
Evaluator 14	6	6	0	0
Evaluator 15	5	5	0	0
Total	102	92	3	7

Table 5. Statistics of queries posed by the evaluators.

answerable or not answerable were analyzed. Out of the unanswerable questions many were found to be out of the scope of the ontology. For example, "What is the socio-political scenario of the place?" or "What are the fields that an educational institution located at a place specializes in?" Queries expressed using spatial relations like "nearby," "adjacent," etc., were either unanswerable or were partially answerable. During the initial conception of the tourism ontology, only geo-coordinates were included for locating a place. But, since users tend to express queries using the natural language sense of distance, hence it was decided that the spatial relations must be included. Some of the spatial relations, as has been mentioned in Dutta et al. (2011), that need to be considered are: directional (north, south, north-east, south-west, etc.), internal (inside, central, etc.), external (adjacent, nearby, etc.), position with respect to a border (overlap, opposite, etc.), longitudinal (behind, towards, etc.), sideways (left, right, etc.) and relative (up, below, etc.).

The class visualization of the entities in the tourism ontology has been done using the ProtégéVOWL (http://vowl.visualdataweb.org/protegevowl.html) visualization tool. Figure 3 partially shows the hierarchy of the tourism ontology and the class visualization on the left and right side of the figure respectively. The connected entities and the visualization were shown to the evaluators. The key terms from the evaluators' queries that were visible in the visualization were pointed out to the evaluators.

For query visualization and analytics, the tourism model was deployed using GraphDB (http://graphdb.ontotext.com) by OntoText, an enterprise-ready semantic graph database, compliant with W3C standards. Figure 4 depicts how the food named "panipuri," an instance of AsianCuisine is related with the country India by the relation isBasedFrom. On the right side of Figure 4, description, type and rank of this namedIndividual are available. Similarly, Figure 5 shows that "pizza," an instance of EuropeanCuisine isBasedFrom the country Italy.

Then they were asked to comment expressing their concerns and suggestions. The reports given by the evaluators once again acted as a tool to determine the usefulness of the tourism ontology and how it could be improved further. This strategy of evaluation ingrained in the HCFOC methodology aims to increase the exhaustiveness of the ontology. After taking into consideration the comments of the evaluators and with a view of expanding the scope and coverage of the ontology many concepts have been noted down. As of now, the tourism ontological model does not contain background knowledge on transport booking system, price or tax havens (for business tourism where people may seek information for taking advantage of lower prices (for example, booze cruise) or people trying to take advantage of tax loopholes), social welfare systems (for benefit tourism where people move to take advantage of welfare schemes), law (for people moving to take advantage of the legal system for filing lawsuits), birthright (for tourists with the purpose of giving birth in the destination), hotel amenities (for example, number of suites, air-conditioning), accessibility of disables (website accessibility, trained staff availability for dealing with accessibility issues, well-adapted

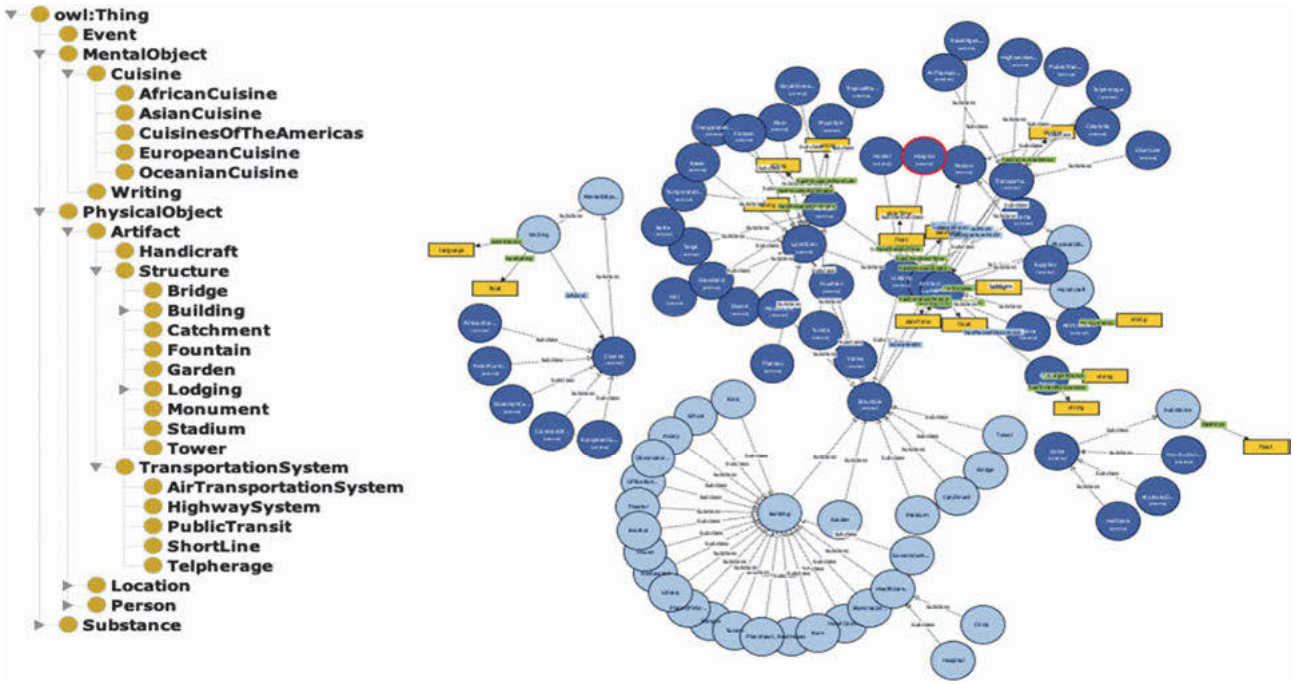

Figure 3. Connected entities.

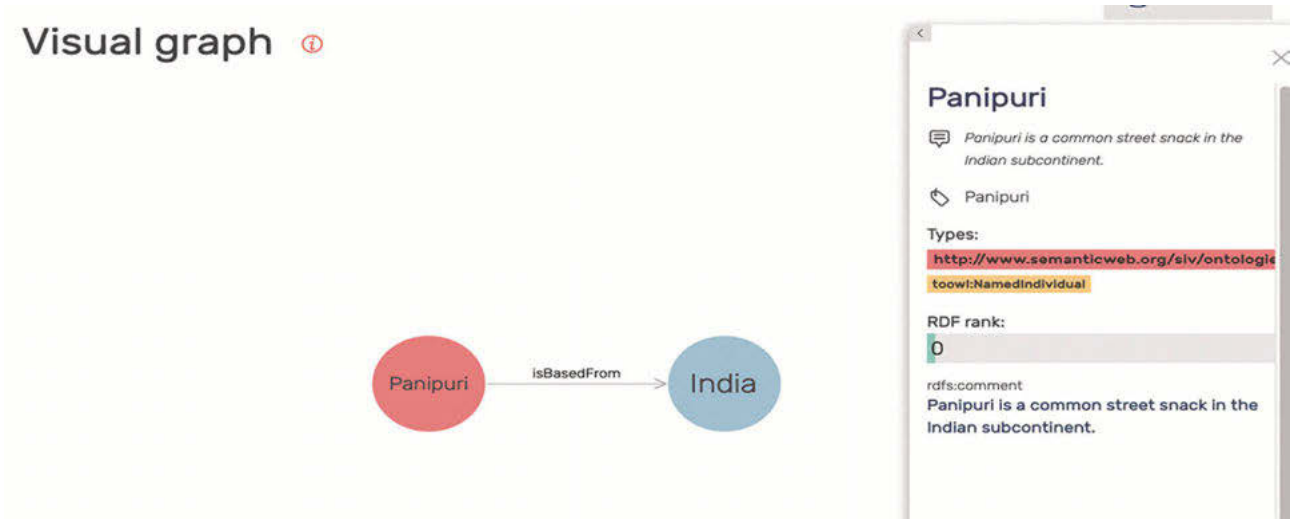

Figure 4. Visualization of individuals panipuri and India.

hotel rooms, technical aids and disability equipment such as wheelchairs, bath chairs and toilet raisers, accessible restaurants and bars, adapted toilets in restaurants and public places, accessible streets and sidewalks, a specific attraction's level of accessibility), agri-tourism events (for people willing to participate in cattle drives or ranches), adventure sports (for example, rambling, climbing, biking, horseback riding, caving, hiking, trekking, snowshoeing, ski mountaineering, diving, rafting), drug tourism, fashion tourism, genealogy tourism, halal tourism (pork and alcohol free flights and hotels, separate spa and swimming pools for men and women, announcement of prayer timings and religious programs), kosher tourism (for orthodox Jews requiring kosher foods, accommodations within walking distance from synagogues, flights with kosher meals), literary tourism (dealing with places and events from fictional texts and their authors' lives, for example, Tolkien tourism by fans of *The Lord of the Rings*), romance tourism (for people travelling in search of relationship), sex tourism (for people travelling to have sex), set-jetting (for people traveling to destinations first seen in movies), medical tourism or wellness tourism (for people travelling to obtain medical treatment or improve health focusing on prevention), suicide or euthanasia tourism (for people traveling to commit

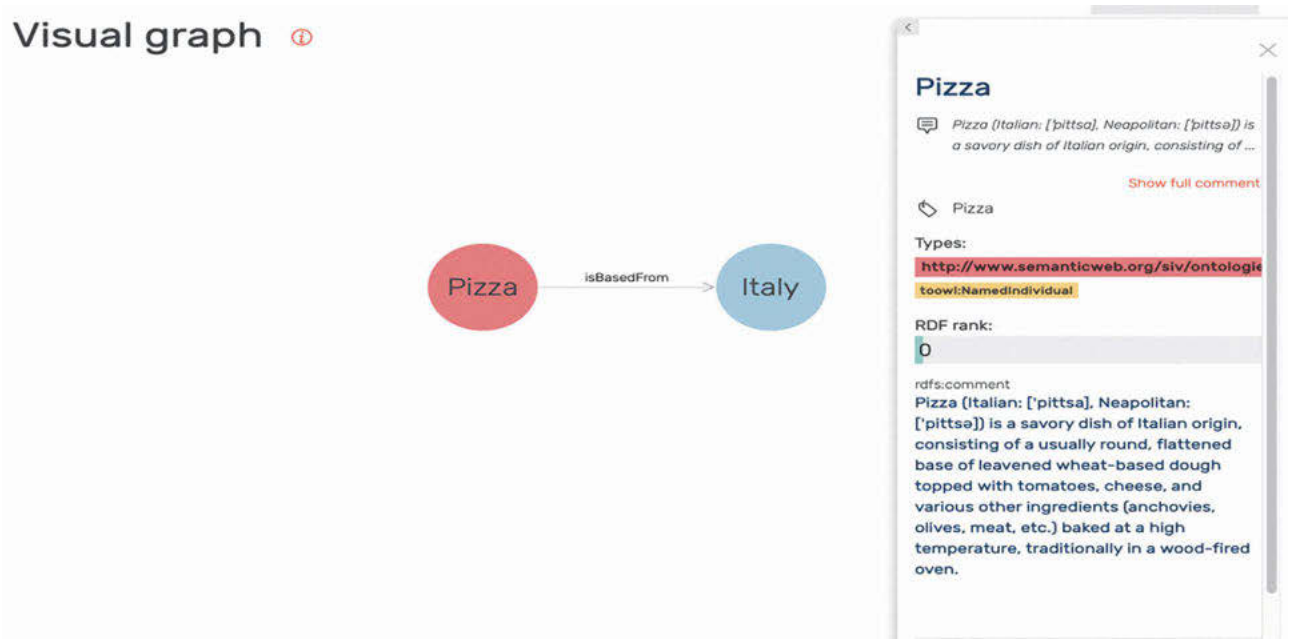

Figure 5. Visualization of individuals pizza and Italy.

suicide or assisted suicide), 3DVT (or 3D virtual tourism, for people willing to explore physical places without physical travel) or tombstone tourism (for people travelling to visit cemeteries, epitaphs, etc.).

The evaluation strategy also presented some questions that will catalogue the guidelines for any tourism information system that may be built on top of this ontology. For example, "How much money is required for this tour?" or "Which mode of transport will be cheaper?" To answer such questions, the information system behind which the tourism ontology will be instrumental, must have capabilities of drawing inference. It would be advantageous to make use of the resources made available by the LOD (linked open data) project rather than populating the ontology with individuals that would be time-consuming. Such an effort has been seen in Dastgheib, Mesbah and Kochut (2013), where the mOntage framework has been introduced, which allows populating the ontology from selected LOD sources. Prototype system architecture has been provided below:

We envisage that the demonstrated tourism ontology constructed following the HCFOC methodology will be able to support any tourism information retrieval system. A prototype system architecture has been provided in Figure 6. In our future work, we intend to implement a prototype system using GraphDB by OntoText (http://graphdb.ontotext.com). It is an enterprise-ready semantic graph database, compliant with W3C standards. Semantic graph databases (also called RDF triple stores) provide the core infrastructure for solutions where modelling agility, data integration, relationship exploration and cross-enterprise data publishing and consumption are important. The "connected" graph is the final implementation of the model in the GraphDB platform. Figure 4 and 5 depicts snap shots of the connected graph of HCFOC ontology. From Figure 4 and 5 we can easily understand how one individual is connected with other related entities. Where the same color nodes represent entities, which belong to the same class, and directed arrows depict how they are connected. GraphDB also supports queries based on simple structured query language (SQL) as well as semantic similarity.

4.0 Conclusion and future work

The HCFOC methodology demonstrated here depicts the detailed modus operandi followed for building the ontology. The efficacy of the human-centric context modeling and the faceted approach ingrained within the HCFOC methodology were explicitly visible while building the tourism ontology. The tourism ontology has been built considering the tourist as the primary information seeker in this domain and thus the tourist was at the centre of conception of the ontology. 90% of the queries could be answered just after the initial phase of the ontology construction. It is contemplated that after inserting the concepts succeeding the initial evaluation phase, the precision of answers returned will further increase. Owing to the advent of the faceted approach, it will be extremely effortless and straight forward to insert concepts into the ontology, in the near future. For example, tombstone tourism can be

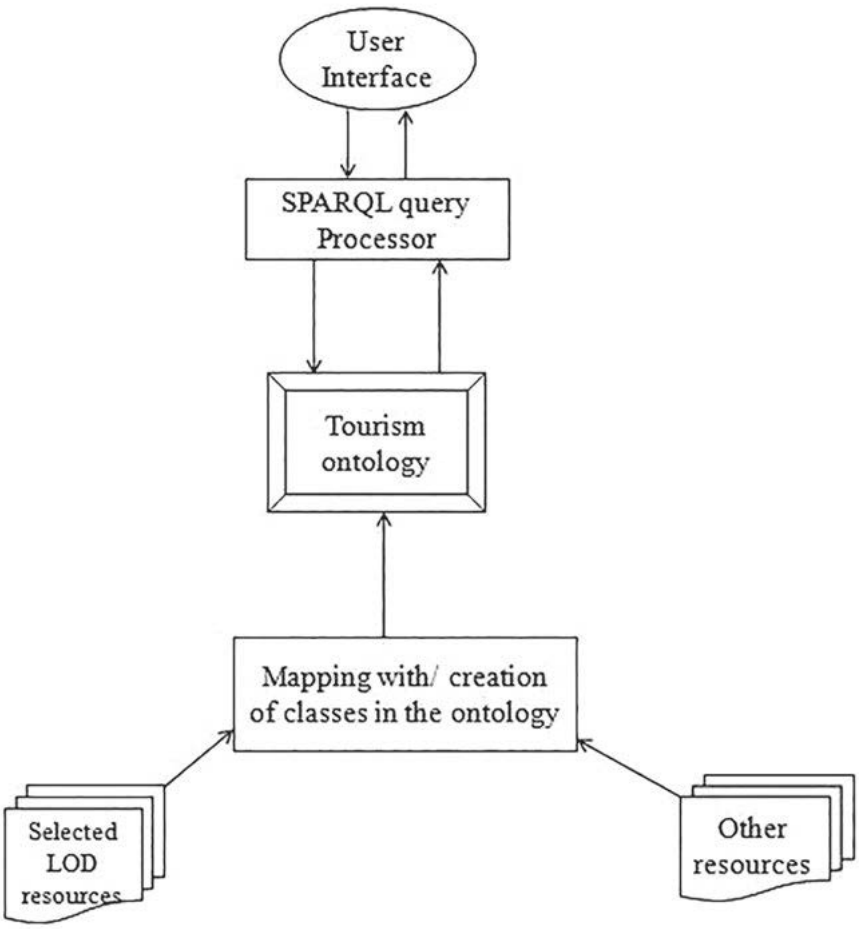

Figure 6. Prototype system architecture.

inserted while considering tombstone under structure and listing down its attributes. Similarly, agri-tourism events, adventure sports can be listed under events. Thus, faceted approach ingrained in the HCFOC methodology accounts for the scalability of the tourism ontology. The HCFOC methodology is non-domain specific and future work involves using it to build ontologies for other domains. The tourism ontology is to be expanded and it is also to be found out whether parts of the tourism ontology can be reused. We intend to take up the work of demonstrating the advantages of HCFOC in respect to the other existing domain ontology models or framework. A detailed comparative and comprehensive study between HCFOC and the ontology construction methodologies covered in our literature review is being planned at a very rudimentary level. In that future work, we wish to include any new ontology construction methodology that emerges, followed with a discussion on how other usual ontology development methodologies can improve by taking insights from HCFOC.

References

Abacha, Asma Ben, Marcos Da Silveira and Cédric Pruski. 2013. "Medical Ontology Validation through Question Answering." In *Artificial Intelligence in Medicine: 14th Conference on Artificial Intelligence in Medicine, AIME 2013, Murcia, Spain, May 29 - June 1, 2013. Proceedings,* ed. Niels Peek, Marin Morales, Luis Roque and More Peleg. Lecture Notes in Computer Science 7885. Berlin: Springer, 196-205. doi:10.1007/978-3-642-38326-7_30

Al-Hassan, Malak, Haiyan Lu and Jie Lu. 2015. "A Semantic Enhanced Hybrid Recommendation Approach: A Case Study of e-Government Tourism Service Recommendation System." *Decision Support System* 72: 97-109. doi:10.1016/j.dss.2015.02.001

Bezerra, Camila, Fred Freitas and Filipe Santana. 2013. "Evaluating Ontologies with Competency Questions." In *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Atlanta, GA, USA, Nov 17-20, 2013.*

Washington, DC: IEEE Computer Society, 284-5. doi: 10.1109/wi-iat.2013.199

Bhattacharyya, Ganesh. 1975. "Fundamentals of Subject Indexing Languages." In *Ordering Systems for Global Information Networks: Proceedings of the Third International Study Conference on Classification Research, Bombay, India. January 6-11, 1975,* ed. A. Neelameghan. Bangalore: Sarada Ranganathan Endowment for Library Science, 83-9.

Bhattacharyya, G. 1981. "Subject Indexing Language: Its Theory and Practice." In *Proceedings of the DRTC Refresher Seminar-13: New Developments in LIS in India, October 14-17, 1981*. Bangalore: Indian Statistical Institute. PAGE

Chu, Yan, Hongbin Wang, Liying Zheng, Zhengkui Wang and Kian-Lee Tan. 2016. "TRSO: A Tourism Recommender System Based on Ontology." In *Knowledge Science, Engineering and Management: 9th International Conference, KSEM 2016, Passau, Germany, October 5-7, 2016, Proceedings*. ed. Franz Lehner and Nora Fteimi. Lecture Notes in Computer Science 9983. Cham: Springer, 567-79. doi:10.1007/978-3-319-47650-6_45

Dahlberg, Ingetraut. 1978. "A Referent-Oriented, Analytical Concept Theory for INTERCONCEPT." *Knowledge Organization* 5: 142-51. doi: 10.5771/0943-7444-1978-3-142

Das, Subhashis and Sayon Roy. 2016. "Faceted Ontological Model for Brain Tumour Study." *Knowledge Organization* 43: 3-12. doi:10.5771/0943-7444-2016-1-3

Dastgheib, Shima, Arsham Mesbah and Krys Kochut. 2013. "mOntage: Building Domain Ontologies from Linked Open Data." In *ICSC 2013: 2013 IEEE Seventh International Conference on Semantic Computing: Proceedings, 16-18 September, 2013, Irvine, California, USA,* ed. Randall Bilof. Los Alamitos, CA: IEEE Computer Society, 70-77. doi:10.1109/icsc.2013.21

Dutta, Biswanath, Fausto Giunchiglia and Vincenzo Maltese. 2011. "A Facet-Based Methodology for Geo-Spatial Modeling." In *GeoSpatial Semantics: 4th International Conference, GeoS 2011, Brest, France, May 12-13, 2011; Proceedings,* ed. Christophe Claramunt, Sergei Levashkin and Michela Bertolotto. Lecture Notes in Computer Science 6631. [Berlin]: Springer, Berlin, Heidelberg, 133-50. doi:10.1007/978-3-642-20630-6_9

Ferrandez, Oscar, Christian Spurk, Milen Kouylekov, Iustin Dornescu, Sergio Ferrandez, Matteo Negri, Ruben Izquierdo, et al. 2011. "The QALL-ME Framework: A Specifiable-Domain Multilingual Question Answering Archtecture." *Web Semantics: Science, Services and Agents on the World Wide Web* 9*:* 137-145. doi:10.1016/j.websem.2011.01.002

Fodor, Oliver and Hannes Werthner. 2005. "Harmonise: A Step Toward an Interoperable E-Tourism Marketplace." *International Journal of Electronic Commerce* 9: 11-39. doi:10.1080/10864415.2005.11044324

Frikha, Mohamed, Mohamed Mhiri, Mounir Zarai and Faiez Gargouri. 2016. "Using TMT Ontology in Trust Based Medical Tourism Recommender System." In *IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), Agadir, Morocco, Nov 29 – Dec 02, 2016*. Piscataway, NJ: IEEE. 1-8. doi:10.1109/aiccsa.2016.7945768

Ghosh, Shiv Shakti and Sunil Kumar Chatterjee. 2019. "Assessment of Digital Reference Tools/Websites Related to Tourism Information Using Webometrics and Qualitative Analysis." *SRELS Journal of Information Management* 54: 26-31. doi:10.17821/srels/2019/v56i1/137324

Giunchiglia, Fausto and Biswanath Dutta. 2011. "DERA: A Faceted Knowledge Organization Framework." http://eprints.biblio.unitn.it/2104/

Giunchiglia, Fausto, Biswanath Dutta and Vincenzo Maltese. 2009. "Faceted Lightweight Ontologies." In *Conceptual Modeling: Foundations and Applications*, ed. Alexander T. Borgida, Vinay K. Chaudhri, Paolo Giorgini and Eric S. Yu. Lecture Notes in Computer Science 5600. [Berlin]: Springer, Berlin, Heidelberg, 36-51. doi:10.1007/978-3-642-02463-4_3

Giunchiglia, Fausto, Biswanath Dutta and Vincenzo Maltese. 2014. "From Knowledge Organization to Knowledge Representation." *Knowledge Organization* 41: 44-56. doi:10.5771/0943-7444-2014-1-44

Gregor, D., S. Toral, T. Ariza, F. Barrero, R. Gregor, J. Rodas and M. Arzamendia. 2016. "A Methodology for Structured Ontology Construction Applied to Intelligent Transportation Systems." *Computer Standards & Interfaces* 47: 108-19. doi:10.1016/j.csi.2015.10.002

Hjørland, Birger. 2017. "Domain Analysis." *Knowledge Organization* 44: 436-64.

Kashevnik, A. M., A. V. Ponomarev and A. V. Smirnov. 2017. "A Multimodel Context-Aware Tourism Recommendation Service: Approach and Architecture." *Journal of Computer and Systems Sciences International* 56: 245-58. doi:10.1134/s1064230717020125

Lamsfus, Carlos, Aurkene Alzua-Sorzabal, David Martín, Zigor Salvador and Alex Usandizaga. 2009. "Human-Centric Ontology-Based Context Modelling in Tourism." In *International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KEOD), Madeira, Portugal, October 6-8, 2009,* ed. Jan L. G. Dietz. Setúbal: SciTePress, 424-34. doi:10.5220/0002300704240434

Lee, Chin-I, Tse-Chih Hsia, Hsiang-Chih Hsu and Jing-Ya Lin. 2017. "Ontology-Based Tourism Recommendation System." In *4th International Conference on Industrial Engineering and Applications (ICIEA), Nagoya, Japan, April 27-29, 2017*. ed. Felix T.S. Chan. Piscataway, NJ: IEEE. 376- 9. doi:10.1109/iea.2017.7939242

Luhn, H. P. 1961. "Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment." *American Documentation* 12: 131-8. doi:10.1002/asi.5090120209

Mathur, Archana, Akshatha K, Apoorva Shastry and Anitha J. 2015. "A Survey on Existing Tourism Ontologies." *International Journal of Research in Engineering and Technology* 4: 20-3. doi:10.15623/ijret.2015.0426005

Miller, George A. 1995. "WordNet: A Lexical Database for English." *Communications of the ACM* 38, no. 11: 39-41

Nguyen, Thi Thanh Sang and Haiyan Lu. 2016. "Domain Ontology Construction Using Web Usage Data." In *AI 2016: Advances in Artificial Intelligence: 29th Australasian Joint Conference, Hobart, TAS, Australia, December 5-8, 2016, Proceedings,* ed. Byeong Ho Kang and Quan Bai. Lecture Notes in Computer Science 9992. Cham: Springer, 338-44. doi:10.1007/978-3-319-50127-7_29

Prieto-Díaz, Rubén. 2003. "A Faceted Approach to Building Ontologies." In *Proceedings Fifth IEEE Workshop on Mobile Computing Systems and Applications*, *Monterey, California, October 9-10, 2003*. ed. DeeberAzada. Los Alamitos, CA: IEEE.458-65. doi:10.1109/iri.2003.1251451

Qiu, Jing, Lin Qi, Jianliang Wang and Guanghua Zhang. 2017. "A Hybrid-Based Method for Chinese Domain Lightweight Ontology Construction." *International Journal of Machine Learning and Cybernetics* 9: 1519-31. doi:10.1007/s13042-017-0661-0

Ranganathan, S. R. 1989. *Colon Classification.* 7th ed., rev. and ed. M.A. Gopinath. Bangalore: Sarada Ranganathan Endowment for Library Science.

Ranganathan, Shiyali Ramamrita. 1937. *Prolegomena to Library Classification*. Madras: Madras Library Association.

Schema.org. 2019. "Thing." Accessed October 24. https://schema.org/Thing

She, Keng-Seong, Su-Cheng Haw, Yu-Gene Loh and Fang-Fang Chua. 2018. "AK Tourism: A Property Graph Ontology-based Tourism Recommender System." In *Knowledge Management International Conference (KMICe), July 25 -27 2018, Miri Sarawak, Malaysia.* ed. Fauziah Baharom Assoc, Yuhanis Yusof, Mazida Ahmad and Rohaida Romli. Changlun, Malaysia: UUM College of Arts and Sciences, 83-8. http://repo.uum.edu.my/25195/

Smiraglia, Richard P. 2015. *Domain Analysis for Knowledge Organization: Tools for Ontology Extraction*. Waltham, MA: Chandos.

Staab, Steffen, Christian Braun, Ilvio Bruder, Antje Düsterhöft, Andreas Heuer, Meike Klettke, Günter. Neumann, et al. 1999. "GETESS: Searching the Web Exploiting German Texts." In *Cooperative Information Agents III: Third International Workshop, CIA'99 Uppsala, Sweden, July 31–August 2, 1999 Proceedings,* ed. Matthias Klusch, Onn M. Shehory and Gerhard Weiss. Lecture Notes in Computer Science 1652. [Berlin]: Springer, Berlin, Heidelberg, 113-24. doi:10.1007/3-540-48414-0_7

Studer, Rudi, V., Richard Benjamins and Dieter Fensel. 1998. "Knowledge Engineering: Principles and Methods." *Data & Knowledge Engineering* 25: 161-97. doi:10.1016/s0169-023x(97)00056-6

Suárez-Figueroa, Mari Carmen, Asunción Gómez-Pérez and Mariano Fernández-López. 2015. "The NeOn Methodology Framework: A Scenario-Based Methodology for Ontology Development." *Applied Ontology* 10: 107-45. doi:10.3233/ao-150145.\

UNWTO (World Tourism Organization). 2001. *Thesaurus on Tourism & Leisure Activities (Trilingual: English, French, Spanish).* [Madrid]: World Tourism Organization. https://www.e-unwto.org/doi/pdf/10.18111/9789284404551

UNWTO (World Tourism Organization). 2019. "Why Tourism? Accessed October 25. https://www.unwto.org/why-tourism

Yang, Yuehua, Yuan Ping, Junping Du and Hui Ma. 2017. "A Domain Ontology Construction Method with Ontology Modification Effort Assessment." In *Proceedings of 2017 Chinese Intelligent Automation Conference, Tianjin, China, June 2-4,* ed. Zhidong Deng. Lecture Notes in Electrical Engineering 458. Singapore: Springer, 473-80. doi:10.1007/978-981-10-6445-6_52