An Overview of Building Blocks of Semantic Web

Shweta Shrivastava
Research Scholar
AISECT University
Bhopal,India

Sanjay Agrawal, PhD
Dean R&D ,Prof Computer Sc. Dept.
NITTTR BHOPAL
Bhopal, India

ABSTRACT
The world wide web expanded day by day ,many website (avg 51 million website)added to the web every year. Almost all organization support open data and make their data available over the web, which increase innovation. The Semantic Web is an evolution and extension of the existing Web that allows computers to manipulate data and information Semantic web is based on the content oriented description of digital documents with standardized vocabularies that provide machine understandable semantics .Basic building block for Semantic web are Ontology ,RDF/OWL ,SPARQL. Semantic web vocabulary can be considered as a special form of ontology. Semantic web provide connection between human and computer by making the computer think more like a human ,It is artificial intelligence which can intelligently learn and understand the semantic. . Semantic web is also understand by Web 3.0 which is the executable and read/write Web. The idea of the Semantic Web is still undergoing research and development.

Keywords
Semantic Web, Ontology, RDF/OWL, SPARQL, Web 3.0 Search. ,Machine Understandable

1. INTRODUCTION
Evaluation in Web Searching Technology

As we know that the www is a way of accessing information over the medium of the Internet. Firstly Tim Berners Lee invented Web 1.0 which include static web pages and provide read only information. According to human need Web1.0 switches Web 2.0 read/write information .Web 2.0 is a ability to share view in a interactive manner. Everything famous today in the www from face book to YouTube is a example of Web 2.0. It expanding access to the Internet through Mobile .Drawback with Web 2.0 is that ,it is only readable by human and not understand by machine. So researcher suggested switching over to Web 3.0[3] in order to make content machine understandable. Web 3.0 is read/ write, executable web which is also known as Semantic Web . The Semantic Web [7] is an extension of the existing World Wide Web. It provides a standardized way of expressing the relationships between web pages, to allow machines to understand the meaning of hyper linked information. The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning. The semantic web is based on the content oriented description of digital documents with standardized vocabularies that provide machine understandable semantics. Web 2.0 is focus on people ,where Web 3.0 focuses on machine. Today's search engines can not search more precise, may be the main reason is that the structure and size of current Web do not allow to make search more precise and efficient. The second reason Web contains now a huge number of documents and this number has a strong tendency to double each one or two years. The structure of documents and Web itself, probably, can be changed in "a better – machine process able way" .Semantic web[13] is being to be developed to overcome the following problems for current web.-

- The web content lacks a proper structure regarding the representation of information.
- Ambiguity of information resulting from poor interconnection of information.
- Automatic information transfer is lacking.
- Usability to deal with enormous number of users and content ensuring trust at all levels.
- Incapability of machines to understand the provided information due to lack of a universal format.

2. SEMANTIC WEB
The semantic Web is a mesh of information linked up in such a way as to be easily process able by machines , on a global scale. In Semantic Web knowledge[8] will be organized in conceptual space with its meaning. Automated tools will support maintenance by checking for inconsistencies and extracting new knowledge .Researches using Semantic Web technology to enable machines to infer new facts from exiting facts and data[10].On the Semantic Web ,Agents and other automated processes will produce more information faster and at a finer level of access and semantic granularity that can be shared via web services.

Advantages of Semantic Web approach

- Computer can make decision like people can do.
- Machine can operate automatically ,it is the so called "Internet of Things".
- Data sharing can be done more easily due to distribution of data warehouse.
- Assemble information , extract knowledge according to users need.

Characteristics of Semantic Web

- **Intelligence-:** It can directly do intelligent analysis and then provide optimal output with less user intervention.
- **Personalization-**: Personal or individual preference would be considered during different activities such as information processing ,searching, forming a personalized portal on the web.
- **Security-**: Defining who may view certain parts of information for security purpose.
- **Application-**: Semantic Web allows searching ,integration, connection and analysis ,pattern finding , discovery, visualization etc.

- **Integration** -: It is capable to integrate data which collected from heterogonous sources. It would be able to run on many types of computers ,microwaves devices ,mobile ,TV etc.

3. ONTOLOGY

Overview about Ontology

Ontology is an explicit specification of conceptualization. Ontology[1] is a body of knowledge describing some domain, typically common sense knowledge domain. Ontology is a description of the concepts and relationship that can exist for an agent on a community of agents. Ontological commitments allow a number of agents to meaningfully communicate about a domain without necessarily operating on a globally shared theory individually.

Ontology means "explicit specification of conceptualization" .Specification is a form of the definition of representation vocabulary(class ,relation and so on). Conceptualization is contains the objects concepts and other entities that are assumed to exist in other area of interest and the relationships that hold them .Every knowledge base ,knowledge based system or knowledge level agent is committed to any conceptualization (explicitly or implicitly).

Ontology Scope

- It provide machine-interpretable definition of basic concepts in the domain and the relationship among them.
- When Ontology[9] must be formulated in some representation language ,it is intended to be a semantic level specification ,i.e. it is independent of data modeling strategy or implementation.
- Due to the unstructured and semi-structured nature of Web pages, it is a challenging task in categorizing and extracting content from the Web

Ontology is being represented as a setof concepts and their inter-relationships relevant to some knowledge domain. The knowledge provided by ontology is extremely useful in defining the structure and scope for mining Web content.

Switching from Database to Ontology

Being a software engineer developing the front end and back end of a web application is no more a big deal but a challenge is certainly faced when it comes to making the system more intelligent ,the best way would be to use is ontology. Ontology is a description of a world view using a linked or networked graph structure. Major problem with database[6] is that it represented in various formats and it is difficult to integrate these database. This is the area where ontology gains weightage. Separating domain knowledge from operational knowledge and enabling their reuse ,sharing, a common understanding of the structure of information among software agent are some of the important goal implemented through the medium of ontology.

Taxonomy Vs Ontology

Taxonomy is a backbone of Ontology. Taxonomy represent a classification of things in a hierarchical form. It is represented in a form of tree or a lattice that express natural relationship, where's ontology generally include descriptive terms in order to specify the vocabulary of the node contents. Taxonomies generally do not include descriptive keywords for each item and much simpler than ontology. Taxonomies tend to be quite lenient about the kind of relationship that exits between parent and child node. The difference between taxonomy and ontology is in two important contexts:

1. An Ontology has a richer internal structure as it includes relations and constraints between the concepts.
2. An Ontology claims to represent a certain consensus about the knowledge in the domain. This consensus is among the intended users of the knowledge.

Types of Ontology

1. Classification according to purpose
   A. Application Based:- Ontology which is used in specific application.
   B. Reference Based:- Ontology Which used during development of applications.
2. Classification according to expressiveness
   A. Heavyweight Ontology: - It is heavily axiomatized
      and thus represent an ontological commitment
      explicitly.
   B. Lightweight Ontology:- It is simple taxonomic
      structure of primitive or composite terms together
      with associated definitions.
3. Classification According to specificity
   A. Generic Ontology:- Define concepts such as state
      event, process etc.
   B. Domain Ontology :- Express conceptualization
      that are specific for a specific universe of
      discourse.
   C. Core Ontology :- Define Concepts which are
      generic across a set of domain.

Need Of Ontology

- To share a common understanding of the structure of information among people or software agents.
- To enable reuse of domain knowledge.
- To make domain assumption explicit.
- To separate domain knowledge[2] from the operational knowledge.
- To analyze domain knowledge.

Ontology Development life cycle

1. Requirement Specification :- Identifying the intended scope and purpose of the ontology [4]is the initial stage of ontology development life cycle
2. Knowledge Acquisition:- It is a process of acquiring domain knowledge . Different possible scenarios are identified and clubbed together.
3. Identifying Key Concepts:- Conceptualization can be defined as a process of identifying the key concepts

that exist in the domain ,their properties and the relationships that hold between them.

4. Coding :- Coding is the process of representing the conceptualization in some formal language for example frames ,objects model or logic.
5. Integration :- Integrate specific exiting concepts .Use or specialize in an existing ontology.
6. Documentation :- It include informal and formal complete definitions, assumptions and examples that are essential to promote the appropriate use and reuse of ontology.

Advantage of Ontology

1. Ontology promote coherent navigation ,it means movement from one concept to another concept in the ontology structure.
2. In Ontology there are many entry points. Ontology can be traced and related to all of its associated concepts.
3. In Ontology connection highlight related information and aid prompt discovery without requiring prior knowledge of the domain or its terminology.
4. Major advantage with ontology is that it has ability to represent any form of information unstructured[2] , semi structured , structured.
5. Ontology has a ability to match concept of same idea for example happy and glad both referring to the concept of a pleasant state of mind .
6. Inference is possible through Ontology, we can also Integrate external content by proper matching ,mining [5] and mapping of these concepts.

Limitation of Ontology

1. Ontology makes the abstract model of a particular domain it does not define the boundaries of the model.
2. If the number of instances is increased to a large extent then it become very hard to manage manually and currently there is no mechanism to manage it automatically.
3. Manual ontology generation process sometimes becomes very complex and time consuming.

4. RESOURCE DESCRIPTION FRAMEWORK (RDF)

Overview about RDF

RDF[12] specially design for a metadata model .It is used to represent conceptual description or modeling of information that is implemented in web resources , using verity of syntax notation and data serialization formats.RDF decomposed knowledge into small pieces, one can write down those pieces in any number of ways and still preserve the original information and structure like human can communicate in different language but meaning or concept is same in every language. RDF is not a language but a model for representing a data about “things on the web “ . The things are resources in RDF vocabulary.

RDF can be defined using three simple rules :-Subject, Predicate and Object which is known as name of entities. Name of entities can be global and local. RDF provide a general ,flexible method to decompose any knowledge into small pieces called triples with some rules about the semantics (meaning) of those pieces. In RDF triples the statements is broken into three parts.

Subject-------□ Object--------□ Predicate of statements

<rdf: Description rdf :about=”Subject”>

<predicate rdf : resource= “object”>

<predicate> value </predicate>

<rdf :Description>

RDF Basics

There are three basics use in RDF construction

1. Resources :- Resource is an object ,a “thing” we want to talk about, for ex it may be authors ,books ,publishers, places, hotels, search queries etc. Every Resource has a URI, it can be URL or any other unique identifier.
2. Properties :- Is is a special kind of resources. It is also identified by URIs, for ex-book written by, title of the book ,no of pages etc.
3. Statements :- Statement assert the properties of resources. A statement is an object, attribute value triple, consisting of a resource ,a property and a value.

5. QUERY RETRIEVAL LANGUAGES

Different type of Query Retrieval Languages

1. RDF The Resource Description Framework is a family of World Wide Web Consortium (W3C)[12] specifications originally designed as a metadata data model. It has come to be used as a general method for conceptual description or modeling of information that is implemented in web resources. An RDF query language is a computer language, specifically a query language for databases, able to retrieve and manipulate data stored in Resource Description Framework format.
2. RDQL RDQL was developed by Hewlett Packard and submitted to the W3C in January 2004. RDQL is a query language for RDF in Jena models. The idea is to provide a data-oriented query model so that there is a more declarative approach to complement the fine-grained, procedural Jena API. It has been implemented in a several RDF systems including Jena, RDFStore, Sesame, PHP XML Classes, 3 Store and RAP-RDF API for PHP . RDQL was derived mainly from SQUISH, an earlier language. The syntax of RDQL is similar to a SQL-like select pattern where the select clause allows the projection of the variables.
3. SPARQL [7] is a query language designed specifically to query RDF databases. SPARQL(Simple Protocol and RDF Query Language) queries are sent from a client to a service known as a SPARQL end point using the HTTP protocol. The interaction between the client and the endpoint is defined in a machine-friendly protocol that is not intended to be interpreted by humans, so use of SPARQL requires an interface that allows the user to enter the queries and to display the results in a meaningful way. As with traditional database languages such as SQL, those interfaces are

commonly constructed so that the queries are constructed and launched through forms that do not require the human user to have any knowledge of RDF and SPARQL.

4. SeRQL ("Sesame RDF Query Language", pronounced "circle") is a new RDF/RDFS query language that is currently being developed by Aduna as part of Sesame. It combines the best features of other (query) languages (RQL, RDQL, N-Triples, N3) and adds some of its own. This document briefly shows all of these features. After reading through this document one should be able to write SeRQL queries.

6. CONCLUSION

Matching and integrating ontologies has been a desirable technique in areas such as data fusion, knowledge integration, the Semantic Web and the development of advanced services in distributed system. Unfortunately, the heterogeneous of ontologies[15] cause big obstacles in the development of this technique. Most of the database research self assessment reports recognize that the thorny question of semantic. Heterogeneity (consisting of many different kind of things), that is of handling variations in meaning or ambiguity in entity interpretation, remain open. In this thesis, focus on a kind of semantic technologies namely ontology matching.

Technically alignment are used in such a setting for query expansion, standard matching technology can be widely reused while the spatial and temporal counterparts that constitute the specificity of GI applications have not received enough attention so far in the ontology matching field. We believe this can be done through addressing specifically promising challenges that we identify as:

- Large-scale matching evaluation,
- Efficiency of matching techniques,
- Matching with background knowledge,
- Matcher selection, combination and tuning,
- Explanation of matching results,
- Social and collaborative matching, and
- Alignment management: infrastructure and support

7. REFERENCES

[1] P Shvaiko, J Euzenat "Ontology Matching: State of the Art And Future Challenges IEEE Transaction on Knowledge and DataEngineering", Vol. 25, No1. January 2013 Page(s): 158 – 176.

[2] Farrag, Tamer Ahmed ,Saleh, Ahmed Ibrahim ; Ali, Hesham Arafat " Toward SWSs Discovery: Mapping from WSDL to OWL-S Based on Ontology Search and Standardization Engine"IEEE Transaction on Knowledge and Data Engineering, Vol. 25, No5. May 2013 pages 1135-1147

[3] A Telang, C Li, S Chakravarthy "One Size Does Not Fit All: Toward User and Query-Dependent Ranking for Web Databases" IEEE transaction on knowledge and data engineering vol 24, no9, September 2012 pages 1671 - 1685 .

[4] P Kremen, Z Kouba "Ontology-Driven Information System Design" IEEE transaction on System ,Man and cybernetics-PartC: Applications and Reviewes ,Vol. 42, No. 3,May 2012 pages 334 – 344.

[5] P Panov, S Dzeroski, L Soldatova "OntoDM: An Ontology of Data Mining" 2008 IEEE International Conference on Data Mining Workshops Page(s): 752 - 760 .

[6] Dimitrios A. Koutsomitropoulos, Ricardo Borillo Domenech, Georgia D. Solomou "A Structured Semantic Query Interface for Reasoning Based Search and Retrieval "The Semntic Web : Research and Applications Lecture Notes in Computer Sc. Volume 6643,2011,pp17-31.

[7] J Zhai, K Zhou - Information Science and Management "Semantic Retrieval for Sports Information Based on Ontology and SPARQL" Published in: Information Science and Management Engineering (ISME), 2010 International Conference of (Volume:1)Date of Conference:7-8 Aug. 2010 Page(s): 395 - 398 .

[8] Z Li, YL Zheng, SN Li, WW Liang "A Knowledge SharingConvergence Platform Based on OWL-S and Semantic Relations" Published in Software Engineering (WCSE), 2010Second World Congress on (Volume:1) 19-20 Dec. 2010Page(s): 65 - 68 .

[9] J. Euzenat and P. Shvaiko, "Ontology Matching" Springer, 2007.

[10] F. Giunchiglia, M. Yatskevich, and P. Shvaiko, "Semantic Matching: Algorithms and Implementation," on Data Semantics, vol. 9, pp. 1-38, 2007.

[11] Web Services Description Language (WSDL), W3C Note, http:// www.w3.org/TR/wsdl, 2001.

[12] Web Ontology Language for Services (OWL-S), W3C Member Submission, http://www.w3.org/Submission/OWL-S/, 2004.

[13] B. Di Martino, "Semantic Web Services Discovery Based on Structural Ontology Matching," Int'l J. Web and Grid Services, vol. 5, no. 1,2009 pp. 46-65.

[14] OWL, http://www.w3.org/2004/OWL/, 2004.

[15] Z Yun, S Huayou, Q Hengnian "A Semantic Web Services discovery mechanism design and implementation based on OWL Ontology " Educational and Network Technology (ICENT), 2010 International Conference on 25-27 June 2010 Page(s): 139 - 143.