International Journal of Engineering Sciences Paradigms and Researches (IJESPR)
Vol. 47, Issue 04, Quarter 04 (October-November-December 2018)
(An Indexed, Referred and Impact Factor Journal)
ISSN (Online): 2319-6564
www.ijesonline.com

Semantic Web Technologies towards an Intelligent Web

Nikita Malik[1] and Sanjay Kumar Malik[2]

[1]**Research Scholar, USIC&T, Guru Gobind Singh Indraprastha University**
Dwarka, New Delhi, India
nikitamalik92@gmail.com

[2]**Associate Professor, USIC&T, Guru Gobind Singh Indraprastha University**
Dwarka, New Delhi, India
sdmalik@hotmail.com

Publishing Date: October 01, 2018

Abstract

The current web is based on keyword searches and does not understand the user's needs completely. Semantic web, the dream of Sir Tim Berners Lee, is envisioned as the smart web which enables better interaction of humans with the web by allowing the machines to understand the data on the web. The capabilities of Semantic web are tremendous, which make it an upcoming research area. But implementing and realizing the goal of semantic web is the greatest challenge because of its technologies as its research concerns. There is a need to explore the various concerned technologies as research aspects associated with this intelligent web, in the form of a complete tour from past review to current progress. In this paper, first, evolution of web and the need of semantic web for an efficient information search has been presented; second, the architecture for semantic web has been revisited; third, various technologies associated with web have been explored, discussed and revisited, and finally, various other research issues have been discussed along with the future scope in this direction.

Keywords: *Semantic Web Technologies, Ontology, RDF, Research Concerns, SPARQL, Intelligent Web.*

1. Introduction

Evolution of Web:

The World Wide Web (WWW), or the Web, is the global information space which combines the content on the web pages in an interactive manner. Tim Berners Lee invented web and presented it in the first International WWW Conference in 1989. It was termed as Web 1.0, where content was created by the producers and presented to the users to read, search and share. Lee and his collaborators made efforts to lay the foundation for designing open standards for the Web including HTML (Hypertext Markup Language), URI (Uniform Resource Identifier), HTTP (Hypertext Transfer Protocol).

The server and client side scripting languages, blogging sites and the social media sites were created in late 1990s and surged in early 2000s. This entire phase of web evolution was termed Web 2.0 and is the web version that is currently in use, where content creation and sharing by the users form the core basis. Web 2.0 has become quite advanced but yet not reached Web 3.0.

The next phase of web evolution, Web 3.0 or Semantic Web, was conceptualized by Sir Tim Berners Lee in 1999, and is still in its initial stages of development, and may become the next IT (Information Technology) revolution as extension of the web to a smarter web [1]. He expressed his vision of an efficient web when he published the roadmap for future web design in 1998, in his book, 'Weaving The Web' [5] and his speech at XML 2000 Conference [6]. His Scientific American article 2001 [4] was a breakthrough in providing the emergence of the concept of semantic web and its architectures' significant components.

Need of an Intelligent Web:

The need for Semantic Web was felt owing to the shortcomings of the Web version 2.0, which is primarily keyword based. The data on the web is largely present in unstructured form. The search results for user queries display a lot of irrelevant content since the current web does not understand the meaning and context of this information. With the introduction of Semantic Web, and its technologies that build beyond the basic HTML web pages, the data on the web can be understood by the machines such that a better cooperation between humans and machines is developed. The dream tried to achieve through this smart web is of creating an open and linked Web of Data where a common knowledge is shared among all applications across different domains.

2. Semantic Web Architecture

Architecture is a set of components that form the system, describing its structure, along with the defining properties, relationships and interfaces allowing interaction. An architecture for semantic web that offers a foundation for other models is Tim Berners Lee architecture which was first presented in 2000 [6]. This was followed by the next version proposed in on-going activities of W3C (World Wide Web Consortium) as part of the SIIA Summit in 2003 [7-9], then version 3 at WWW 2005 [10], and finally introduced version 4 at AAAI's keynote address in 2006 [11]. Berners Lee's latest proposed layered architecture (Version 4) for Semantic Web is represented in Figure 1.

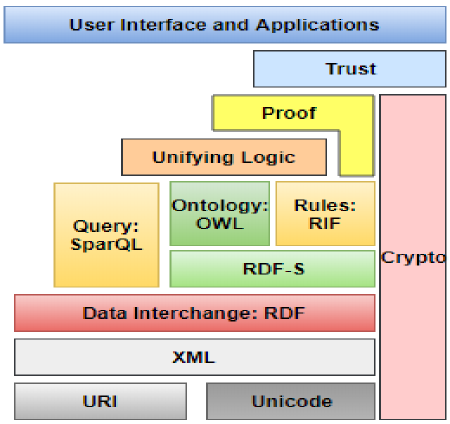

Figure 1: Version 4 of Tim Berners Lee's Semantic Web Architecture (*Source: Lee, 2006*)

The above layered architecture/ layered cake comprises of several layers which refer to various semantic web technologies which are also the key research issues or concerns like XML, RDF, Ontology, SPARQL etc.

3. Foundation Semantic Web Technologies as Research Concerns

SWTs have the capability to incorporate semantics in the vast, heterogeneous and decentralized content on the web, such that it is machine-understandable and forms a web of linked data, using the technologies of RDF, OWL, SPARQL XML, URI for effective automation, reuse and integration across various applications. Figure 2 shows the basic technologies that are required for semantic web, which are also the key research concerns.

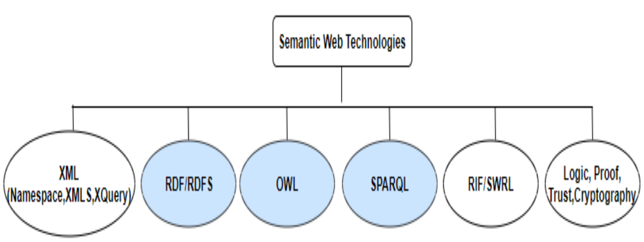

Figure 2: Basic Technologies that forms the basis of Semantic Web

3.1 URI and Unicode

Uniform Resource Identifier is the basic syntax for strings used for identifying uniquely a resource that may be a physical or abstract thing. A resource associated with URI means that a representation of that resource may be retrieved, and it may be referred to or linked with [31]. URIs can be classified as a name (URN), or as a locator (URL) which describes the primary access mechanism for locating a resource on the web (http://tools.ietf.org/html/rfc3986).
URI is generalized as IRI (International Resource Identifier), which supports the Unicode character set to facilitate representation/ encoding of the documents available in different languages. By supporting various document formats, it transcends the limitations of the traditional encoding systems.

3.2 XML

In 1996, W3C developed Extensible Markup Language (https://www.w3.org/TR/1998/REC-xml-19980210) as a standardized and easy to use syntax for storing data which describes its structure and contents. XML plays a role in realizing the semantic web by serving as a language for exchanging data that allows users to specify tags of their own. It also allows users to define terms and express relationships among them and, assert constraints for well-formed data. Unlike HTML, XML acts as the foundation that provides basic structure and rules for developing other markup languages for developing human and machine-understandable web pages [2]. However, it lacks in offering any semantics and logic capabilities on the web.

3.3 XMLS

A richer language than XML is XMLS- XML Schema (http://www.w3.org/XML/Schema) for defining in detail the content, structure and semantics of XML documents and for expressing the shared vocabularies. XMLS language provides the means for specifying the constraints on the XML documents, describe the structure of the data, encode exchanges among systems and, enable machines to execute rules made by people [40].

3.4 XQuery

The XML Query (XQuery) language uses intelligently the XML structure to express queries across kinds of data. It provides to the semantic web stack, a common query layer for discovering data, logic, documents and rules by allowing developers to use a familiar grammar (http://www.w3.org/XQuery).

It functions great with content publishers and native XML databases. Also, it is very useful for information intensive applications and for finding data in tree representations. XQuery is now becoming part of SQL standard too.

3.5 RDF

Resource Description Framework is responsible for introducing semantics into web data. The first working draft came in 1997 for defining metadata, and soon RDF became a W3C recommendation (http://www.w3.org/RDF/). It is a framework defined on top of XML for expressing the description of concepts in the form of subject-predicate-object triples. RDF is essentially a data model with the following building blocks:

Subject: Refers to the resource described by the statement.
Predicate: Refers to the subject's property talked about in the statement.
Object: Refers to the value of the subject's property in the statement.

The sentence 'The paper is written in English' for example has 'paper' as the subject, 'English' language as the object and 'written as' the predicate. These triples form RDF graphs that model the data/objects and the relationships between them, where each element is graphically represented as a node and has an associated URI, and is connected to others via relationships [40]. RDF thus provides powerful models for publishing and linking data. Example of a RDF graph model is shown in Figure 3 [33].

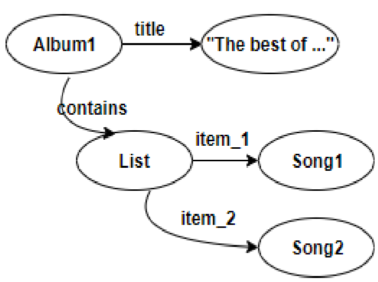

Figure 3: RDF Graph Model Example

RDF/XML (http://en.wikipedia.org/wiki/RDF/XML) is an XML based syntax which may be used to represent this standard exchange syntax for serialization of RDF. It also allows for exchanging and recording RDF relationships across web applications.

3.6 RDFS

RDF Schema extends the RDF to define vocabulary of RDF data models. It specifies the properties and taxonomy of classes of RDF-based resources. RDFS may be referred to as a semantic extension of RDF (http://www.w3.org/TR/rdf-schema/) which depicts the structure of RDF knowledge and provides mechanism for grouping related resources, and is therefore an ontology description language. Some relevant primitives for defining metadata vocabularies, as provided by the RDFS specification are [33]:

- type: this property relates a resource to a class to which it belongs. The resource thus possesses the characteristics of this class as it gets categorised as a member of it.
- Class: represents a set of things which have a common conceptual abstraction and share some characteristics.
- subClassOf: the taxonomical relations among classes are held by this property, i.e., if class C is a subclass of class S, then class C will have all the typical characteristics of class S, along with some additional characteristics that distinguish it from S [33].

3.7 OWL

In 1993, Gruber defined an ontology as "a formal, explicit specification of a shared conceptualisation" generated with the objective of knowledge sharing [18]. An ontology is basically a knowledge base that consists of a set of instances of classes. It is responsible for developing an understanding of a domain commonly shared by all web applications, and plays a key role in information exchange and re-use of knowledge across various applications [31].
Web Ontology Language (OWL) is an ontology language that captures the ontologies. It provides added constructs over RDFS and further extends the RDF vocabulary with richer typing of properties, cardinality enumerated classes etc. OWL fetches the reasoning and expressive capability of Description Logic (DL) to semantic web and has three forms that have different expressiveness degrees-OWL Lite, OWL DL and OWL Full (http://semanticweb.org/wiki/Ontology). On merging with a reasoning tool, OWL also provides logical abilities for inference and reasoning, subsumption being the chief reasoning concept, which is the subclass relationship amongst concepts or concepts' properties. Consider an example- The 'name' property of countries may be expressed by a string value as follows:

```
<rdfs:Classrdf:ID="Country">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onPropertyrdf:resource="#name"/>
            <owl:allValuesFrom
Rdf:resource=http://www.w3.org/2001XMLSchema#string/>
        </owl:Restriction>
    </rdfs:subClassOf>
</rdfs:Class>
```

3.8 SPARQL

SPARQL Protocol and RDF Query Language has emerged as the query language for manipulating and accessing RDF data. W3C's RDF Data Access Work Group (DAWG) developed SPARQL in 1998 as the standard over semantic web for retrieving data. It works over the graph patterns of RDF and also may be employed to accessing over OWL and RDFS. It is not just a query language, but also serves as a protocol for RDF Specifications- describing the remote protocol to fire queries and receive results (http://www.w3.org/TR/sparql11-query/). Example of a SPARQL query to search all names stated in Lee's FOAF file is:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name
WHERE {
    ?person foaf:name ?name .
}
```

3.9 RIF and SWRL

To specify rules beyond the available constructs, like relations that can't be described using an ontology description language's logics, RIF (Rule Interchange Format) or SWRL (Semantic Web Rule Language) languages may be used. This rules' layer facilitate proofs without the full logic machinery. To perform certain tasks on the web, agents on the web communicate with others, for which ontologies or metadata may be utilised. RIF develops an interchange format between the existing rule systems to allow such exchanges [41]. SWRL, based on a previous RuleML (Rule Modelling Language) initiative, covers the entire spectrum of rules- from transformation and derivation to reaction rules. Therefore, it may specify mappings between web ontologies along with queries and inferences in them, and dynamic behaviours of agents and services [33].

3.10 Logic, Proof and Trust

The Ontology language is further enhanced by the Logic layer for allowing application-specific declarative knowledge to be written. This layer can export the rules' codes and can validate proofs, based on its functioning on the first order predicate logic principles, such that the displayed information on the web is accurate.

The Proof layer encompasses the representation of proofs in web languages, the actual deductive processes on the basis of complex properties, as well as proof validation. This layer in semantic web aims at creating smarter content that is machine-understandable, such that assertions can be made by the machines on the data, to provide added new information, and an inference engine can provide justifications or proofs for the conclusions arrived at [41].

The trust layer sits on top of the semantic web technology stack, which aims at subjectively evaluating the trustworthiness of information by each information consumers. It then excludes information providers that are unrated or those which don't publish in a particular way the trust relevant information. Much work hasn't been done in these layers so far, and there is a need of investigating standards for applying logic, proof and trust mechanisms in semantic web.

3.11 Cryptography

Along all the technologies used in semantic web implementation, cryptography plays a significant role. Through mechanisms of digital signatures, encryption schemes, access control models, this layer aims at providing security in the semantic web- by restricting information access to authorized agents only, digitally signing and validating documents for maintaining integrity and verifying sources, encrypting exchanged information for maintaining privacy etc.

4. Other Significant Technologies as Semantic Web Research Concerns

Figure 4 shows some other significant technologies that are associated with semantic web research concerns which are discussed as below:

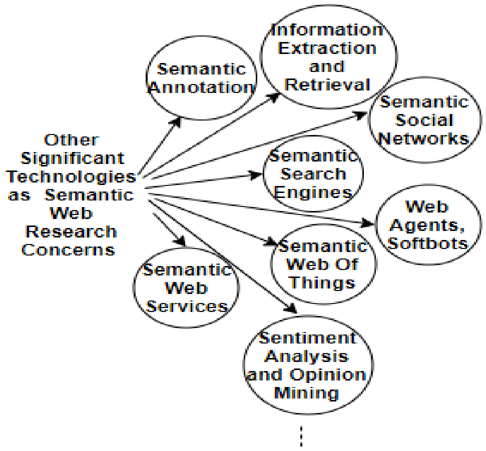

Figure 4: Other significant technologies as Semantic Web Research Concerns

4.1 Semantic Web Services

A Web Service is a program accessible through web, which directly, or through a third-party, provides access to a web-based application to a user. For aligning the huge mass of incompatible web services semantically, it is required to annotate and formalize their interfaces and relations between them; and this metadata may then be logically processed based on their semantic meaning to interpret the services' descriptions/ functionalities non-ambiguously [38]. The types of semantics associate with web services are data, non-functional, functional and execution semantics [28]. Semantic web services (SWS) were introduced in 2001, with the goal of providing a shared, common knowledge coming from heterogeneous and disparate sources. SWS allow software agents to exploit the user's request for automatic web services' discovery, composition, selection, invocation, monitoring and interoperation. Their immediate advantage also includes dynamic integration and smooth cooperation of different platforms. SWS are built upon the WSDL (Web Services Description Language) and SOAP (Simple Object Access Protocol) technologies [27]. Several languages and frameworks for describing (marking-up) a web service formally exist presently, such as OWL-S (Web Ontology Language for Services), SAWSDL (Semantic Annotations for the Web Services Description Language), WSMO (Web Service Modelling Ontology) etc. [38].

Various challenges are also introduced by SWS, quality being the foremost. The service developer needs to perform tests to ensure that it is correctly implemented and its quality from the consumer's perspective also needs to be assured by testing. Compared to the web services which are just syntactically described, SWS's characteristics like dynamic service composition further raise the challenges of testing, where tests needs to be performed over the semantic layer [36].

The SWS field has been identified as one of the most promising emergent research area in semantic web, which presents both wide-ranging commercial potential as well as substantial attention from the research community [38]. The future scope of SWS is to develop SWS concept based real distributed applications that gain high flexibility, reuse and interoperability levels [27].

4.2 Information Extraction and Retrieval

Information Retrieval (IR) refers to the method of intelligently filtering out relevant information from the enormous volumes of data on the web. Special extraction rules, known as 'wrappers' handle the extraction of meaningful information from unseen collection of pages, such that time is saved and results retrieved are accurate. Information Extraction (IE) is a subfield of IR and refers to the process of extracting structured data from the semi-structured or unstructured sources, i.e., process the text in natural language based on the user's information need, and present the retrieved information in the form of ontologies (classes and relationships among them). Different IE techniques include semantic annotation, web usage mining, web scrapping [32].

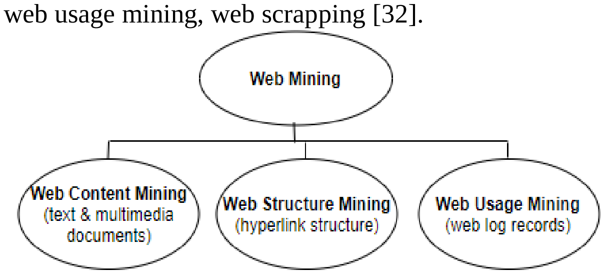

Figure 5: Types of Web Mining

Web Mining refers to the application of technique of data mining on the web, with the objective of extracting useful information from the patterns discovered in either the web structure, web content or web usage records, as shown in Figure 5. Web Scrapping is another technique of extracting web data automatically to retrieve useful information from it by parsing the web pages with special coded programs that carry out manipulations like converting the format of the web page etc. This is similar to web indexing, and much work hasn't been explored in this direction yet.

An IR model provides the mathematical framework to define the search process. The search task in IR however faces some issues like: i) large scope of search task, which involves huge collection of data and requires results speedily and with accuracy, ii) expression of the user's query in an efficient form, formulated using operators and taking the user's context and linguistic features into consideration, iii) effective evaluation of retrieved results by ranking/ordering them according to relevance to user's needs using performance measures like recall, precision etc. [32]. These issues are open research areas in IR.

Including semantics in IR involves the role of ontologies, which may help in expanding the user queries by matching the query terms with the ontologies' concepts. User profile ontologies can support personalized information systems and can return different results for the same query terms based on the user's personal interests. The ontologies may also aid in the learning of retrieved results. This provides another direction of research where ontologies may be designed for formalizing the concept relationships and identifying associations among the concepts. These correspondences can help broaden the search by rewriting the user's query, and thus help fetch better, relevant results [32].

4.3 Semantic Annotation

Semantic Annotation (SA) is the process of enriching content with additional attached information which is processable by machines. Semantics can be added to various concepts on the web in semi-structured or unstructured documents. It is the technique of tying together natural language and semantic models by assigning to the entities and their relations, the links to their semantic description in an ontology [25]. Annotations that are based on a common ontology may provide a common framework for integrating information from heterogeneous sources. The basic steps followed by SA of resources involves: i) entity identification, ii) entity disambiguation, iii) relationship extraction, iv) entity annotation and, v) semantic graph database indexing and storing. Besides the content's semantic tagging, various applications also provide features to support usage of annotation, access APIs, user interfaces, annotations' storage. The techniques for information extraction used by these applications vary from grammar rule based, to discovering patterns or using machine learning [28]. Different SAPs (Semantic Annotation Platforms) for manual or semi-automated SA are based depending on the kind method used for annotation-pattern based or machine learning based [29]. Although SA is an error prone and difficult task, its automatic implementation enables various applications such as categorization, highlighting, unambiguous resource discoveries, and much better search capabilities [28]. The creation of fully automatic SA remains an unsolved issue. SAPs that are created with extensible architectures can acclimate to the technology as it progresses. For realizing semantic web, the continued evolution of SAPs by extending the existing annotations along with providing new and better features is vital [29].

4.4 Semantic Search Engines and Browsers

The traditional search engines are keyword based and face various limitations like ambiguity in keywords resulting in retrieval of irrelevant data, synonymous words resulting in missing out some relevant documents, incapability of recognizing and satisfying semantic constraints. These issues may be resolved by semantic search by allowing word alternatives and recognizing context and specified conditions [39]. Semantic search engines thus provide an improved search on the web where both the structure and meaning are extracted from the user's web query and the web content. It is trivial to correctly interpret the semantics incorporated in documents and user queries to avoid negative impact of misinterpretations [13]. Such semantic search involves semantic web technologies for interpreting search queries and evaluation using reasoning over the web. They also allow search on the web using natural language queries.

Instead of isolated evaluations of semantic search technologies, comprehensive evaluations should be performed using initiated series like semantic search evaluations (SEALS) or SemSearch challenges etc. which foster development and research. In addition, an agreement on the approaches for evaluating semantic search tools is also a key aspect for standardization, since user-centric activities such as search are critical for user-satisfaction [13].

Semantic Web Browsers can also be built as standalone systems or as lightweight extensions to the existing web browsers to allow browsing the semantic web structured and linked data in the form of RDF along with heavy dependence on metadata, instead of only HTML documents with URIs on the previous web. For this reason, such browsers are also termed as 'hyperdata browsers'. Examples include Marbles, Ripple etc., but there is still a need to work on scalable and efficient semantic web browsers.

4.5 Intelligent Web Agents, Crawlers and Softbots

A web agent is an intelligent software system that, with some degree of autonomy, carries out operations on behalf of the user/other program in a goal-directed manner by employing knowledge and/or interacting with other agents. Intelligent agents perceive changes and may respond to them in a timely fashion, thus they are flexible and reactive, and may also proactively provide smart support [30]. Other characteristics of web agents for accomplishment of tasks include accuracy of provided information, cooperation and negotiation with other agents, accessing domain knowledge, composition of

meaningful messages, reasoning about tasks, decision-making, and learning and improving performance over time. The semantic analysis based on web agents also enables documents in natural language to be understood by machines [30] [33].
Web Crawlers are the software programs that traverse and fetch information from information space of WWW in an automated way by following the hyperlink structures using HTTP protocol. They are required to maintain high freshness of the pages in collection.
Softbot (software robot) or Internet robot is an artificially intelligent agent that uses knowledge and inferences to satisfy user's request. It acts as an intelligent personal assistant and can handle expressions of objectives in first-order logic. Softbots provide threefold advantages of providing an integrated interface to the WWW, dynamically deciding the facilities to be invoked, and backtracking fluidly the information collected at run-time from one facility to another. Softbots can further tolerate errors, ambiguity and omissions [15].

4.6 Sentiment Analysis and Opinion Mining

Sentiment Analysis refers to the identification, classification and understanding of sentiments published on social media in the form of user reviews, comments, opinions, and emotions. Taking text as input and outputting whether it is opinionated, what the opinion is about, its degree of strength etc. is provided by various opinion mining tools. While sentiment analysis is used for polarity detection, i.e. whether the sentiment being analysed is negative or positive, opinion mining is more generic, however, the terms may be used interchangeably. Opinion mining tools can be employed in various industries for predictive analysis targeting political opinions, or stock market predictions or social analytics to draw important inferences. These fall in the scope of natural language processing (NLP) since the user text is expressed using natural language. This gives rise to challenges of understanding the complex linguistics-implicit associations of data, sarcasm, conditional sentences, context, assumptions, unusual terms etc. [26].
To go beyond analysis of words and instead focus on approaches of semantic analysis which are based on linked data, ontologies and other semantic resources, refers to concept-level sentiment analysis. This approach also relies on natural language concepts' implicit features. An example of one such resource is SentiWordNet, which adds semantic information to every word and thus allows variants and synonyms to be found by linking of sentiment words [26].
It is trivial to adapt tools that deal with emotions expressed on social media in order to drive research in this field. The development of such tools is hampered by the need to integrate the sentiment analysis' linguistic resources with semantic concepts. An example of one such initiative in this context is the LLOD (Linguistic Open Data Cloud) [26].

4.7 Semantic Social Networks

Huge social networks are being formed by the increasing Web 2.0 popular sites. For the study of networks, a specialization of Social Network Analysis (SNA) is used for providing a mathematical and visual analysis of these online networks among users. SNA uses graph theory methods to identify user interactions and manage the lifecycle and predict evolution of social networks. The network's nodes represent people while their relationships are represented by the links between nodes [14]. One of the key issues is of locating the prominent nodes in the network, for which SNA uses different centrality measures as presented in Figure 6.

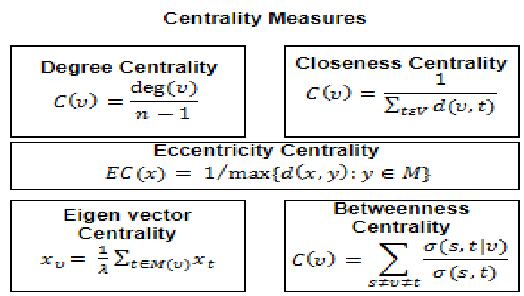

Figure 6: Various Centrality Measures in Semantic Social Networks (*Source: https://en.wikipedia.org/wiki/Centrality*)

By leveraging the social data semantics in a format that is machine understandable, new perspectives of SNA open up and online social experiences are enhanced. Social networks are fundamental in realizing a web of trust by enabling trustworthiness and credibility estimates for information in the context of semantic web [14]. Semantic social networks associate ontology engineering, microformats (OWL, RDF, RIF etc.) and semantic web languages with dynamic or static SNA for producing relevant knowledge related to characterising communities and people within a social network. With only a few significant works in this field, there is a need for various innovative applications that make SNA outlines operational using RDF graphs, ontologies and semantic web languages [37].

4.8 Semantic Web of Things

The first phase of moving towards a semantic web of things (SWoT) is the interconnection of everything to the internet. Next, a semantic interoperability is to be enabled among these existing heterogeneous entities using IPv6 (Internet Protocol version 6) and providing connectivity using solutions like 6LoWPAN (IPv6 over Low-Power Wireless Personal Area Networks). This creates the IoT (Internet of Things), which then needs connecting these things to the web using standard web solutions to conceive the web of things (WoT). It allows the various systems and things to interact, even constrained devices using lightweight protocols like CoAP (Constrained Application Protocol), thus allowing more complex services and solutions to be composed [21]. Powering the IoT with web technologies offers not just web access, but also web linking for description of resources, resource discovery, resource allocation, as well as security, and is therefore a vehicle which provides communication along with interoperability and integration. The major driver for this interoperability and better ability to understand things is driven by semantics [35]. Semantic Web of Things is a seamless extension to IoT which allows integration of the digital entities and smart objects and ensures a common understanding, providing web-scale re-use and sharing of things, with the fusion of IoT, WoT concept, REST (Representational State Transfer) architecture, and semantic web technologies. The vision of SWoT enables autonomic knowledge-based systems for information management, storage, and providing access to sources of information transparently [21].

The huge amounts of data generated from the IoT/WoT may be exploited through SWoT using semantics based capabilities of intelligent data analysis and mining [21]. This integration of semantic technologies in IoT also benefits the representation of the things' capabilities and semantic annotation of the data produced by these things [35]. However, some challenges are associated with this transition from WoT/IoT to SWoT such as defining a universally understandable common description, creation of semantic annotations that are extensible and, agreement on the ontologies. Nevertheless, WoT, and consequently SWoT will be the key drivers for IoT convergence [21].

5. Research Challenges of Semantic Web

Various issues are faced in the implementation of different technologies linked with semantic web, which form the open research challenges for future work, some of which are shown in Table 1 and discussed below.

Table 1: Some significant Research Challenges associated to Semantic Web Technologies

TECHNOLOGIES	CHALLENGES
·RDF, Ontology and Information·Access	·Scalability of Semantic Web Content
·Ontology	·Availability, ·Development and ·Evolution
·Web Services	·Semantic Web Services' Implementation
·Information Retrieval	·Open Knowledge Extraction
·Opinion Mining	·Semantic Sentiment Analysis
·Trust, Proof, Cryptography	·Security Issues
·Query	·Question Answering over Linked Data
·RDF Stores/Knowledge Bases	·Massive Storage
·RDF, Ontology, Query and·Services	·Multilingual Access
·SPARQL	·Query Performance
·RDF, OWL, Rules, Query	·Programming and ·Mathematical Representation

- Scalability of Semantic Web Content:

With the heavy dependence on web in present times, the semantic web content is also growing. It needs to be organized in a scalable manner to be able to provide efficient search and management of the shared web content. Also, for easy development of applications, it is necessary to aggregate the content. However, the underlying semantics of the ontology-based annotated pages can't be exploited fully by the hyperlinked configuration of the existing web, and the usage of semantic indexes for grouping of semantic web content according to topics may be useful. Aggregation of such huge content on a global scale remains a difficult task [2].

- Ontology Availability, Development, and Evolution:

Ontologies form the building blocks of semantic web, since they are the carriers of knowledge. However, it is not easy to agree to a common ontology for a particular domain, nor is it to design or reuse and maintain existing ontologies. Standard tools for designing ontology need to be developed and the developers need to be trained and skilled for utilizing these tools for creating effective technologies. All this implies extended development times and higher costs as compared to the original web [2].

- Semantic Web Services' Implementation:

SWS are self-described services and can get together data and programs by means of automated discovery, invocation, maintenance. There exist numerous different languages and frameworks to formally describe a web service, varying from ontology modelling to semantic annotations; however, developing a standard which is uniformly accepted by venders remains a challenge [2].

- Open Knowledge Extraction (OKE):

Providing open extraction of knowledge requires defining of a reference evaluation framework for carrying out research on extracting knowledge from text for semantic web by re-defining the typical information and knowledge extraction steps such that they take specific semantic web requirements into account. This needs collaboration between the knowledge extraction and semantic web communities for further investigating the overlaps between them and growing semantic web further. So, achieving OKE is a challenge. The lack of easy-to-use knowledge acquisition interfaces also poses a challenge [12].

- Semantic Sentiment Analysis:

Extensive use of social media provides users a means to express their thoughts and share their opinions on a wide array of subjects. These sentiments can be analysed and exploited to derive useful insights. Lexicon-based techniques, machine learning etc. are typical approaches implemented by sentiment analysis engines. However, approaches for semantic sentiment analysis, that take huge semantic knowledge bases into account, implement best practices of semantic web, and may potentially deliver higher performance, are still under experimental evaluation. Realizing these is a challenge and a core understanding of texts in natural language, along with their related semantics, is required, to be able to analyse the inherent semantics characteristics associated with natural language concepts [12].

- Security Issues:

Concerns related to semantic web security include- maintaining privacy of shared data, controlling level of access for all agents, security of entailments, integrating disparate data with varying security needs, ensuring trust among interacting agents, and proof of reliability of logic and results [24]. The existing works around semantic web security, privacy and policy have very few works that consider the potential of semantic web technologies to address issues of distortion, appropriation, i.e. information dissemination, and don't consider the challenges associated with invasion [22]. To progress the condition of semantic web, it is essential to develop standards which maintain user's faith in the correctness of information. Promoting the semantic web content's trust and proof also takes up time and constantly needs to be updated to remain valid [24].

- Question-Answering over Linked Data:

Querying over linked data requires translation of a user's information need into a structure which, using typical semantic web query processing, optimising and inferencing techniques, can be evaluated. There is growing research on this retrieval of answers to natural language keywords or questions over RDF datasets, to try and shape an interaction paradigm which facilitates the end users to benefit from the semantic web standards' expressive power, while concealing their intricacy behind a user-friendly and intuitive interface [12].

- Massive Storage:

The continual growth in the volume of linked web data raises the need for storage strategies that allow ingestion of enormous data flows, facilitate efficient and interactive querying, and enable browsing and analysis of large-scale industry datasets. Devising systems for applications that are based on linked data which are capable of achieving satisfactory performances on real data loads is thus of prime importance [12].

- Multilingual Access:

There are valuable resources that are written in languages other than English, although English remains the chief language for web documents. Multilingual access has a growing role at the level of metadata, ontologies, and application interfaces, requiring mappings and translations, and needs to be addressed [2].

- Query Performance:

With the humongous content on the web, the SPARQL queries issued on the data need to be processed in a fast and efficient manner. Large, composite queries need to be optimized and checked for performance over distributed, huge data loads, which remains a work in progress.

- Programming and Mathematical Concerns of Semantic web Data:

The nature of the data published by semantic web significantly differs from the kind of data that users deal with in case of established approaches of databases, which poses a challenge for programming with semantic web data. Programming approaches like semantic search or graph programming partially account for the assumptions and challenges faced with dealing with semantic web data, and while aspects of selection, fetching and programming can be handled using peculiar techniques, there is a need of approaches that completely consider the characteristics of semantic web data and aim to reduce the resisting mismatches between data engineering and the programming approaches, for the semantic web to reach its full potential [34].
Integration of mathematical sources with associated metadata (annotations) opens up new possibilities which allow better inferencing, more powerful search, similarity search capabilities, and possibility of algebraic manipulation etc. However, various fields of mathematics still haven't been realized as proper ontologies in semantic web, and there is a need to identify significant aspects of mathematical knowledge which can be integrated with the mainstream semantic web technologies to achieve this [23]. Programming and mathematical concerns are the foundation of semantic web which need to be explored further.

6. Conclusions and Future Scope

This paper gives a useful insight into the semantic web and its technologies, which are the concerns for research, towards the goal of achieving an intelligent web.

Despite the identified use of varied technological aspects of semantic web, substantial room for growth still remains. Although there have been huge progresses on the part of Facebook's Open Graph, Schema.org and others, the vision of a complete interoperable web of data hasn't still been realized. Most companies are not aware or have adopted semantic web technologies, except for a few select frontrunners. Also, the learning curve is steep for users who are new to the concepts, since limited resources are available currently which discuss how and when to apply the semantic web technologies to real world scenarios. It is believed that over the course of next few years, the success in the discussed semantic web technological aspects as open research challenges is vital for the web of data and its widespread adoption across enterprises and among end users, which means that various semantic web technologies act as foundation and backbone for pursuing research in semantic web, which is a challenge for reaching an intelligent web.

References

[1] Aghaei, S., Nematbakhsh, M. A., &Farsani, H. K. (2012). Evolution of the world wide web: From WEB 1.0 TO WEB 4.0. International Journal of Web & Semantic Technology, 3(1), 1.

[2] Alesso, H. P. and Smith, C. F. (2008). Thinking on the Web: Berners-Lee, Godel and Turing. Wiley-Interscience.

[3] Al-Feel, H., Koutb, M. A. and Suoror, H. (2009). Toward An Agreement on Semantic Web Architecture. International Journal of Computer and Information Engineering, 3(1), pp. 806-810.

[4] Berners-Lee, T., Hendler, J. and Lassila, O. (2001). The semantic web. Scientific American, 284(5), 28-37.

[5] Berners-Lee, T. and Fischetti, M. (1999). Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor.Harper& Brothers Publishers.

[6] Berners-Lee, T. (2000). Semantic Web – XML 2000. W3C Web site (Last accessed 11/8/2006). URL: http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html

[7] Berners-Lee T. (2003). The Semantic Web and Challenges. W3C Web site Slideshow (Last accessed 10/10/2006). URL:http://www.w3.org/2003/Talks/01-sweb-tbl/Overview.html

[8] Berners-Lee T. (2003). Standards, Semantics and Survival. SIIA Upgrade, pp. 6–10.

[9] Berners-Lee T. (2003). WWW Past and Future. W3C Web site (Last accessed 10/10/2006). URL: http://www.w3.org/2003/Talks/0922-rsoc-tbl/slide30-0.html

[10] Berners-Lee T. (2005). WWW 2005 Keynote. W3C Web site (Last accessed 10/10/2006). URL: http://www.w3.org/2005/Talks/0511-keynote-tbl/

[11] Berners-Lee T. (2006). Artificial Intelligence and the Semantic Web.AAAI 2006 Keynote, W3C Web site (Last accessed 12/10/2006). URL: http://www.w3.org/2006/Talks/0718-aaai-tbl/Overview.html

[12] Blomqvist E., Solanki M. and Dragoni M. (2017). Semantic Web Challenges. Proceedings of 4th Semantic Web Evaluation Challenge at ESWC, Springer.

[13] Elbedweihy, K. M. et. al (2015). An overview of semantic search evaluation initiatives. Web Semantics: Science, Services and Agents on the World Wide Web, 30, 82-105.

[14] Erétéo, G., Gandon, F., Corby, O., and Buffa, M. (2009). Semantic social network analysis.arXiv preprint arXiv:0904.3701.

[15] Etzioni, O., and Weld, D. (1994). A softbot-based interface to the internet. Communications of the ACM, 37(7), pp. 72-76.

[16] Gerber, A., Merwe, A. and Barnard, A. (2007). Towards a Semantic Web Layered Architecture. IASTED International Conference on Software Engineering, pp. 353–362.

[17] Gerber, A., Merwe, A. and Barnard, A. (2008). A functional semantic web architecture. The Semantic Web: Research and Application, pp. 273-287.

[18] Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. Presented at the Padua workshop on Formal Ontology, March 1993, later published in 1995 International Journal of Human-Computer Studies, 43(4-5), pp. 907- 928.

[19] Harb, I. H., Abdel-Mageid, S., Farahat, H. and Farag, M. S. (2010). Enhanced semantic web layered architecture model. In Proceedings of the 10th WSEAS international conference on applied informatics and communications, and 3rd WSEAS international conference on Biomedical electronics and biomedical informatics, pp. 341-347.

[20] Horrocks, I., Parsia, B., Patel-Schneider, P. and Hendler, J. (2005). Semantic Web Architecture: Stack or Two Towers? PPSWR, 3703, pp. 37-41. Springer.

[21] Jara, A. J. et.al (2014). Semantic web of things: an analysis of the application semantics for the iot moving towards the iot convergence.

International Journal of Web and Grid Services, 10(2-3), 244-272.

[22] Kirrane, S., Villata, S., and d'Aquin, M. (2018). Privacy, security and policies: A review of problems and solutions with semantic web technologies. Semantic Web, (Preprint), pp. 1-10.

[23] Lange, C. (2010). Integrating Mathematics into the Web of Data. Linked Data in the Future Internet, Future Internet Assembly, pp. 12-16.

[24] Malik, N. and Malik, S.K. (2018). Security in web semantics: A revisit. In 5th International Conference on Computing for Sustainable Global Development.IEEE.

[25] Malik, S. K. and Rizvi, SAM. (2014). An intelligent web framework based on ontology design. Ph.D Thesis, Guru Gobind Singh Indraprastha University.

[26] Maynard, D., Bontcheva, K., and Augenstein, I. (2016). Natural language processing for the semantic web. Synthesis Lectures on the Semantic Web: Theory and Technology, 6(2), 1-194.

[27] Nacer, H., and Aissani, D. (2014). Semantic web services: Standards, applications, challenges and solutions. Journal of Network and Computer Applications, 44, 134-151.

[28] Nagarajan, M. (2006). Semantic annotations in web services. In Semantic Web Services, Processes and Applications, Springer, pp. 35-61.

[29] Reeve, L., and Han, H. (2005). Survey of semantic annotation platforms.In Proceedings of the 2005 ACM symposium on Applied computing, ACM, pp. 1634-1638.

[30] Rudowsky, I. (2004). Intelligent agents.The Communications of the Association for Information Systems, 14(1), 48.

[31] Shadbolt, N., Berners-Lee, T., and Hall, W. (2006). The semantic web revisited. IEEE intelligent systems, 21(3), pp. 96-101.

[32] Singh, A., Dey, N., Ashour, A. S., and Santhi V. (2017). Web Semantics for Textual and Visual Information Retrieval, IGI Global.

[33] Sreeja, M. U. (2011). Agent Based Semantic Web. Seminar Report, CUSAT. http://dspace.cusat.ac.in/jspui/handle/123456789/3620

[34] Staab, S., Scheglmann, S., Leinberger, M., &Gottron, T. (2014, May). Programming the Semantic Web.In European Semantic Web Conference, Springer, pp. 1-5.

[35] Szilagyi, I., and Wira, P. (2016, October). Ontologies and Semantic Web for the Internet of Things-a survey.In Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE, pp. 6949-6954.

[36] Tahir, A., Tosi, D., and Morasca, S. (2013). A systematic review on the functional testing of semantic web services. Journal of Systems and Software, 86(11), pp. 2877-2889.

[37] Thovex, C., and Trichet, F. (2013). Semantic social networks analysis. Social Network Analysis and Mining, 3(1), 35-49.

[38] Wang, H. H., Gibbins, N., Payne, T., Patelli, A., and Wang, Y. (2015). A survey of semantic web services formalisms. Concurrency and Computation: Practice and Experience, 27(15), 4053-4072.

[39] Zitzelberger, A. J., Embley, D. W., Liddle, S. W., and Scott, D. T. (2015). HyKSS: Hybrid keyword and semantic search. Journal on Data Semantics, 4(4), 213-229.

[40] S. Decker, D. Fensel, et. al, "The semantic web: The roles of XML and RDF", IEEE Internet computing, Vol. 4(5), pp. 63-73, 2000.

[41] G. Antoniou, F. von Harmelen: A Semantic Web Primer, MIT Press, Cambridge, 2004.