North Sumatera tourism spots searching system based on semantic web

A Amalia, Herriyance, and R Angriani

Department of Computer Science, Universitas Sumatera Utara, Jl. dr. Mansur No. 9
Kampus USU Medan 20155

E-mail: amalia@usu.ac.id

Abstract. North Sumatera Province in Indonesia has many amazing tourism spots to visit. Unfortunately, information about these places is not organized well yet. Not all of these tourism spots have official websites and that made information with specific parameters according to user needs cannot be fulfilled if only depend on a search engine on the internet. The limitation of ordinary search engines cannot generate complex queries that involving relevant background knowledge of these places such as tourist locations, facilities offered, culinary and so on. Therefore in this paper, we offer North Sumatera Tourist sports searching system based on the semantic web. We utilized ontology as a source of knowledge that describes tourism in the province of North Sumatera. We observed and scrapped information about tourism spots in North Sumatera from trusted traveling websites. The next stage is the design of the Resource Description Framework (RDF) and ontology to describe relations between parameters. The result of this study produces a system that can help and understand the desires of tourists to get information about tourism objects in North Sumatera. Test results with various relevant semantic keywords parameter show the success of this system is above 90%.

1. Introduction

North Sumatera is one of the biggest provinces in Sumatera which has some unique tourism spots. Some of them are Lake Toba, Maimun Palace, Lumbini Natural Park and else. Unfortunately, there is still limited information about tourism spots in North Sumatera. Only limited tourism spots that already have official websites. Tourists have to do extensive searching in a search engine to collect information on tourism spots such as the location, accommodations, food, and many more. The drawbacks of search engine based on keyword, it will generate many documents that not all are relevant to the tourism objects. The searching result still not as expected. The limitations of this kind of search engine that are not able to find tourism information based on the query. Therefore, in this paper, we implemented semantic web technologies that able to search tourist information based on the query. Semantic web is a collection of technologies that allow computers to understand the meaning of an information based on metadata, rather than depending on the results of the word [1], this technology facilitates the sharing of the web knowledge [3], not only for human but also for machine/computer [2]

Semantic web is an evolution of the WWW (World Wide Web) coined by Tim Berners-Lee in 1999 [2], he has a vision for the future of the World Wide Web where the information has a logical meaning and connection of terms that will facilitate the machine to integrate data and process information on the web [4]. In the semantic web, the logical meaning and connection terms are represented in Resource Description Framework (RDF) serialization like RDF+XML or Turtle notation. To describes broader semantic relationships we can deploy an ontology schema in ontology language like OWL. Ontology is an extension of RDF. Unfortunately, there are still minimal websites about tourist spots in North Sumatera even less for RDF/OWL. This study aims to design a more optimal search for tourism spots in North Sumatera using semantic web technology, including generating its RDF serialization.
The paper is divided into six sections. Section 2 describes related works from the previous researcher about the semantic web. Section 3 describes the methodology of this study. Section 4 presents the result and discussion and Section 5 concludes this study.

2. Related Work
Several previous studies concern about semantic web for the examples the research by [5] [6] [7] and [8]. Research conducted by [5] states that the mechanism in semantic web technology allows the development of new applications for commerce in the field of tourism. While in research [6] states that e-tourism can be the perfect application in the application of semantic web technology because it has more sophisticated. Both studies utilize semantic web technology in affairs in North Sumatera. Then in the research [7] provides information on the intended tourism spots and information on tourism objects in the vicinity. The research conducted by [8] also uses an ontology-based tourism recommendation system that provides information about location, transportation, and accommodation. The difference with this research is the application of the semantic web that provides tourism information and location, the nearest hotel and the facilities offered as well as the nearest restaurant and the cuisine it provides in North Sumatera.

3. Methods
In this study, the system consisting of several steps. First, this research collected the related data to tourist spots, hotels, and restaurants in North Sumatera. After the information has been collected, then the next step is creating an RDF and ontology. RDF and ontology is a data model to describe the relationship between tourism spots, hotels, and restaurants. After the data is completed in the form of ontology, the research continues to create an interface to facilitate the user in finding the desired information about tourism spots in North Sumatera. The diagram architecture of this study can be seen in Figure 1.

3.1 Observing Related Websites and Other Sources
In this stage, we build the vocabularies or terms of the tourism domain, these terms which then be described in an RDF serialization. We retrieve the information about North Sumatera Tourism Spots from some valid and reliable websites like TripAdvisor.com, Traveloka.com, and Booking.com. We also retrieve the information about tourism spots in North Sumatera from other sources like books and magazines.

3.2 Scraping
Web scraping is a process to extract essential pages from the websites. In this study, we only take the needed pages that contain information about name, location, the closest hotel of tourism spots. We also retrieve the pages that contain information about hotel facilities and the closest restaurants from tourism spots. This process generates HTML pages.
3.3 Convert to XML
This stage is a process to extract the keyword from generated HTML files. The keywords of this study are the name of the tourism spots, location, nearby hotel and so on. All of this keyword, then described in XML format. XML stands for eXtensible Markup Language. XML was designed to store and transport data and could be able to be read both by human and machine.

3.4 Convert to RDF
This stage is to convert XML into RDF format. RDF is a standard model for data interchange on the web. The RDF information is described as a sentence consisting of subject, predicate, and object. These three elements are known as triple. In the RDF, a subject or a resource is described as Uniform Resource Identifier (URI), the predicate uses the property that connects the subject and object and the object is a URI or literal [9]. In this study, we deploy RDF+XML as RDF serialization. The example of RDF files can be seen in Figure 2.

```xml
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:resto="http://localhost/semantik/obyekwisata/sumut/Resto/#">
  <rdf:Description rdf:about="http://localhost/semantik/obyekwisatasumut/Resto/#JabuBerastagi">
    <resto:name>Jabu Berastagi</resto:name>
    <resto:location>Jl. Perwira no. 68, Berastagi 22152, Indonesia</resto:location>
    <resto:offers>Eropa</resto:offers>
    <resto:offers>Asia</resto:offers>
    <resto:offers>Indonesia</resto:offers>
    <resto:offers>Sesuai Untuk Vegetarian</resto:offers>
  </rdf:Description>
</rdf:RDF>
```

Figure 2. The Example of RDF

Figure 2 is an RDF example that contains several triples. It can be seen that this RDF describes data models in triples which are subject, predicate, and object. For example JabuBerastagi as subject, location as predicate and "Jl Perwira No. 68, Berastagi 22152" as an object. We implemented vocabularies from several namespaces to describes this triple. Every tag that starts with rdf or resto comes from the vocabulary contained in the namespaces prefix. While the rdf tag: Description is used to describe the object that is explained in the rdf: about attribute.

3.5 Ontology
Ontologist gives the right vocabulary with the knowledge that can be represented. This vocabulary allows us to determine which entities will be represented, grouped, and relationships among them [10]. The more precise the ontology is, the higher the potential for understanding how data can be used [11]. There are several steps to build an ontology.

3.5.1 Determining domain
The domain in this study is tourism spots in North Sumatera, which includes the closest hotels and restaurants to these tourist spots.

3.5.2 Determining class and sub-class
The design for ontology can be seen in Figure 3, where a class is determined based on the most common data, and SubClass is determined based on more specific (special) data from the Class. In this study, we deploy ontology by implementing it using Protégé application. Protégé is a tool to create an ontology domain, adjust forms for data entry, and input data.
3.5.3 Determining Properties

In Table 1 it is shown the description of properties. There are two types of properties, they are datatype property and object property. Datatype property used to connect individuals with its data value while Object Property is used to connect individuals with individuals.

| Property   | Function of Property                                      | Type Property     |
|------------|----------------------------------------------------------|-------------------|
| name       | Link the value string name of tourism spots, hotels, and restaurants | DataType Property |
| location   | Link the value string location of tourism spots, hotels, and restaurants | DataType Property |
| based_near | Link the tourism spots with hotels and restaurants around | Object Property   |
| offers     | Link the value string of restaurants and the foods       | DataType Property |
| facilities | Link the value string of hotels and the facilities       | DataType Property |

3.5.4 Determining Individuals

Individuals (instances) are the most basic real forms of classes associated with properties. The individual of this study can be seen in Figure 4.

An RDF / OWL can describe a relationship between subjects. This relationship can be described as a semantic net graph. The semantic net between Siantar WaterPark attractions and Avalokisvara Vihara can be seen in Figure 5.
Figure 5 Graph Siantar WaterPark dan Vihara Avalokisvara

Figure 5 shows the relevance of Siantar WaterPark and Vihara Avalokisvara. By this semantic net, we can find the location of these spots, and we also can find nearest restaurants and cuisine type of these restaurants. From the picture and explanation above, it can be concluded that if the user wants to find a restaurant that serves Asian cuisine between Siantar WaterPark and Vihara Avalokisvara, the results that come out are the Massa Kok Tong restaurant and Mie Pangsit Awai.

In this study, we created an RDF/OWL that describes the relevance of many triples in domain North Sumatera Tourist Spots. Based on this RDF, the user can search any information not only by keyword but also with query and questions. We deploy an interface that can extract this information with utilized SPARQL syntax. SPARQL is a language that can retrieve data written using RDF / OWL or XML by the query. In this study, we deploy Apache Jena Fuseki as triple stores to store and also as a query tool. An example of a query with SPARQL syntax can be seen in Figure 6, this syntax is a query to retrieve information about the closest hotel to Siantar Water Park.

Figure 6. SPARQL To Extract Information The Nearest Hotel to Siantar Water Park

| o | ᵙ | ᵗ |
|---|---|---|
| o:obyek_wisata:Hotel/SiantarHotel | “Siantar Hotel” |
| o:obyek_wisata:Hotel/HorisonSiantarHotel | “Horison Siantar Hotel” |
| o:obyek_wisata:Hotel/Pamamalin | “Pamamalin” |
| o:obyek_wisata:Hotel/GrandMegaHotel | “Grand Mega Hotel” |
| o:obyek_wisata:Hotel/SapadisHotelSiantar | “Sapadis Hotel Siantar” |

Figure 7 SPARQL Results for The Nearest Hotel to Siantar Water Park
4. Result and Discussion

The result of this study is a system that can retrieve information about North Sumatera Tourism Spots. A text box is provided for users to input queries or questions that they want to retrieve. The Interface of this system that shows the example of a query and the result generated by the system can be seen in Figure 8.

![System Interface](image)

**Figure 8. System Interface**

To check the accuracy of the system, we evaluated this system with 20 various queries. We confirm the result with "true" if the system generates the expected result and "false" if the system does not generate the expected result. The excerpts of these sample queries can be seen Table 2.

| NO | Input                                                                 | System                                                                 | Result |
|----|------------------------------------------------------------------------|------------------------------------------------------------------------|--------|
| 1  | Hotel terdekat dengan candi bahal apa ya? (in English: What hotel is closest to the Bahal temple?) | Mega Permata Hotel, Pia Hotel, Hotel Sitam diang, Natama Hotel         | TRUE   |
| 2  | hotel terdekat dari Siantar Waterpark? (in English: the closest hotel to Siantar Waterpark?) | Siantar Hotel, Horison Siantar Hotel, Pariama Inn, Grand Mega Hotel, Sapadia Hotel Siantar | TRUE   |
| 3  | hotel apa yang ada fasilitas penitipan anak nya? (in English: What hotels are there for child care facilities?) | The Hill Hotel Resort Sibolangit, Sapadia Hotel Siantar, Hotel Grand Antares, Pariama Inn, Grand Mega Hotel, Graha Kardop Hotel | TRUE   |
| 4  | hotel Grand Inna Medan menyediakan fasilitas apa saja? (in English: Grand Inna Medan hotel provides any facilities?) | -                                                                     | FALSE  |
| 5  | di mana lokasi restoran Simpang tiga? (in English: where is the location of the Simpang Tiga restaurant?) | -                                                                     | FALSE  |
| 6  | restoran apa saja yang terdekat dengan pulau berhala (in english: any restaurant that is closest to the Berhala island) | Mi Aceh Titi Bobrok, Gaboh Burger, Rumah Makan Sipirok, Lembur Kuring Restaurant | TRUE   |
| ...| ...                                                                     | ...                                                                    | ...    |
| 20 | lokasi obyek wisata yang bernama pulau malau itu di mana ya? (in English: Where is the tourism object called Malau Island?) | Samosir Indonesia                                                      | TRUE   |
From the 20 sentences have been tested, 18 sentences or 90% display expected responses according to the knowledge base.

5. Conclusions

There are still minimal websites about tourist spots in North Sumatera even less for RDF / OWL. Therefore, in this study, we created an RDF/OWL that describes the relevance of many triples in domain North Sumatera Tourist Spots. The step of this study contains collecting data from trusted websites, books, magazines and else, then the research continues by creating an RDF and OWL to describe the relationship between tourism spots, hotels, and restaurants. The research continues to create an interface to facilitate the user in finding the desired information about tourism spots in North Sumatera. Based on this RDF / OWL, the user can search any information not only by keyword but also with query and questions. We deploy an interface that can extract this information with utilized SPARQL syntax. The result of this study produces a system that can help and understand the desires of tourists to get information about tourism objects in North Sumatera. Test results with various relevant semantic keywords parameter show the success of this system is 90%. The result of this research shows that semantic web technologies implementation in the searching system shows a better search result rather than conventional search engine as it removes spam results that could be shown in the conventional search engine and this system can be a reference for government tourism office in North Sumatera.

Reference

[1] Guha, Ramanathan., McCool, Rob., & Miller, Eric. 2003. Semantic Search. Proceedings of the 12th International Conference on World Wide Web, pp. 700-709.
[2] Afuan, Lasmedi, Azhari. 2016. Penerapan Semantik Web Pada Ontologi Learning Resource Repositori. Yogyakarta: Universitas Gadjah Mada.
[3] Roy, Raktim Kumar. 2017. Weather Phenomenon prediction Using Semantic Web Technology. Dhaka, Bangladesh: East West University.
[4] Shekhar, Monica, Saravanaguru Ra. K. 2013. Semantic Web Search Based On Ontology Modeling Using Protégé Reasoner. India: VIT University Vellore, Tamil Nadu.
[5] Maedche, Alexander and Staabb, Steffen. 2002. Applying Semantic Web Technologies for Tourism Information Systems. (DOI: 10.1007/978-3-7091-6132-6_32).
[6] Siricharoen, Waralak V. 2008. Learning Semantic Web from E-Tourism. Vol 4953, (ISSN 0302-9743), (DOI 10.1007/978-3-540-78582-8_52), pp. 516-525.
[7] Zendrato, Roslein Ria Putri, Wahyudi, Adhie Tri, AdhiWicaksana, Bagus Ismail. 2016. Implementasi Semantic Trip Planning Dalam Perancangan Aplikasi Mobile Perencanaan Perjalanan Wisata di Wilayah Ekksaresidenan Surakarta. Techno Vol 17(1), (ISSN 1410 - 8607), pp. 001 – 007.
[8] Lee, Chin-I, et al. 2017. Ontology-Based Tourism Recommendation System. 2017 4th International Conference on Industrial Engineering and Applications, pp. 376-379.
[9] Yu, Liyang. 2011. A Developer’s Guide to the Semantic Web. Springer Berlin Heidelberg: Berlin.
[10] Mouhim, S, et al. 2010. Towards a Knowledge Management System for tourism based on the Semantic Web Technology. Morocco: Faculty of sciences & ENCG.
[11] Segaran, Toby, Colin Evans, Jamie Taylor. 2009. Programming The Semantic Web. United States Of America : O’reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, Ca 95472.