The 2nd China-ASEAN International Conference 2020
The 2nd International Conference on Tourism, Business, & Social Sciences 2020
Insight to China and ASEAN's Wellness, Tourism, & Innovation
https://www.dpu.ac.th/caiconf2020/
DOI: 10.6947/caicictbs.202004.0059

Development of Semantic Search for Local Tourism Website in Case of Southern Isan of Thailand and Northern of Cambodia

Sarim Touch[1] Jongkol Janruang[2*]
[1]Master Student [2*]Assistant Professor,
Department of Information and Communication
Faculty of Sciences and Liberal Arts, Rajamangala University of Technology Isan
[1]sarim.touch.rupp@gmail.com [2*]jj@sci.rmuti.ac.th

Abstract

Today, tourism web data are growing every day and storing in relational database, Thai and Cambodian tourists use internet to find tourism information every day. Absolutely, the problem of information that not preferable what they are looking for with mess of data. Therefore, ontology plays an important role in semantic web search to solve the problem of tourist finding information. In this research, the main point is to adopt semantic web to the local tourism website allowing users to search information from web search engine to get relevant information. The method of development has the 6 modules as follow: (1) removing stop word from user query, (2) construction tourism ontology OWL model, (3) SPARQL query semantic related word with ontology, (4) converting relational database into RDF dataset, and (5) SPARQL query construction to query data from RDF dataset and (6) display result. Additionally, the propose approach supported the tourism information system which can search all necessary information from tourism domain such as locations, services, attractive places, activities…etc. Metric-based approach is used to evaluation ontology schema and knowledgebase. The result shown schema and knowledgebase design are preferable quality. The two systems searching show that system use only keyword to search information get a few relevant data with 12% and system use ontology keyword-based get more relevant information with 88% . It indicate that the system searching OWL model can help tourist to find information which they are looking for without waste of time and system better handling of data.

Keywords: Semantic Web, Ontology, Evaluation, Tourism.

1. Introduction

Thailand and Cambodia are neighbored country, have a long history together, culture relations and languages. Southern Isan of Thailand and Northern of Cambodia has a lot of attractive tourist places such as ancient temples, pagodas, mountains, beautiful resorts, hotels, culture...etc. Today, Thai and Cambodian people are crossing the border for doing business, learning and tours. Development of Information technology and communication has transformed landscape of society and tourism industry of both countries. For encouraging both people crossing the border for tours the web is becoming the most advantageous information for both Thai and Cambodian people looking for everyday where they want to find, visit and service needing. The information of web data stores in relational database and data is growing every day and searching information from web data face the problems of the user queries. Information present in the old web data are semi-structure, unstructured, hard to integrate and people not readable. User try to find information from the internet they face the problems are, first issue with keyword search in general, it's more of information which can impact the effectiveness of the search in terms of the quality of the search results and second is SQL provide keyword-based search but if words mismatch, the user cannot get relevant information from web data.

Semantic search tries to improve searching accuracy by understanding the context of terms and

user's intent. Semantic web is the web with meaning and computer can be readable and understanding. The W3C, Web Ontology Language (OWL) is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things. The Semantic Web introduces the Resource Description Framework (RDF) is a standard model for data representation and interchange on the Web, and uses SPARQL as the query language for RDF data. Ontology is a formal explicit of a shared conceptualization of a domain interest. Concepts and relationships are basic of ontology component. Ontology has a capacity to solve many problems in tourism [5]. Implementation of semantic web search engine requires understanding of technologies such as OWL, RDF, SPARQL, JENA API and JAVA web technologies.

Yunzhi [13] and Ayepeh & Janruang [1] , semantic web is a semantic and keyword web based technique of information retrieval and relevant keyword matching. The development tourism ontology is based on tourism information of Lambok Island in Indonesia. The top-down approach is used to design ontology class hierarchy which related to often question of tourist such as kind of tourism, facilities and locations [12]. The ontology construction tourism in Africa, to the demand with 7 steps for ontology domain analysis, conceptual classification and ontology construction. The Classes design with the most very important information includes attractions, countries, services, cities and other concepts. Ontology construction contains 34 classes, 40 kinds of object properties, 22 data properties and a number of individual with large content of ontology [14]. The ontology of information for tourism development planning in Thailand comprised of 15 main classes such as natural resource, cultural attraction, road situations, local product, histories, legends, etc., [2].

In this research, semantic web search for local website is created and adopted ontology as keyword-based to establish tourism ontology that integrate with RDF dataset and SPARQL query language. System enables user to request information with significant and give benefit to user with joyful.

2. Research Objective

2.1. Develop tourism ontology as keyword-based for local tourism website
2.2. Develop semantic web search engine for support local tourism website
2.3. To evaluate the efficiency and effective of the system and ontology quality

3. Scope of Research

This research was conducted to develop tourism ontology domain as keyword-based for semantic web search engine to search information from local website (www.phimai-angkorwat.com) and this website cover tourism information of southern Isan of Thailand which includes provinces are Nakhonratchasima, Buriram, Surin, Sisaket, Ubonratchathani and Northern Cambodia which include provinces are Siemreap and Odomeachey. This region has a lot of tourism destinations, which will interest tourist from both countries and has many beautiful places for tourist.

4. System framework of the research project

This research aims to propose ontology-based, keyword-based information searching for local tourism website. The system have six modules are; (1) removing stop words, (2) developing tourism ontology, (3) calculating semantic similarity with ontology (SPARQL query related words), (4) converting RDB to RDF dataset , (5) writing SPARQL to query data (SPARQL query construction) with Jena API and (6) sending result to display. These modules shown in

figure 1.

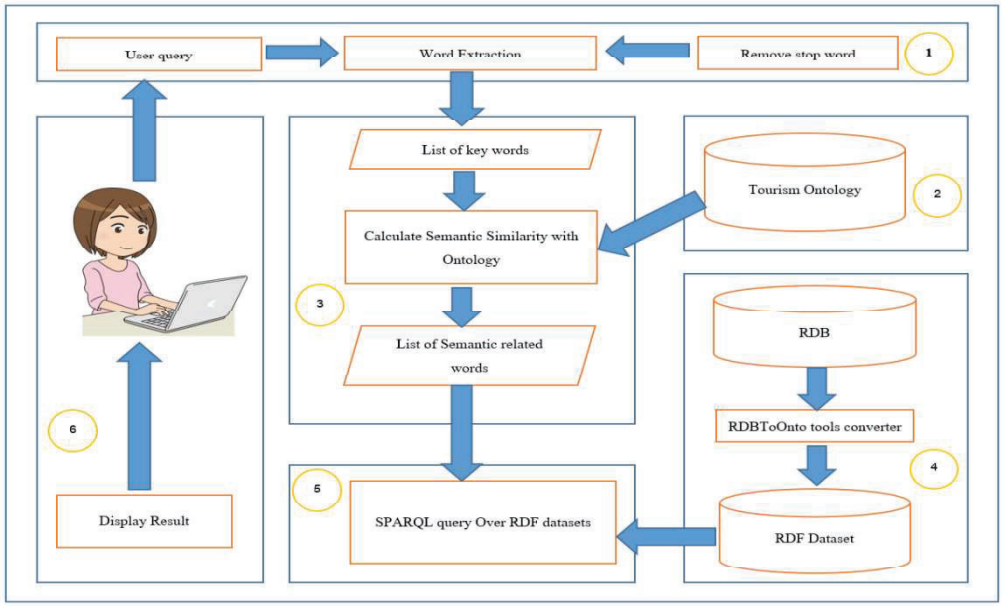

Figure 1 Proposed System Framework

5. Research Methodology

According to figure 1, proposed system implementation and evaluation, ontology construction and evaluation approach, tool, statistic report of ontology and system searching result, describe below are :

5.1 Removing stop word processing,

It is to remove the word that has no meaning in the sentence from user query. Stop word such as, a, and, so, what, how etc.

5.2 Tourism ontology development

According to Lee *et al*. [7] and Staab & Studer [10], the difficulty of construction ontology domain is related to the complexity of the problem and the knowledge of the respective domain such as biology, geography, tourism..etc. For constuction of tourism ontology domain, local tourism web site is used as the resoureces, the main classes include acommodation,shopping, event, attraction, transportation .. etc. Property define is isLocatedIn property, the property description is shown in Table 1. Each class comprised of sub-classes, their relationships were direct properties (Is-a) of the class. For example class of the "Attraction" class include: shopping, entertainment, market, world heritage, food etc.

Figure 2 Ontology main class

Table 2 Property description

No	Domain	Range	Description
isLocated In	Attraction, Shopping, Activity, Accomodation, and Event.	Destination	specify that Destination is belongin to Domain

Table 3 Total superclasses, subclasses and instances

Total Superclasses	Total Subclasses	Total Instances
8	45	60

5.3 5.3 Calculating Semantic Similarity (SPARQL query semantic related word)

Calculating semantic similarity is processed to finding related word in ontology. Example the word "Phimai" has related to word "Temple" and has related to word "World Heritage", and finding synonym of words.

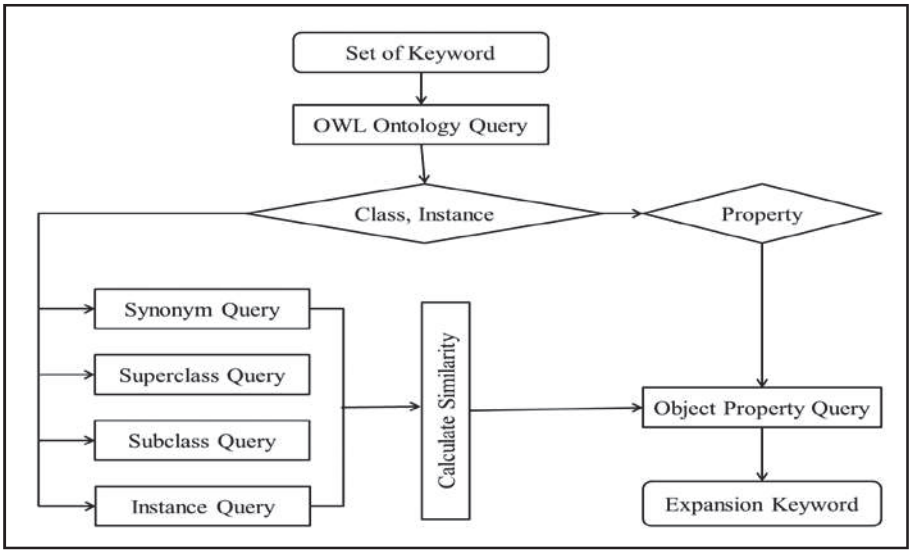

Figure 3 Keyword Expansion

According to Figure 3, Set of Keywords is a group of query words that are removed stop words. Synonym Query is process that program finds the words with the same meaning and connect with WordNet database dictionary. Superclass, Subclass and Instance query are process that program uses SPARQL to queries superclass, subclass and instance with keywords then return related words. Calculate Similarity is process that use SPARQL query language to find relationship between instance to subclass and superclass shown in figure 4. Object property query is used specific object property to query instances from ontology with specific two instances shown in figure 5. Expansion Keywords is process to compute or arrange group of words together. The results of SPARQL query semantic related word shown in table 3.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT? Class ?sub   ?instance
WHERE {  ?Class rdfs:subClassOf  ?Object. ?sub rdfs:subClassOf  ?Class.
          ?instance rdf:type owl:NamedIndividual. ?instance rdf:type ?sub.
(3)  FILTER (regex (str(?Class), "thailand", "i") || regex (str(?Class), "phimai", "i") )
(4)  FILTER (regex (str(?sub), "thailand",  "i") || regex (str(?sub), "phimai", "i") )
(5)  FILTER (regex (str(?instance), "thailand",  "i") || regex (str(?instance), "phimai", "i") )  }
```

Figure 4 SARQL Query semantic related word

```
PREFIX tour: <http://sarim.com/newisancambodia.owl#>
SELECT DISTINCT ?x
WHERE { (6) ?x tour:isLocatedIn tour:southern_isan (7) ?x tour:isLocatedIn tour:northern_cambodia }
```

Figure 5 Object property query with specific instance name

User enter text into searching box "I go to Phimai in Thailand". The process is shown below:

Table 4 Table sparql content query and WordNet

①	.Removing stop word	
1	Phimai, Thailand	
②	.Synonym words by WordNet	
2	Phimai	not found in WordNet database
3	Thailand	Thailand, Kingdom of Thailand, Siam
③	.SPARQL Superclass query	
4	Phimai	Not found because Phimai is an instance.
5	Thailand	Not found because Thailand is an instance.
④	.SPARQL Subclass query	
6	Phimai	Not found because Phimai is an instance.
7	Thailand	Not found because Thailand is an instance.
⑤	.SPARQL Instance query	
8	Phimai	world heritage, temple, Phimai
9	Thailand	destination, country, Thailand
⑥	.SPARQL southern_isan instance query	
10	Buriram, Nakhonratchasima, Sisaket, Surin, Muang Tum, Phanom Rung, Phanom Wan..etc.	
⑦	.SPARQL northern_cambodia instance query	
11	angkowat, bayon,odomeachey, Cambodia noodle, palm sugar…etc.	
⑧	.Expansion keywords	
12	Thailand, Kingdom of Thailand, Siam, world heritage, temple, Phimai, angkorwat, Buriram, Nakhonratchasima, Sisaket, Surin, Muang Tum, Phanom Rung ..etc.	

5.4 Converting Relational Database to RDF database

According to Ramathilagam & Valarmathi [9] identified a few mapping rules to conversion database schema to ontology schema by converting table to ontology class, column to data type property, primary key to functional data type property, and foreign key to object type property. Depending on mapping rules, a new table is designed by selecting, grouping and querying the only very important information into new tables for conversion to RDF dataset and follow the mapping rule. RDBToOnto tool allow generating ontology from relational database or transferring data from table format to RDF dataset format. In this project our database is MySQL Server. For converting MySQL data, the RDBToOnto tool require to input some information are Input Database, User, Password, Ontology File Name and Namespace.

5.5 SPARQL Query Construction

There are 4 processes to construct SPARQL query, (1) Set of keywords and set of RDF graphs (RDF is a collection of triples, each consisting of a subject, a predicate and an object, predicates is a property that connect subject with object), (2) Mapping keywords with RDF predicates, this process uses keyword to compare with value of object, (3) Convert Mapping to SPARQL query syntax is process that writing SPARQL query language to query data from RDF database and (4) SPARQL engine with JENA API will send result back to the user. User entered text into searching box "I go to visit Phimai in Thailand". The searching system will process as table 3.

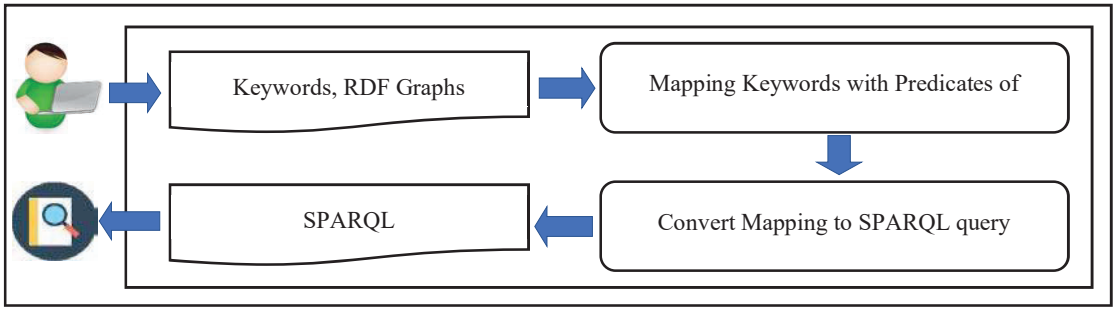

Figure 6 SPARQL query construction

5.6 Display result to the user by using SPARQL Query

The semantic related words (expansion keywords) are being sent to SPARQL query engine and JENA API to retrieve relevant information from RDF dataset. Java web application (java servlet, JSPs, and bootstrap) is developed to display result to the user. The search result shown in figure 8.

```
Algorithm: Keyword Search over RDF database
   Input keywords Q= {k1, k2,…, Kl) where l is number of keywords.
   Define set of nodes N = {n1,n2,…,ni) where i is number of node in RDF graph.
   Each node of RDF have predicate to access or get data from node.
   SET Result=NULL
     Foreach  Ni      ∈      N
              Foreach          Kl     ∈      K
                       IF value Ni isContainKeyword Kl Then  Result.add (Value of Ni)
                       End IF
              END
       END
     Return Result.
```

Figure 7 Algorithm Keyword Search over RDF database

6. Evaluation Methods, Tools and Result

6.1 Ontology evaluation approach and tool

According to Fahad et al., [3], ontology evaluation is the most potential part for developing new ontology for semantic web search engine. According to Gomez-Perez (2001), the goal of the evaluation process is to determine what the ontology defines correctly, does not define, or even defines incorrectly. There is an approach which used as the tools to evaluate ontology quality as bellow:

6.1.1 Metric-based approach

According to Tartir et al. [11] and Jain & Valerie [6], metric-based (feature-based) techniques to evaluate ontologies present a quantitative aspect of ontology quality. These techniques show different types of statistics about the knowledge presented in the ontology. These techniques consider the ontology schema graph and domain knowledge of ontology. The distribution of instances on the classes of the schema might also give specification on the quality of ontology.

6.1.2 Schema Metrics

It covers the ontology design. We cannot certainly know correctly the ontology domain knowledge design. The metrics specify the richness and inheritance of an ontology schema design. (1). Inheritance Richness (IR) is the average number of subclasses per class. It describes the classes place into the different levels of the ontology inheritance tree classes and differentiate of knowledges grouped into the different classes in ontology. The lowing value of IR present the ontology domain with detailed. A high value of IR specify the ontology represents a huge of general knowledge with a low level of detail. (2). Attribute Richness (AR) defined as the average number of attributes per class. It evaluates quality of ontology design and the number of information instances. More attributes present more knowledge in ontology. A high value of AR is more desirable.

6.1.3 Knowledgebase Metrics

The way data is placed within an ontology is also the most potential measure of ontology quality because it can specify the effectiveness of the ontology design and the amount of real-world knowledge represented by the ontology. (1) Average Population (AP) measure the total of instances compared to the total of classes. The average result present how data extraction process that performed to populate the knowledgebase. A lowing value specify, the instances extracted into the knowledgebase is insufficient to represent of knowledge in schema. Noting that some of the schema classes might have a little number or a very high number by the nature of what it is representing. (2) Class Richness (CR) compute the average of the total of classes contains instances with total classes in ontology classes. A lowing value of CR represent the ontology not have enough knowledgebase concept in the schema. A high value of CR represent the ontology have enough knowledgebase concept in the schema.

6.1.4 OntoMetrics ontology evaluation tool

OntoMetrics is online web tool use to validate ontology by displays with statistics report (https://ontometrics.informatik.uni-rostock.de/ontologymetrics).

6.2 System Evaluation

The system evaluation, focused on the result that system be able to find more result from RDF dataset. For evaluation system searching performance with relevant information from system. First system use only keywords without ontology to searching information, and second system use keyword combine ontology keyword-based to searching information.

6.3 Evaluation Results

In this section describe the result of the research with two parts, first ontology evaluation with statistic report and second system searching performance that show the different of searching by using keyword and ontology keyword-based search.

6.3.1 OntoMetric Ontology evaluation result

Table 5 OntoMetric evaluation statistic report

Inheritance Richness (IR)	Attribute Richness (AR)	Average Population (AP)	Class Richness (CR)
0.98	0.2	1.2	0.30

According to the table 4 show that, the lowing value of IR present the ontology domain with detailed. A high value of IR specify. More attributes present more knowledge in ontology. A high value of AR is more desirable. A lowing value of CR represent the ontology not have enough knowledgebase concept in the schema. A high value of CR represent the ontology have enough knowledgebase concept in the schema. A lowing value specify, the instances extracted into the knowledgebase is insufficient to represent of knowledge in schema. According to Pak & Zhou [8], it is not possible for only one approach to ontology evaluation to work well for applications context. Instead, the selection of an evaluation approach should examine many factors, domain in which the ontology to be used and the aspect of the ontology to be evaluated.

6.3.2 System Evaluation and Searching Result

The main interface of our system is shown by figure 8 below. By typing text or sentence " I go to Phimai in Thailand "into the search box , click search and system response result back to the users.

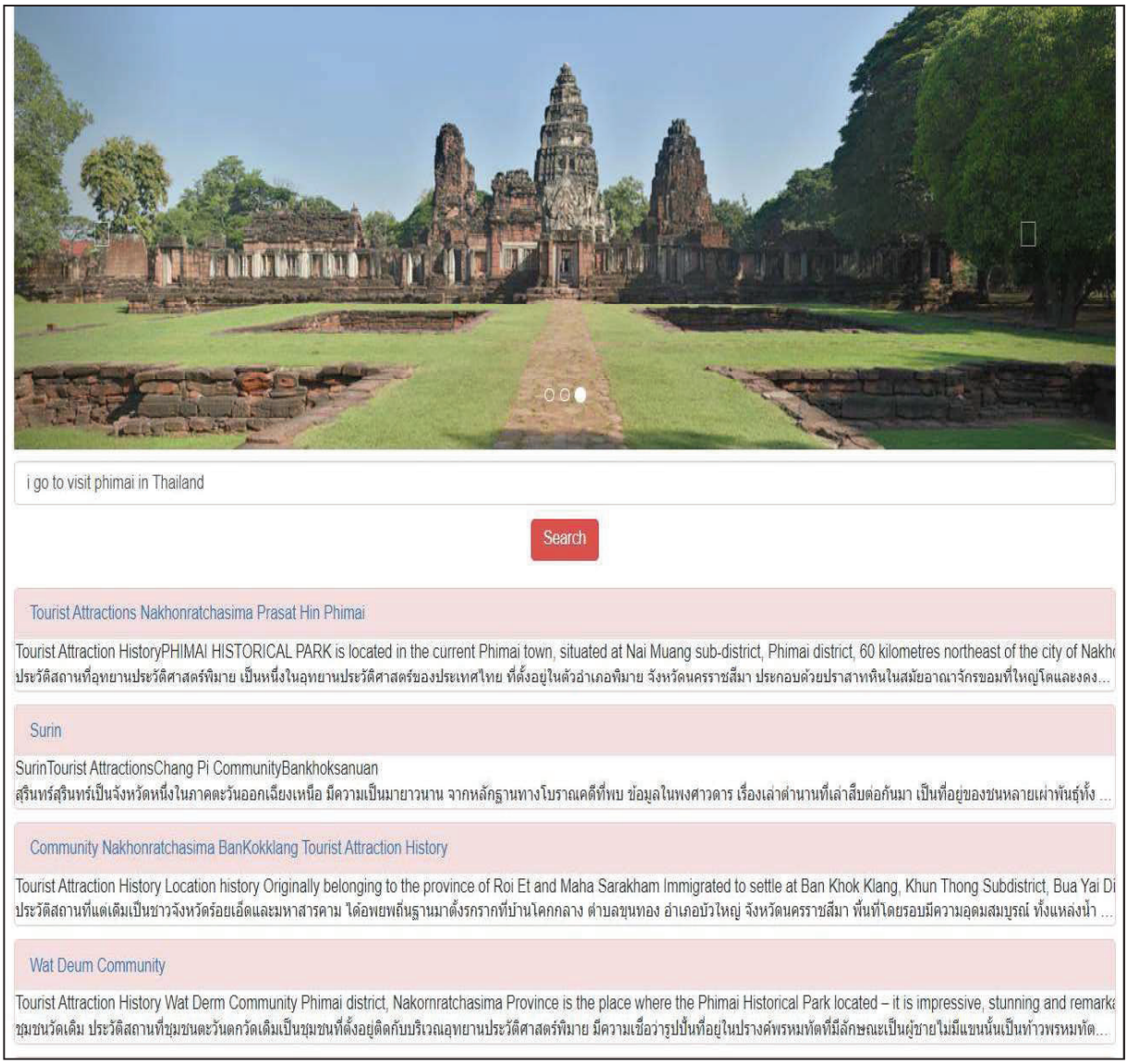

Figure 8 User interface and searching result

The system evaluation, focused on the result that system be able to find more result from RDF dataset. For evaluation system searching performance with relevant information from system. First system use only keywords without ontology to search information, and second system use keyword combine ontology keyword-based to search information.

Table 6 Keyword and user searching result

No	Only Keywords	Record Found	keyword with ontology	Record Found
1	surin	2 Links	surin	12 Links
2	Cambodia, Thailand	1 Links	Cambodia, Thailand	13 Links
3	Angkorwat	1 Links	Angkorwat	11 Links
4	Phimai	5 Links	Phimai	12 Links
5	Korat City	0 Links	Korat City	12 Links
6	Korat Zoo	0 Links	Korat Zoo	12 Links
7	Moeung Tam	0 Links	Moeung Tam	12 Links
8	Nakhonratchasima	6 Links	Nakhonratchasima	12 Links
9	siemreap	0 Links	siemreap	12 Links
10	buriram, province	1 Links	buriram, province	12 Links

Note. User query of ten sentences has removed stop words ready and the keyword is shown on table 5 above. These keywords will use to query information from two searching systems. You can see the process on Table 3 sparql content query and WordNet. The keywords of table 5 will use to combine with ontology keyword-based (expansion keyword) to search relevant information from the semantic web search engine system.

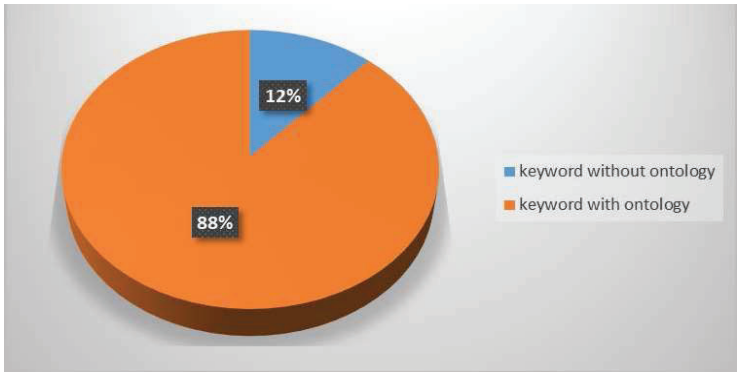

Figure 9 Searching result

According to table 5 and figure 9 above, the different result searching information between keyword and ontology keyword-based searching shown that Ontology keyword-based search can find much relevant information from RDF dataset. The searching system use only keyword to search information from web search get a few relevant information with 12%. The searching system use keyword combine with ontology keyword-based to search information from semantic web search engine get more relevant tourism information with 88%. It indicate that semantic web use ontology as keyword-based can get more searching relevant information and joyful.

7. Conclusion

This research presents a new semantic web search which used ontology tourism construction process, calculating semantic related word by using SPARQL to query semantic related word of superclass, subclass and instance in tourism ontology, object property query with two specific

instances and keyword expansion process. Also, keywords search over RDF dataset with SPARQL query language are applying to search the similar meaning of the words. According to evaluation, the proposed system can help to search the similar words on https://www.phimai-angkorwat.com/tour/frontend which consisted various local data of Southern Isan of Thailand and Northern of Cambodia. The proposed system based on semantic web technology and shows that user can specify the preference of needing of a vocation and get relevant sources of information from web such as culture, festival, activities, attraction places, shopping centers, accommodations and so on. For the future work construction of tourism ontology will add more classes, subclass, and instances and object property. Construction new algorithm for calculating semantic similarity word and keyword scoring algorithm. Use more approach or technique to valuate ontology for satisfying result.

8. References

[1] Ayepeh, T., & Janruang, J. (2017). The Development of Ghana Semantic Tourism Guide System. The 41st National and 5th International Graduate Research Conference. 18.

[2] Chinnapatjeerat, R., Tuamsuk, K., & Supnithi, T. (2016). Ontology of information for tourism development planning in Thailand. *2016 5th International Conference on Computer Science and Network Technology* (ICCSNT), 856–878. https://doi.org/10.1109/ICCSNT.2016.8070281

[3] Fahad, M., Qadir, M. A., & Noshairwan, M. W. (2007). Semantic Inconsistency Errors in Ontology. *2007 IEEE International Conference on Granular Computing* (GRC 2007), 283–283. https://doi.org/10.1109/GrC.2007.153

[4] Gómez-Pérez A. (2004) Ontology Evaluation. In: Staab S., Studer R. (eds) Handbook on Ontologies. *International Handbooks on Information Systems* (pp. 251-273). Springer, Berlin, Heidelberg

[5] Grimm, S., Abecker, A., Völker, J., & Studer, R. (2011). Ontologies and the Semantic Web. In J. Domingue, D. Fensel, & J. A. Hendler (Eds.), *Handbook of Semantic Web Technologies* (pp. 507–579). Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007/978-3-540-92913-0_13

[6] Jain, S., & Valerie, M. (2018). Evaluation and Refinement of Emergency Situation Ontology. International *Journal of Information and Education Technology*, 8(10), 713–719. https://doi.org/10.18178/ijiet.2018.8.10.1127

[7] Lee, C.-I., Hsia, T.-C., Hsu, H.-C., & Lin, J.-Y. (2017). Ontology-based tourism recommendation system. *2017 4th International Conference on Industrial Engineering and Applications* (ICIEA), 376–379. https://doi.org/10.1109/IEA.2017.7939242

[8] Pak, J., & Zhou, L. (2010). A Framework for Ontology Evaluation. In R. Sharman, H. R. Rao, & T. S. Raghu (Eds.), *Exploring the Grand Challenges for Next Generation E-Business* (Vol. 52, pp. 10–18). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-17449-0_2

[9] Ramathilagam, C., & L. Valarmathi, M. (2013). A framework for OWL DL based Ontology Construction from Relational Database using Mapping and Semantic Rules. International *Journal of Computer Applications*, *76*(17), 31–37. https://doi.org/10.5120/13341-0776

[10] Staab, S., & Studer, R. (Eds.). (2009). *Handbook on ontologies* (2nd ed). Springer.

[11] Tartir, S., Arpinar, I. B., & Sheth, A. P. (2010). Ontological Evaluation and Validation. In R. Poli, M. Healy, & A. Kameas (Eds.), *Theory and Applications of Ontology: Computer Applications* (pp. 115–130). Springer Netherlands. https://doi.org/10.1007/978-90-481-8847-5_5

[12] Wardhana, H., Mustofa, K., & Sari, A. K. (2018). Utilization of Semantic Web Rule Language for Tourism Ontology. *2018 Third International Conference on Informatics and Computing* (ICIC), 1–5. https://doi.org/10.1109/IAC.2018.8780474

[13] Yunzhi, C., Huijuan, L., Shapiro, L., Travillian, R. S., & Lanjuan, L. (2016). An approach to semantic query expansion system based on Hepatitis ontology. *Journal of Biological Research-Thessaloniki*, *23*(S1), 11. https://doi.org/10.1186/s40709-016-0044-9

[14] Zhao, X., Liu, L., Wang, H., & Song, W. (2015). Ontology Construction of the Field of Tourism in Africa. *2015 8th International Symposium on Computational Intelligence and Design* (ISCID), 47–50. https://doi.org/10.1109/ISCID.2015.180