Semantic Search System for Common Cultural Traditions of the Greater Mekong Subregion

Suwannee Hoaihongthong¹, Kanyarat Kwiecien²*

¹,²Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand,

*Corresponding author: Kanyarat Kwiecien (Email: kandad@kku.ac.th)

Abstract

This article presents the development of a semantic search system for the common cultural traditions of the Greater Mekong Subregion, which has been studied using common cultural tradition data from Thailand, Laos, and Cambodia. Our system presents the implementation of semantic search to solve the problem of semantic gaps in common cultural characteristic concepts for information access and retrieval. The field of cultural heritage management represents a growing area of research that contains valuable information for users, researchers, and domain experts. Finding information on traditions or cultures of interest among unstructured materials requires a laborious search process. In this paper, we demonstrate how to extract semantic entities and relationships represented in the information using the theory of information science, specifically the Knowledge Organization (KO) approach, which involves determining the concepts and semantic relations of a particular issue and establishing their relations with relevant concepts. It serves as a foundation for knowledge structure development and ontology development, which contribute to effective information searching and knowledge access. The system was developed based on systematic ontology using the Hozo Ontology Editor and Database and ontology mapping using the Ontology Application Management (OAM) framework, which helps distinguish between relevant and irrelevant data, thereby leading to search effectiveness. To cut down on the time wasted in browsing and searching, reduce associated user frustration, and support the research of domain experts, more selective user access is needed.

Keywords: Common culture, Data analysis, Knowledge organization, Ontology, Semantic search system, Traditions common culture.

1. Introduction

Common cultures are an underlying strength of the Greater Mekong Subregion (GMS) countries. Regional cultures not only portray the prosperity of society in each locality but also serve as a vital foundation of the culture as a whole. The globalized cultural trends are important for adaptation, appeal, and the genuine character of each community and culture. As
As a result, the realization of diverse local cultural values can support long-term social and cultural development in line with current circumstances [1, 2].

The term “common culture” refers to a shared cultural basis or similar cultural traits that are shared by many parties or transcend ownership [3]. This form of culture refers to cultural features shared among countries, such as history, values and principles, purposes and sense of mission, and symbols and boundaries [4]. Thailand, the Lao People's Democratic Republic, and the Kingdom of Cambodia share cultural similarities. Simply put, they share cultural roots due to geographical factors, cultural foundations, religion, and similar ways of life. Furthermore, they continue to transmit culture to one another through visits, trade, and relocation, resulting in similar practices and trends. The common culture is regarded as an important component of public diplomacy policies and is used to strengthen international relations to maintain benefits for each country. Culture is thus used to establish international relations, carry out tourism policies, and develop strategies for tourism promotion, particularly cultural tourism, in order to generate revenue for countries [5]. Aside from that, a common culture is associated with economic policies; that is, attention is given to the promotion of the creative and cultural services industries, such as food, films, fashion, fighting, and festivals, as well as Thai cultural products. Most importantly, the ASEAN Socio-Cultural Community Blueprint was developed to foster interaction, awareness, and pride in the regional identity and culture [6].

Despite various cooperation policies and the cultural cooperation framework, no international organizations have been clearly assigned the task of conducting research and storing and organizing knowledge about the GMS countries’ common culture. Access to cultural knowledge is the responsibility of each country, which each have its own organization tasked with collecting cultural information about that specific country [7]. For example, the access to cultural information is similar in Laos and Cambodia; the information is kept at the Ministry of Culture and Fine Arts, and the Heritage and Arts Information Center’s website provides information services. However, the presentation of this information lacks knowledge organization, links, and search tools [8]. A significant amount of knowledge is available from a variety of sources and genres. An automatic system is needed to retrieve meaningful information from a vast volume of data. Material summarization systems, for instance, help in content reduction by keeping the vital information and filtering out the irrelevant text, as well as by encoding the semantic relations between words, which is critical to getting relevant information [9]. Previous research on the search systems used for knowledge about common culture among GMS countries has revealed that the general search systems are rather restrictive. The system primarily uses character comparison methods to find data or documents. The system does not take into account the precise concepts behind or meanings of search queries. Some search terms have inappropriate attributes or do not represent the content. These shortcomings are mostly due to the semantic gap, the difference between what a computer can understand and what humans understand. The semantic gap distinguishes between two descriptions of an object using different linguistic representations, such as when traditions are searched for by name. Different languages and dialects may refer to the tradition differently, such as Boon Pha Wet, Bun Pha Wet, Bun Phawet, Boun Phavet, The Traditional Mahajati Preaching, Mahachat, and so on. A semantic search system that uses a string matching method and also considers synonyms and related terms would provide a solution to this problem. A search for Naga beliefs, for example, could be used to find traditions based on cultural attributes. The search results would display the traditions associated with that belief. For this reason, the current study has used ontology concepts to develop a structural model of knowledge about traditions in the GMS. This ontology will be used in the future development of semantic search systems.

One frequently employed tool for describing how knowledge is represented is ontology. The benefits of an ontology include its interoperability to share common understanding among humans or software agents; it permits the reuse of domain knowledge and facilitates explicit norms [10]. Ontology is one of the key technologies underlying the Semantic Web. On the Semantic Web, ontologies are exchanged and reused by various developers in a distributed context. Collaboration among numerous developers is required to build large-scale ontologies. Therefore, one of the most important problems is the distributed and collaborative development of ontologies [11]. Digital cultural heritage ontologies are one of the most challenging methods in information management. The digital cultural heritage field has gained wide interest, both from the general public and researchers, and boasts considerable heterogeneity in terms of content and potential uses [12]. This study was based on Hjørland’s concept of knowledge organization [13]. Specifically, categories of knowledge were identified by determining concepts and semantic relations as well as establishing their relations with relevant concepts. The results were organized in the form of classification schemes by displaying the data of the main concepts and roughly sorting the data hierarchy according to groups of related contents. The structure of the data was later developed into ontology. Domain ontology development was grounded on Uschold and King’s concept [14], using three processes: determining the purpose and scope, ontology development, and ontology evaluation. Specifically, Noy and McGuinness’s seven steps for ontology development [15] were applied, and the ontology editor used in this study was the Hozo Ontology Editor (Hozo is a Japanese graphical ontology editor specifically created to produce heavy-weight ontologies and developed through a partnership between the Department of Knowledge Systems (Mizoguchi Laboratory), ISIR-Osaka University, and Enegate Co., Ltd.) [16]. The ontologies were created and evaluated by domain experts and application-based evaluation methods.

The goal of this research was to apply digital humanities research concepts to develop ontologies for the common cultural traditions of the GMS. The ontologies focused on intangible cultural heritage and were intended to be used as information resources and for information retrieval, i.e., the development of a semantic knowledge search concerning the common cultural traditions of the GMS countries. Specifically, it could be used as a model search system for other types of cultural information to facilitate information search in specific fields. Lastly, researchers can develop similar search systems for other issues, which would serve as a means of storing and retrieving knowledge.
2. Literature Survey

An ontology is a commonly used tool for explaining the representation of knowledge. The advantages of an ontology include its interoperability to share common understanding among people or software agents; also, it enables the reuse of domain knowledge and facilitates explicit conventions. As mentioned, digital cultural heritage ontologies are one of the most challenging methods in information management. The digital cultural heritage field has gained wide interest, both from the general public and researchers, and boasts considerable heterogeneity in terms of content and potential uses [12]. In this study, we examined an existing ontology to present a relevant academic discussion on information regarding common cultural traditions in the context of the GMS. The purpose of developing an ontology for cultural heritage is the potential for integrating it with semantic applications. An application provides easy, quick, and intelligent access to project information through a user-customized definition search [17]. Previous studies on cultural heritage ontologies have developed ontologies for the representation of Twelve-month Isan Merit-Making Traditions [18, 19]. However, the context and scope of those studies were narrower than those of the current research. Another study on event ontology developed a vocabulary for describing events that can be used to describe traditions in terms of time and place, but it cannot explain other tradition-related details [20]. Other research has developed semantic search systems for cultural events [21]. This work extracts spatiotemporal and semantic entities and relationships from cultural event news reports and searches within the information, using Natural Language Processing (NLP) and automatic ontology population. A semantic search engine enables us to answer where and when an event will take place. Finally, a study by Tuamsuk, et al. [12] on cultural heritage ontologies focused on the ontologies of folktales in the GMS. The folktales domain, as a part of literature, is vast in scope and divergent in terms of concepts and conceptual relations and is also related to traditions [22]. The disparity between this study and the previous research lies in the different research focuses. This study examined the common cultural traditions of the GMS, which represent a wider scope than in previous research. In addition, the content used in the present study more explicitly references traditions, and it can be described as encompassing the coherence of cultural traditions in the GMS. More significantly, an ontology and semantic search system of the knowledge of common cultural traditions in the GMS has yet to be created.

3. Method
3.1 Designing Ontology Structures with the Hozo Ontology Editor

Ontology research is a way to add semantics and structure to web documents. Ontologies semantically model and describe details of domain-related entities (e.g., tradition, culture) in a fully integrated way and can represent a large amount of information using a small number of axioms (classes and properties) [23].

In this study, we developed an ontology of the common cultural traditions of the GMS and constructed the knowledge domain using the knowledge classification approach. The list of terms and their relations were input, defined, and confirmed by experts. This section describes how an ontology of the common cultural traditions of the GMS was developed using the Hozo Ontology Editor.

The ontology editor used in this study was the Hozo Ontology Editor (a Japanese graphical ontology editor specifically created to produce heavy-weight ontologies and developed through a partnership between the Department of Knowledge Systems (Mizoguchi Laboratory), ISIR-Osaka University, and Enegate Co, Ltd., Hozo is a tool for creating and using ontologies. We present the architecture, module functionalities, interface, and some experiences of designing and using ontologies. Hozo is an environment that allows users to work on these things, Hozo includes an API, HozoAPI ver 1.15, which accesses existing ontologies and makes the slot definition option available. Moreover, inheritance information is clear and easy, and the software provides the facility to correct errors at the time of validating the ontology [16].

The scope of our ontology development mainly focused on intangible cultural heritage and was intended to be used for information retrieval and as an information resource. The properties of the “common cultural traditions” class make up the main ontological class with some defined properties, such as history, belief, purpose, location, ritual, activity, literature, values, place, time, principle, person, equipment, and ethnic group. This allows the system to retrieve relevant information. The defining condition or criterion for the validation of class properties was that the classes of common cultural traditions had relationships with all other classes in the ontology, which helped link the data in the semantic search for common cultural traditions of the GMS countries. Based on the definitions of the class properties or relationships between concepts, it was found that 73 concepts had the “is-a” property, while 15 concepts had the “part-of” property, and 52 had the “attribute-of” property, as illustrated in Figure 1.
This paper presents the development of an ontology to represent the knowledge of the common cultural traditions of the GMS. The ontology was carefully designed to specify the concepts relevant to intangible cultural heritage and the relations among them. The knowledge domain in this work focuses on the cultural context and implicit attributes of the traditions of Thai, Laos, and Cambodia as an initial case study. To further illustrate the potential of the developed ontology, a semantic search application was implemented and then evaluated by experts. In the evaluation process, several complicated queries were designed to fully utilize the relations among ontological classes, and the results were returned accurately. The evaluation proved that the ontology was well defined in terms of its hierarchical structure and relations.

Table 1.
Evaluation of knowledge structure for system results.

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Average</th>
<th>Meaning (Level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determine the Scope</td>
<td>4.22</td>
<td>High</td>
</tr>
<tr>
<td>Define Classes / Concepts</td>
<td>3.83</td>
<td>High</td>
</tr>
<tr>
<td>Define Properties</td>
<td>4.11</td>
<td>High</td>
</tr>
<tr>
<td>Create Instances</td>
<td>4.00</td>
<td>High</td>
</tr>
<tr>
<td>Application to Ontology Development</td>
<td>4.17</td>
<td>High</td>
</tr>
<tr>
<td>Total</td>
<td>4.07</td>
<td>High</td>
</tr>
</tbody>
</table>

Ontology evaluation is conducted to verify the structure and descriptions of an ontology. The structure of the ontology was evaluated by three ontology experts, see the results in Table 1. Also, prototyping and hypothesis-testing application-based ontology evaluations were conducted, see Figures 4 and 5.

3.2 Database to Ontology Mapping

The Ontology Application Management (OAM) framework serves as a development platform for ontology-based Semantic Web applications, with a focus on applications for semantic search and recommender systems. It supports RDF data publishing from databases, as well as building knowledge-based systems, and supports DBMS and MySQL. This application framework is designed for ontology apps as it creates a middle layer between user applications and the coding and development environment. It emphasizes support for ontology-based data publishing, as well as access, abstraction, and interoperability. Using the framework's reusable and customizable data and application templates, users can create ontology apps without any programming experience [24, 25].

When we mapped the database using the ontology developed using the Ontology Application Management (OAM) framework created by NECTEC as a software management platform, we planned for it to have a data storage structure that
was consistent with the ontology structure. The implementation of the OAM Framework consisted of 1) preparing the OWL ontology file, 2) creating a database in MySQL, 3) database to ontology mapping configuration, 4) search application configuration, and 5) starting the search application. These steps are illustrated in Figure 2.

![Figure 2](image-url)

**Figure 2.** The implementation of the OAM framework.

### 4. Results and Discussion

#### 4.1. Semantic Search System for Common Cultural Traditions of the GMS

Semantic search is an improved form of contextual search since it has the ability to understand the connections between ontological schema and contextual meanings. A general search system mainly searches for data or documents using character comparison methods. Such systems do not consider the exact concept or meaning of the search query. Some search terms contain inappropriate attributes or do not represent the content. These challenges are mainly due to the huge gap between what the computer can interpret and what humans understand, known as the semantic gap [24]. In this study, the semantic search system's knowledge basis was the common cultural traditions of the GMS. For ease of access, the system was developed as a web-based application. To deliver meaningful search results, the system takes several factors into account, such as the search context, word variations, synonyms, generalized and specialized searches, concept matching, and natural language queries. The software can recognize the subtleties of intangible cultural heritage and provide the most pertinent findings. Only a semantic approach can ensure these benefits since it features a contextual grasp of the meaning of the information. For example, a tradition may have different names in different languages and dialects, such as Boon Pha Wet, Bun Pha Wet, Bun Pwawet, Boun Phvett, The Traditional Mahajati Preaching, Mahachat, etc. Thus, a semantic search system offers a solution to this problem because it uses a string-matching method and also considers synonyms and related terms.
For example, to search for traditions based on cultural attributes, a search for traditions regarding the Naga belief can be used. The search results will then show the traditions associated with that belief.

The semantic search system for common cultural traditions of the GMS has the following properties: 1) The semantic search engine can be programmed via a web browsing program. 2) The system can search by class properties or by traditional name or keyword. The search methods include simple search and semantic search. The prototype of the semantic search system is presented in Figure 3.

Figure 3.
Prototype of a semantic search system for common cultural traditions of the Greater Mekong Subregion.

4.2. System Evaluation

In this research, we used application-based ontology evaluation to conduct an ontology assessment and measure the effectiveness of information retrievals using the semantic search prototype. To develop this prototype, we used the OAM framework. This application provides reusable and configurable application templates and can build prototypes. Moreover, the application template is ideal for rapid prototyping and hypothesis testing. The framework provides Web API to support more advanced application development. The ontology can subsequently be evaluated using the information retrieval measures of precision, recall, and F-measure. We evaluated the system’s performance by calculating: 1) the precision value, to determine the fraction of relevant retrieved documents, 2) the recall value, to determine the fraction of relevant documents that are retrieved, and 3) the F-measure, using an equation. The performance of the keyword search of the semantic search system had precision, recall, and F-measure values of 0.90, 0.88, and 0.85, respectively (see Figure 4), and the evaluation results with the advanced search were 1.00, 1.00, and 1.00, respectively (see Figure 5).

Figure 4.
The performance of the semantic information retrieval system with keyword search.
5. Conclusion

The study and development of a semantic search system for the common cultural traditions of the GMS countries were carried out in compliance with the policy of managing the area’s intangible cultural heritage [26, 27]. As stipulated in the policy, it is necessary to develop a list of intangible cultural heritage, promote research, and collect related documents or evidence to ensure appropriate safeguarding measures and facilitate education. In addition, the knowledge organization in this study aligned with Thailand’s policy for the Management of Intangible Cultural Heritage [27]. That is, a proposal was made to build a database of Thailand’s cultural heritage with accurate and high-quality data verified by experts. It was also proposed that data be systematically collected and organized to create a knowledge base for education and research purposes; this data should be linked and constantly updated. The dissemination of knowledge and a knowledge base must be effective and comprehensive; information sources should be accessed with ease via a knowledge base. These processes are regarded as fundamental approaches to the management of intangible cultural heritage [28-30]. This study of the common cultural traditions of the GMS countries was compliant with the aforementioned policy; that is, the common cultural traditions of the GMS countries are a form of intangible cultural heritage, which we systematically organized and presented. Based on the results of the study, the developed semantic search system of knowledge about the common cultural traditions of the GMS is effective. Moreover, it could be used as a model search system for cultural information, facilitating information searches in specific fields. Finally, researchers can develop similar search systems centered on other issues, which would serve as a way to store and retrieve knowledge. From this perspective, the research presented here opens up new possibilities for further research on the use of ontologies to support the conservation of cultural heritage and, in particular, web-based information search system development using a semantic network to develop a sustainable cultural heritage information system.

References