Domain Models Schemas: A Semantic Web Perspective

THABET SLIMANI

Computer Science Department, Taif University ,KSA.

ABSTRACT

In Semantic web, domain model is an abstract image of a small part of the world. It serves to capture the common understanding of the domain to create a basis for clear communication. This paper performs a study clarifying the different domain modeling types to help the individual interested by this topic to gain new insights and gudlines. In this paper, after a general introduction about the basis of ontologies and its components, a description of the most common domain modeling schema and a comparison between them has given.

Key words: Ontologies, Ontology classes, Ontology families, Ontology languages, Ontology modularization, Semantic web.

INTRODUCTION

Domain Modeling is a difficult intellectual effort which requires thoughtful understanding of the domain modeled in several manners. Three widespread domain modeling schemes are adopted in several applications which are: taxonomy, thesaurus and ontology. We have compared these three schemes among others because they are of great interest to practitioners. Generally there is confusion as to how these schemes differ from each other and what they are best used for. A taxonomy is defined as a hierarchical structure (tree) modeling a domain from abstract to specific. A thesaurus is a controlled vocabulary defining each term by three types of relationships: hierarchical, associative and equivalent. An ontology is the most formal model which defines an individual by the meaning of concepts (modeling constraints that restrict the number of possible interpretations). The following section presents two different definitions of ontologies:

- In computer science, an ontology is an attempt to make a complete and rigorous conceptual schema within a specified domain. Typically, an ontology is a hierarchical data structure including all the significant elements and their relationships and regulations within the domain [1].
- In AI field, an ontology is an explicit specification of a conceptualization [2, 3]. The universe of discourse of an ontology, is the concept names (e.g. classes, relations, axioms) accompanied with a description of what the concepts mean, and their formal axioms.

Several types of ontologies have been enumerated in academia. Depending on context, the word "ontology" can designate different computer science objects. For example, an ontology has a distinguished naming which depends on the domain:

- in the field information retrieval is named thesaurus
- in the field of linked data is a model represented by OWL format
- in the field of databases is an XML schema
- etc.

These described three types of models presented above, among other models, differ principally in their degree of precision. If the model is more precise, it offers more features and the effort goes into making it is more extensive.

Although the use of ontologies suggests a common questions amongst information architects, domain modelers and practitioners. The main questions that need answers are as follows:

- Which are the different components of a domain modeling?
- Which problems do these features and what exactly is gained with increasing precision?
- These domain models are especially suited for which problems?

However, to perform a study clarifying the different domain modeling types to help the individual whom interested in it to choose the appropriate domain modeling is increasingly important. Additionally, it is important to define precisely the specificity and characteristics of each domain modeling.

The rest of this paper is organized as follows: In the next section, we describe the fundamental concepts of ontologies, their definitions and the principle of domain modeling schema. Then ontology components are presented. In the following section, some domain modeling schemas most recognized in Semantic Web and their degree of formality are described. After that, the common

languages and ontologies for semantic web modeling are explained. And finally, this paper is enclosed with a conclusion.

Fundamental Concepts of Ontologies and Domain Modeling Shema

An ontology is an explicit conceptualization of a domain which enables their comparison and analyzes. Gruber [2] defines an *ontology* as an explicit specification of a conceptualization. It can be viewed as a description of knowledge-level [4] where the representational formalism is independent [5]. Additionally, ontology is a representation of the entity's type, their relations, and their constraints [4]. It consists of a set of classes, relations, instances, functions and axioms ordered hierarchically. Formally, an ontology is a description of data that remains constant over various data/knowledge bases in a certain domain [6].

The ontologies categorization based on their generality level is presented in the classification described by Guarino in [7]. Ontologies can be classified according to the conceptualization subject (content). Very general things such as time, space, insubstantial or concrete objects, and so on can be covered by the *top-level Ontologies*, independent to the domain usage. The construction of either *domain* or *task* Ontologies can be done based on these top-level Ontologies. The first category includes Ontologies dedicated to covering a given domain (medical or university, for example) independently to the task that uses the ontology. The second category includes ontologies specified for a generic mission (content annotation or situation recognition, for example) irrelevant of usage domain. In conclusion, the development of application Ontologies helps particular tasks to be solved within particular domains, and consequently often requires both domain and task Ontologies for reusability.

An ontology may be classified as follows, based on the scope of the ontology:

- Upper/top-level ontology: it describes general knowledge (i.e. What time is and what space is).
- Domain ontology: it describes the domain (medical domain, personal computer domain

or electrical engineering domain).
- Task ontology: it is suitable for a specific task (assembling parts together).
- Application ontology: it is developed for a specific application (assembling personal computers).

For instance, upper ontology could includes modules for real numbers, time, and space (parts of an upper ontology, generally are called generic ontologies). Upper level ontologies could be imported by ontologies at lower levels and adding them specific knowledge. Domain and task ontologies may be independent and are combined to be used for application ontology. As an example, the FOAF ontology defines classes and properties to describe people.

A domain model is defined as an abstract representation of a small part of the world (special case of ontologies). The domain model components include concepts, relationships between these concepts, and properties of the concepts and relationships. The definition of a concept in the context of other concepts is done by relationships. The specification of the characteristics of a concept is realized by properties. By modeling a domain, the knowledge about it is captured and the assumptions on which the domain is built are made explicit [8]. However, a domain model is adopted for a common understanding of the domain capturing with the aim to create a basis for unambiguous communication. Each individual in the world has a unique personal conceptual model. It is difficult to model a domain, because the individual conceptual models of
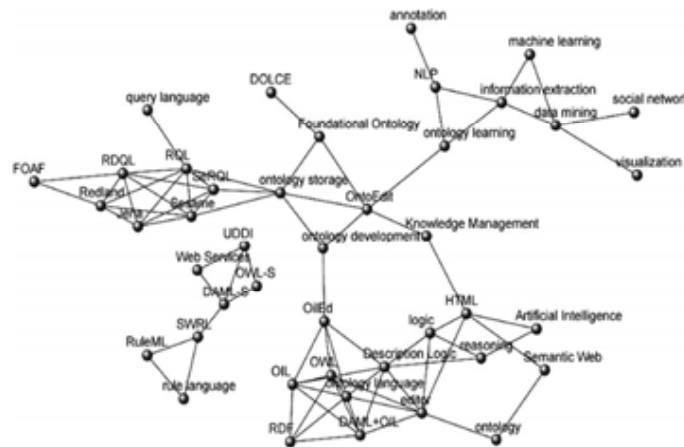

Fig. 1: Example of ontology of research topics (An automatically generated ontology on the research area Semantic Web on the website of Flink)
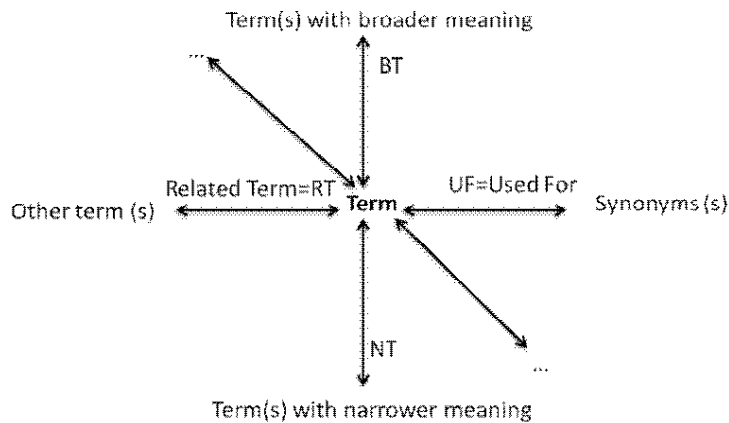

Fig. 2: Example of Thesauri related to information searching

people need to be extracted in a first time, and then reconciled within a single model in a second time.

Ontology Components

Ontology components can be represented by specific ontologies. For example, if we focus on concepts, as a key component of ontologies, they can be represented based on different behaviors:

- Textual Definitions: As instance, it defines concept by a sentence ("University", "Person"….)
- Properties set: As instance the concept "University" has the property "name", "address" and "creation date".
- Logical definition: Constituted by several formulas. For example, in Quran ontology,

the Earth, Sun, and the Moon are classified under "*Astronomical Body*":

- Instance Set: Constitutes a set of instance that belongs to a concept. For example, "firdous paradise" is an instance of "afterlife-location" concept of Quran ontology.

Ontology components such as concepts (things, events…), instances and properties are represented by one or more symbols denoting terms rapidly understood and readable by humans. The connection between all these ontology components is realized through relations. Semantic relation is a connection that link only concepts together: for instance the location relationship indicates that student concept is enrolled in a university concept. The link that connects only instances is denoted

Table. 1: Comparison between ontologies, taxonomies and thesauris

	Ontology	Thesauri	Taxonomy
Components	Classes, properties, formal logical rules, class instances.	Concepts and terms, concept labels, concept translations, concepts/terms synonyms and descriptions / definitions.	Hierarchically organized categories.
Uses	Is used to instantiate and reason	Used to index content or resources with keywords and search (with the same keywords)	Used to classify, to store content or resources
Degree of Logical formalism	Very formal	Few formal	Medium formal (more formal than a thesaurus , because the hierarchy of inputs can follow certain constraints.)
Proximity level with natural language	Far from natural language (uses techniques IDs to abstract natural language)	Close to natural language (gives linguistic equivalents of each entry, translations in other languages)	Not related to natural language
Types of used relationships	Inclusion, (class / subclass), union, intersection, exclusion, domain arrival set, transitivity, inverse properties, etc.	Hierarchical and associative; alignment relationships (eventually)	Hierarchical only
Representation of the data on standard web	OWL	SKOS	SKOS

by instance relations which are in turn instances of semantic relations. It is difficult to generalize the relation between all instances of their concept, because some relations between instances are contextual. As an example of instance relation the student instance named "John" is enrolled in the university instance named "Stanford". An example of a contextual instance relation can be that the professor instance named "Jeffry" is localized in the university instance named "Stanford". The relationship that terms can have is expressed by the terminological relations. For example the term "university" is synonym to the term "education".

According the type of structure and the amount of their use, Sowa [8] distinguishes two main categories: terminological (lexical) and axiomatized (formal).

Domain modeling schemes and Their Degree of formality

In the context of Semantic Web, ontologies can be distinguished according to their degree of formality. The degree of formality of an ontology determines its degree of axiomatization by means of logical statements about the domain. Only a few or without axioms constraining the use of the entities in their signature are included in lightweight ontologies. Conversely, heavyweight ontologies are characterized by wide axiomatization for interrelating the signature elements in a sophisticated manner.

Concept Schemas

Concept schemes often evolve, in a similar manner to the informal semantic networks of interlinked conceptual nodes, as a result of shared tagging activities in a Web context. Tag taxonomies [9] and informal hierarchies are examples modeled in SKOS. Their low expressivity and very limited possibilities for an axiomatization makes them well-suited to the fairly uncontrolled environments within a larger community of uncoordinated knowledge contributors.

Thesaurus

The thesaurus is the least formal form of an ontology. According to lexical criteria, a thesaurus provides the possibility to organize the words used in a certain domain. Language-specific dictionaries that also encode information about synonyms or a classification of medical terms for diseases are examples of a thesaurus. The Thesauri expressivity is low in terms of logic-based knowledge representation and is restricted to lexical relations between words, such as synonymity or homonymity (Example of wordnet). Figure
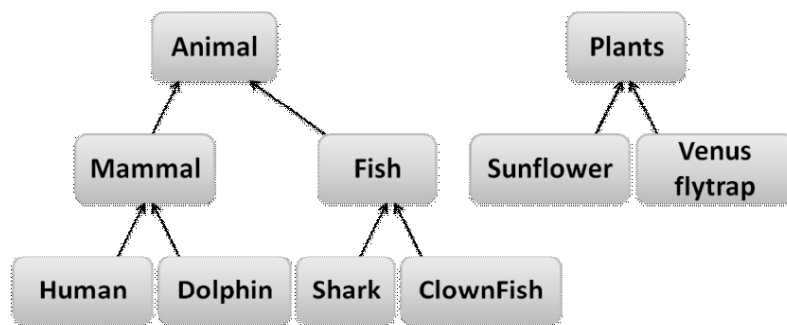

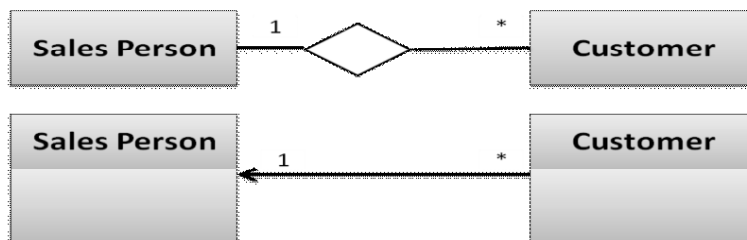**Fig. 3: Example using a simple biological taxonomy**

Fig. 4: A Conceptual data model that uses Entity-Relationship and UML

1 illustrates an example of a thesaurus related to information searching. Several examples of Thesaurus using SKOS and/or RDF are described in the following section:

AGROVOC Thesauri

AGROVOC is a multilingual agricultural thesaurus designed to cover the terminology of all subject fields in agriculture and several other environmental domains (forestry, fisheries, environmental quality, pollution, etc.). The AGROVOC thesaurus is used to develop the Agricultural Ontology Service (AOS). As presented in (AGROVOC), it consists of descriptors (indexing terms consisting of one or more words) and numerous (10,758 in English) non-descriptors (synonyms or terms helping the user to find the appropriate descriptor(s)), available in different languages and controlled by relationships, used to identify or search resources. Fig.4 schematizes an example of the theme list presented by AGROVOC.

HEREIN Thesauri

The HEREIN project is a European Heritage Network information system which focuses on cultural heritage, especially on architectural and on archaeological heritage. HEREIN is intended to gather governmental services in charge of heritage protection within the Europe Council. The multilingual thesaurus related to the HEREIN project aims to offer a terminological standard for national policies that deals with architectural and archaeological heritage (for more clarification, see http://www.european-heritage.net/sdx/herein/index. xsp).

GEMET Thesauri

short for General Multilingual Environmental Thesaurus and developed as an indexing, retrieval and control tool for the EEA. GEMET is the reference vocabulary of the European Environment Agency (EEA) and its Network (Eionet). Additionally, GEMET was designed as a "general" thesaurus, aiming to define a common general language, a foundation of environment general terminology. Several languages are used in GEMET (for more clarification, see http://www.eionet.europa.eu/ gemet).

URBAMET Thesauri

URBAMET is a bibliographic data bank designed for the French library. It covers thematic fields on urban development, housing and accommodation, town planning, public facilities, architecture, transport, local authorities etc. This data bank is created in 1986, and since this date the hierarchical organization of all these topics gave place to the construction of URBAMET thesaurus. The thesaurus is available in French, Spanish and English. The URBAMET application allows the search, display and circulation of these topics through the different tables. Regularly updated and revamped in 2001, today this thesaurus contains 4250 descriptors divided into 24 tables. Each descriptor belongs to a single semantic table and its acceptation arises from the position which he has been assigned in this organization.

Taxonomies

The notion of class hierarchies based on the subsumption concept is the base of the taxonomies. Taxonomies are frequently used for a formalized hierarchical organization of domain knowledge. A catalog of product categories that constructed based on a strict subsumption hierarchy of product classes is an example of taxonomy. The main characteristic of taxonomies is their strict hierarchical category of classes. Consequently, the subsumption relationship formalization is typically realized logically. The example in Figure2 presents a simple biological taxonomy.

A taxonomy should be visualized as a tree, because there is no common approach for domain modeling as a taxonomy, nor is there a commonly understood definition of taxonomy. We illustrate the domain modeling constructs of such a structure with the example taxonomies. As instance, Wikipedia categories are used to organize Wikipedia entries.

Conceptual Data Models

In Computer Science, the Conceptual Data Model (CDM) is the most abstract form of data model. A CDM is often adapted toward designing information systems or database management systems. Entity relationship diagrams or UML diagrams used for domain modeling are examples of CDM. A CDM includes the expressivity for

structuring a domain for the data employed in a software system by means of concept subsumption hierarchies and domain class properties and attributes. CDM is typically used for checking constraints on the conceptual model to detect faulty data situations, if any logical formalization occurs. The example of Figure 3 presents a CDM that is rendered using two of the notations supported by Enterprise Architect.

Rule and Fact Stores

In some applications where the source of data is knowledge bases, rule or fact bases serve to handle a large numbers of individuals with various basic reasoning. The description logic A-Boxes and RDF(S) graphs for facts querying with simple reasoning over class and property hierarchies can be examples. Additional examples were logic programming rule bases that derivates, instantiates and asserts axioms. Typically, these ontologies have the expressivity for instances interrelating and typing and for a rule-based facts derivation based on the logic programming mechanisms.

Abstract Logical Theories

The ontology having most formality and expressivity is an ontology that uses general logical theory, where the represented domain has a high degree of axiomatization and expressed with a logic-based knowledge representation formalism (first-order predicate logic or even higher-order or modal logics). The formal specification of an upper-level ontology with a wealthy axiomatization for very general notions in modal logic axioms form is an example of general logical theories. Additionally, a general logical theory captures a rich axiomatization about classes and properties which allows the illustration of the conclusions related to general situations in the domain under the format of complex axioms.

Comparison between some domain modeling schema

This section provides a comparison between three languages of knowledge representation (ontologies, thesauri and taxonomies). The ontology is useful to describe the world as it is; the thesaurus is adopted to facilitate access to content; the taxonomy is constructed to classify resources in folders and categories. The content, data or

knowledge access systems combine and articulate these three organizational systems to describe the world, index, and categorize content. The comparison illustrated in Table 1 take into accounts the properties of domain modeling components, degree of logical formalism, the proximity level with natural languages, the types of relationships used and an example of data representation.

Common Languages for Semantic Web Domain Modeling

Different languages can be used for semantic web domain modeling as follows:

RDF: stands for Resource Description Framework [10], was developed by the W3C to describe Web resources and allows the specification of the semantics of data based on XML in a homogeneous, interoperable manner. It also provides mechanisms to clearly represent services, processes, and business models, while allowing recognition of information not clear. RDF has a number of features, including:

- Containers and Collections to group Resources
- Reification to describe RDF statements themselves
- Structured values to support describing relations between more than two Resources
- Defining SubProperties of Properties

RDFS

stands for RDF Schema [11] and was built by the W3C as an extension to RDF with frame-based primitives. RDF(S) is obtained by the combination of both RDF and RDF Schema. RDF(S) just allows the representation of concepts, taxonomies of concepts and binary relations for that reason it is not very expressive. Three additional languages have been developed as extensions to RDF(S) and described in the following section (OIL, DAML OIL and OWL).

XML

it is a W3C recommendation stands for EXtensible Markup Language [12], was built in 1996 much like HTML and designed to describe data and not to display data. As an effect, XML has been

used to modify SHOE syntax and subsequently, additional ontology languages were built on the XML syntax.

Dublin Core

The Dublin Core Schema is a small set of vocabulary terms that can be employed for web resources description (video, images, web pages, etc.), as well as physical resources such as books or CDs, and objects like artworks [13]. Dublin Core Metadata may be used for several purposes, from simple resource description, to combining metadata vocabularies of different metadata standards.

OWL

OWL short for Web Ontology Language is a recommendation of W3C designed to be used by content information processing applications, instead of information presenting to human. OWL makes easy the interpretability of Web content by machines than that supported by RDF, XML, and RDF Schema (RDF-S) by offering additional vocabulary providing along with a formal semantics. OWL includes three increasingly-expressive sublanguages:

OWL Lite

is a simple sub-language of Owl intended for quick reasoning and programming simplicity that supports those users that need a classification hierarchy and simple constraint features. The expressiveness of OWL Lite is limited to classification hierarchy and a simple constraints functionalities 0 or 1 (functional relations) providing a quick migration path for thesauri and other taxonomies. For instance, a person has a single address, but may have one or more names; OWL Lite does not allow its representation.

OWL DL

DL short fort for description logic. OWL DL has an increased expressiveness and supports those users who want the maximum expressiveness without to lose computational completeness (all inferences will be taken into account) and decidability (all calculations will be completed in a finite time) of reasoning systems. OWL DL includes all the constructs of OWL language with some restrictions such as type separation (a class cannot be an individual or a property, and a property cannot be an individual or class).

OWL Full

OWL Full is intended for users who want to increase expressiveness and the syntactic autonomy of RDF with no computational guarantees. OWL Full includes a full compatibility with RDF/RDFS. Additionally, in OWL Full the reasoning is often complex, slow, incomplete and unsolvable. For instance, OWL Full treat a class simultaneously as a collection of individuals and as an individual in its own right. An additional significant difference from OWL DL is that an owl:DatatypeProperty may be an owl:InverseFunctionalProperty. OWL Full permits an ontology to expand the meaning of the pre-defined (RDF or OWL) vocabulary.

Common Ontologies for Semantic Web Domain Modeling

There are several popular, typical manners to model data, some of them have emerged later than others. It is useful best to make a review of some of the approaches to modeling data that have already been established. Standard formal ontologies that represent terms within a knowledge domain are already available from several organizations devoted to creating standard vocabularies for a number of subjects. Below are some examples:

- Dublin Core Metadata Initiative (DCMI): DCMI Creates ontologies for a variety of subjects, mainly focusing on common, everyday terms and terms important in the media.
- Friend Of A Friend (FOAF) : FOAF is an RDF based schema intended to describe persons and their social network in a semantic way. It focuses on the standard vocabulary/ontology development for social networking purposes.
- OpenCyc - An ontology of everyday, common sense terms.
- Card: vCard [14] is a specification developed by the IETF intended for the description of people and organizations (describe Personal information including address). VCard has been significantly updated to Version 4 as documented in [RFC6350]. Typically, vCard objects are encoded in its own defined text-based syntax or XML renderings.

- Basic Geo Vocabulary (BGV): BGV is an RDF vocabulary which provides the Semantic Web community a namespace for latitude/longitude representation and other information about spatially located things [15].
- Notation 3 (N3): is a syntax (inspired from semantic web) that aims to optimize the data expressiveness and logic in the same language. N3 allows RDF to be expressed and rules to be easily integrated with RDF. Additionally, it allows quoting so that statements about statements can be prepared and aims to be as readable, natural, and symmetrical as possible [16].
- SIOC (the Semantically-Interlinked Online Communities): SIOC is a vocabulary of terms and relationships modeling social web data spaces (discussion forums, blogs, mailing lists, shared bookmarks, feed subscriptions and image galleries) [17].

- Sindice: Sindice is a search engine intended for ontologies, documents, terms and data published in Semantic Web formats. It uses a system of crawlers for RDF documents discovering and embedded RDF content [18].

CONCLUSION

There are several approaches for domain modeling. This paper presented some domain modeling schemas, but focused mostly on those used in Semantic Web (taxonomy, thesaurus and ontology). The intention is to support the information architect for the appropriate choice making between one or more of these schemes. In addition, this paper has provided a comparison between the three most commonly used in order to facilitate the understanding of the individual interested by it.

REFERENCES

1. Wikipedia. Ontology (computer science) From Wikipedia, the free encyclopaedia. 2006 [cited 8 June 2006]; Available from: http://en.wikipedia .org/wiki/Ontology_%28computer_science%29.
2. Gruber. T.R. A translation approach to portable ontology specification. In Knowledge Acquisition. ACM Knowledge Acquisition, Special issue: Current issues in knowledge modeling, 1993.
3. Gruber. T.R. Toward principles for the design of ontologies used for knowledge sharing. In Proceedings of International Workshop on Formal Ontology in Conceptual Analysis and Knowledge Representation. Padova, Italy: Kluwer Academic Publishers, Deventer, The Netherlands, 1993.
4. A. Newell. The Knowledge Level. *Artificial Intelligence,* **18** (1), 1982.
5. Van Heijst G , Schreiber A.T., Wielinga B.J. Using explicit Ontologies in KBS development. *International Journal of Human-Computer Studies,* **46**, no.(2-3), pp.183 – 292 (1997).
6. Guarino,N., Giaretta, P. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In N. J. I. Mars (ed.), Towards Very Large Knowledge Bases. Amsterdam, The Netherlands, pp.25-32. IOS Press, 1995.
7. Guarino, N. editor. Formal Ontology in Information Systems: Pro ceedings of the First International Conference (FOIS'98). Ios Press Inc, 1998.
8. Natalya F. Noy and Deborah L. McGuinness. Ontology development 101: A guide to creating your first ontology. Technical Report KSL-01-05, Knowledge Systems Laboratory, Stanford University, March 2001.
9. Golder, S., Huberman, B.A.: Usage patterns of collaborative tagging systems. *J. Inf. Sci.* **32**(2), 198–208, (2006).
10. Lassila O, Swick R. Resource description framework (RDF) model and syntax specification, W3C Recommendation (1999), http://www.w3.org/TR/REC-rdf-syntax/.
11. Brickley D, Guha R V. RDF Vocabulary Description Language 1.0: RDF Schema, W3C Working Draft, 2002. Available from <http://www.w3.org/TR/PR-rdf-schema>.
12. Bray T, Paoli J, Sperberg-McQueen CM, Maler E. Extensible Markup Language (XML)

1.0, second ed., W3C Recommendation, 2000. Available from <http://www.w3.org/TR/REC-xml>.

13.   "DCMI Metadata Terms". Dublincore.org. Retrieved 13 December 2015.

14.   http://www.w3.org/TR/vcard-rdf/. Retrieved 13 December 2015.

15.   Basic Geo (WGS84 lat/long) Vocabulary http://www.w3.org/2003 /01/geo/.

16.   Notation 3. <http://www.w3.org/DesignIssues/Notation3.html>.

17.   SIOC Core Ontology http://rdfs.org/sioc/spec/.

18.   Sindice search engine <http://sindice.com/>