International Journal of Advanced Research in IT and Engineering

ISSN: 2278-6244

INTRODUCTION TO SEMANTIC WEB AND ITS FUTURE

Hedayat Bahadori*

Sara Bahadori*

Amin Samimi Behbahan*

Milad Samimi Behbahan*

Shabnam Azari*

Abstract: *Semantic web is considered as the most important research subject in web realm which allocates many researches in recent years. One of the main implementation for accomplishing web is to promote the knowledge of present web and add semantic metadata to web pages or on the other hand, is semantic annotation. Regarding to the high value of web pages and remarkable subject variability of its present contents, it can be realized that annotation of web page based on present ontology is one of the most important challenging problems of confronting semantic web. Whereas annotating semantic web and developing ontology necessitate information extraction and discovering knowledge, it can be realized that various methods of data mining such as assortment, clustering, mining association rules, and text mining play effective roles in removing foresaid challenge.*

*Department of Computer Engineering, Omidiyeh Branch, Islamic Azad University, Omidiyeh, Iran

1. INTRODUCTION

Digital world grew astonishingly during 50 years after naming AI in Dortmund Conference and produce information substrate that it was less than predicted. Today, World Wide Web links more than tens billion pages. Search engine can deduce subjects which include these links.

Inventor of World Wide Web presented a kind of semantic web as an article with James Hendler and Ora Lassila [1]. Semantic web creates structures for web pages that people can do complicated responsibilities by computer. The goal of research about semantic web is to pass the documents and to design for realizing and to process by machine.

Semantic web tries to develop documents of web as meaningful interchangeable.

Semantic web has broad applications in various fields such as management of knowledge, web-based training, bioinformatics, E-government, digital libraries and etc. to ensure identical realizing of machines of metadata labels and communication and collaboration between them, a subscription repository is necessary for description of labels. Ontology is used as such repository of meanings in semantic web. The term of ontology attributes to show the knowledge from special amplitude in realm of computer science. The set of objects and relationships between them are defined by dictionary in ontology. Ontology is a tool for describing added meanings to web documents and provides the possibility of using these meanings for intelligent software agents and web users [2].

2. SEMANTIC WEB

Future web is realizable and processed by computer in addition to human from point of view of Brenner-lee known as father of web. The main goal of semantic web is a web that is able to converse to machine [1]. Semantic web is substrate for many various applications which can be developed. Semantic web contains of operational information, the Information which should be obtained by semantic theory for interpretation of symbols. Semantic theory develops logic connections of terms with capability of collaborations between systems. Necessary key for semantic web is capability of collaboration. Among definitions of semantic web, we can mention following options [3]:

- Semantic web is a project with the goal of creating world media for exchanging information to be realizable for computer

- Semantic web is network of information on a global scale how it is easily to process by machines
- Web semantic includes intelligent data of web that is processed by machine

The future of web can be divided two sections from point of Breners' view. The first section is to manufacture web with environment with collaboration and communication of users. Second section is to manufacture web that is realized and processed by the machine. Figure 1 shows the point of Breners' view about presenting web [4]. The language HTML that is used in present web is not able to state objects and their relationships. Another language has been created called RDF to use in semantic web.

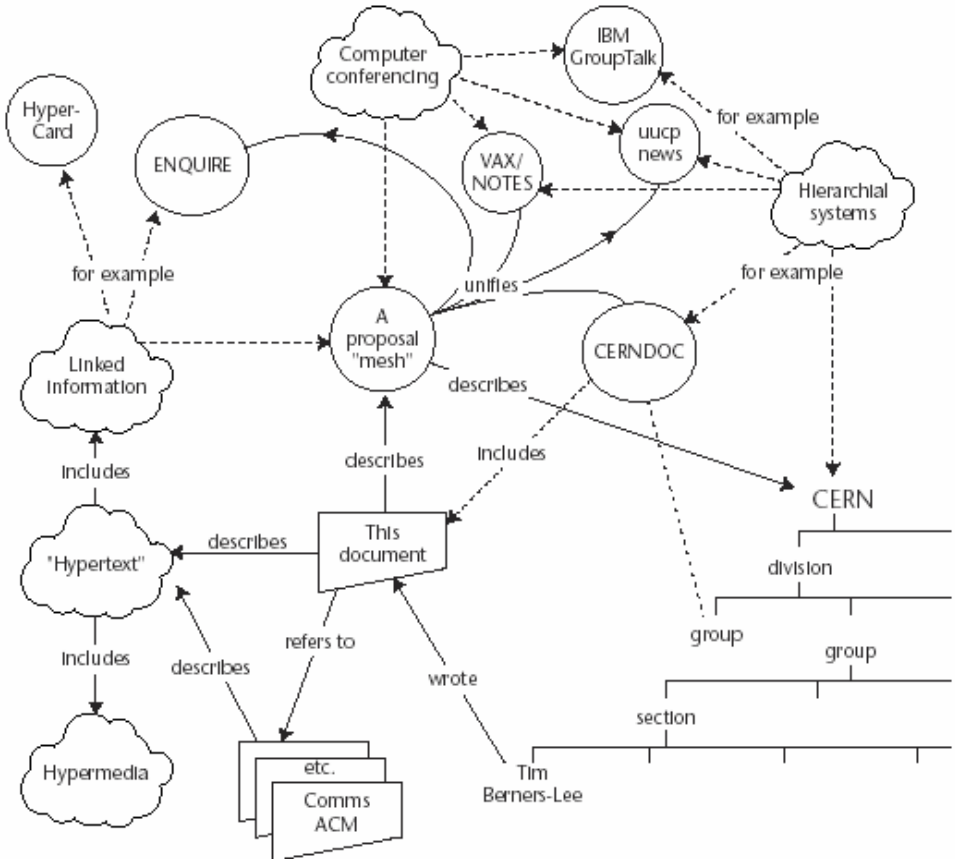

Figure 1: suggestion of Mr. Tim Berners Lee about web

The first step is to change the thought about data. Data should be more intelligent in semantic web for being realized by machines. More concepts should be saved with data how the machines can realize them [3]. Various levels of intelligent data have been showed in figure 2.

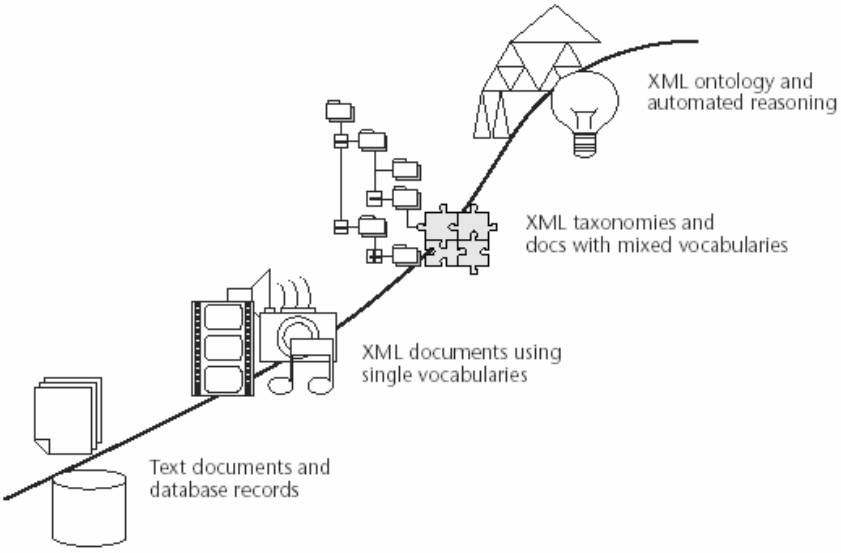

Figure 2: Various Levels of Intelligent Data

Different architectural layers of semantic web are shown in figure 3 RDF, basic language used in semantic web is established based on XML [1]. XML has been established based on Unicode and URL and backs up different languages. URL is used for determining concepts in semantic web. For instance, URL is used as a kind of URL for determining reservoirs in web. The main part of semantic web is ontology. Ontology communicates between labels of documents of semantic web and real objects. By using rules we can obtain new knowledge from present knowledge. Manufacturing standard framework for present rules, we can achieve proof. Obtained proofs are communally used in different applications. So, trust web is created as final level of semantic web.

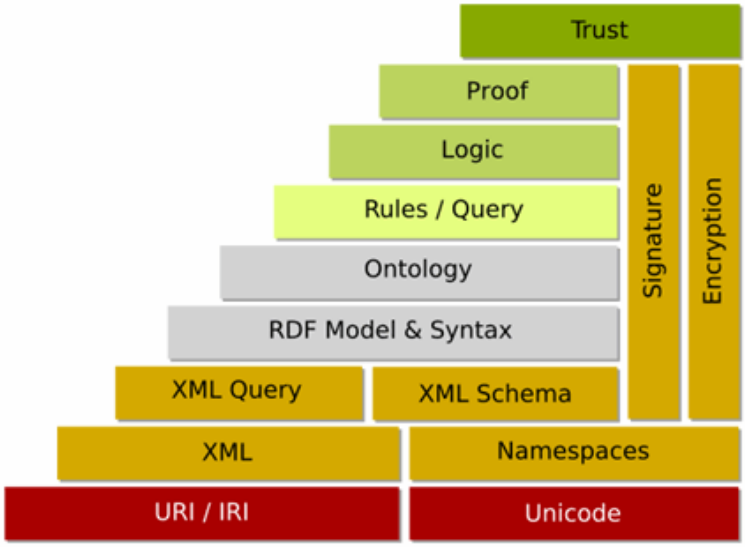

Figure 3: Architecture of Semantic Web

3. SEMANTIC RELATIONS

Ontology and semantic webs create semantic relations between demonstrated existences. Semantic relations are meaningful dependencies between two or more concepts, existences, or set of existences.

Display systems of type information are divided into three general sections of terms list, classification and categories and relations between these lists [Zhang, 2007. Terms list is list of terms are usually shown by their definitions. Classifications and categories create sets of objects. Relationship lists specify the connections between terms and concepts. This list is summarized systems in following classification:

- Term Lists
  - Authority Files
  - Glossaries
  - Dictionaries
  - Gazetteers
- Classifications and Categories
  - Subject Headings
  - Classification Schemes
  - Taxonomies
  - Categorization Schemes
- Relationship Lists
  - Thesauri
  - Semantic Networks
  - Ontologies

Terms are listed with specific meaning (term list) like in alphabetical order in the first category. Terms are easily processed in necessary time. Dependency of these terms is separated from their alphabetical order. The meaning of the term is not related to previous term. Second section (classification & categories) terms and concepts are arranged as hierarchical manner. Hierarchy has been determined as specific type of relation between terms and concepts. Present terms are placed in higher level, higher classification or wider categories. These terms usually have wider meaning than present terms in lower levels.

Hierarchy lists show the relation of class-subclass as dependent to meaning between terms and concepts. In third section (relation lists), relations are shown between terms and concepts. Semantic relationship is more created and stated in relation lists. Terms and concepts are linked together as meaningful, for example, hierarchy, horizontal order (synonym), reverse order (antonym) or causal order [5]. Showed different Realization of semantic relation creates relation lists in term lists. Relation lists provide useful framework for understanding difference and similarity between ontology and the other forms of knowledge representation. Ontology depends to one or more traditional library display like reservoir, classification, category pattern, controlled dictionary and even a dictionary. Dependency of traditional library display models and ontology are shown by arranging them based on following category [5]:

Term Lists

Controlled vocabulary

Dictionary

Hierarchical lists

Classification Scheme

Taxonomy

Relationship Lists

Thesaurus

Ontology

4. TYPES OF MEANINGS

Summary of three meanings have been presented in this section [5]:

- Implication: this type of meaning is implicit of data patterns and it is not displayed explicitly in manner of being processed by machine. Tools and applications were developed for implications and it is easily available. Learning techniques of machine like clustering, hidden markov model, artificial neural network and so on apply implication.
- Formal Meaning: human beings often communicate by language. Natural language is based on syntax rules and it is ambiguous in concept. Computers are lack of removing ambiguities and perception of complicated natural language. Natural

language is not used as a tool for communicating among machines. Concepts or facts must be stated by being processed by computers. Formal meanings are displayed the good way form.

- Description Logic: description logic is highlighted form of knowledge presentation. Formal meanings in DLs are based on theory of sets. Relations between concepts or roles are stated as multiplied vector from amplitude.

5. THE ABILITY OF PROCESS OF MACHINE

To develop a web with meanings, reservoirs on web must be displayed with realizable descriptions for machine. Descriptions are as standard concepts and relations among concepts. Annotating web contents must be done using dictionaries and structures which have been defined with a domain of ontology explicitly and officially. The ability of process of machine severely depends to accessibility and increasing ontology. Creating ontology is manually tedious task, time-consuming and prone to error. Manufacturing automatic tool for creating ontology of available information extraction is primary step of implementation. Semi-automatic ontology extraction method is considered as applied solution. Some steps are necessary for re-application of ontology online in creating new ontology. These steps include ontology search, ranking, segmentation, mapping and merging, annotating and assessing [5].

6. CHALLENGES OF INFORMATION SYSTEMS ABOUT SEMANTIC WEB

Superior semantic web is an approach of collecting simple words in search engines and is related to meaning of text. Subscription information, information extraction, maintenance and coincidence are easily possible in search based on key word [6].

Following cases have been stated as challenges related to semantic web in information systems in [6]:

- Challenge of developing ontologies domain
- Challenge of mapping, installing and merging ontology
- Challenge of interpretation management
- Challenge of retrieving information based on ontology

6.1. CHALLENGE OF DEVELOPING ONTOLOGIES DOMAIN

Ontologies play important roles in semantic web. To develop ontologies, experts are appointed in the same domain. Each of experts may have different opinions (a social challenge). A development environment of collective ontology is necessary (a technical challenge).

A collective development environment must be able to control model, proposal, releasing and supporting coordination and collaboration. Developing such environment is a challenge. Today, most of the tools of ontology development like individual ontology editors are lack of this usefulness [6]

6.2. CHALLENGE OF MAPPING, INSTALLING AND MERGING ONTOLOGY

Using a unique ontology is so difficult for any data reservoirs. Users should apply ontologies related to themselves for interpretation of data reservoirs. Then if they intend, they create added mapping for designing them as standard domain ontology (central). Standard domain ontology is useful to back up the ability of collaborating data and inquiring. This mapping is a challenge among ontologies due to the possibility of dissimilarity in content, schema and meaning among ontologies. The process of mapping may not only include adjustment of ontology for create coherent ontologies. Ontologies are merged with central ontology for adding new words [6].

6.3. CHALLENGE OF INTERPRETATION MANAGEMENT (ANNOTATION)

First challenge is to allocate interpretation tool for every domain based on presented ontology and requirement of users. Second challenge of world development is interpretation tool. Allocated interpretation tool for several domains needs an open architecture and backup operational interpretation for different domains. The third challenge is the capability of shared interpretation. Shared interpretation needs backup concurrency control. When the users interpret separated data sections, shared data interpretation is effective. Suppose that a person interprets the announcer's movements and the other specifications of discourse linguistic. Two ontologies can be used for this purpose. Each person may be familiar with one of them. One of projects of guiding shared interpretation is Annotea [6].

6.4. CHALLENGE OF RETRIEVING INFORMATION BASED ON ONTOLOGY

Interpreted data that can be searched by its interpretation is useful. One of affirmations of semantic web is high precision. Search engines should apply semantic reasoning and available ontologies in order to obtain accurate results and determine semantic relations between them. New opportunities need new approaches for refining search and strategies of connector of users. Main challenge of search is through data collection. These data collections were interpreted by various ontologies. Several ontologies can be existed. Ever domain can be schemed by several ontologies. A domain may need application of several ontologies. Therefore, not only mapping of ontology, but also mapping of user search is necessary [6].

7. ONTOLOGY

Ontology was created by two words onto means existence and Logia means study. Briefly, ontology defines as following: ontology is specification and definition of conceptualization [3].

Nowadays, ontologies include many similarities of structures regardless the language in that is schemed. Most of ontologies describe individuals (instances), classes (concepts), properties and relations. Ontologies are basic element in creating semantic web. If the meaning of data is displayed explicitly on web, interaction of meaning level is possible among users of web. Data of web should be realizable regarding domain and content for machine [7].

7.1. LANGUAGES OF ONTOLOGY

There are two types of languages for creating ontology. Languages perform based on graph. Like Topic maps، UML،RDF Semantic networks. Languages which perform based on proposition logic. Like (RuleML, LP/Prolog) Rules، First Order Logic. Some of languages stating ontologies include [3]:

- RDF(S) (Resource Description Framework (Schema))
- OIL (Ontology Interchange Language)
- DAML+OIL (DARPA Agent Markup Language + OIL)
- OWL (Ontology Web Language)
- XOL (XML-based Ontology Exchange Language)

- SHOE (Simple HTML Ontology Extension)
- OML (Ontology Markup Language)

7.2. THE MANNER OF CREATING ONTOLOGY

Two manual and automatic methods are used for creating ontology. Automatic methods use different approaches for extracting ontology from available texts. Most of these methods use process techniques of natural language. Automatic method that can produce plausible ontology, has not introduced yet. Therefore, manual or semi-automatic methods are used for creating ontology. There is not standard method for creating ontology. Different methods are used based on application of ontology. One of these methods is shown in figure 4 [3].

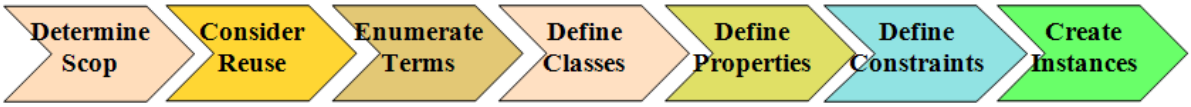

FIGURE 4: METHOD OF CREATING ONTOLOGY

The process of creating ontology is recursive. It may need to create amendment and recursive to previous step in each step. Steps of this method include [3]:

1- Realizing range: first study is done for creating ontology. Range that should be covered, type and manner of using ontology are determined.
2- Re-use: comprehensive study is done on available ontologies. It is investigated in this step whether available ontologies are useful for creating mentioned ontology.
3- Words diagnosis: mentioned words scope and their specifications are determined.
4- Diagnosis of classes and their structures: using obtained methods, their class and subclass are determined. For this reason, we can use bottom-up method or up-bottom or both of them.
5- Defining attributes: internal structure of classes is described. Usually, most of obtained words in step 3 that were not used in step 4 create mentioned attributes.
6- Defining ranges: ranges that every attributes can be formed are determined.
7- Creating objects: after creating ontology, mentioned instances can be created based on them.

7.3. ONTOLOGY EXTRACTION

Ontology engineer has various responsibilities such as editing, assessing, writing, annotating, regulating, merging, re-application and extracting ontology. Responsibility of editing presents an editor for manually manufacturing ontology [8] .

Extraction may includes techniques related to linguistic, statistical techniques, machine learning, and composite techniques affiliated to technology of retrieving information. Approach of machine learning refers to sets of techniques and algorithms for obtaining knowledge to an automatic method. Table 1 has briefly been stated different approaches of extracting ontology [8]:

Table 1: Different Approaches of Extracting Ontology

Name	Method	Language	Auxiliary source
Agirre et al.	Statistical	Unstructured	Ontology
Arasu and Garcia-Molina	Machine learning	Semi-structured	None
Buttler et al.	Machine learning	Semi-structured	None
Craven et al.	Joint method	Semi-structured	Both
Crescenzi et al.	Machine learning	Semi-structured	None
Davulcu et al.	Machine learning	Semi-structured	Samples
Faatz and Steinmetz	Statistical	Unstructured	Ontology
Faure and Poibeau	Linguistics	Unstructured	Both
Heyer et al.	Statistical	Unstructured	Samples
Jiang and Tan	Statistical	Unstructured	Both
Kietz et al.	Joint method	Unstructured	Both
Maddi et al.	Statistical	Unstructured	Samples
Maedche and Staab	Joint method	Unstructured	Both
Navigli and Velardi	Joint method	Unstructured	Both
Shamsfard and Abdollahzadeh	Linguistics	Unstructured	Both
Han and Elmasri	Machine learning	Semi-structured	Both
This study	Machine learning	Semi-structured	None

8. CONCLUSION

Semantic web and some of affiliated technologies have been described. The idea of semantic web is enrichment of web content with semantic metadata. Semantic web is developed copy of current web that cause to promote cooperation between human and computer by adding well-defined semantics.

REFERENCES

[1] Berners-Lee T., Hendler J., and Lassila O. "The Semantic Web", the Scientific American, 2001.

[2] Hassanzadeh H., Keyvanpour M.R., “Data mining methods for classification and evaluation methods for semantic annotation and ontology development”, iran dataminig conference 4th, 2009.

[3] Al-ahmad A., “Introduction to the Semantic Web”, TEHRAN University, faculty of electronic and computer Iran, 2006.

[4] Daconta M., Obrst L., Smith K., “The Semantic Web: A Guide to theFuture of XML, Web Services, and Knowledge Management.”, (2006)

[5] Zhang J., “Ontology and the Semantic Web.” Proceedings of the North American Symposium on Knowledge Organization., 2007, Vol. 1. Available: http://dlist.sir.arizona.edu/1897/

[6] Chebotko A., Lu S., Fotouhi F.,”Challenges for Information Systems towards the Semantic Web”, Department of Computer Science Wayne State University, 2004

[7] Casevie D., Dguric D., devedzic V.,” Model Driven Architecture and Ontology Development”, chapter 3, Springer 2006

[8] Timon C., Du A. , Li B.F., King I.C., “Managing knowledge on the Web – Extracting ontology from HTML Web”, Decision Support Systems 47 (2009) 319–331,Elsevier.