MyGeo-EXPLORER: A SEMANTIC SEARCH TOOL FOR QUERYING GEOSPATIAL INFORMATION

Subashini Panchanathan[1], Abba Lawan[2] and Abdur Rakib[2]
[1]Malaysian Center for Gespatial Data Infrastructure (MaCGDI), Putrajaya, Malaysia
[2]School of Computer Science, University of Nottingham, Malaysia Campus
E-Mail: khyx3alw@nottingham.edu.my

ABSTRACT

We propose a semantic search approach for geospatial information systems to resolve the issue of semantic heterogeneity in metadata catalogues. Using the Malaysian geospatial data infrastructure service portal, MyGDI Explorer, as a comparative case study, a semantic search tool developed in this work is based on the model-view-controller (MVC) approach, which comprises three major components, an ontology component modelled using the Malaysian Geospatial Standard (MS 1759) as a domain, a java-based middleware component using Apache-Jena API and a query-based semantic search engine developed in Java EE. In this paper, we present MyGeo-Explorer, a web application coupled with a semantic search engine for exploring geospatial information and illustrate the use of the tool with some common usage scenarios. We show how ontology-based semantic search provides more relevant search results as compared to traditional keyword search and helps to deal with the problems of semantic heterogeneity inherent in keywords definition.

Keywords: semantic search engine, ontology, geospatial information, semantic web, knowledge management.

1. INTRODUCTION

Geospatial information is an important facet of Geographical Information System (GIS), universally employed for location based decision-making and planning. One of the main infrastructures provided for the discovery of geospatial information is the metadata catalogue. Searching for information through metadata catalogues are usually achieved using keywords. However, due to the semantic heterogeneity of the keywords offered by different data providers, searching these catalogues can be tedious and end-users may not always get efficient search results. Access to accurate geospatial information is vital to every geographical information systems, which are crucial for effectively managing environmental, social, as well as economic situations. Spatial Data Infrastructures (SDIs) facilitate the discovery and sharing of relevant geospatial information in a region, country or continent through the infrastructures provided. One of the main infrastructures that can be provided for the discovery of geospatial information is the metadata catalogue. Metadata is data about data and in the context of this paper; it refers to any data describing geographical location data. End users can use keywords to search for the geospatial information through metadata catalogue services.

This is the case of MyGDI Explorer, an open access metadata catalogue developed by the Malaysian Centre for Geospatial Data Infrastructure (MaCGDI) – a department under the Ministry of Natural Resources and Environment, Malaysia (MyGDI, 2012). MaCGDI develops the Malaysian Geospatial Data Infrastructure (MyGDI) initiative that comprises of policies, data and standards, etc. to promote sharing and dissemination of geospatial information among government agencies, private and the public sectors (Arshad & Hanifah, 2010). Under the MyGDI initiative, the MyGDI Metadata Catalogue or MyGDI Explorer was developed in 2002, which enables searching, accessing and sharing of Malaysian geospatial information. Data providers who wish to share their geospatial data or services are allowed to publish their metadata in MyGDI Explorer and end users can search the online catalogue either based on keywords, data category, modified date or the content type. However, due to the semantic heterogeneity in the keywords supplied by different data providers, users may not always get relevant search results. Moreover, as Bernstein and Klein explained keyword-based searches in general, are known to produce results that are low in precision and recall (Bernstein & Klein, 2002). Moreover, language barrier is another issue with data searching, sharing and utilization. As such, information retrieval methods based solely on keyword searching are inefficient in getting desired search results.

To overcome this issue, a semantic search technique based on ontology can be considered, where the search engine first uses domain ontology to define the search keyword and its relationship with other concepts before exploring the metadata catalogue. This helps to ensure that only relevant data may be returned as search results. Ontologies explicitly define and explain (using semantic annotations) the relevant terms in a domain with the relationships that exists between them (Gruber, 1993). We discuss more on ontology and its languages in Section 2.2. In essence, the proposed solution discussed in the remainder of this paper consists of developing a domain ontology describing the geospatial data as described in MS1759 – a standard for geospatial data features and attribute coding in Malaysia, and a semantic search engine that will utilize the domain ontology to provide relevant search results. The framework to develop the semantic search engine is an open source solution comprising four components, viz. ontology, convertor, middleware, and a search engine component. The ontology component consists of a global ontology developed in Web Ontology Langauge (OWL) (Patel-Schneider, 2004), and local

VOL. 10, NO. 23, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

ontologies developed in RDFS – an acronym for Resource Description Format Schema and a predecessor to OWL. OWL ontologies may describe relevant domain concepts as classes, properties or roles between the classes, and their individual assertions or instances. The convertor component is a semi-automated tool used to convert the metadata in XML to RDFS. The middleware component, Apache Jena (Jena, 2014), is used for the integration of ontology component and the web-based search engine. While the scope of the ontology is restricted to the MS 1759 as the domain, the overall framework is however domain independent as the semantic search approach can be used to improve any keyword-based search engine by explicitly modeling the domain of discourse using ontology. Thereby improving search efficiency with higher precision and recall.

The remainder of the paper is structured as follows: Section 2 describes the background literature and Section 3 discusses the methodologies employed in the development of the ontology and the semantic search engine. Implementation of the MyGeo Explorer tool is presented in Section 4, while Section 5 discusses and analyzes the results with comparison to the existing keyword based search engine. We then conclude in Section 6 with a highlight on the future works.

2. LITERATURE REVIEW

a) Geospatial Information

Spatial data is also known as geographic information that identifies geographic location of features and boundaries on Earth. It can be mapped and is stored as coordinates and topology. Spatial data infrastructures (SDI) are set up by countries, region or continent to enable access to geospatial information. Some of the SDIs established are the Canadian Geospatial Data Infrastructure (CGDI), European Spatial Data Infrastructure, US National Spatial Data Infrastructure and the Malaysian Geospatial Data Infrastructure (MyGDI) (Arshad & Hanifah, 2010). Basic components of SDIs are data, access network, standards, policy and people. However, as (Smits *et al.*, 2007) conferred, the most important part of an SDI is the catalogue services, which enables users to search and retrieve distributed geospatial information. This information is represented as metadata, which can be queried and evaluated by humans as well as software (ISO, 2014).

Geospatial information in metadata catalogues usually describes the content, title, format, abstract, data providers and related information. Metadata should comply with certain standards for easier description and comparison. The three main metadata standards developed for metadata description are the Content Standard for Digital Geospatial Metadata, CEN Pre-standard and ISO 19115/19139 Standard (Douglas, 2004). The metadata standard for Malaysia follows the ISO 19115. This standard defines the metadata sections, entities and elements including the schema for describing them. An implementation standard called ISO-19139 was developed to validate the XML structure of ISO-19115. Extensible Markup Language (XML) describes data objects in XML documents and behavior of computer programs that processes them. It is a restricted form of SGML or Standard Generalized Markup Language. It was developed in 1996 under the World Wide Consortium (W3C). This language is used to describe the geospatial information in metadata catalogues to enable transfer and sharing of information across heterogeneous platforms and supports interoperability. However, XML has a complicated data structure and therefore, cannot express semantics or be scaled globally (Yu, 2007). As such, it has to be extended with ontology models to add semantic meaning to documents.

b) Ontology and Semantic Web

The Semantic Web is defined as an extension to the current web of documents. It is also known as Web 3.0 or web of linked data (Lee, 2001). Semantic Web, which uses ontology as its knowledge model, can connect data from different sources and enables computers to understand it. One major goal of the Semantic Web is to provide semantic markup that enables access to information from various sources in the Web (Lei, Uren, & Motta, 2006). The Semantic Web is more concerned on the meaning of data compared to the World Wide Web (WWW) or Web 1.0, which is more concerned with human readable structure. Figure-1 shows the Semantic Web Stack.

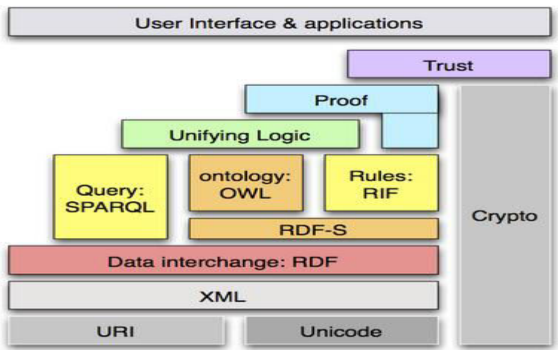

Figure-1. The Semantic web stack. Image source: http://www.w3.org/2009/Talks/0120-campus-party-tbl/#(14).

Ontology on the other hand, is a term borrowed from philosophy and is widely quoted as "an explicit specification of a conceptualization" (Gruber, 1993). Conceptualization is any simplified version of knowledge that we wish to represent for some purpose. Ontology is also defined as any common vocabulary employed and adopted to share information in a given domain (Noy & Mcguinness, 2000). By adding semantics to domain concepts, ontologies help to model a knowledgebase that is both human understandable as well as machine readable. When developing ontologies, a reasoner needs to be invoked to detect inconsistencies. Commonly used reasoners include Pellet, HermiT, Fact++, KAON2, Cerebra Engine and RACER among others (Gennari *et al.*, 2003).

Implementation of ontology in SDI can improve data sharing and information retrieval. According to (Lacasta, *et al.* 2007), ontologies can be applied in three main areas of SDI. Firstly, they can be used for data sharing and developing GIS systems by defining meaning of geospatial data features, secondly, they can help in classifying resources and for information retrieval, and thirdly, ontologies can help in the management of metadata by profiling a metadata and also provide interoperability between different metadata standards. Although metadata catalogue services support the discovery, organization and access of geographic information, they cannot solve semantic heterogeneity problems (Klien *et al.*, 2004) – a problem frequently faced by metadata consumers. In order to overcome this problem, ontologies can be used to identify and associate semantic concepts within a metadata.

i. Types of Ontologies

Ontologies are often divided into upper level ontologies, domain ontologies, task ontologies and application ontologies. Upper level ontologies explain generic concepts, domain ontologies explain vocabulary of a scoped knowledge area, task ontologies describes activities and application ontologies describe a specific application's vocabulary in a domain (Lawan *et al.* 2014). For this work, ontology for the geospatial domain was developed.

ii. Ontology Languages

Resource Description Framework (RDF), its schema, RDFS and the Web Ontology Language (OWL) are the common ontology languages for the semantic web. In principle, RDF is an XML-based language to identify and describe information in a web page or any object on the web. It is lightweight and flexible. The basic elements of RDF are resource, property and statement. The Web Ontology Language (OWL) on the other hand, is the standard ontology language for the semantic web as recommended by the W3C. Various upper levels as well as domain ontologies based on OWL are developed and publicly available for reuse (Lawan *et al.* 2014). It is considered as an improvement over its predecessor, the RDF, with much added expressiveness and management tools. The current version OWL 2 is highly expressive and divided into three sub profiles, viz. the OWL 2 Rules Language (OWL2RL), OWL 2 Query Language (OWL2QL) and OWL 2 Expressive Language (OWL2EL). Figure-2 (Djuric, 2006), shows various languages arranged in layers of syntax and semantics.

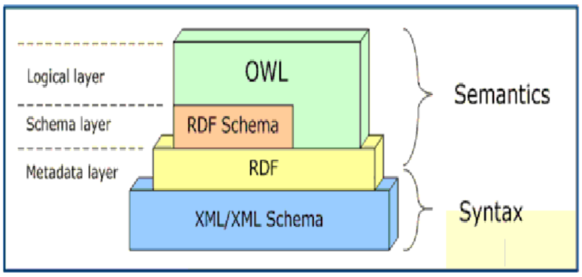

Figure-2. Layered architecture of ontology languages.

c) Ontologies and Knowledge Management

Information Science and knowledge management practices, involves the development of tools and techniques for acquisition, representation, usage, preservation, as well as evolution of human knowledge. However, in order to use existing information to create knowledge, knowledge engineers need to understand and generate semantic relationships that are bound to exist between various terms, keywords and facets of domain knowledge. This can be made easier through the use of ontologies, which provides an explicit specification of terms, keywords or concepts in a given domain (Gruber, 1993). These semantics added to keywords, using an ontology language, are human readable as well as machine-processable and thus gives an edge to using ontologies as a tool for knowledge management.

Moreover, in his article, "Biomedical ontologies in action: Role in knowledge management, data integration and decision support", (Bordenrider O., 2008) describes further the roles of ontologies in knowledge management systems (KMS) to include: annotation or indexing of resources, retrieval of data and information, data exchange and integration, providing semantic interoperability among domain concepts, as well as knowledge discovery. While biomedical ontologies are specifically mentioned for domain referencing, these roles are indeed applicable to other domain ontologies, including geographical information ontologies. In this paper, we develop the domain ontology for MS 1759 to provide semantic relationships between terms and further integrate them with other metadata sources in RDFS to enhance the information extraction through the semantic-based search engine, called MyGeo-Explorer. We present the relevant works in the section 2.5 and provide a detailed description of MyGeo-Explorer in section 4.

d) Semantic Search

Semantic search is an application under the Semantic Web where search engines try to understand the meaning of search terms before exploring the knowledge base for relevant results. By adding semantic tags into documents, and using standard definition of domain concepts – usually provided by ontologies, a semantic search is able to return precise search results as it understands the meaning of search keywords and queries performed by the end users. While the general idea of a semantic web is to allow users to search for information from various sources and domains, the concept of semantic search is used to define intelligent searching of information from a single domain (Cambridge, 2014).

e) Relevant Works

According to a study done by (Sudeepathi *et al.* 2012), various semantic search engines have been previously developed, such as Hakia, DuckDuckGo, SHOE and Swoogle. Swoogle is a metadata and search engine for the semantic web, which is able to discover documents, extract their metadata and construct a semantic relationship between them (Ding *et al.* 2004). Whereas, Simple HTML Ontology Extensions (SHOE) is another search engine that allows users to make queries that retrieves information from different semantic sources (Heflin, & Hendler, 2000). In the geospatial domain, the technique of semantic search for information retrieval has also been widely adopted. Spatial Web Portals is used by the Earth Science Community for sharing, exchange and interoperation of geospatial data, metadata and web servers. Moreover, Spatially-Aware Information Retrieval on the Internet (SPIRIT) is another spatial search engine that can be used to retrieve geographical information on the web and search for web documents with spatial contents (Jones *et al.*, 2002).

Previous works on using semantic search to enhance metadata discovery have been presented in (Berkley, 2009; Shanming, 2008; & Xu, 2008) and as summarized in (Singh, 2013), a comparison study between traditional search engines such as Yahoo, Google etc. and semantic search engines such as Bing and DuckDuckGo shows that more relevant documents are retrieved from the semantic search engines. For this research, a semantic search engine for searching and querying geospatial information, based on MS 1759 ontology, have been developed.

3. METHODOLOGIES

As described earlier, the proposed solution to the limitations of traditional keyword-based searching in MyGDI explorer is to enhance the knowledge retrieval process through Semantic Search approach. This involves developing a domain ontology describing the semantic relationships that exists between keywords and further integrate the keywords with other metadata sources in RDFS to act as the knowledgebase. A semantic web-based search engine, called MyGeo-Explorer is then developed to serve as the interface for enhanced searching and information retrieval. In this section, we describe the methods and procedures used in the development of the domain ontology and semantic search engine. The architecture of the overall framework is based on open source solution and consists of four major components described below.

a) Ontology Modeling

For scalability of the knowledgebase, a hybrid ontology method for data integration was employed, where the ontology model is separated into a global ontology consisting of shared concepts and various local ontologies each representing a metadata source.

i. Global Ontology

As stated earlier, the global ontology is based on MS 1759 as a domain. We chose MS 1759 as it is currently the Malaysian Standard for sharing geospatial data among users and data providers in Malaysia. In MS 1759, geospatial data is categorized into twelve data categories, which are Built Environment, Hydrography, Hypsography, Soil, Aeronautical, Demarcation, Transportation, Geology, Utility, Vegetation, Special Use and General. The global ontology was modelled based on these concepts to solve the issue of semantic heterogeneity in data sources from different data providers. A conceptual model of the global ontology is shown in Figure-3.

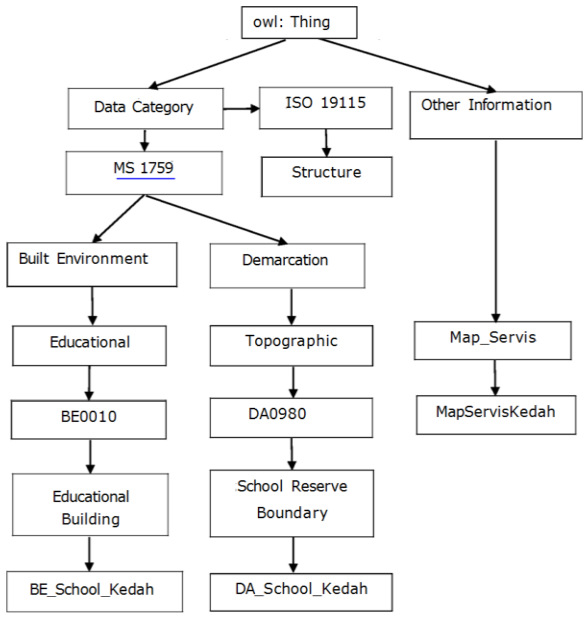

Figure-3. A fragment of the MS1759 ontology.

In the root of the global ontology is the owl:Thing, a top-level ancestral concept of OWL ontologies. This is followed by the domain concept Data Category subclass, which also contains the MS1759 and ISO19115 concepts as subclasses. Two important subclasses defined in the MS1759 class are Built Environment and Demarcation classes, which categorize geographical boundaries. Other relevant information that does not fall directly under the domain concepts, such as the Map_Servis class, is defined in Other Information subclass of the root concept.

However, the global ontology fragment shown in figure-3 only specifies the conceptual hierarchy of two data categories and detailed semantic relationships between these concepts are defined in the local ontologies in RDFS. We use the Protégé ontology editor for developing the ontology model in OWL 2. Protégé is a desktop application for developing ontologies and ontology based systems (Gennari *et al.* 2003).

ii. Local Ontologies

Using real metadata downloaded from MyGDI Explorer, the local ontologies are generated based on the metadata sources. The metadata, in XML format, is converted to RDFS using the convertor component so that the instances are represented as RDF triples in subject, predicate and object. The RDF structures give semantic meaning to the metadata. We chose to develop the local ontologies in RDF due to its simplicity. While the relationships between concepts are defined in the global ontology, the local ontologies provides the basic semantic representation of the metadata so that it can be linked with the concepts or classes in the global ontology through their Uniform Resource Identifiers (URIs) – an annotation that can be read by the Jena middleware during searching.

b) XML to RDF Conversion

With the metadata available in XML format we generate the RDF triples using an external component, the XML2RDF converter (Cao *et al.* 2009), which is a free generic converter that uses the Extensible Stylesheet Language for Transformations (XSLT) to convert XML documents to RDF. This converter avoids blank nodes, supports one-to-one mapping for inverse transformation and preserves the XML schema. For a detailed discussion on XML to OWL conversion and application, we direct interested reader to (Rakib *et al.* 2015). The RDF assertions generated from the metadata are triples describing the basic elements of the metadata, such as: Content Information, Dataset Category, West Bound Longitude, East Bound Longitude, South Bound Latitude, North Bound Latitude, Keyword, etc.

c) Middleware Component

In order to achieve the query processing of the ontology knowledgebase from the search engine interface, a mediator is needed for integrating the knowledge model and the user interface. We use Apache Jena, an open source Java framework for developing applications for the Semantic Web. It provides application programmable interfaces (APIs) to create and manipulate RDF graphs. Jena is originally developed to support RDF specifications and provided an API that was easy to use (Mcbride, 2001). Apart from allowing the manipulation of ontologies through java codes, Jena also provides inferencing using the inbuilt Reasoner. In this work, we chose Pellet reasoner due to its compatibility with the Jena framework.

d) Semantic Search Engine Component

The semantic search engine is designed based on the Model-View-Controller (MVC) software design paradigm. This is because it complies with industry standard, provides easy code management and helps to develop more secure applications. Similarly, the Java Enterprise Edition follows the MVC paradigm and thus it is used for the development of the web based search engine. JSP and Servlets are the components used in the development of the web based semantic search engine. Jena and Pellet API's are downloaded and stored in the Java library. It is invoked in the JSP code when integration is done with the ontology to provide results for semantic search. Apache Tomcat Server version7 is used to host the search engine. The architecture of the semantic search engine is shown in Figure-4.

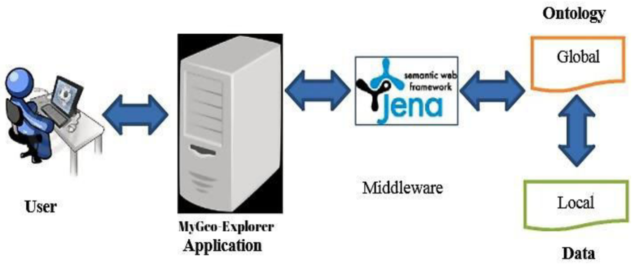

Figure-4. Architecture of MyGeo-Explorer.

To develop the search engine, we use the Java Enterprise Edition or JEE due to its web-based development support through components such as Servlets and Java Server Pages (JSP). JEE was also selected for ease of integration with our Jena API and is suitable because applications developed in JEE are scalable and can be deployed on any platform. Moreover, many semantic and ontology based applications have been previously developed using the JEE technology, including those presented by (Bahreini & Elci, 2007, & Xu *et al.* 2008).

4. IMPLEMENTATION AND RESULTS

a) The Domain Ontology

With a collaborative, top-down approach to the ontology development, we develop the domain ontology using Protégé 4.3 desktop application. The MS1759 class and ISO 19115 are implemented as sibling classes, each with its corresponding data categories as subclasses. Other relevant classes and their relations are also defined. For brevity, we focus our discussion on two important concepts to highlight the semantic heterogeneity of the terms used by the different data providers. The first case is for the data category concept, where the keyword Built Environment is used as data category of the MS 1759, which we found to be equivalent to Structure keyword in the data category of ISO 19115. The second case involves the language barrier issue, where the concept of School is provided in English at one data source and Sekolah in 'Bahasa Melayu' in another. While this can be easily solved by providing the language annotation in ontologies, it is not easily resolved with the existing structure of metadata. As such, an owl:SameAs construct is used to denote that the two concepts are not only similar but equivalent in meaning. In this case, the Keyword School is equivalent to Sekolah for features related to School in the state of Kedah. There are four data sources related to School in different classes of the global ontology. The equivalent concept relation in Protégé is shown in Figure-5.

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

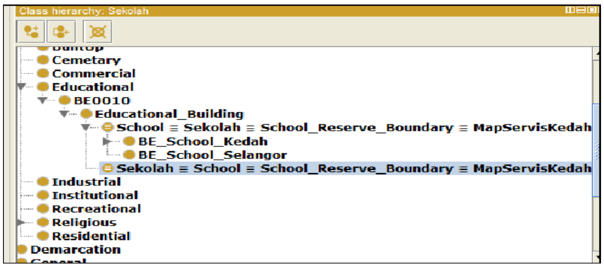

Figure--5. Implementation of 'equivalent concepts' role for School and Sekolah.

The global ontology is saved in RDF/XML serialization because Jena is designed to only support serialization in this format. To get instances and properties from the local ontology, the global ontology will link to the RDFS data source using a URI. This URI is also declared in the RDF data sources or local ontology so that the URIs are the same, allowing Jena middleware to create the linked-data. This is easily implemented using the 'Change Entity URI' option available in Protégé.

b) Query Formulation

Apache Jena version 2.12.0 was used as the middleware to integrate the ontologies (OWL and RDFS) with the search engine. The search engine provides results of matching between keyword and related concepts and returns values from each RDF file as a navigation link. Jena provides the APIs and a SPARQL query engine to accept the search keyword and construct the queries we embedded in a java code to return relevant search results. These APIs are stored in the Java library and invoked in the OntologyManager class to access the ontology and provide inferences. While Jena has its own API for reasoner, we use the Pellet reasoner version 2.3.1 to provide inference for ontology management and searching. As stated earlier, a SPARQL query language is used to provide search results based on class, entity and the map area. The SPARQL query to query the classes in OWL is shown below:

"PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#> "+ "PREFIX rdf<http://www.w3.org/1999/02/22-rdf-syntax-ns#>"+ "PREFIX xsd:<http://www.w3.org/2001/-XMLSchema#> " + "PREFIX gco:<http://www.isotc211.org/-2005/gco#> " + "PREFIX bpkpoint: <http://www.isotc211.org/2005/gmd/BPKPoint#> " + "PREFIX bpkpoly: <http://www.isotc211.org/2005/gmd/BPKPolygon#> " + "PREFIX mskedah: <http://www.isotc211.org/-2005/gmd/MSKedah#> " + "PREFIX suba: <http://www.nottingham.edu.my/ontologies/2014/6/suba/ Ontology-Nottingham-Suba#> " + " " +
SELECT ?subClass //select all subclasses.
WHERE { " + " ?subClass rdfs:subClassOf suba:" + className + " . " + " " + " } ; //where the role subclassOf is found.

In this query, the prefixes are used to define the short names for the schemas and ontologies to be queried. These definitions are then followed by the SELECT statements, which in this case is to select all subclasses, where the rdfs schema subclassOf is defined for any class in the suba ontology. Similarly, the SPARQL to query for distinct data from RDF files is shown below. The keyword is embedded in the queryString object and the query tries to find a match with the RDF subjects in the local ontologies. The result is then ordered by subjects.

… <list of prefixes>
SELECT DISTINCT ?subject ?property ?object " + "
WHERE { " +" ?subject a suba:"+queryString +" . " + " ?subject ?property ?object."+" }
ORDER BY ?subject ; // results ordering by subject.

c) The Semantic Search Engine

The semantic search engine is known as MyGeo-Explorer and the interface is designed similar to that of the MyGDI Explorer. For consistency, searching is also designed to begin with a keyword search. The search keyword is then matched against the domain ontology to retrieve similar keywords and their relationships. These concepts are then used to explore the local ontologies or RDF data for semantic similarity. Where one is found, the title of the dataset is returned as the search results. The current version of MyGeo-Explorer allows for semantic searching of keywords previously asserted as class names or as related concepts in the global OWL ontology. The detailed assertions can then be retrieved from the RDF data sources linked to it through the local ontologies. Four metadata related to Schools in 'Kedah' are used for this purpose. They are: SGDC Kedah-Kawasan Bangunan Pendidikan 2009 (Polygon) and the SGDC Kedah-Lokasi Bangunan Pendidikan 2009 (Point) under the Built Environment class, SGDC Kedah-Kawasan Rizab Sekolah 2009 (Polygon) under Demarcation class and Area of Interest (AOI) Kedah, which is a map service under Other Information class. These metadata are converted to RDFS and termed as local ontologies. Where end-users search for known keyword or directly search the metadata name, if known, a list of results categorized under the subject name of each metadata is provided as links for further selection as shown in Figure-6.

Figure-6. Metadata Search Results.

 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

A map to view the location of the data is also provided in the interface and Eclipse Java EE IDE for Web Developers version Indigo is used as a programming platform to build the entire search engine. For implementation, the code to build the search engine can be packaged into a war file and deployed easily in other machines. When users click on a particular metadata, they can see the metadata details that are derived from subject and object pairs in the RDF statements of the local ontology representing the particular metadata. A view of the area where the data is located is also provided – Figure-7.

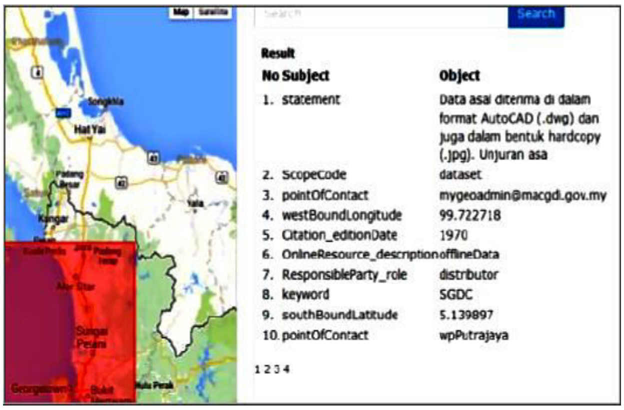

Figure-7. Metadata details and map view.

The location information or map view, is extracted from the east bound longitude, north bound latitude, south bound latitude and west bound longitude elements in the metadata. Google maps API version 3 was used as a base map for the area to be located on map.

5. ANALYSIS AND DISCUSSION OF RESULTS

In order to validate the new semantic search engine to verify that its search returns relevant results while solving the semantic heterogeneity problems associated with keyword searches, we compare the new system with the existing system based on the same data sources. We use two search concepts, as earlier explained in Section 4.3, which are: (i) School vs Sekolah and (ii) Built Environment vs Structure.

For the first concept, search for School and Sekolah in the MyGDI Explorer provides different results as shown in Figure-8 (a) and (b). The search for School provides only one result while the search for Sekolah provides more than one results as shown in Figure-8 (center). This is due to the use of different keyword syntaxes to describe same data by different data providers. However, this issue can be solved using MyGeo-Explorer semantic search, where the search for School or Sekolah provides the same set of results as shown in Figure-8 (c).

Similarly, in the second conflict, where the concept Built Environment is equivalent to the Structure data category, current search results for the Built Environment keyword displays irrelevant results in MyGDI because most of the data providers don't use the exact keyword to categorize their data. As the majority prefer to use the ISO 19115 data category instead, we simply create a semantic annotation using the owl:sameAs construct to map the concepts between the two data categories. Thereby enabling the semantic search engine to realize that the two keywords are simply equivalent and search data containing any of the two keywords is then returned as search result. Similar approach was done for other conflicting concepts and others involve the use of annotations such as labels to denote aliases.

6. CONCLUSIONS AND FUTURE WORKS

From the works presented in this paper, it can be concluded that semantic searches based on domain ontology can provide more relevant search results and can be used to solve semantic heterogeneity issues found in keyword-based searches. In the process of developing the semantic search engine, we show that ontology models can be developed to represent geospatial information based on a domain. The development of the MS1759 domain ontology is something new and based on available literature, not attempted by other researchers.

Figure-8. Comparison of Search Results. **a).** MyGDI Explorer's search results for School. **b).** MyGDI Explorer's search results for Sekola.

ARPN Journal of Engineering and Applied Sciences

Figure-8. c). MyGeo-Explorer's search results for School or Sekolala.

The ontology is available online and can be downloaded from the web-protégé site through the URL . We present the background literature, introducing the concept of semantic search, ontologies and the semantic web and further discuss the methods and procedures employed in the development of the ontology model and the semantic search engine.

Implementation details of MyGeo-Explorer's components are presented followed by evaluation results showing a simple comparison analysis between the existing keyword based search engine and the new semantic search engine. No doubts, the development of semantic search engine for geospatial information can benefit other Spatial Data Infrastructures (SDI) around the world, where semantic search mechanism can be developed for efficient discovery of geospatial information from various data providers and across various language barriers. Moreover, the domain ontology developed based on the Malaysian Geospatial Standard (MS 1759) can be reused by other researchers and the effort can promote the use of MS 1759 as geospatial information standard in Malaysia. However, as highlighted in Section 1, the scopes of the ontology is restricted only to the MS 1759 standards, as such more standards needs to be added to the knowledgebase for generality. Moreover, with regards to the literature and ontology developments, the paper only discusses the role of semantic technologies, especially ontologies, in knowledge management and provides a brief overview on Ontologies. For textbook details on ontology development, we refer interested reader to (Noy F. & Mcguinness, 2000; Steffen S. & Rudi S. 2009).

In the future, we would like to extend the semantic web search with remote RDF data available in SPARQL end points to retrieve geospatial information directly from data provider's database or websites during query answering. We also hope to extend the ontology model with other data categories under the MS 1759 and equip the semantic search engine with additional indexing and retrieval algorithm for more precise search results. For the map view, we would like to integrate the current ontology with other location-based datasets and ontologies such as FAO's geo-ontology to display more accurate data location.

REFERENCES

[1] Apache Jena (2014) https://jena.apache.org/index.html.

[2] Arshad, N. H., & Hanifah, F. A. B. U. (2010). Issues and Challenges in NSDI Implementation. Proceedings of the 9th WSEAS International Conference on System Science and Simulation in Engineering (pp. 65-70). USA.

[3] Bahreini, K., & Elci, A. (2007). Enterprise Semantic Information Search System Based on New Music and Audio Ontology Integrating Existing Ontologies, (Aow), 7–13.

[4] Berkley, C., Bowers, S., Jones, M. B., Madin, J. S., & Schildhauer, M. (2009). Improving Data Discovery for Metadata Repositories through Semantic Search. 2009 International Conference on Complex, Intelligent and Software Intensive Systems, 1152–1159. doi:10.1109/CISIS.2009.122

[5] Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American pp. 29–37 (May 2001)

[6] Bernstein, A., & Klein, M. (2002). Towards High-Precision Service Retrieval The Challenge: High Precision Service Retrieval.

[7] Bodenreider, O. (2008). Biomedical Ontologies in Action: Role in Knowledge Management, Data Integration and Decision Support. Yearbook of Medical Informatics, 67–79.

[8] Cambridge Semantics (2014). Retrieved from: http://www.cambridgesemantics.com/semantic-university/semantic¬search-and-the-semantic-web.

[9] Cao, Y., Klamma, R., Khodaei, M., & Informatik, L. (2009). A Multimedia Service with MPEG-7 Metadata and Context Semantics. In CEUR Workshop Proceedings.

[10] Douglas D. Nebert, Technical Working Group Chair, G. (2004). The SDI Cookbook (p. 171).

[11] Ding, L., Finin, T., Joshi, A., Pan, R., Scott Cost, R., Peng, Y., Reddivari, P., Doshi, V.C., Sachs, J.: Swoogle: A Search and Metadata Engine for the Semantic Web. In: 13th ACM Conference on Information and Knowledge Management, Washington D.C. (2004)

[12] Djuric, D. Gaševic, D., Devedžic V. (2006). The Tao of Modeling Spaces, Journal of Object Technology, Vol. 5, No. 8.

[13] Gennari, J. H., Musen, M. a, Fergerson, R. W., Grosso, W. E., Crubézy, M., Eriksson, H., Tu, S. W. (2003). The evolution of Protégé: an environment for knowledge-based systems development. International Journal of Human-Computer Studies, 58(1), 89–123.

[14] Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications by A Translation Approach to Portable Ontology Specifications, 5(April), 199–220.

[15] J. Heflin and J. Hendler (2000). Searching the Web with SHOE. In Artificial Intelligence for Web Search. Papers from the AAAI Workshop. WS-00-01, pages 35–40. AAAI Press,.

[16] Klien, E., Einspanier, U., Lutz, M., & Hübner, S. (2004). An Architecture for Ontology-Based Discovery and Retrieval of Geographic Information, 179–188.

[17] Lacasta, J. Nogueras-Iso, J., B´ejar, R., P. R. M.-M., & Zarazaga-Soria, F. J. (2007). A Web Ontology Service to facilitate interoperability within a Spatial Data Infrastructure: applicability to discovery, 63(June), 947– 971.

[18] Lawan, A. Rakib, A., Alechina, N., Karunaratne A. (2014). Advancing Underutilized Crops Knowledge Using SWRL-enabled Ontologies - A survey and early experiment. In Proceedings of workshop on Linked Data and Ontology in Practice (JIST-WP 2014), CEUR Workshop Proceedings, ISSN 1613-0073, available from: http://ceur-ws.org/Vol-1312/ldop2014_paper2.pdf

[19] Lei, Y., Uren, V., & Motta, E. (2006). SemSearch : A Search Engine for the Semantic Web.

[20] Mcbride, B. (2001). Jena: Implementing the RDF Model and Syntax Specification 2. Interpreting the RDF Model. Proceedings of the Second International Workshop on the Semantic Web SemWeb2001.

[21] MyGDI: Malaysia, K. Garis Panduan Perkongsian Dan Penyebaran Maklumat Geospatial Melalui Infrastruktur Data Geospatial Negara (MyGDI) (2012).

[22] Noy, N. F., & Mcguinness, D. L. (2000). Ontology Development 101: A Guide to Creating Your First Ontology, 1–25.

[23] Patel-Schneider, P. F., Hayes, P., Horrocks, I.: OWL Web Ontology Language Semantics and Abstract Syntax, W3C Recommendation, World Wide Web Consortium, 10 February, (2004)

[24] Rakib, A., Lawan, A., Walker, S. (2015). An Ontological Approach for Knowledge Modeling and Reasoning over Heterogeneous Crop Data Sources. In: Abraham, A., Muda, A. K., Choo, Y. H. (eds.): Pattern Analysis, Intelligent Security and the Internet of Things. In: Advances in Intelligent Systems and Computing, Volume 355, 2015, pp 35-47

[25] Shanming, W. U., & Jianjing, S. (2008). Ontology-Based Framework for Geospatial Web Services. 2008 International Symposium on Information Science and Engineering, (1), 107–110. doi:10.1109/ISISE.2008.65

[26] Singh, J. (2013). A Comprative Study Between Keyword and Semantic Based Search Engines, 130–134.

[27] Steffen Staab , Rudi Studer, (2009). Handbook on Ontologies (2nd ed.), Springer Publishing Company, Incorporated.

[28] Sudeepathi, G., Anuradha, G., Babu, P.M.S.P. (2012): A survey on Semantic Web Search Engine. International Journal of Computer Science Issues 9, 241–245.

[29] Xu Quan-Lia, Yang Kun, Wang Juna, Peng Shuang-Yunb, Y. J.-H. (2008). A Study on Ontology-Driven Geospatial-Information Retrieval. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B2. Beijing 2008, (50).

[30] Yu, L. (2007). Semantic Web and Semantic Web Services (2007th ed., pp. 1–50). Boca Rota: Chapman & Hall/CRC.