ABSTRACT

Data that is generally hidden away in HTML files is often useful in some contexts, but not in others. The problem with the majority of data on the Web is that it is difficult to use on a large scale, because there is no global system for publishing data in such a way as it can be easily processed by anyone. So the Semantic Web can be seen as a solution. Two search strings with same syntax will have different meanings.

If searching page is a less popular topic which happens to have a same syntax of popular topic which is totally irrelevant to the searching topic, the finding will take time and ingenuity. To overcome this, newly introduced concept is semantic web which semantically enable the search engine to find desire page. Semantic web use ontology for displaying a user's desired web page. Ontology is used as a backbone for indexing.

The Semantic Web is generally built on syntaxes which use URIs to represent data, usually in triples based structures: i.e. many triples of URI data that can be held in databases, or interchanged on the World Wide Web using a set of particular syntaxes developed especially for the task. These syntaxes are called "Resource Description Framework" syntaxes.

Ontology is defined as a rigorous and exhaustive organization of some knowledge domain that is usually hierarchical and contains all the relevant entities and their relations.

The paper is presented as an a survey of semantic web, which comprises the tools used for developing Semantic Web ontologies and building semantic applications, its architecture, the application areas, the current projects, future work.

KEYWORD

INTRODUCTION - SEMANTIC WEB

The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming. A concept proposed by World Wide Web inventor Tim Berners-Lee.

The Semantic Web is World Wide Web inventor Tim Berners-Lee’s idea that the Web as a whole can be made more intelligent and perhaps even intuitive about how to serve a user's needs. Berners-Lee observes that although search engines index much of the Web's content, they have little ability to select the pages that a user really wants or needs. He foresees a number of ways in which developers and authors can use self-descriptions and other techniques so that context-understanding programs can selectively find what users want.

The Semantic Web is focused on machines. The Web requires a human operator, using computer systems to perform the tasks required to find, search and aggregate its information. It's impossible for a computer to do these tasks without human guidance because Web pages are specifically designed for human readers. The Semantic Web is a project that aims to change that by presenting Web page data in such a way that it is understood by computers, enabling machines to do the searching, aggregating and combining of the Web's information — without a human operator.

Tim Berners-Lee originally expressed the vision of the semantic web as follows

I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web — the content, links, and transactions between people and computers. A ‘Semantic Web’, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The ‘intelligent agents’ people have touted for ages will finally materialize.

— Tim Berners-Lee, 1999

The Semantic Web is not about links between web pages. It describes the relationships between things and the properties of things. The Semantic Web is not a separate entity from the World Wide Web. It is an extension to the Web that adds new data and metadata to existing Web documents, extending those documents into data. This extension of Web documents to data is what will enable the Web to be processed automatically by machines and also manually by humans. To do this RDF (Resource Description Framework) is used to turn basic Web data into structured data that software can make use of. RDF works on Web pages and also inside applications and databases.

Ontology
Ontology is a rigorous and exhaustive organization of some knowledge domain that is usually hierarchical and contains all the relevant entities and their relations. The main thread of ontology in the philosophical sense is the study of entities and their relations.

The subject of ontology is the study of the categories of things that exist or may exist in some domain. The product of such a study, called ontology, is a catalog of the types of things that are assumed to exist in a domain of interest $D$ from the perspective of a person who uses a language $L$ for the purpose of talking about $D$.

An informal ontology may be specified by a catalog of types that are either undefined or defined only by statements in a natural language. A formal ontology is specified by a collection of names for concept and relation types organized in a partial ordering by the type-subtype relation. Formal ontologies are further distinguished by the way the subtypes are distinguished from their super types: an axiomatized ontology distinguishes subtypes by axioms and definitions stated in a formal language, such as logic or some computer-oriented notation that can be translated to logic; a prototype-based ontology distinguishes subtypes by a comparison with a typical member or prototype for each subtype. Large ontologies often use a mixture of definitional methods: formal axioms and definitions are used for the terms in mathematics, physics, and engineering; and prototypes are used for plants, animals, and common household items.

The term ontology means a specification of a conceptualization. That is, ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly. Ontology is an explicit specification of a conceptualization. Ontology is an explicit specification of a conceptualization.

The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies, and is endorsed by the World Wide Web Consortium. This family of languages is based on two semantics: OWL DL and OWL Lite semantics are based on Description Logics, which have attractive and well-understood computational properties, while OWL Full uses a novel semantic model intended to provide compatibility with RDF Schema. OWL ontologies are most commonly serialized using RDF/XML syntax. OWL is considered one of the fundamental technologies underpinning the Semantic Web, and has attracted both academic and commercial interest.

The data described by OWL ontology is interpreted as a set of "individuals" and a set of "property assertions" which relate these individuals to each other. OWL ontology consists of a set of axioms which place constraints on sets of individuals (called "classes") and the types of relationships permitted between them. These axioms provide semantics by allowing systems to infer additional information based on the data explicitly provided.

**OWL Characteristics**
OWL provides the capability of creating classes, properties, defining instances and its operations. User-defined classes which are subclasses of root class owl: Thing. A class may contain individuals, which are instances of the class, and other subclasses. For Example, Employee is the subclass of class owl: Thing while Dealers, Managers, and Labors are all subclass of Employee.

Properties

A property is a binary relation that specifies class characteristics. They are attributes of instances and sometimes act as data values or link to other instances. There are two types of simple properties:

1. Data type properties are relations between instances of classes and RDF literals or XML schema data types.
2. Object properties are relations between instances of two classes. Properties may possess logical capabilities such as being transitive, symmetric, inverse and functional. Properties may also have domains and ranges.

Instances

Instances are individuals that belong to the classes defined. A class may have any number of instances. Instances are used to define the relationship among different classes. For example, Mr. Smith is the instance of Manager Class and Mr. Shah is the instance of Dealer class then Mr. Smith is connected to Mr. Shah. By customer-dealer relationship. This is how OWL helps to establish various relationships among classes and instances of web content.

Operations

OWL supports various operations on classes such as union, intersection and complement. It also allows class enumeration, cardinality, and disjointness.

ONTOLEGY ARCHITECTURES:
Different ontology architecture for information integration

a. Single Ontology approach uses global ontology providing a shared vocabulary for the specification of the semantics. All information sources are related to the global ontology. Such approaches can be applied to problems where all information sources to be integrated provide a very similar view on a domain.

b. Multiple ontology approach, each information source is described by its own ontology. Each of these application ontologies can be a combination of several other ontologies and it can be developed independently. Lack of a common vocabulary makes it difficult to compare different application ontologies.

c. Hybrid approach is similar to multiple ontology approaches in that the semantics of each source is described by its own ontology. But in order to make the local ontologies comparable to each other they are built from a global shared vocabulary.

Semantic Web Architecture

The first layer, URI and Unicode, follows the important features of the existing WWW. Unicode is a standard of encoding international character sets and it allows that all human languages can be used (written and read) on the web using one standardized form. Uniform Resource Identifier (URI) is a string of a standardized form that allows to uniquely identifying resources (e.g., documents). A subset of URI is Uniform Resource Locator (URL), which contains access mechanism and a (network) location of a document - such as http://www.example.org/. Another subset of URI is URN that allows identifying a resource without implying its location and means of dereferencing it. The usage of URI is important for a distributed internet system as it provides understandable identification of all resources. An international variant to URI is Internationalized Resource Identifier (IRI) that allows usage of Unicode characters in identifier and for which a mapping to URI is defined. In the rest of this text, whenever URI is used, IRI can be used as well as a more general concept.
Extensible Markup Language (XML) layer with XML namespace and XML schema definitions makes sure that there is a common syntax used in the semantic web. XML is a general purpose markup language for documents containing structured information. A XML document contains elements that can be nested and that may have attributes and content. XML namespaces allow specifying different markup vocabularies in one XML document. XML schema serves for expressing schema of a particular set of XML documents.

A core data representation format for semantic web is Resource Description Framework (RDF). RDF is a framework for representing information about resources in a graph form. It was primarily intended for representing metadata about WWW resources, such as the title, author, and modification date of a Web page, but it can be used for storing any other data. It is based on triples subject-predicate-object that form graph of data. All data in the semantic web use RDF as the primary representation language. The normative syntax for serializing RDF is XML in the RDF/XML form. Formal semantics of RDF is defined as well. RDF itself serves as a description of a graph formed by triples. Anyone can define vocabulary of terms used for more detailed description. To allow standardized description of taxonomies and other ontological constructs, a RDF Schema (RDFS) was created together with its formal semantics within RDF. RDFS can be used to describe taxonomies of classes and properties and use them to create lightweight ontologies.

More detailed ontologies can be created with Web Ontology Language OWL. The OWL is a language derived from description logics, and offers more constructs over RDFS. It is syntactically embedded into RDF, so like RDFS, it provides additional standardized vocabulary. OWL comes in three species - OWL Lite for taxonomies and simple constrains, OWL DL for full description logic support, and OWL full for maximum expressiveness and syntactic freedom of RDF. Since OWL is based on description logic, it is not surprising that a formal semantics is defined for this language. RDFS and OWL have semantics defined and this semantics can be used for reasoning within ontologies and knowledge bases described using these languages. To provide rules beyond the constructs available from these languages, rule languages are being standardized for the semantic web as well. Two standards are emerging - RIF and SWRL.

For querying RDF data as well as RDFS and OWL ontologies with knowledge bases, a Simple Protocol and RDF Query Language (SPARQL) are available. SPARQL is SQL-like language, but uses RDF triples and resources for both matching part of the query and for returning results of the query. Since both RDFS and OWL are built on RDF, SPARQL can be used for querying ontologies and knowledge bases directly as well. Note that SPARQL is not only query language; it is also a protocol for accessing RDF data.
It is expected that all the semantics and rules will be executed at the layers below Proof and the result will be used to prove deductions. Formal proof together with trusted inputs for the proof will mean that the results can be trusted, which is shown in the top layer of the figure \ref{sw-layers}. For reliable inputs, cryptography means are to be used, such as digital signatures for verification of the origin of the sources. On top of these layers, application with user interface can be built.

**SEMANTIC WEB TOOLS:**

It contains the information on RDF and OWL tools that used to be listed on the home pages of the RDF and OWL Working Groups at W3C. Keeping such lists up-to-date is obviously a problem when the number of Semantic Web tools increases every day. Day by day, large amount of tools being created in the community. Here the some tools are listed below

<table>
<thead>
<tr>
<th>NAME (URL)</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>3store</td>
<td>A core C library that uses MySQL to store its raw RDF data and caches, forming an important part of the infrastructure required to support a range of knowledgeable services</td>
</tr>
<tr>
<td>4Suite 4RDF</td>
<td>The 4Suite 4RDF an open-source platform for XML and RDF processing implemented in Python with C extensions</td>
</tr>
<tr>
<td>Active RDF</td>
<td>Active RDF is a library for accessing RDF data from Ruby programs. It can be used as data layer in Ruby-on-Rails. You can address RDF resources, classes, properties, etc. programmatically, without queries</td>
</tr>
<tr>
<td>Adaptiva</td>
<td>A user-centred ontology building environment, based on using multiple strategies to construct an ontology, minimizing user input by using adaptive information extraction</td>
</tr>
<tr>
<td>Aduna Metadata Server</td>
<td>The Aduna Metadata Server automatically extracts metadata from information sources, like a file server, an intranet or public web sites. The Aduna Metadata Server is a powerful and scalable store for metadata</td>
</tr>
<tr>
<td>AeroText</td>
<td>Entity extraction engine from Lockheed Martin</td>
</tr>
<tr>
<td>AJAX Client for SPARQL</td>
<td>AJAX Client for SPARQL is a simple AJAX client that can be used for running SELECT queries against a service and then integrating them with client-side JavaScript code</td>
</tr>
<tr>
<td>AKT Research Map</td>
<td>A competence map for members of the AKT project</td>
</tr>
<tr>
<td>AKT-Bus</td>
<td>An open, lightweight, Web standards-based communication infrastructure to support interoperability among knowledge services.</td>
</tr>
</tbody>
</table>

**SEMANTIC WEB APPLICATION**

The Semantic Web is dissimilar in many ways from the World Wide Web. There have been a number of small scales Semantic Web applications like library system, medical research field, tour and travel etc. Semantic web application for e-Commerce, e-Government and e-Learning has been improved. E-Commerce is more useful for the user because the required information
is available in a comparable format and useful for decision making. Making e-Commerce to its full potential requires a Peer-to-Peer approach.

Projects

Dbpedia

DBpedia is an effort to publish structured data extracted from Wikipedia: the data is published in RDF and made available on the Web for use under the GNU Free Documentation License, thus allowing Semantic Web agents to provide inferencing and advanced querying over the Wikipedia-derived dataset and facilitating interlinking, re-use and extension in other data-sources.

FOAF

A popular application of the semantic web is Friend of a Friend (or Foaf), which describes relationships among people and other agents in terms of RDF.

SIOC

The SIOC Project - Semantically-Interlinked Online Communities provides a vocabulary of terms and relationships that model web data spaces. Examples of such data spaces include, among others: discussion forums, weblogs, blogrolls / feed subscriptions, mailing lists, shared bookmarks, image galleries.

Open GUID

Aimed at providing context for the Semantic Web, Open GUID maintains a global Identifier repository for use in the linked web. Domain-specific Ontologies and content publishers establish identity relationships with Open GUIDs.

SIMILE

Semantic Interoperability of Metadata and Information in unLike Environments

SIMILE is a joint project, conducted by the MIT Libraries and MIT CSAIL, which seeks to enhance interoperability among digital assets, schemata/vocabularies/ontologies, Meta data, and services.

NextBio

A database consolidating high-throughput life sciences experimental data tagged and connected via biomedical ontologies. Nextbio is accessible via a search engine interface. Researchers can contribute their findings for incorporation to the database. The database currently supports gene or protein expression data and is steadily expanding to support other biological data types.

Linking Open Data
Class linkages within the Linking Open Data datasets. The Linking Open Data project is a community-led effort to create openly accessible, and interconnected, RDF Data on the Web. The data in question takes the form of RDF Data Sets drawn from a broad collection of data sources. There is a focus on the Linked Data style of publishing RDF on the Web. See #Triplify for a small plug-in to expose data from your Web application as Linked Data. The project is one of several sponsored by the W3C's Semantic Web Education & Outreach Interest Group (SWEO).

Insemtives

Insemtives is a European Seventh Framework Program (FP7) -funded project with the objective to bridge the gap between human and computational intelligence for the semantic content authoring.

GETTING THE NEEDED WEB PAGE:

By using semantic web, user can get the desired web page. To do this a table structure is created. Table presents the information in rows and columns. Data in row/column usually belongs to same category. Labels of the row describe this category. The table structure may be simple or complex. Table structure is also nested based on the problems complexity. Table has to be matched with the user entered information for retrieval of user specified web page.

CONCLUSION

The user access World Wide Web for collecting details of need. Since search engine look for syntax a fetch a page which satisfies the user specified one. The fetched page may be of entirely different to the page needed by the user. To overcome this, semantic web came into existence. This makes the search engine to analyze the semantics of the request posted by the user. Semantic web uses ontology for making this semantic analysis easier. To fetch a page a table structure is used to store the data in row and column format. An algorithm in used for matching table and fetching data.

REFERENCE

http://www.w3.org/DesignIssues/
http://infomesh.net/2001/swintro/
http://www.webopedia.com
http://www.w3schools.com/semweb
http://searchsoa.techtarget.com
http://www.clickz.com
http://wordnetweb.princeton.edu
http://www.jfsowa.com/ontology
http://www.ontology.com
http://www.formalontology.it