

Article

A Digital Thesaurus of Ethnic Groups in the Mekong River Basin

Wirong Chansanam ¹, Kanyarat Kwiecien ¹, Marut Buranarach ² and Kulthida Tuamsuk ^{1,*}

¹ Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen 40002, Thailand; wirach@kku.ac.th (W.C.); kandad@kku.ac.th (K.K.)

² National Electronics and Computer Technology Center, Pathumthani 12120, Thailand; marut.buranarach@nectec.or.th

* Correspondence: kultua@kku.ac.th; Tel.: +66-89-8614696

Abstract: This research was aimed at constructing a thesaurus of the ethnic groups in the Mekong River Basin that is a compilation of controlled vocabularies of both Thai and English language, with a digital platform that enables semantic search and linked open data. The research method involved four steps: (1) organization of knowledge content; (2) construction of the thesaurus; (3) development of a digital thesaurus platform; and (4) evaluation. The concepts and theories used in the research comprised knowledge organization, thesaurus construction, digital platform development, and system evaluation. The tool for developing the digital thesaurus was the Tematres web application. The research results are: (1) there are 4273 principle words related to the ethnic groups that have been compiled and classified by the terms for each of the eight deep levels, 2596 were found to have hierarchical relationships, and 6858 had associative relationships; (2) the digital thesaurus platform was able to manage the controlled vocabularies related to the Mekong ethnic groups by storing both Thai and English vocabularies. When retrieved, the vocabulary, details of the broader term, narrow term, related term, cross reference, and scope note are displayed. Thus, semantic search is viable through applications, linked open data technology, and web services.

Keywords: digital thesaurus; ethnic groups; Mekong river basin; digital humanities; linked open data; web services



Citation: Chansanam, W.; Kwiecien, K.; Buranarach, M.; Tuamsuk, K. A Digital Thesaurus of Ethnic Groups in the Mekong River Basin. *Informatics* **2021**, *8*, 50. <https://doi.org/10.3390/informatics8030050>

Academic Editor: Roberto Theron

Received: 14 July 2021

Accepted: 7 August 2021

Published: 9 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The term 'ethnic group' is generally understood in the humanities as a social group or a category of the population who cluster together as part of a bigger society, which can be a community, a country, or a region. Due to common characteristics, frequently taken from biological and physical aspects, and relations in terms of kinship, language, culture, religion, and belief, an ethnic group possesses common key cultural heritage that is different from other groups or communities. Each member is able to perceive the differences and each has a means of communication and interaction in the group. The members of an ethnic group have lifestyles and social activities that differ from one area to another, which also depend on the government's policy, the environment, and socio-economic development carried out in their domicile [1–4]. It can be said that an ethnic group is the population that has become part of a bigger multi-cultural society. An ethnic group is an important element behind the country's development because the members' cultural identities and beliefs are hard to change or demolish. Thus, national development that brings impact on the cultures and beliefs of ethnic groups requires profound knowledge and understanding, especially in a region where a high number of different ethnic groups live. In short, the cultural multiplicity and differences of ethnic groups are significant elements when setting national development strategies [5,6].

Research on the social and developmental aspects of ethnic groups has been conducted in a great number, e.g., studies on the influence of ethnic groups and cultural identities

on conflicts in the United States [7–12], studies on various public policies, for example, education, public health, and population movement, which have an impact on the way of life and cultures of ethnic groups [13,14], and studies on risk factors of minority groups residing in different areas, which include health, consumption, way of life, and cultural change [15,16]. There are also a number of studies that are aimed at classifying ethnic groups, with in-depth analysis or comparison of genetic traits, classification of the colors of skin, eyes, and other characteristics [17–19]. It can be understood that the ethnic group issue is linked to many fields of study other than humanities. There are studies of ethnic groups conducted under the fields of history, archaeology, politics and government, demography, social development, and medical science.

When looking at the field of information sciences, which covers analysis, categorization, system management, and construction of access tools to serve users, research involving ethnic groups is few. As for systematic classification of knowledge that enables access to information in libraries or databases, such as the Dewey Decimal System and American Cabinet System, ethnic groups have been categorized in subclasses and subdivisions of the social science category and/or sociology and humanities, without branching out contents or related details at the ethnic group level (subclass 305 in the system: DDC [<https://www.oclc.org/content/dam/oclc/webdewey/help/300.pdf> (accessed on: 11 March 2021)], subclass HT in the system: LC [https://www.loc.gov/aba/cataloging/classification/lcco/lcco_h.pdf (accessed on: 2 February 2021)]). Research on the spoken languages of ethnic groups, e.g., in Thailand, classified ethnic groups into five categories: Austro-Asiatic, Tai dialect, Sino-Tibetan, Hmong-Mien, and Austronesian [20]. The ethnic groups in Southeast Asia have been classified into four groups: Austro-Asiatic, Tai-Kadai dialect, Sino-Tibetan, and Malayo-Polynesian [21]. Chaikhambung and Tuamsuk additionally conducted a study and categorized the ethnic groups in Thailand in order to build an ontology for semantic search purposes, which was based on knowledge organization. The ethnic groups in Thailand have been divided into 12 classes and 51 subclasses [22,23]. Later, Chansanam et al. [24], extended the finding by arranging the Thai ethnic groups ontology in an accessible form using linked open data technology. Categorization of ethnic groups in different countries in the Mekong Basin appears in the research by, for instance, Pholsena [25], who reported that the first Prime Minister of Lao PDR, Kaysone Phomvihane, changed the vocabulary and ethnic group classification in Lao PDR by dividing them into Lao-Thai, Mon-Khmer, Hmong-Mien, and Sino-Tibetan. In addition, Mackerras [26] mentioned the classification of the minority groups in China in his study. These works, however, did not analyze the categories or organize knowledge based on the principle of information science, nor did they have the objective to apply the results to knowledge management or semantic search, and hence they were not able to provide open access via a web service.

A thesaurus is a knowledge resource that compiles vocabularies that have been classified into groups based on semantic similarities and word relationships. Schütze and Pedersen [27] defined a thesaurus as the matching of one word to another that is related. The list of words in a thesaurus is in the alphabetical order, with priorities given to the classes and the broader term (BT), related term (RT), and cross reference that states both the use and use-for cases [28,29]. It can be said that a thesaurus is a tool that assists users to understand the overall vocabulary under any one domain. Moreover, it is important as a retrieval tool for information in various databases. This is because a thesaurus organizes knowledge that represents the concept and vocabulary used in practice or found in the natural language. It also provides additional meaning and relationships of words in its corpus. A thesaurus is thus used as an efficient and precise retrieval index as required by users [30]. In this research, we defined theoretical concepts of a “word” as “a unit of language that native speakers can identify” and a “term” as “a word or expression used for some particular thing.”

A thesaurus in information science is the controlled vocabulary developed, in terms of both structural and grammatical methods, to compile the words used and is used as the

tool to provide descriptive keywords to the document. A thesaurus is also used to assist in selecting the keyword when searching information. It is thus the controlled vocabulary, it is not only a set of vocabulary that assists the index maker or the searcher to select a word that best represents the topic needed, but it also enables the user to see the holistic picture of all words and arrive at the elements of the topic being searched, which cover all relevant aspects. Compared to subject headings, which is also a controlled vocabulary, the controlled vocabulary of a thesaurus seems similar to subject headings, but with a more complicated structure. The symbols representing the relationship also have clearer specific forms and meaning than subject headings [31–33].

From studying related digital thesauruses or online thesauruses, no digital thesauruses in the ethnic group domain was found. The existing, internationally and widely known thesaurus, for instance, the UNESCO Thesaurus [<http://vocabularies.unesco.org/browser/thesaurus/en/> (accessed on: 2 February 2021)], is the controlled vocabulary in the field of education, culture, natural science, humanities, social sciences, communication, and information. It was found that most vocabularies related to ethnic groups were of the ethnic groups in America and Africa, but it had inadequate details to be used as a management or specific access tool for ethnic groups. Other cultural digital vocabulary corpuses include Yale University [<https://hraf.yale.edu/> (accessed on: 16 April 2021)] and The Getty Research Institute [<https://www.getty.edu/research/tools/vocabularies/aat/> (accessed on: 2 February 2021)], in which limited ethnic vocabulary is found that has not been put in the form of a thesaurus.

The Mekong is the life source for over 60 million people who live around its basin in the areas of six countries including Southern China, Myanmar, Laos, Thailand, Cambodia, and Vietnam. The basin is the source of multiplicity of cultures and the domicile of over 95 different ethnic groups, whose lifestyles still follow the beliefs and roots of their former cultures, although changes can be traced owing to the influence of environmental and technological development [34]. The Mekong River Basin is “the river of economy”, owing to its major role in the socio-economic development in Southeast Asia. It is an upstream resource of agricultural systems, energy production, food security, ecological system, and human well-being. Thus, attempts to enable the Mekong River Basin to acquire its roles in regional economic propulsion are being investigated at an international level. However, besides political issues and interrelations among the countries, the development process also involves the problem of understanding the ethnic groups in the area, which is of great importance [35]. Therefore, research studies for knowledge and understanding of ethnic groups are unavoidable.

Besides compiling the vocabularies related to the ethnic groups in the Mekong River Basin, the development of a thesaurus for the ethnic groups was carried out based on the concept of knowledge organization in the classification of the information on the ethnic groups, such that there is linked open access to all related contents, for example, dialects, languages, beliefs, attires, rituals, or social systems. Thus, the study did not only involve the organization of the names of the ethnic groups, but the thesaurus can also be used as a resource for studying the relationships of the content about the ethnic groups and as a tool to access and retrieve knowledge about the ethnic groups in the database and on the Internet. Furthermore, the results obtained can be used in semantic search and open access for an international standard of data exchange.

2. Research Objectives

This research was aimed at constructing a thesaurus of the ethnic groups in the Mekong River Basin, which compiles the controlled vocabularies both in Thai and English languages, with a digital platform for managing the thesaurus in terms of semantic search and linked open data.

3. Methodology

The Research and Development Method was applied for the research, consisting of four steps: (1) analysis, synthesis, and knowledge organization; (2) construction of the thesaurus; (3) development of a digital thesaurus platform for the ethnic groups in the Mekong River Basin; and (4) evaluation of the digital thesaurus platform for the ethnic groups in the Mekong River Basin.

1. Analysis, synthesis, and knowledge organization: These processes were performed by means of document analysis and knowledge organization as follows:
 - 1.1. Data resources for the analysis: The information related to the ethnic groups in the Mekong River Basin was compiled from various sources, namely: (1) the domestic and international databases of information resources, in which a lot of collections exist in the fields of humanities and socio-cultural aspects of the Mekong River Basin, thus the research emphasized the studies conducted in Thai and English; (2) the information corpus or international databases, in which cultural vocabularies have been collected: Yale University [<https://hraf.yale.edu/>] (accessed on: 14 July 2021), The Getty Research Institute [<https://www.getty.edu/research/tools/vocabularies/aat/>] (accessed on: 16 April 2021), and UNESCO [<http://vocabularies.unesco.org/browser/thesaurus/en/>] (accessed on: 2 February 2021); (3) the Internet information resources in the ethnic groups in the Mekong River Basin, including the database of the ethnic groups in Thailand, [<https://www.sac.or.th/databases/ethnic-groups/ethnicGroups/>] (accessed on: 11 March 2021); and (4) the classification of the ethnic groups in Thailand from the research work by Chaikhambung and Tuamsuk [22] as shown in Table 1.
 - 1.2. Compilation of data: The researcher stipulated the keywords for retrieval of information, which included; ethnic group, ethnicity, and the Mekong River Basin, and retrieved information from the different databases stated in 1.1. from the retrieval channel of each data source, which were the topics, keywords, subject headings, abstracts, or descriptions. The data was then downloaded and the documents were filed systematically on a cloud drive.
 - 1.3. Extraction and screening of data: The researcher extracted the keywords or vocabulary appearing in the collected data in the cloud drive by considering the vocabulary with specific meaning related to the ethnic groups. Next, the vocabulary was screened and selected by counting the frequency of the same word that appeared, removing repetitive words, synonyms, and ambiguous words, and obtained 4069 words related to the ethnic groups in the Mekong River Basin.
 - 1.4. Word classification: The researcher classified the vocabulary according to the fundamental criteria for categorization and justification for knowledge organization based on domain-specific criteria [36,37]; starting from high-frequency down to low-frequency words, placing words with the same meaning together, words with close meanings next to one another, separating words with different meanings, checking the correctness and avoiding ambiguity of meanings based on an online dictionary in English (WordWeb, <https://www.wordwebonline.com/>] (accessed on: 11 March 2021), and finally recording the word groups that had been arranged using the TemaTres 3.1 Program [38]. The outcome is the structure of vocabularies convenient for use and development of further thesauruses. The completed process provides arrangements of 12 vocabulary groups of the ethnic groups in the Mekong River Basin: language groups, social organization, costume, art works and entertainment, general name, demography and residential, history, customs and rituals, social dynamics, economic system, way of life, and religion and beliefs (Figure 1). In each group, there are subgroups of different levels on the

same topic, or close topics, as the example in the language groups shown in Figure 2.

Table 1. The information corpuses of international databases.

Domestic and International Databases	Information Corpus or International Databases	Internet Information Resources	Classification of the Ethnic Groups
Khon Kaen University Library	Yale University	Princess Maha Chakri Sirindhorn Anthropology Centre (Public Organisation)	Knowledge Classification on Ethnic Groups in Thailand
Chiang Mai University Library Mahasarakham University Library Mahidol University Library Naresuan University Library	The Getty Research Institute UNESCO Thesaurus		

Title: Ethnic Groups in Mekong River Basin (MRB)
 Author: Wirapong Chansanam
 Keywords: digital thesaurus; ethnic groups
 Scope: The purpose of this project is to collect and create the digital thesaurus
 URI: <https://www.thesaurus.asiana.net/vocab/>
 Powered by: TemaTres 3.1

Ethnic groups (กลุ่มชาติพันธุ์)

- . กลุ่มภาษา (Language groups)
- . องค์กรชุมชน (Social organization)
- . ธรรมเนียม (Costume)
- . งานศิลปะและการบันเทิง (Art works and entertainment)
- . ชื่อกลุ่มชาติพันธุ์ (general Name)
- . ประชากรศาสตร์และถิ่นที่อยู่อาศัย (Demography and residential)
- . ประวัติศาสตร์ความเป็นมา (History)
- . ประเพณีและพิธีกรรม (customs and rituals)
- . พลวัตทางสังคม (Social dynamics)
- . ระบบเศรษฐกิจ (Economic system)
- . วิถีชีวิต (Way of Life)
- . ศาสนาและความเชื่อ (Religious and beliefs).

Figure 1. Key groups of vocabularies on ethnic groups in the MRB.

Title: Ethnic Groups in Mekong River Basin (MRB)
 Author: Wirapong Chansanam
 Keywords: digital thesaurus; ethnic groups
 Scope: The purpose of this project is to collect and create the digital thesaurus
 URI: <https://www.thesaurus.asiana.net/vocab/>
 Powered by: [TemaTres 3.1](#)

Ethnic groups (กลุ่มชาติพันธุ์)

- . กลุ่มภาษา (Language groups)
 - . กลุ่มภาษา (Language family)
 - . ทอน (Tones)
 - . วัน (Day counting)
 - . ก (Ga)
 - . ข (Khat)
 - . ค (Khet)
 - . ด (Dap)
 - . ง (Mung)
 - . ต (Tat)
 - . ฮ (Huang)
 - . ฉ (Tao)
 - . ช (Piak)
 - . ช้าง (Khiam Mueang)
 - . ช้างใต้ (Southern Thai accent)
 - . ช้างเหนือ (Nakhon Si Thammarat accent)
 - . ชื่อกลุ่มภาษา (Language name family)
 - . Austroasiatic (ออสโตรเอเชียติก)
 - . กลุ่มภาษานิโคบาร์ (Nicobarese subgroup)
 - . กลุ่มภาษาเขมร-มอน (Mon-Khmer subgroup)
 - . กลุ่มภาษาเขมร (Khmuic subgroup)
 - . ภาษาเขมร (Katuic)
 - . ภาษาเขมร (Katang language)
 - . ภาษาเขมร (Katu language)
 - . ภาษาเขมร (Katuic language)
 - . ภาษาเขมร (Kui language)
 - . ภาษาเขมร (Suai language)
 - . ภาษาเขมร (Ta'o'ih language)
 - . ภาษาเขมร (Bru language)
 - . ภาษาเขมร (Phuang language)
 - . ภาษาเขมร (Arak language)
 - . ภาษาเขมร (Loe language)
 - . ภาษาเขมร (Khasian)
 - . ภาษาเขมร (Khmuic)

Figure 2. Example of groups of language groups of vocabularies in ethnic groups in the MRB.

2. Construction of the thesaurus: The approaches in thesaurus construction were investigated from many concepts [39–42] and the following steps were followed:

- 2.1. The classified vocabularies were checked for correctness, relationships among the words in the same group, repetition, and the standard use of Thai and English words according to the accepted references, i.e., the terminology given by the Royal Institute (<https://coined-word.orst.go.th/> (accessed on: 10 March 2021)), the Thesaurus of ERIC descriptors (<https://eric.ed.gov/> (accessed on: 16 April 2021)), the UNESCO thesaurus (<http://vocabularies.unesco.org/browser/thesaurus/en/> (accessed on: 2 February 2021)), and the Online Thai Subject Headings [43].
 - 2.2. The relationships of words in each group and between groups were prioritized according to the relationship structure of the thesaurus, which comprised broader term (BT), narrow term (NT), and related term (RT).
 - 2.3. All of the vocabularies were recorded according to the structure of the thesaurus stipulated in 2.2 using the TemaTres 3.1 Program (<https://www.vocabularyserver.com/> (accessed on: 2 February 2021)) [38].
 - 2.4. Cross referencing was done, i.e., USE and UF (use for), to link the words with the same meaning or the words that can be used interchangeably. Scope notes were next added to the words having broader term to make the thesaurus complete.
 - 2.5. The thesaurus word list was verified and evaluated by specialists including two information scientists who have expertise in knowledge organization and thesaurus construction and three academics in anthropology and sociology who have expertise in the ethnic groups of the Mekong River Basin. The snowball technique was used in selecting the experts, beginning from the first and the second information science experts, followed by the third, fourth, and fifth ethnic group experts. The vocabularies were adjusted following the experts' opinions and suggestions before arriving at the thesaurus structure for the ethnic groups in the Mekong River Basin.
3. Development of a digital thesaurus platform for the ethnic groups in the Mekong River Basin: The platform was developed as a system for digital vocabulary management that allows semantic search and open access, which are useful in information usage and broad information exchange related to ethnic groups at the international level. The development of the digital thesaurus platform was conducted following the constructed architecture of the platform for the thesaurus (Figure 3) using the Tematres 3.1 Program that worked on a cloud service host.

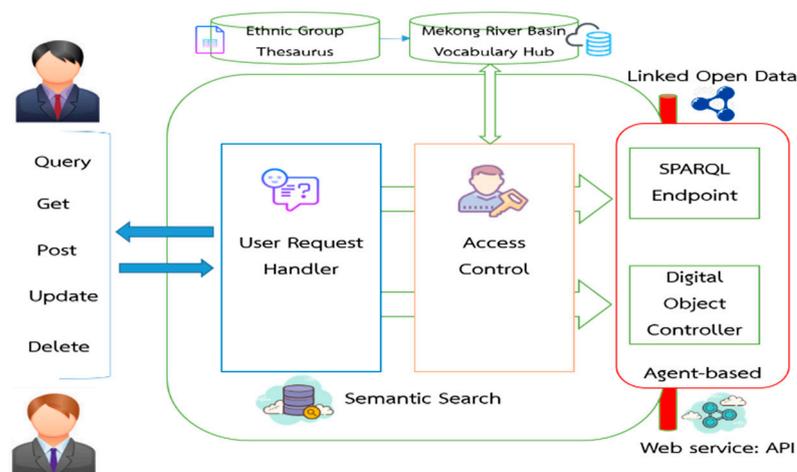


Figure 3. Architecture of the digital platform for the thesaurus of ethnic groups in the MRB.

4. Evaluation of the digital thesaurus platform for the ethnic groups in the Mekong River Basin: Evaluation of the digital thesaurus for the ethnic groups completed the four objectives of thesaurus development, namely; translation, consistency, indication

of relationship, and retrieval [44]. The evaluation was performed following the steps below:

- 4.1. Query selection of the term in order to show the system efficiency in terms of the stored corpuses was done by two experts in ethnic studies. In this research, 15 sets of vocabularies were sampled for retrieval from 160 corpuses as shown in Table 2 in order to find the “precision” and “recall” values in the next step.
- 4.2. Evaluation of retrieval is very important in measuring the efficiency and effectiveness of the system [45]. In this research, the system efficiency was measured by precision and recall [46], as shown in Table 3, and the F-measure was used to test the precision.

Table 2. Query selection and search results.

No of Query	Queries	Search Results	
		Relevant	Irrelevant
1	Phlong Karen, Phlong, Phlong, Su, Karen	30	130
2	Khun, Tai Khun, Tai Khoen	8	152
3	Kui, Kuoy	9	151
4	Khamu, Kammu, Ta Moi	4	156
5	Tai, Kon Tai, Tai Long, Tai Luang, Tai Yai, Tai Luang	18	142
6	Nyahkur, Nyah Kur, Lawa, Chao Bon	6	154
7	Phu Tai, Phu Tai	9	151
8	Phu Yoi, Yoi, Tai Yoi, Yoi	8	152
9	Meo, Hmong, Miao	8	152
10	Nyo, Yor, Yo, Yo	3	157
11	Mon, Raman, Khanon, Mon people	8	152
12	Lue, Tai Lue, Tai, Thai Lue	11	149
13	Lua, Lavua, Lavua, Lawa, Htin, Mal, Plai	7	153
14	Lao Song, Phu Lao, Tai Dam, Tai Song Dam, Thai Song Dam	15	145
15	Viet, Yuan, Kaew	10	150

Table 3. Experimentation result—percentages of precision and recall.

No of Query	Total Relevant in the collections	Total Retrieved	Relevant Retrieved	Precision (%)	Recall (%)
1	30	33	28	90.91	93.33
2	8	9	5	88.89	62.50
3	9	10	7	90.00	77.78
4	4	6	4	66.67	100.00
5	18	19	11	94.74	61.11
6	6	7	6	85.71	100.00
7	9	11	8	81.82	88.89
8	8	10	8	80.00	100.00
9	8	11	8	72.73	100.00
10	3	5	3	60.00	100.00
11	8	8	4	100.00	50.00
12	11	14	8	78.57	72.73
13	7	8	4	87.50	57.14
14	15	15	14	100.00	93.33
15	10	12	9	83.33	90.00
Average				84.06	83.12
Convert to integer				0.8406	0.8312

In Table 3, the means of the precision and recall values obtained in this study were 84.06% and 83.12%, respectively. When transformed to full integers, i.e., 0.8406 and 0.8312,

as compared to 1, the system and corpuses prove to have high effectiveness in retrieval and completeness [47].

F-measure was used to measure the system efficiency, according to the following formula:

$$F - measure = \frac{2PR}{(P + R)} \quad (1)$$

$$F = \frac{2(0.8406)(0.8312)}{0.8406 + 0.8312}$$

$$F = 0.8358$$

The *F-measure* value obtained from this study was $F = 0.8358$, which indicates that the system is efficient and the precision in retrieval is high. Thus, searching the information related to the ethnic groups in the Mekong River Basin by the developed thesaurus with the controlled vocabulary can be of use.

4. Results of Research

1. The developed thesaurus of the ethnic groups in the Mekong River Basin contains two languages (Thai and English), with 4273 principle words. These words have been categorized based on the depth of eight levels as follows: 4 words at Deep Level 1, 22 words at Deep Level 2, 137 words at Deep Level 3, 1486 words at Deep Level 4, 527 words at Deep Level 5, 181 words at Deep Level 6, 47 words at Deep Level 7, and 22 words at Deep Level 8. There are 2596 words with hierarchical relationships, or broader and narrower terms, whereas 6858 have associative relationships or related terms, as shown in Figure 4.

Title: Ethnic Groups in Mekong River Basin (MRB)	
Author: Nirapong Chansanam	
Keywords: digital thesaurus; ethnic groups	
Scope: The purpose of this project is to collect and create the digital thesaurus of ethnic groups	
URI: https://www.thesaurus.asiana.net/vocab/	
Powered by: TemaTres 3.1	
<hr/>	
Akha (อักษะ)	BT : ชื่อชนชาติ (Ethnonym)
Alak (อละก)	BT : ชื่อชนชาติ (Ethnonym)
Anu (อานุ)	BT : ชื่อชนชาติ (Ethnonym)
	RT : จีน (Chin)
Anun (อานุณ)	BT : ชื่อชนชาติ (Ethnonym)
	RT : จีน (Chin)
Arak (อระก)	BT : ชื่อชนชาติ (Ethnonym)
	RT : จีน (Chin)
Artisan fisheries (อาชีพประมงพื้นบ้าน)	BT : ชื่ออาชีพ (Occupation)
	RT : คนใต้ (Southern people)
Austroasiatic (ออสโตรเอเชียติก)	BT : ชื่อกลุ่มภาษา (Language name family)
	NT : ภาษากลุ่มภาษาอินโดนีเซีย (Nicobarese subgroup)
	NT : ภาษากลุ่มภาษาอินโดนีเซีย-เขมร (Mon-Khmer subgroup)
	NT : ภาษากลุ่มภาษาอินโดนีเซีย (Munda subgroup)
	NT : ภาษากลุ่มภาษาอินโดนีเซีย (Monic subgroup)
	RT : ภาษากลุ่มออสโตรเอเชียติก (Austroasiatic languages)
B'li (เบลี)	BT : ชื่อกลุ่มภาษา (Language name family)
B'ri (เบรี)	BT : ชื่อชนชาติ (Ethnonym)
Bamar (บามาร์)	BT : ชื่อชนชาติ (Ethnonym)

Figure 4. Example of the thesaurus of ethnic groups in the MRB.

2. The digital thesaurus of the ethnic groups in the Mekong River Basin is a digital platform that manages the controlled vocabulary related to the ethnic groups in the Mekong River Basin. It contains both Thai and English vocabularies, and when searched, it will display the result of the word with its broader term, narrower term, related term, cross reference, and scope note (for a certain word). The digital thesaurus enables semantic search via an application based on WWW technology at <https://www.thesaurus.asiana.net/vocab/> accessed on 14 July 2021 (Figures 5 and 6), open data via SPARQL endpoint at <https://www.thesaurus.asiana.net/vocab/sparql.php> accessed on 14 July 2021 (Figure 7), and a web service by means of an application programming interface (API) at <https://www.thesaurus.asiana.net/vocab/services.php> accessed on 14 July 2021 (Figure 8).

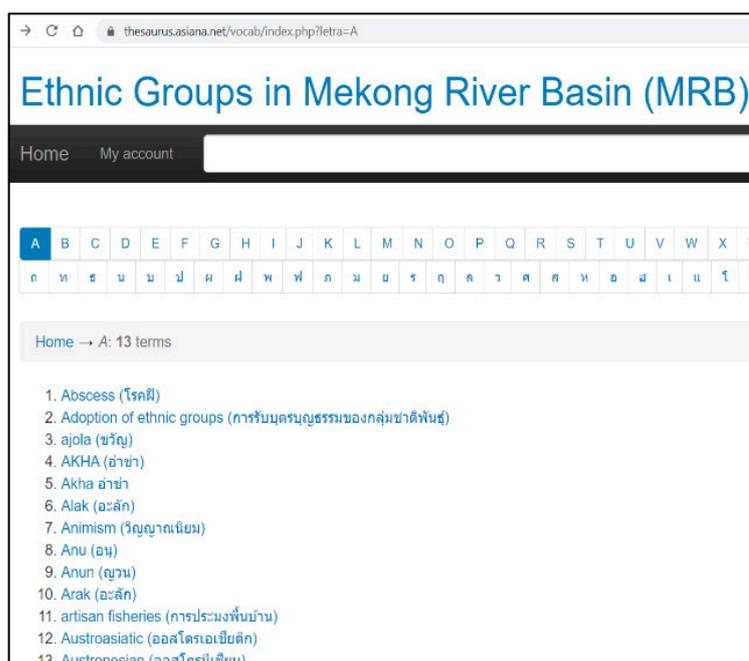


Figure 5. Web access of the digital thesaurus of ethnic groups in the MRM, showing a list of terms begin with A.

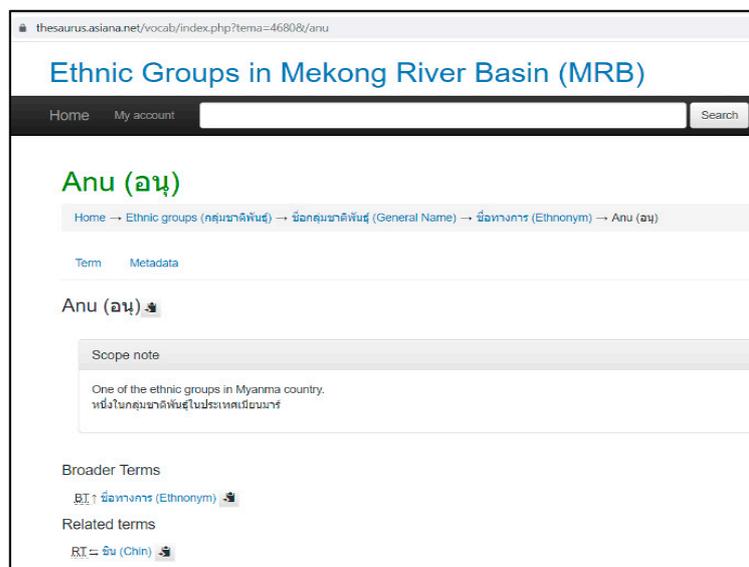


Figure 6. Search results for the term “Anu” with scope note, BT, and RT.



Figure 7. Screen shot of the SPARQL endpoint of the digital thesaurus of ethnic groups in the MRB.

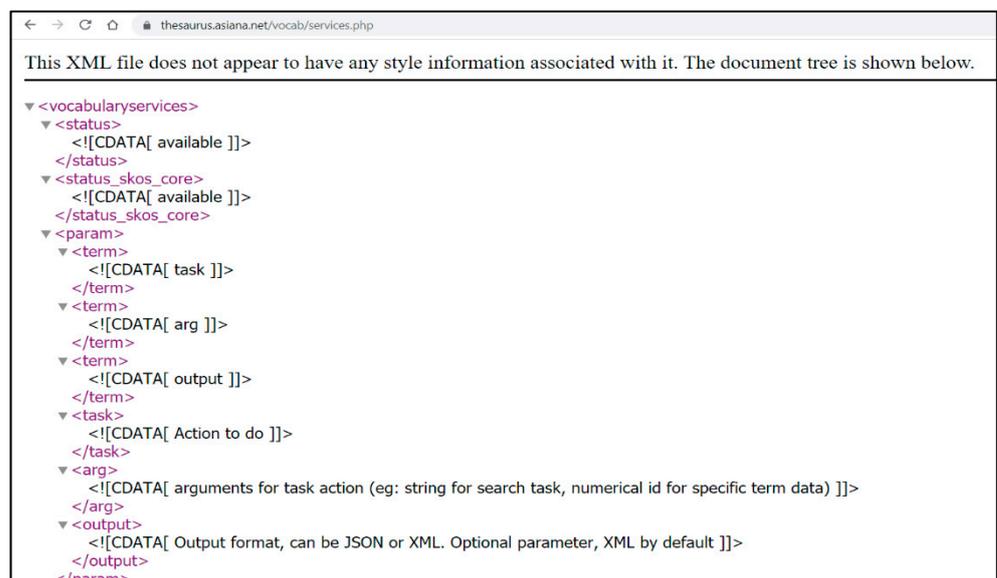


Figure 8. Screen shot of the API web service for the digital thesaurus of ethnic groups in the MRB.

5. Discussion

This research is the continuation of former work [22–24], which organized knowledge and constructed the channel for exchanging information on the ethnic groups in Thailand in the form of ontology taxonomy and linked open data. This research differs from the previous work as demonstrated in Table 4, especially in the scope of knowledge that has been expanded to cover the ethnic groups in the Mekong River Basin and the thesaurus construction with broad terms, narrow terms, related terms, and vocabulary relationships as well as cross references. This structure helps point out the connections of information and knowledge of ethnic groups in different dimensions, i.e., languages, social structure, marriage, art and entertainment, demography, history, tradition, beliefs, lifestyles, socio-economic movements, etc. In addition to the relationships of the content, other relationships can be seen in the Mekong Basin’s ethnic groups that are found to be from the same root. The thesaurus in this research, moreover, can manage and provide digital platform access on the Internet, with functions that serve semantic search and linked open data. Therefore, interested academics and researchers can use the thesaurus as a tool to link with the existing databases of resources already existing in various aspects related to the ethnic groups, for instance, folk tales, folk music, folklore, beliefs, etc. based on the international standard.

Table 4. Comparison of the research works on ethnic group knowledge organization and management.

Research Works	Scope	Resources	Study Approach
Chaikhambung & Tuamsuk (2017a, 2017b)	Ethnic groups in Thailand	Reference resources, books, universities, collections	Content analysis, classification, ontology development.
Chansanam et al. (2020)	Ethnic groups in Thailand	Database of the Princess Maha Chakri Sirindhorn Anthropology Centre Chaikhambung & Tuamsuk (2017a, 2017b); Chansanam et al. (2020); and Databases of research resources in universities' libraries.	LOD
This research	Ethnic groups in the MRB		KO, Digital thesaurus, LOD, Web service

The digital thesaurus of ethnic groups in the MRB was developed in a two-language version: English and Thai. As the thesaurus was aimed at offering an access tool to the source of knowledge of the ethnic groups in the Mekong River Basin, standard controlled vocabulary and open access were essential. When compared to the existing and internationally known thesaurus, the UNESCO Thesaurus [<http://vocabularies.unesco.org/browser/thesaurus/en/> accessed on 14 July 2021] developed in 1977, which offers a list of controlled vocabulary that is useful for content analysis and documentary search in culture, natural science, humanities, social science, communication, and information, and is at present continuously improved, it is found that the UNESCO Thesaurus is outstanding in presenting vocabularies in many languages including English, Arabic, French, Russian, and Spanish and is thus called a multilingual thesaurus. This is probably the limitation of our research that can be expanded in the next step because Tematres does not support multilingual thesaurus development. However, when considering the vocabularies related to the ethnic groups, the vocabularies found are of the ethnic groups in America and Africa. Few words are offered for Asia, with inadequate details such that it cannot be specifically used as a management tool for access to the ethnic groups. As demonstrated in Figure 9, there are 12 items related to ethnic groups with narrower concepts. If “Asian” is selected, only one ethnic group, i.e., Indians, is shown. Other capacities are provided such as linked data by searching a screenshot through a web browser, linked open data, or web services (API). It can be concluded that the outcome of this research is a tool to access standard and international thesauruses other than The UNESCO Thesaurus. The thesaurus from this research has the specificity of the ethnic groups in the Mekong River Basin that enables knowledge management and digital information resources in humanities with high coverage and completeness.

This development of a digital thesaurus platform is the start towards more expansion of contents in humanities and cultural heritage and other issues such as development of a thesaurus of digital humanities, information analysis, storing and semantic search, and development of information resources for open access to other information sources. One important issue is to develop other digital platforms on online social networks by presenting vocabularies. Besides group discussions to find the conclusion in considering vocabularies, opportunities should be provided for academics or interested individuals to take part and experiment on real usage by construction or set contents on online platforms through the presentation of research results that link with the internet networks in a wide circle. With the aim that users can have access to all research works at any place and time. Interested individuals can also take part in the consideration of words and provide their opinions, to enhance the value of the research works.

The figure consists of two screenshots of the UNESCO Thesaurus interface. The top screenshot shows the 'Ethnic groups' entry. The preferred term is 'Ethnic groups'. Under 'NARROWER CONCEPTS', 'Asians' is circled in red. The bottom screenshot shows the 'Asians' entry. The preferred term is 'Asians'. Under 'NARROWER CONCEPTS', 'Indians' is circled in red. Both screenshots show a hierarchical tree on the left and various relationship categories on the right.

Figure 9. Screen shot of The UNESCO Thesaurus from <http://vocabularies.unesco.org/browser/thesaurus/en/index/E> accessed on 14 July 2021.

Author Contributions: Conceptualization, W.C., K.T.; methodology, W.C.; software, W.C.; validation, W.C., K.T., K.K., M.B.; formal analysis, W.C.; resources, W.C.; data curation, W.C.; writing—original draft preparation, W.C., K.T.; writing—review and editing, K.T.; visualization, W.C.; supervision, M.B.; project administration, K.T., K.K.; funding acquisition, K.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research is supported by the Office of National Higher Education Science Research and Innovation Policy Council, Thailand.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Barth, F. Introduction: Ethnic groups and boundaries. In *The Social Organization of Culture Difference*; George Allen & Unwin: London, UK, 1969; pp. 9–38.
2. Hale, H.E. Explaining ethnicity. *Comp. Polit. Stud.* **2004**, *37*, 458–485. [[CrossRef](#)]
3. Seol, B.S. A critical review of approaches to ethnicity. *Int. Area Rev.* **2008**, *11*, 333–364. [[CrossRef](#)]
4. Ethnic Group. *Encyclopedia Britannica*. 2017. Available online: <https://www.britannica.com/topic/ethnic-group> (accessed on 6 June 2021).
5. Bitzer, J.; Gören, E. Measuring Capital Services by Energy Use: An Empirical Comparative Study, Oldenburg Discussion Papers in Economics, No. V-351-13, University of Oldenburg, Department of Economics, Oldenburg. 2013. Available online: <https://www.econstor.eu/bitstream/10419/105039/1/V-351-13.pdf> (accessed on 11 March 2021).

6. Okpanocha, O.S.; Nwankwo, I.U. Ethnicity, ethnic identity and the crisis of national development in Nigeria. *Int. J. Health Soc. Inq.* **2019**, *5*, 61–81.
7. Freeberg, A.L.; Stein, C.H. Felt obligation towards parents in Mexican-American and Anglo-American young adults. *J. Soc. Pers. Relatsh.* **1996**, *13*, 457–471. [[CrossRef](#)]
8. Rhee, E.; Uleman, J.S.; Lee, H.K. Variations in collectivism and individualism by ingroup and culture: Confirmatory factor analyses. *J. Personal. Soc. Psychol.* **1996**, *71*, 1037–1054. [[CrossRef](#)]
9. Gaines, S.O., Jr.; Marelich, W.D.; Bledsoe, K.L.; Steers, W.N.; Henderson, M.C.; Granrose, C.S.; Barajas, L.; Hicks, D.; Lyde, M.; Takahashi, Y.; et al. Links between race/ethnicity and cultural values as mediated by racial/ethnic identity and moderated by gender. *J. Personal. Soc. Psychol.* **1997**, *72*, 1460–1476. [[CrossRef](#)]
10. Ting-Toomey, S.; Yee-Jung, K.K.; Shapiro, R.B.; Garcia, W.; Wright, T.J.; Oetzel, J.G. Ethnic/cultural identity salience and conflict styles in four US ethnic groups. *Int. J. Intercult. Relat.* **2000**, *24*, 47–81. [[CrossRef](#)]
11. Coon, H.M.; Kimmelmeier, M. Cultural orientations in the United States: (re)examining differences among ethnic groups. *J. Cross Cult. Psychol.* **2001**, *32*, 348–364. [[CrossRef](#)]
12. Hamer, K.; McFarland, S.; Czarnańska, B.; Golińska, A.; Cadena, L.M.; Łuzniak-Piecha, M.; Jułkowski, T. What is an “ethnic group” in ordinary people’s eyes? different ways of understanding it among American, British, Mexican, and Polish respondents. *Cross Cult. Res.* **2020**, *54*, 28–72. [[CrossRef](#)]
13. Randi, H. Archaeological classification and ethnic groups: A case study from Sudanese Nubia. *Nor. Archaeol. Rev.* **1997**, *10*, 1–17.
14. Pablo, M.; Alex, S.; Paul, L. Uncertainty in the analysis of ethnicity classifications: Issues of extent and aggregation of ethnic groups. *J. Ethn. Migr. Stud.* **2009**, *35*, 1437–1460.
15. Gilbert, P.A.; Khokhar, S. Changing dietary habits of ethnic groups in Europe and implications for health. *Nutr. Rev.* **2008**, *66*, 203–215. [[PubMed](#)]
16. Platt, L.; Warwick, R. At Greater Risk: Why COVID-19 Is Disproportionately Impacting Britain’s Ethnic Minorities. 2020. Available online: http://eprints.lse.ac.uk/104918/1/politicsandpolicy_covid19_ethnic_minorities.pdf (accessed on 2 June 2021).
17. Poulsen, M.F.; Johnston, R.J.; Forrest, J. Is Sydney a divided city ethnically? *Aust. Geogr. Stud.* **2004**, *42*, 356–377. [[CrossRef](#)]
18. Aud, S. *Status and Trends in the Education of Racial and Ethnic Groups*; National Center for Education Statistics, Institute of Education Sciences: Washington, DC, USA, 2010.
19. Huang, T.; Shu, Y.; Cai, Y.D. Genetic differences among ethnic groups. *BMC Genom.* **2015**, *16*, 1093. [[CrossRef](#)] [[PubMed](#)]
20. Ratanakul, S.; Premsrirat, S.; Dawwatanahong, L.; Wannadee, W. *Research Report on Comprehensive Knowledge of the Ethnic Minorities in Thailand*; Institute of Language and Cultural Research for Rural Development, Mahidol University: Bangkok, Thailand, 2000.
21. LeBar, F.M.; Hickey, G.C.; Musgrave, J.K. *Ethnic Groups of Mainland Southeast. Asia*; Human Relations Area Files Press: New Haven, CT, USA, 1964.
22. Chaikhambung, J.; Tuamsuk, K. Development of semantic ontology of the knowledge on ethnic groups in Thailand. *TLA Res. J.* **2017**, *10*, 1–15.
23. Chaikhambung, J.; Tuamsuk, K. Knowledge classification on ethnic groups in Thailand. *Cat. Classif. Q.* **2017**, *55*, 89–104.
24. Chansanam, W.; Tuamsuk, K.; Chaikhambung, J. Linked open data framework for ethnic groups in Thailand learning. *Int. J. Emerg. Technol. Learn.* **2020**, *15*, 140–156. [[CrossRef](#)]
25. Pholsena, V. Nation/representation: Ethnic classification and mapping nationhood in contemporary Laos. *Asian Ethn.* **2002**, *3*, 175–197. [[CrossRef](#)]
26. Mackerras, C. What is China? Who is Chinese? Han-minority relations, legitimacy, and the state. In *State and Society in 21st-Century China: Crisis, Contention, and Legitimation*; Gries, P.H., Rosen, S., Eds.; Routledge Curzon: New York, NY, USA, 2004; pp. 216–234.
27. Schütze, H.; Pedersen, J.O. A cooccurrence-based thesaurus and two applications to information retrieval. *Inf. Process. Manag.* **1997**, *33*, 307–318. [[CrossRef](#)]
28. Sokal, R.R.; Sneath, P.H.A. *Principles of Numerical Taxonomy*; W.H. Freeman & Co.: New York, NY, USA, 1963.
29. Sowa, J.F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*; Brooks/Cole Publishing Co.: Pacific Grove, CA, USA, 2000.
30. Bates, M.J. After the Dot-Bomb: Getting Web Information Retrieval Right this Time. *First Monday.* 2002. Available online: <https://firstmonday.org/ojs/index.php/fm/article/view/971/892> (accessed on 10 February 2021).
31. Redmond-Neal, A.; Hlava, M.M.K. (Eds.) *ASIST Thesaurus of Information Science, Technology, and Librarianship*, 3rd ed.; Information Today: New Jersey, NJ, USA, 2005.
32. Prajayayothin, N. *Thesaurus in Information Storage and Retrieval Context*; Apichart Printing: Maha Sarakham, Thailand, 2013.
33. Nakayama, K.; Hara, T.; Nishio, S. A thesaurus construction method from large scale web dictionaries. In Proceedings of the 21st IEEE International Conference on Advanced Information Networking and Applications, Niagara Falls, ON, Canada, 21–23 May 2007; pp. 932–939.
34. Greater Mekong Subregion Environment Operations Center. People and Cultures. 2012. Available online: <http://www.gms-eoc.org/uploads/resources/149/attachment/3.Peoples-of-the-Greater-Mekong-Subregion.pdf> (accessed on 16 April 2021).
35. Pegasys Consulting. *Mekong River in the Economy*; WWF Greater Mekong Programme: Hochiminh City, Vietnam, 2016.

36. Moine, M.P.; Valcke, S.; Lawrence, B.N.; Pascoe, C.; Ford, R.W.; Alias, A.; Balaji, V.; Bentley, P.; Devine, G.; Callaghan, S.A.; et al. Development and exploitation of a controlled vocabulary in support of climate modelling. *Geosci. Model. Dev.* **2014**, *7*, 479–493. [[CrossRef](#)]
37. Tuamsuk, K.; Chansanam, W.; Chaikhambung, J.; Kaewboonma, N. *Digital Humanities Research*; Klangnanawittaya Printing: Khon Kaen, Thailand, 2018.
38. Gonzales-Aguilar, A.; Ramírez-Posada, M.; Ferreyra, D. TemaTres: Software para gestionar tesauros. *Prof. De La Inf.* **2012**, *21*, 319–325. [[CrossRef](#)]
39. American National Standard Institute. *Guidelines for Thesaurus Structure, Construction and Use*; ANSI: New York, NY, USA, 1974.
40. Aitchison, J.; Gilchrist, A.; Bawden, D. *Thesaurus Construction and Use: A Practical Manual*, 4th ed.; Fitzroy Dearborn Publishers: Chicago, IL, USA, 2000.
41. Broughton, V. *Essential Thesaurus Construction*; Facet Publishing: London, UK, 2006.
42. Prajayayothin, N. Vocabulary control. In *Information Organization and Retrieval*; Sukhothai Thammathirat Open University Press: Nonthaburi, Thailand, 2017; pp. 41–57.
43. *Online Thai Subject Heading*; Task force for Information Resources Organization, Thai Academic Libraries Consortium, 2020. Available online: <https://webhost2.car.chula.ac.th/thaicweb/main.php> (accessed on 2 June 2021).
44. Fayen, E.G. Guidelines for the construction, format, and management of monolingual controlled vocabularies: A revision of ANSI/NISO Z39.19 for the 21st century. *Inf. Wiss. Und Prax.* **2007**, *58*, 445.
45. Singhal, A. Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.* **2001**, *24*, 35–43.
46. Saini, B.; Singh, V.; Kumar, S. Information retrieval models and searching methodologies: Survey. *Inf. Retr.* **2014**, *1*, 20.
47. Kelbessa, I.W. The effects of having lists of synonyms on the performance of Afaan Oromo text retrieval system. *arXiv* **2021**, arXiv:2103.02900.