Pertanika J. Sci. & Technol. 21 (1): 217 - 226 (2013)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Multi-Format Information Fusion through Integrated Metadata Using Hybrid Ontology for Disaster Management

Che Mustapha Yusuf, J.[1,3] *, Mohd Su'ud, M.[2], Boursier, P.[3] and Muhammad, A.[1]

[1]*UniKL-Malaysian Institute of Information Technology, Kuala Lumpur, Malaysia*

[2]*UniKL-Malaysia France Institute, Bangi, Selangor, Malaysia*

[3]*Laboratoire L3i Université de La Rochelle, La Rochelle, France*

ABSTRACT

Finding relevant disaster data from a huge metadata overhead often results in frustrating search experiences caused by unclear access points, ambiguous search methods, unsuitable metadata, and long response times. More frequently, semantic relation between the retrieved objects is neglected. This paper presents a system architecture that makes use of ontologies in order to enable semantic metadata descriptions for gathering and integrating multi-format documents in the context of disaster management. After a brief discussion on the challenges of the integration process, the Multi-format Information Retrieval, Integration and Presentation (MIRIP) architecture is presented. A specific approach for ontology development and mapping process is introduced in order to semantically associate user's query and documents metadata. An ontology model approach was designed to follow inspirational and collaborative approaches with top-down to bottom-up implementation. A prototype of the integrated disaster management information system is currently under development, based on the architecture that is presented in this paper.

Keywords: Ontology Engineering, hybrid ontology, multi-format document, metadata integration, disaster management

INTRODUCTION

In managing disasters at all stages, a large amount of information within multiple media documents is produced and collected. The information contained in spatial and non-spatial documents, which were collected before and after disasters, is composed of collections of features, coverage and high-resolution imageries, snap photos, text report, video and audio clips. Even though

Article history:
Received: 31 March 2012
Accepted: 31 August 2012

E-mail addresses:
jawahir@miit.unikl.edu.my (Che Mustapha Yusuf, J.),
mazliham@unikl.edu.my (Mohd Su'ud, M.),
patrice.boursier@univ-lr.fr (Boursier, P.),
muhammad.unikl@gmail.com (Muhammad, A.)
*Corresponding Author

ISSN: 0128-7680

the documents can be in multiple different formats, multiple characteristics and are available from different sources, their contents may tell an equivalent semantic of objects, the calamity story, share the space-time extension, and may be closely related to each other (see Fig.1).

As the description outside the packaging describes information about such product, metadata also describes the context and elementary contents of a document. With metadata, access to information at the first level to get the document of information can be found by searching. However, extracting relevant information from a massive number of available metadata still remains a challenge. Users often get frustrating search experiences caused by unclear access points, ambiguous search methods, unsuitable metadata, and long response times (Larson *et al*., 2006), especially when metadata is used in various forms to describe different document formats such as spatial and non-spatial metadata (e.g. multimedia metadata). Hence, understanding the semantics of metadata is a good way to combine different sources of information while ensuring effective integration and access to information. Current research into semantic metadata integration still lacks of focus to combine between spatial and non-spatial metadata which hold descriptions of both document types. Most of the works consider only a single format type or a single context type (e.g. texts, images, spatial) as in Gagliardi *et al*. (2005), Hurtienne *et al*. (2008) and Olteanu *et al*. (2008), when implementing schema and/or instance matching to align terminology between different metadata and user's query terms. Such effort is insufficient to solve matching problem within the environment with high structural and semantic heterogeneity.

In the presented context, there is a need to establish and combine users' and metadata conceptualization, as well as to provide machine automatism to semantic metadata integration.

For this purpose, this paper presents a system architecture using ontology to enable semantic metadata descriptions to gather and integrate multi-format documents. An approach for semantic metadata integration using ontology modelling of the available resources is currently being studied in the context management of national disaster and relief in Malaysia.

Challenges to Information Integration: Malaysian Disaster Management

Malaysia has a good mechanism in managing disasters through Malaysian public agencies, particularly amongst the local authorities, police, fire brigades, and medical agencies. The committee is established at Federal-level, State-level and District-level, under the administrative

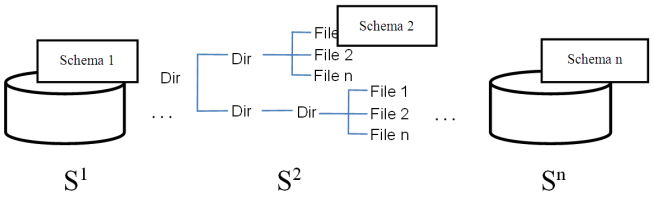

Fig.1: Different document formats that are closely related to each other

control of the National Security Division Secretariat to coordinate all the activities related to disaster. This mechanism is well stated in "Policy and Mechanism in the Management of National Disaster and Relief in Malaysia" or "National Security Council (NSC) Directive No. 20" (Umar, 2011). These agencies perform their own daily work routines and maintain their own information, either manually or in digitized form (e.g., file-systems and Geographic Information System (GIS) and/or non-GIS databases).

During disaster events, a vast amount of information is acquired from different sources to be disseminated amongst them. Various challenges emerge to enforce knowledge sharing, which makes the information integration more difficult. The required datasets are not only difficult to obtain from the system network but also lack automated data coordination at operational level such as during counter-disaster, rescue and relief activities. The major challenge is in the system, structural/schematic and syntactic differences. A diverse distributed storage system is used to store the information that ranges from databases and file-systems. For example, collection of images is stored as Binary Large Object (BLOb) at data provider 1, but as a file-system at data provider 2. The differences in software platform, file formatting and data models certainly add the challenges to interoperability. Furthermore, some data providers provide metadata database to manage data about data they have but some are not. Obviously, the current environment has no composition and consumption of metadata. In order for documents to be identifiable, the metadata should be produced and stored in a format that allows its efficient management. Another challenge arises if metadata system is utilised. Each agency may use different terminologies to refer to similar data, and also different document formats to store spatially and semantically related information.

It is important to note that semantic integration to group and combine data (metadata) from different sources of various agencies involved in a disaster management is necessary. For this purposed, semantic integration has to ensure that only data related to the same real-world entity are merged. Ontology is the current best practice to resolve semantic conflicts in these diverse information sources. Gruber (2007) states that ontology is an enabling technology (a layer of the enabling infrastructure) to enforce knowledge sharing and manipulation. The authors strongly believe that an appropriate ontology development approach should depend on the current organisational environment of Malaysian management of disasters. Majority of the current systems and metadata standards hold less explicit semantics of information (Fig.2), and this makes data fusion tasks difficult (Halevy *et al*., 2006; Haslhofer & Klas, 2010; Haas, 2007). Modern information system is encouraged to embed more semantics in their systems so as to allow a better information integration and this can be achieved by using ontology. This research opens up significant opportunities to achieve more flexible and adaptable ways to start employing ontology within disaster management agencies.

A System Architecture for Multi-format Information Fusion through Metadata Integration Using Hybrid Ontology

The advent of Semantic Web technologies and ontology engineering facilitates the idea to enable semantic metadata integration within various metadata sources. For this purpose, an ontology-based architecture for information retrieval, integration and presentation is formed. The

designed system aims at providing users with data input, discovery and access to multiple media documents via rich ontology-based metadata describing them. The architecture comprehends the semantic matchmaking between user's query and metadata of multi-format information using ontologies. This system architecture is called MIRIP (Multi-format Information Retrieval, Integration and Presentation), conformant to Service Oriented Architecture (SOA) standards (Sprott & Wilkes, 2011). A conceptual notion of the system is depicted in Fig.4. There are three sections involved in the overall process model, namely, provider, mediator and client sections. Details pertaining to the processing components in the architecture are as Fig.2.

Provider section. This section comprises distributed data repositories, which are administered by different data owners. The data repository can be in the type of file-systems and special purpose databases (administration, GIS and multimedia). Information is stored as a set of files containing multiple media types such as GIS vector files, high-resolution aerial photographs and satellite images, snap photo images, audio-video clips and text documents. Metadata exist for managing the various media formats. To provide the semantic descriptions of each metadata, local (source) ontology is constructed according to bottom-up ontology principle. Concurrently, the local ontology is aligned with the upper-ontologies representing the common concepts and its relations.

Mediator section. This section provides a searchable repository of data service descriptions, thus enabling data providers to publish their data and data requestors to search for these data. Semantic metadata-base permits the storage of metadata from various sources. Ontology component provides a common vocabulary to define the relationships between object classes once new metadata are (automatically) registered to the system. The ontology facilitates the retrieval system to identify the relevant media information through metadata, which are semantically annotated and matched to the terms specified in the query.

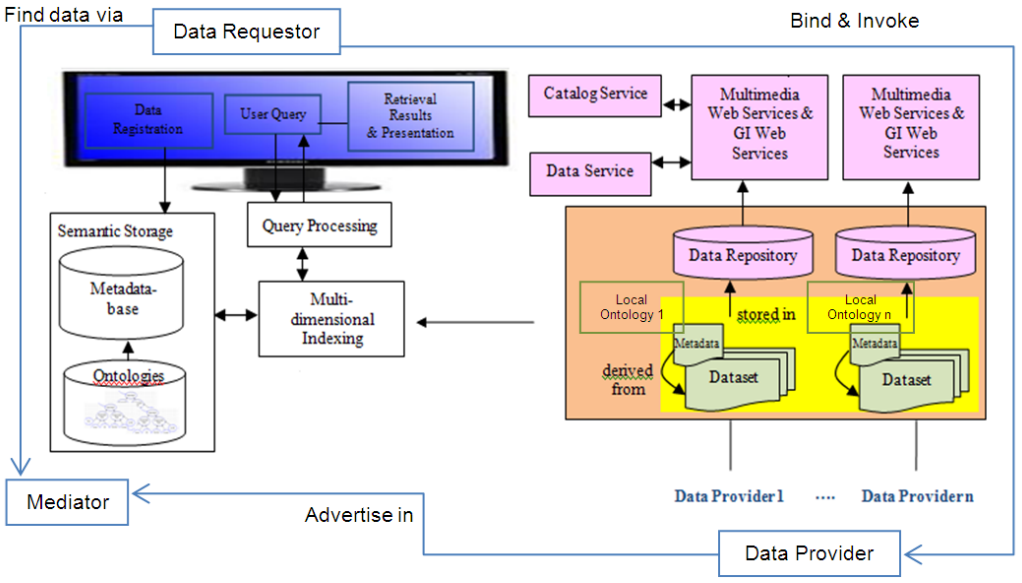

Fig. 2: The MIRIP (Multi-format Information Retrieval, Integration and Presentation) architecture

Client section. This section enables human-machine interaction at registering data services, querying, navigating and accessing the information. By using the *data registration* interface (dRI), data providers can register their available metadata (push) to the catalogue service as data sources. The metadata catalogue is useful to facilitate data requestors to locate and evaluate the data services before it automatically makes a request to access the data. Once data services registration is made, service description is maintained by the mediator. *The user query interface* allows users to provide input to the query to automatically search the required information service. When the request method is posted, a user's request will be sent to the mediator to search for matching values. If a data provider is found, the connection will be bonded between the data provider and the data requestor. *The retrieval results and presentation interface* allow search results from metadata catalogue to be presented in a ranked-list. These search results are delivered to the user through a combination of graphical and textual (e.g. descriptions) elements.

Ontology Modelling for Semantic Integration

In this work, ontology development takes advantages of the hybrid ontology with the implementation of top-down and bottom-up ontology designs (see Fig.4). The upper-ontology used in the hybrid ontology approach transfers the burden of information correlation and filters the query processing system (Mena *et al*., 2000). Following up the top-down design, the set of top-level ontology is firstly provided. Common terms are specified at a very abstract extent, so that a new source ontology can be easily mapped to the upper-ontology. If the new source contains a local concept that is not described in upper-ontology, the common concept that matches with the local concept will be established in the upper shared-vocabulary. Secondly, the source ontologies that contain more specific terms are extended from the primitive terms in the set of top-level ontology. Terms at both levels are comparable easily because the source ontologies only use the vocabulary of top-level ontology. Based on the bottom-up ontology design, the existing source schema and its instances are extracted to generate the source ontologies which contain more high level data descriptions. Then, the source ontologies of all disparate sources are mapped to the abstract concepts of top-level ontology which has been constructed earlier.

In developing the ontology for the current environment and to enable bottom-up ontology mapping, each source must have at least one common concept. Some uncommon concepts that are considered important for query purpose will be declared as sharable in global ontology to avoid data loss. For instance, Source 1 holds concept 'channels' that describe the number of channels represented in the waveform data, such as 1 for mono or 2 for stereo. The concept, however, is not common to another source. Thus, the concept along with its possible sub-concepts will be added in the shared vocabulary. The participating data sources in the integration process have no pre-existing ontologies. Thus, local ontology for each data source will be created with reference to shared-vocabulary. The body of the local ontology is extended to list more specific entities and properties. With no pre-existence of ontology, data sources still have the autonomy to maintain its own name concepts.

The shared-ontologies (vocabularies) include top-level ontology to describe the primitive concepts and domain-specific disaster management ontology. At domain-specific level,

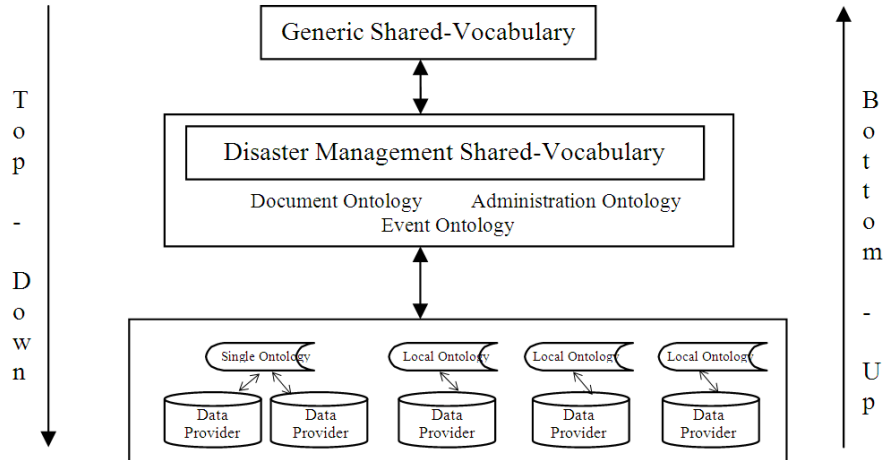

Fig. 3: Ontology model for disaster management domain

document ontology along with administration ontology that is presented by Xu and Zlatanova (2007) are generated to represent the concepts associated with the documents dataset and the providers. The top-level ontology contains common concepts that are associated with space, time and theme. The domain-specific ontology is created to capture spatial and non-spatial concepts related to disaster management. A specific case study of landslide disaster in Malaysia is drawn to demonstrate the applicability of the proposed ontology. So, the ontology also captures details of the landslide concepts which are described as the sub-concepts of natural disaster in the event ontology.

Ontology building methodology, proposed by Uschold and King (1995) and Uschold and Gruniger (1996), is used as basis steps to build the ontology components. The following steps are adhered to properly design the ontologies under provision; 1) Identification of ontology purposes & scopes, 2) Ontology conceptualization, semi-formal specification and formalization, 3) Ontology evaluation and 4) Ontology documentation. The sets of ontology that are still at the development stage play a key role to semantically associate user's query and the document metadata. Consequently, the ontology components here support the classification of resources and retrieval to the resources. Semantic Web ontology languages such as Resource Description Framework Schema-RDFS (Antoniou & Harmelen, 2008) and Web Ontology Language-OWL (W3C, 2009) are utilized along with the Protégé and the Jena API (Protégé, 2011) as an automatic development tool in this work.

Ontology Mapping Methodology

In the modelling ontology, each class and property is assigned with *primary identifier* as in Parts LIBrary (PLIB) ontology (Pierra, 2004) to map between concepts. The *primary identifier* is used to indicate the similarity or different concepts between participating data sources and it upper-vocabularies. Fig.5 depicts the top-down to bottom-up mapping implementation with the use of *primary identifier*. An example of the text document concepts is presented. In the , local

ontology is defined based on the schema of the local data. Data owners will decide their own definition of the local ontology concepts. The concepts that are rational to be disclosed will be pulled out to domain-shared list. Meanwhile, concealed concepts (shaded in Fig.5) will not be shared but can be accessed locally or may be shared (right away or later) in a different domain.

Generic and domain shared-vocabulary are the list of shared concepts for all participating data sources. In this approach, the design of shared-vocabulary begins with inspirational approach (Holsapple & Joshi, 2002). At the initial stage, the specification of generic and domain shared-vocabulary, that are substantially potential to be shared with the group of the data owners, are initiated. Concerned with the importance of information sharing, the data owners may collaboratively (Holsapple & Joshi, 2002) use the existing shared-vocabulary as the anchor and supportively extend it if necessary. However, the data source owners will not be attentive to each other's data. This is important for most of the intelligence systems that are confidentiality-related.

MIRIP PROTOTYPICAL IMPLEMENTATION

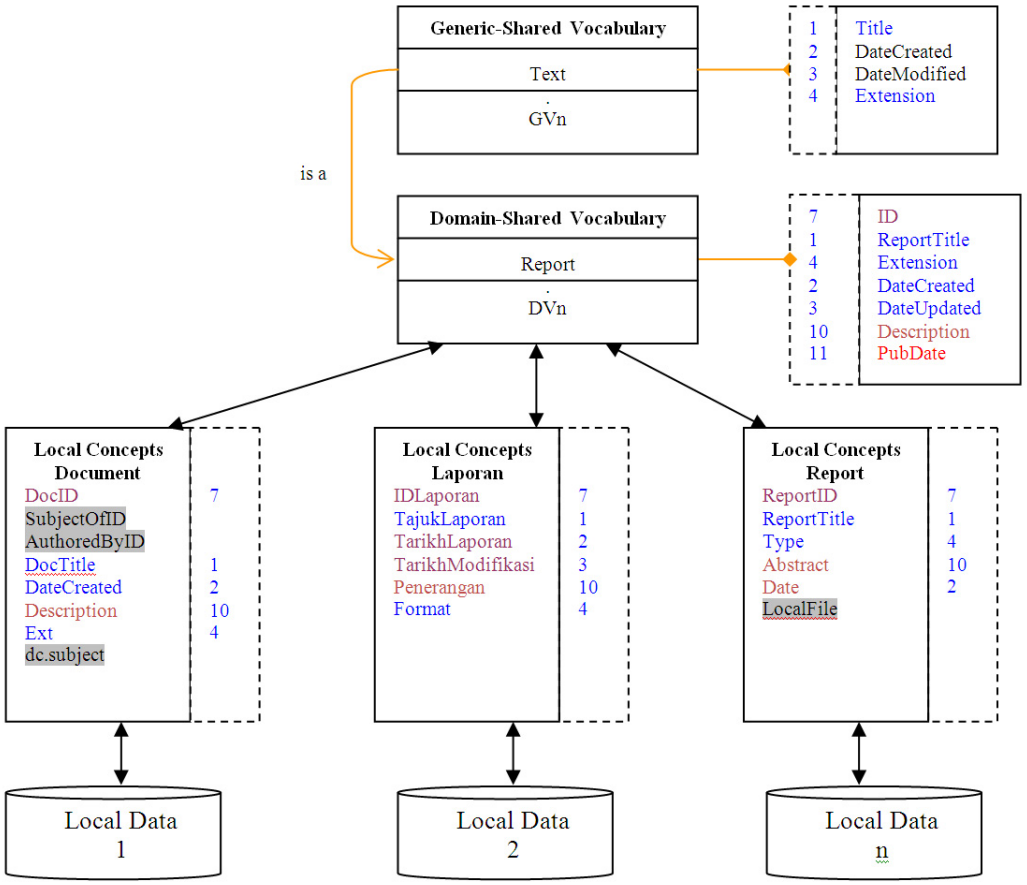

Fig. 4: Top -down to bottom-up ontology mapping

An implementation plan to demonstrate the interaction (data services registration and discovery) within MIRIP is illustrated in disaster management domain. In MIRIP, data providers are from different mapping agencies and institutions involved in disaster management. They present a collection of data services while mediator is employed with a set of data registry services, and client (data requestor) is equipped with a set of client applications.

Data services registration involves data listing process in catalogue services. Data providers can use the MIRIP dRI not only for submitting their new data (metadata), but they can take out data from the catalogue and update some specific data. The following case shows how the data service registration works. Let's assume that the Public Work Department of Malaysia (PWD) is cooperatively giving out a landslide investigation report (in MIRIP geo-portal). Thus, PWD must use the concepts enumerated in ontologies to generate the data service descriptions. With the help of the dRI to access the relevant concepts from ontologies, PWD will perform multiple steps to select an appropriate disaster event (in this case, landslide). In the proposed event ontology shown in Fig.6, 'landslide' is enumerated as a sub-concept of geological natural hazard. After the landslide concept has been chosen, PWD is navigated to stipulate an important attributes (i.e. format, reference date, abstract etc.) for the published data. In the case of publishing spatial dataset, more detail spatial representation properties are required. Once the new metadata registration is submitted, it is stored in the metadata-base (enriched with ontology descriptions).

Data services discovery always concerns with the identification of service descriptions that match a data service requested by the client. From the previous case, if PWD registers multiple datasets, PWD has provided a service description that defines an important metadata which entirely describes the characteristics of services that are deployed on MIRIP geo-portal. This metadata provides an abstract definition of the information that is essential to deploy and interact with a service. Suppose that Malaysian Public Works Institute (Ikram) is about to request datasets about specific landslide resources. Analogous to registration step, Ikram will perform the request (via query interface with catalogue service) by specifying the properties such as the event type, location, date, etc. Since the information about landslide has been

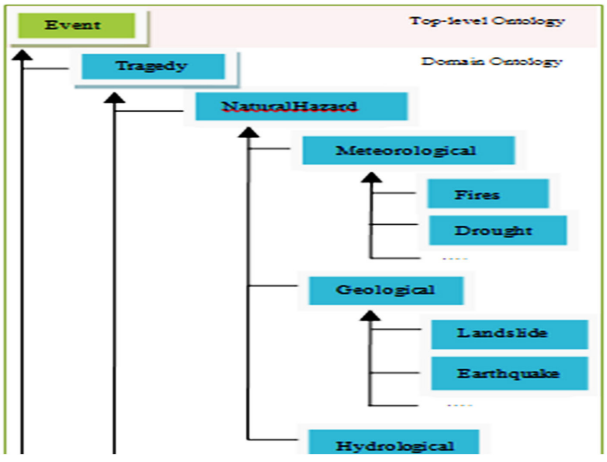

Fig. 5: Snippet of class event and its membership

registered and classified under a specific category and aligned with the ontologies, the query is foreseen to return the relevant result list and ready to be browsed for further discovery (i.e. reiterate search, display and download).

OUTLOOK

This specific system was designed to increase the efficiency in document query and integration, particularly for disaster management domain. The presented ontology-based architecture for a semantic integration of documents via metadata approach was currently developed. Validation upon the designated ontology would be attained with the professional members involved in Malaysian disaster management, particularly with personnel involved with data acquisition and risk analysis. The recent research is still in focus to scrutinize the ontology formation as well as the complex ontology matchmaking process highlighted in this paper to produce ideal mapping between upper and low level ontology. The approach using ontology is foreseeable to help achieve the goal of automatic data search and integration to response a specific query.

REFERENCES

Antoniou, G., & Harmelen, F. V. (2008). *A semantic web primer* (2nd Edition). Cambridge, MA: The MIT Press.

Gagliardi, H., Haemmerlé, O., Pernelle, N., & Saïs, F. (2005). An automatic ontology-based approach to enrich tables semantically. *1st International Workshop on Context and Ontologies: Theory, Practice and Applications*. Pittsburgh, Pennsylvania, pp. 64–71.

Gruber, T. (2007). Ontology of folksonomy: a mash-up of apples and oranges. *International Journal on Semantic Web and Information Systems (IJSWIS), 3*(1), pp. 1-11.

Haas, L. M. (2007). Beauty and the beast: the theory and practice of information integration. *International Conference on Database Theory*, pp. 28-43.

Halevy, A., Rajaraman, A., & Ordille, J. (2006). Data integration: the teenage years. 32nd International Conference on Very Large Databases, September 12-15, Seoul, Korea.

Haslhofer, B., & Klas, W. (2010) A survey of techniques for achieving metadata interoperability. *ACM Computing Surveys (CSUR), 42*(2), 1-37.

Holsapple, C. W., & Joshi, K. D. (2002). A collaborative approach to ontology design. *Communications of the ACM, 45*(2), 42-47.

Hurtienne, J., Weber, K., & Blessing, L. (2008). Prior experience and intuitive use: image schemas in user centred design. *Designing Inclusive Futures*, 107–116.

Larson, J., Olmos Siliceo, M. A., Pereira dos Santos Silva, M., Klien, E., & Schade, S. (2006). *Are geospatial catalogues reaching their goals?* Paper presented at the 9th Agile Conference on Geographical Information Science, Visegrád, Hungary.

Mena, E., Illarramendi, A., Kashyap, V., & Sheth, A. P. (2000). OBSERVER: An approach for query processing in global information systems based on interoperation across pre-existing ontologies. *Distributed and Parallel Databases, 8*(2), 223-271.

Pierra, G. (2004). *The PLIB ontology-based approach to data integration*. Paper presented at the18th IFIP World Computer Congress (WCC'2004), Toulouse, France.

Protégé-Stanford Center for Biomedical Informatics Research. (2011). *Integration of jena in protégé-owl*. Retrieved on June 29, 2011 from http://protege.stanford.edu/plugins/owl/jena-integration.html.

Sprott, D., & Wilkes, L. (2011). *Understanding service-oriented architecture*. Retrieved on August 6, 2011 from msdn.microdsoft.com/en-us/library/ aa480021.aspx.

Olteanu, A., Mustière, S., & Ruas, A. (2008). *Matching imperfect spatial data*. Paper presented at the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.

Umar, C. M. (2011). *Policy and mechanism on national disaster and relief management*. Retrieved on July 23, 2003 from http://spm.um.edu.my/news/disastermanagement 23072008/ talk02.pdf.

Uschold, M., & King, M. (1995). *Towards a methodology for building ontologies*. Paper presented at the Workshop on Basic Ontological Issues in Knowledge Sharing. Montreal, Canada, pp. 1-13.

Uschold, M., & Gruninger, M. (1996). Ontologies: principles, methods and applications. *Knowledge Engineering Review*, *11*, 93-136.

W3C OWL Working Group (2009). *OWL 2 web ontology language document overview. W3C Recommendation*. Retrieved on July 15, 2011 from http://www.w3.org/TR/ 2009/REC-owl2-overview-20091027/.

Xu, W., & Zlatanova, S. (2007). Ontologies for disaster management response. *Geomatics Solutions for Disaster Management*, 185-200.