Present and future of semantic web technologies: a research statement

Article *in* International Journal of Computers and Applications · January 2019

DOI: 10.1080/1206212X.2019.1570666

2 authors:

Archana Patel
Eastern International University

Sarika Jain
National Institute of Technology, Kurukshetra

International Journal of Computers and Applications

ISSN: 1206-212X (Print) 1925-7074 (Online) Journal homepage: https://www.tandfonline.com/loi/tjca20

Present and future of semantic web technologies: a research statement

Archana Patel & Sarika Jain

INTERNATIONAL JOURNAL OF COMPUTERS AND APPLICATIONS
https://doi.org/10.1080/1206212X.2019.1570666

Present and future of semantic web technologies: a research statement

Archana Patel and Sarika Jain

Department of Computer Applications, National Institute of Technology, Kurukshetra, India

ABSTRACT

Semantic web and its technologies have been eyed in many fields. They have the capacity to organize and link data over the web in a consistent and coherent way. Semantic web technologies consist of RDF schema, OWL, and rule and query languages like SPARQL and these technologies will help the various domains to resolve their problems. This review paper starts with analyzing the nature of semantic web and its requirements. We have considered all 10 domains which are closely related to semantic web and its technologies. For a better understanding of the paper, we have separated it into three major contributions. First, we analyze semantic web and those domains that increase the growth of the semantic web. Second, we discuss all domains where semantic web technologies play a vital role. Third, we emphasize those domains that go hand in hand with semantic web technologies. This review paper will be utilized as an unbiased direction for the researchers.

ARTICLE HISTORY
Received 1 November 2018
Accepted 13 January 2019

KEYWORDS
Semantic web; OWL; SPARQL; linked data; ontology

1. Introduction

The Semantic Web means sharing data and facts rather than sharing the text of a page. The thought of a Semantic Web was given by Sir Tim Berners-Lee in 2001. The Semantic web helps build a technology stack to support a 'web of data' rather than a 'web of documents.' The final aim of the web of data is to provide capacity to the computer to do more meaningful tasks and to develop systems that can support trusted interactions over the network. Semantic web technologies (SWTs) include different data interchange formats (e.g. Turtle, RDF/XML, N3, N-Triples), query languages (SPARQL, DL query), ontologies, and notations such as RDF Schema and Web Ontology Language (OWL), all of which are intended to bestow a formal description of entities and correspondences within a given knowledge domain. These technologies are helpful for achieving the overall objective of the semantic web. The heart of the semantic web is the linked data because linked data provide large-scale data integration and reasoning on the data. Linked data become powerful by technologies such as SPARQL, RDF, OWL, and SKOS, but there are also many challenges for linked data which are described by various papers. Ontologies are the backbone for structuring linked data and play a major role in defining links within a dataset and across datasets to other linked data. They enable users to search a schematic model of all data within the applications. By using ontology we can combine deep domain knowledge and raw data and bridge datasets across domains. Ontologies are efforts to more precisely classify parts of the data and to permit communications between the data available in distinct formats.

The universal standard for communicating ontologies and data on the Semantic Web is web ontology language. The database is not openly suitable in the area of the semantic web because it holds a dissimilar data model. Most database benchmarks are designed in the direction of a relational data model. The mathematical idea behind the relational data model is the set theory which is a part of the Cartesian product. The web ontology language data model, on the opposite hand, gives a lot of adaptability. The resource description framework (RDF) is based on the idea of graph theory. Furthermore, web ontology language is based on description logic; it includes description logic (DL) expressions and axioms or restriction. Knowledge graph is also an essential component for the semantic web. The term 'Knowledge Graph' was coined by Google in 2012 and is intended for any graph-based knowledge. There are many types of knowledge graph available such as DBpedia, Freebase, OpenCyc, Wikidata, YAGO and so on. Ultimately, comprehensive knowledge bases like DBPEDIA and WIKIDATA play an essential role in dealing with the problem of information overload. The thought of acquainting semantics to quest on the Web is not clear in an exclusive manner. Other factors like scalability, availability of content, visualization, ontology development and evolution, and multilingualism and stability of semantic web languages are the major challenges for the semantic web, which provides directions for the researcher. The two most ordinary behaviors of semantic web technologies are (1) to understand Web queries and Web resources annotated along with background knowledge defined by ontologies and (2) to look into the organized huge datasets and knowledge bases of the semantic web as an option or a complement to the present web. The vast applications of semantic web technologies make possible the provision of benefits to other domains such as sensor network, big data, cloud computing, Internet of things, and so on.

This paper discusses various problems and research directions for the semantic web. Particularly, it provides a detailed

CONTACT Archana Patel archana_6170005@nitkkr.ac.in Department of Computer Applications, National Institute of Technology, Kurukshetra, India

guide to all the domains related to the semantic web and its technologies. The goal of this work are three-fold: (1) to discuss all the domains which help increase the growth of the semantic web, (2) to introduce those domains where semantic web technologies have been applied, and (3) to describe those domains that go hand in hand with the semantic web. The survey is as self-contained as possible, and thus also serves as a good tutorial for newcomers to this fascinating and highly topical field. The remainder of the paper is structured as follows: Section 2 introduces a review of related works. Section 3 presents semantic web and its technologies. Section 4 describes those domains that increase the efficiency of the semantic web. Section 5 discusses all domains where semantic web technologies have been applied. Section 6 describes those domains that benefit each other. The last section concludes the discussion.

2. Review of related works

Many authors have presented surveys about the domains related to the semantic web, but most of them pick one domain and discuss how that domain relates to semantic web technologies and define all the aspects of that domain. At the end of the paper, they conclude how they resolved the problems of the semantic web by using that domain. In this section, we discuss all the related surveys on the semantic web and its related domains and lastly explore how the art of our survey is different from existing surveys. Many researchers have played a vital role in the field of the semantic web. Bittencourt et al. [1] have described some major research directions for semantic web and education. They have opined that the use of the semantic web in education systems can help in the achievement of 'Anytime, Anywhere, Anybody Learning' because it uses the concept of ontologies which provide information in a computer-understandable manner. The arrival of semantic web technologies in tourism information systems facilitates the management and interoperation of semantically diverse data and provides accurate and flexible information. Markellou et al. [2] have proposed a framework for personalized e-Learning on the basis of aggregate usage profiles and domain ontology. The whole process is divided into two stages: offline stage (data preparation, ontology creation, and usage mining) and online stage (production of recommendations). Semantic web technologies such as RDF, ontologies, and OWL provide flexible representation of knowledge. These technologies are used in geographical information systems for searching and classifying feature. Some authors have used the concept of semantic web technologies for the earthquake recommendation system. They have stored earthquake-related information into ontology and then provided inference over it by using SPARQL queries. Kenekayoro [3] has presented the future of the semantic web by reviewing the technologies of the traditional web, semantic web, and their areas of application. He has pointed out that migration from the current web to the semantic web is a particularly slow process because publishing data for the semantic web is not as simple as publishing in HTML. The main problem is not only accessing and processing information, but also extracting and interpreting. By using ontologies and intelligent services, we can transform the web of information into the semantic web. Abaalkhai et al. [4] have discussed 18 ontologies on the basis of different states of human behavior like mood, emotion, needs, and so on. In their view, ontology represents real-world knowledge in such a format that a computer can easily process. Ontology is an idea that empowers enhanced machine processing because structured documents are present on the web in such a manner that they become understandable by computers.

Bizer et al. [5] have explained the progress of the linked data on the web along with their applications. They provide guidelines for the researcher and illustrate different methods for publishing the data on the web. Encoding, publishing, and interlinking between data are an example of linked data that can be obtained from semistructured sources. Bizer et al. have explained the problems related to the implementation of the global linked data and nontechnical barriers to the implementation of web 3.0. Harth et al. [6] have written a chapter on linked data and the standards of the semantic web. They have explained all the steps of the semantic web stack and provided a comprehensive overview about RDF terms, triples, graphs, vocabulary, syntax, query clause, etc. Semantic approaches arrange all the data over the web in a regimented way. They represent a method by which we map the data over the web via ontology and then access the data through an agent. By using semantic web technologies, we access domain knowledge easily, and semantic annotation is an important step for data mining. Kumar [7] has argued that interoperability among applications is a complex issue. He has described how best we can employ SWTs to accomplish interoperability and security issue among applications. Some authors have presented a study for social network analysis by using big data and semantic web technologies. Social networking services like Facebook deal with huge amounts of dissimilar data that act as an input for social network analysis by researchers. The main aim of their analysis is summarization of all techniques and challenges along with research directions that exist in social network analysis. Various authors have offered an exhaustive survey of the Semantic Web in the fields of Data Mining and Knowledge Discovery along with an outline of semantic web approaches in different platforms of the knowledge discovery process. Dhenakaran and Yasodha [8] have presented a review paper on semantic web mining. The target of these authors is to provide a sketch of web mining, its classifications, and its subtasks and present their point of view to the research community about the potential of applying techniques to mine meaningful patterns. They have additionally provided a study of the recent works and compared traditional web applications with semantic web applications and presented a direction for future research in the field of semantic web mining. Some authors have also provided all ongoing research in semantic web mining and pointed out the obstacles faced by the researcher. Stumme et al. [9] have presented a review paper and potential guidelines for semantic web mining. They have provided scenarios where semantic web mining will take place and inferred how the semantic web can enhance the consequences of web mining. Many researchers have focused on mapping between relational databases to RDF and defined reference architecture. Authors have analyzed all the approaches of semantic web and web mining areas and finally discussed all the mining tools. Sridevi and UmaRani [10] have written a review paper on semantic-based solutions for web mining. They have said that the task of retrieval of relevant documents is difficult because of the availability of heterogeneous information. To access relevant information, machines need additional semantic information and this has been made possible by the semantic web which is an intelligent web. Berendt et al. [11] have presented a whole scenario of web mining and described all semantic web languages that support knowledge representation and reasoning, the entire description of the document content, and standardization.

Huang et al. [12] have used an artificial neural network approach and proposed a new algorithm for ontology matching. They have avoided the existing problems of rule-based and learning based matching algorithms of biological ontologies. Caliusco and Stegmayer [13] have presented a chapter on 'SWTs and Artificial Neural Networks for Web information Source detection.' They have described the advantages of integrating Artificial Neural Networks (ANNs) and SWTs. Chen et al. [14] have discussed the point that semantic web accommodates computational intelligence (CI) which adds Evolutionary Computation (EC), fuzzy logic, and ANN. They have said that we can use Supervised ANNs for Ontology Alignment and Unsupervised ANNs for Ontology Learning. The fuzzy logics are helpful in improving query results in the Semantic Web. Yu and Chen [15] have offered a survey on web scale semantic information processing for cloud computing. They have presented a summary of existing technologies for semantic information processing in a cloud computing environment and described parallel processing, storage, query, and reasoning. The challenges of cloud computing can be overcome by semantic web technologies. Sheth and Ranabahu [16] have discussed those areas where semantic models can support cloud computing. They have described that semantic models can work in three dimensions of cloud computing, namely data modeling, functional and nonfunctional definitions, and service description enhancement. Androcec et al. [17] have presented a review paper on cloud computing ontologies and their applications. They have summarized the selected studies into four main categories: Cloud interoperability, Cloud services discovery and selection, Cloud security, and Cloud resources and service description. At the end of the paper, they have given research directions including those that are closely related to the security and interoperability of Cloud Computing offerings.

Kotis and Katasonov [18] have written a paper on 'Semantic Interoperability on the web of Things (WoT)' and proposed ontology learning and alignment methods. They have specified the requirements of semantic registration, coordination, and the retrieval of things. Zeng et al. [19] have presented a review paper on the WoT and discussed a detailed architecture and some key enabling technologies of the 'WoT' along with recent research results. They have pointed out some open challenging issues in this field. Kotis and Katasonov [20] have offered a paper on 'Semantic Interoperability on the Internet of Things (IoT)' and presented a use case diagram for the alignment of entity. They have opined that ontology is the key component for the abstraction and semantic registration of IoT entities. Zafeiropoulos et al. [21] have reviewed all the techniques of data gathering in the sensor network and provided three layers of architecture for this process. This architecture deals with many problems such as data aggregation, data management, and query answering by using semantic web technologies. Khan et al. [22] have proposed an architecture of data annotation for semantic applications in Virtualized Wireless Sensor Networks (VWSNs). They have stated that semantic applications play a critical role in VWSNs. Compton et al. [23] have described one of the essential technologies in the semantic sensor network, called 'semantic specification of sensors.' Twelve sensor ontologies, and reasoning and search technology developed in conjunction with these ontologies, are reviewed and analyzed for pointing out the range and expressive power of their concepts.

There are various domains which are somehow related to semantic web technologies. We cover a total of 10 domains, namely computational Intelligence, Evolution and Swarm Computing, Knowledge representation of smarter data, Big data, sensor network, cloud computing, Internet of Things, Mining and Analytics, Machine learning, and Natural language processing. Art of review differs from other reviews in three key areas: first, we present three domains that feed the progress of semantic web; secondly we explain five domains that utilize the semantic web technologies, and third, we mention two domains that go hand in hand with semantic web technologies. This paper provides a condensed review of the present and future scenarios of semantic web technologies. The paper closes with a list of 25 references. Thus, this survey should also make a good tutorial for a researcher previously unfamiliar with the semantic web.

3. Semantic web and technologies

The Semantic Web, which is an extended version of the current web, provides a standard structure for data representation and reasoning. Here data are stored in the form of ontologies that provide inference power over the stored data. These days the web is used for three major important tasks, i.e. Searching, Combining of data, and Mining of data. The efficiency of tasks heavily relies on the storage and representation of the knowledge structure. Therefore, various knowledge representation schemes and languages have been developed in such a way that no part of knowledge in the said domain remains uncovered. Figure 1 shows OWL is the topmost knowledge representation language for the semantic web, which is based on the RDF/XML. Ontology, RDF, OWL, and SPARQL are the essential components of the semantic web. They have the capacity to encode semantics and to provide automated reasoning, amalgamation, and sharing of information from various sources and its management. Large-scale data are being published on the Semantic Web, which requires advanced analysis. The Semantic Web collects and represents web documents in a structured style and provides effective reasoning and derives meaningful results. With this thought, ontologies that describe real-world entities play a vital role in making the Semantic Web a reality.

Inference rules and taxonomy are the most typical kind of ontology for the Web. Taxonomy defines the relationship between classes. Many applications use ontologies to relate the knowledge structure and inference rules between pages. Semantic technologies are used for querying, knowledge representation, and storage that can marvel the entire web browsing experience. Figure 2 depicts all domains where SWTs come into the scenario. The incoming arrows toward SWTs show how that domain enhances the growth of the semantic web, the outgoing arrows from the SWTs show the use of SWTs in that particular

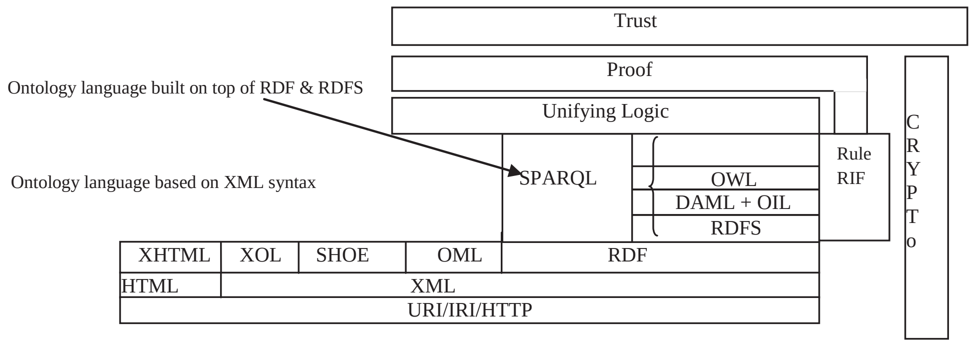

Figure 1. Hierarchy of semantic web stack [3,25].

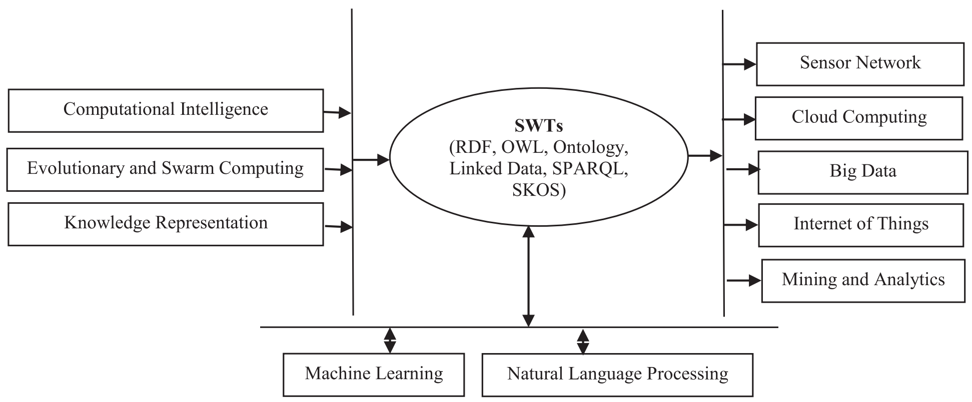

Figure 2. Domains related to semantic web technologies.

domain, and the bidirectional arrows (↔) show that particular domains go hand in hand with SWTs.

The Semantic web uses the concept of text mining for better data processing of the raw data that exist on the web in the form of XML and RDF. The extraction of Meta data is of two types: explicit metadata extraction and implicit metadata extraction. The implicit Meta data extraction implicates semantic information deduced from the material itself, for example, name of entity and associations enclosed in the text. This process essentially takes the help of ontology. Traditional information extraction is based on a flat structure, but we need information in a hierarchical structure for the Semantic Web because we connect semantic metadata along with documents and address the concepts in the ontology.

Therefore, text mining improves progress of the semantic web. The aggregation and amalgamation of diverse kinds of existing data and the mining of relevant knowledge and its storage or representation have become key challenges for researchers of the Semantic Web. Many researchers are working in these areas and have solved many issues, but there are still many challenges in the semantic web which generate further directions for the researcher. These challenges can be tackled by knowledge representation, CI, and Swarm computing.

4. Toward the growth of the semantic web

The Semantic Web data space is distributed, dynamic, incoherent, and very sensitive to privacy issues. There are three domains that help in furthering the growth of the semantic web. The first domain is computational Intelligence, the second one is Evolutionary and Swarm Computing, and the last is knowledge representation methods. With the help of swarm computing we store large-scale data and provide reasoning over the web. Computational intelligence provides the methods for handling vagueness and uncertainty issues. Knowledge representation methods help store the data in a consistent and coherent manner.

4.1. Computational intelligence (CI)

CI is a set of methods that mainly consists of Evolutionary Computation, Fuzzy Logic, and ANNs. The Semantic web has many problems because of vast, incomplete, uncertain, inconsistent, and decentralized information. A variety of CI methods are used to crack the issues of the semantic web. We review these issues from three aspects, where different CI techniques have been used. The first is Vastness and Tractability, the second is Vagueness and Uncertainty, and the third is Divergence and Inconsistence. Researchers use evolutionary computation in case of tractability and vastness issues. The fuzzy logic system is useful for solving the problems of uncertainty and vagueness in Web semantics. ANNs are widely used for solving the problem of inconsistency in data mapping and alignment. ANNs take the advantage of structural and functional views of biological neural networks. They provide a way by which we extract hidden correspondences or patterns from large amounts of data. Researchers have utilized the techniques of ANN in ontology alignment, ontology enrichment, automatic ontology construction and concept mining, and so forth. The ANN is separated into two typical groups: unsupervised ANN used for ontology learning and supervised ANN used in ontology alignment. For both groups, we present two most important works from the text.

4.1.1. Ontology alignment

Supervised ANNs are extensively used between heterogeneous ontologies in semantic mappings. The Recursive Neural Network (RNN) model, intended to process structured data accurately, is used with ontologies that represent structured data. RNN has been widely used in automatic ontology alignment. We provide input to the neural network by using the concept of structures and individual relations in ontology which will be utilized for aligning the concepts of other ontologies. Find an optimal configuration is one issue regarding ontology alignment that can best pamper ontological constraints, such as 'if a concept P_1 map to another concept P_2 is true, then concept P_3 maps to concept P_4 is false.' The Interactive Activation and Competition neural network fetches the global optimal solution which satisfies various ontological constraints. It is intended to solve constraint satisfaction issues in word recognition. The plan is to employ a node in it to characterize an element mapping hypothesis.

4.1.2. Ontology learning

Ontology learning, sometimes called 'ontology generation,' 'ontology extraction,' or 'ontology acquisition,' is the semiautomatic and automatic creation of ontologies. It extracts relationships between the concepts that represent domain knowledge. In the literature, unsupervised ANN is used to find out new entities and instances or individuals from domains, which provides automatic ontology construction. One type of unsupervised neural network is Self-Organization Map (SOM) that can deliver a low-dimensional portrayal of the input space of the training samples, called a 'map.' SOM has been used to enhance concepts and instances of domain ontologies from a domain text corpus. The preliminary step of an SOM is to transfer ontology into a neural representation. Next, it fetches the concept and instance from ontology via the text mining process and represents it. We can provide ontology enrichment by neural network with unsupervised training, which exposing the initialized SOMs and their information fetched from the domain depend on some sorts of similarity metrics. Projective Adaptive Resonance Theory Neural Network is a kind of unsupervised ANN that gives automatic ontology construction from the web. The most illustrative preferred standpoint of CI techniques for the Semantic Web is their ability to handle complicated issues in an extremely decentralized and dynamic way. Because of the decentralized nature of the Web, it is a nontrivial vision for enhancing the web with semantics and empowering web intelligence. The main fact behind this is that there is no central component that checks the uncertain, dynamic, and random behaviors on its constituents. Fuzzy logic improves the quality of the ontology learning process by providing precise information of entities. Vagueness, randomness, and autonomy have been well studied in nature-inspired methods and CI techniques can resolve all these problems.

One major drawback of existing CI techniques is that they are generally focused on only some aspects of a crisis and can solve problems only separately. The problems of ambiguity, autonomy, inconsistence, and uncertainty are available simultaneously in many applications of the semantic web. These scenarios require an integration of various CI methods and empowering them to work collectively.

4.2. Evolutionary and swarm computing

Evolutionary and swarm algorithms are used for optimization problems with a large and dynamic search space. Semantic web applications require scalable, adaptive, and vigorous approaches to store and examine large-scale data. The solution to this problem is to dispense both data and requests onto various computers. Along with storage, the annotation of data with machine-understandable semantics is also very vital for understanding the idea of the semantic web. Reasoning and storage over the web entails the same requirements; therefore, we use the evolutionary approaches for querying and swarm algorithm for entailment. Swarm-based techniques have been revealed to generate optimal solutions for critical issues in an entirely decentralized manner. We utilize a new concept for reasoning inside an entirely spread and self-organized storage system that is rooted in the collective behavior of swarm instances and does not want any schema replication. Thus, we can surely say that the swarm-based methods are useful or helpful in developing a large-scale distributed storage and reasoning system. Manually configuring and operating wide-scale distributed systems is no longer possible because they consist of billions of nodes. Self-organizing distributed systems are capable of working independently and are a solution to the problem of managing distributed systems that give wide-scale storage and investigation for the web of data. We can utilize the swam computing technique for completely distributed storage and reasoning (inference) over Semantic Web data.

The challenge of storing, reasoning, and retrieving wide-scale semantic data in a distributed situation is a job that can be adequately performed only in a truly decentralized manner. This makes swarm-based methodologies fascinating candidates for obtaining the preferred self-organization. However, all distributed systems have to trade-off between various objectives; these properties come at a cost. Swarm-based techniques trade deterministic guarantees to obtain their benefits. Evolutionary and Swarm computing solves many problems of other domains, but still it has many problems that generate future directions for the researcher like Enhancement of algorithms for data clustering [24], Optimization problems for storing, reasoning, and querying, and Exploration and Exploitation search space problem.

4.3. Knowledge representation of smarter data

A major aspect of the development of any structure or applications heavily depends upon the representation scheme. A good scheme that describes the unit of knowledge is the most inevitable requirement of the time. Extended Hierarchical Censored Production Rules (EHCPR), can be used to represent the domain knowledge because it handles problems related to representation, learning, and reasoning. The structure of EHCPR visualizes real-world entities in an effective manner. It has the capacity to store and manage large amounts of data with the help of constraints and default values of entities. Here, properties are divided into two parts, namely defining property and characteristic property. For an individual, the defining property of a concept must be true, whereas characteristic property may or may not be true. The structure of EHCPR is divided into two parts: knowledge base (storage of concept) and databases (storage of instances). The sequence of class is described by generality and specificity operators. Equation (1) describes various operators that work as a unit of knowledge for a given concept.

$$Y\,(:-X\;)\,(\lfloor Z)\,(G\%\,G)\,(\$S)\;\;(@Pa)\,(^{*}Pr)\;\;(\#I) : \gamma,\;\delta, \qquad (1)$$

where $:-$, $\lfloor$, $G\%$, $\$$, $@$, $*$, and $\#$ are used for If, Unless, Generality, Specificity, Has_Part, Has_Property, and Has_Instance operators, respectively and symbols X, Z, G, S, Pa, Pr, and I denote the corresponding information relegated with them [25].

Along with a good knowledge representation scheme, a query language is also required to retrieve only relevant output from the stored knowledge. Therefore, an effective query language is a major requirement for the development of web-based applications. There are various steps for building the semantic web like ontology development languages, editor, query language and reasoner, and so on. The amalgamation of these steps will help achieve the overall goal of the semantic web. Now, the requirement is to integrate all tasks (search engine, NLP, SPARQL converter) under one roof, which helps people develop intelligent applications. An integrated approach should have a good scheme to represent knowledge and an efficient reasoning power. EHCPR is one such integrated approach that fulfills all requirements required for the development of intelligent applications. Knowledge representation provides both opportunities and challenges to the researcher. Although it provides a good structure for the semantic web, there is an urgent requirement to work for the enhancement of knowledge representation language and reasoning, methods for dealing with the heterogeneity of data and knowledge, sophisticated methods for knowledge capturing, making knowledge representation accessible to nonexperts, etc.

5. Domains related to semantic web technologies (SWTs)

Semantic Web technologies involve many areas of computer science and have resolved many issues regarding the representation and extraction of information. In this section, we mention five domains that are closely related to semantic web technologies.

5.1. Sensor network

Sensor networks are utilized in many areas for capturing physical natural events and observe the characteristics of physical objects like temperature, sound, pressure, and so on. A sensor network produces a huge amount of data that require enhanced logical processing and interpretation by machines. Sensor networks have generated a lot of interest today in academia as well as industry. They are a new paradigm for distributed sensing and actuation. The major challenges in this field are coverage, connectivity, and clustering in heterogeneous sensor networks, deployment strategies and topology or neighbor discovery techniques, localization algorithms, secure data aggregation, energy optimization, and the security and quality of services. In real-world application, sensor data will be a mix of dissimilar data that come from various networks. The processing and interpretation of the vast amount of unstructured sensor data and the utilization of a consistent structure for this sensor data are a crucial part of scalable and interoperable sensor network architecture. The existing data exchange format for sensor networks depends on syntactic models that do not give machine-understandable meanings to the data. The semantic technologies provide an extra interoperable structure for sensor data and empower machines to process and interpret the emerging semantics to develop more intelligent sensor networks.

The semantic sensor web is an expansion of the sensor web where sensor nodes swap and process data automatically without any human interference. The major components of the semantic sensor web are ontologies, query languages semantic annotation (comment), and rule languages. Ontologies serve as dictionaries that contain the definitions of all concepts used by sensor web. Semantic Sensor Network Ontology (SSNO) contains sensors, procedures, and their observations. Semantic annotation language, for example RDF and RDFa, is used for annotating the sensor's measurement and observation. Reasoning service provides inferences on existing facts and rules that are defined via SWRL by which we extract additional information. All of the above-mentioned information forms the backbone of the semantic layer. SWTs play a very important role in the sensor network because through them, we infer semantic information from the raw data gathered by sensors. Hence, we can utilize meaningful information in many smart applications like health care, meteorology and environment observation, and so on.

5.2. Cloud computing

Cloud computing is an extended form of Internet-based computing that provides shared resources of computer processing and data to computers and other devices on demand. It delivers computing services like software, servers, networking, storage, and so on over the Internet. Cloud computing is a very vast area for research; therefore, the landscape of cloud computing has significantly changed over the last decade. The most challenging research directions for cloud computing are storage and fault tolerance strategies, peer-to-peer-based cloud workflow system, adaptive and data-driven workload manager for general clouds, service scalability and interoperability over the cloud, the combining of high-performance computing into cloud computing services, scientific services and data management in the cloud, and cloud computing privacy and security preservation in the cloud. We can remove problems of cloud interoperability by storing the information of cloud Resources and Services Description into ontologies with five main layers, namely software infrastructure, software environments, software kernel, hardware, and applications. Each layer holds more than one service if it has similar levels of abstraction. The cloud system for service discovery uses the concept of cloud ontology to compute the correspondences among the services. This system is an agent-based discovery system that utilizes three types of similar reasoning and enables to obtain the correspondences of cloud services. Their cloud ontology has entities of different cloud services for infrastructure as a service, software as a service, and platform as a service. Cloud computing ontologies are mainly used for the selection and discovery of the best service according to users' needs and the description of cloud services resources.

The combination of semantic technologies and cloud computing not only utilizes the cloud services, but also includes the field of distributed computing, which has permitted semantic web technologies to scale to ever bigger data sets. We can say that semantic technologies can be utilized over the entire cloud world to enhance the management and security of the clouds. However, the utilization of cloud computing technologies and services over semantic technologies guarantees improvements in the fields of scalability to large datasets and elasticity in the reaction to changing usage profiles and systems that are more credible and easier to construct and preserve. Clients and designers can exploit the strengths of semantic technologies to understand diverse data on a huge scale without the requirement for its own infrastructure, benefitting from the elastic scalability, adaptability, and the ease of deployment and maintenance in the cloud. A clear restriction of these offerings is the restriction to closed data sets, the shortage of the capability to combine diverse data sets, and to examine the data semantically.

5.3. Big data

'Big Data' is a term that has been used to explain a large amount of data that have been generated over the last 20 years. Facebook, Twitter, and other social media not only create a lot of data; they additionally make it feasible for developers to access. There are many problems in organizing the data that emanate from various sources and in various formats. Various authors have written review papers on the future challenges of big data. Some of the challenging future research guidelines for big data are Scalable Architectures for Massively Parallel Data Processing and Stream Data Processing, Scalable Storage Systems, Security and Privacy Issues in Big Data, Large Scale Data Analysis for Social Networks, Privacy Preserving Big Data Analytics and Adaptation, Big Data Analytics for Business Intelligence and Smart Healthcare, Uncertain Data Management in Big Data, and Big Data Visualization and Semantics. Among them 'Big Data Analytics' has become the center point of research. One of the real issues for Big Data Analytics development is the diversity of web-based information because the collected data are unstructured in nature. This is where 'Semantic Web Technologies' come into the picture. The international community W3C has encouraged a common data format to make the data on the web more reliable and easier to interpret. For Big Data Analytics-based organizations, SWTs assist them in businesses enabling then in making an even better judgment in real time. For business to consumer organizations (led by Google), the semantic web permits them to offer consumers with superior answers and experiences immediately. The approach to address the problem that is connected with big data is to store data in a structured format and characterize the data sets as graphs. This permits software agents to query on the databases. The processing of linked data makes it feasible to find information.

A knowledge graph is a good example of big data on the semantic web. This knowledge graph was added in Google in 2012–2013 and provides an updated algorithm called 'Hummingbird.' The semantic knowledge is used by Google knowledge graph which increases the conventional search engine result pages. Another good example is social media. The Facebook Open Graph protocol empowers any web page to become a rich object in a social graph. Facebook Graph Search is a semantic search engine introduced in 2013 to provide answers to natural language queries of users instead of a list of links. According to Tim Berners-Lee (2009), 'Linked Data is simply about using the Web to create typed links between data from different sources.' Linked data is a method that links other data sets and publishes structured data on the web. It has been proved that linked data provides better data integration when compared with existing data models. The set of SWTs confers an environment where the application can query that data and draw inferences.

5.4. Internet of Things (IoT)

The IoT has required a semantic backbone to thrive. More than 25 billion devices were estimated to be connected to the Internet in 2015 and 50 billion by 2020. The interoperability among the things on the IoT 'is one of the most essential requirements to support object discovery, addressing and tracking in addition to information storage, Security, representation and exchange.' IoT will essentially comprise different sets of devices and different communication strategies between the devices. This type of heterogeneous system should evolve into an additional organized set of solutions, where 'things' are made consistently discoverable, empowered to communicate with different entities, and are strongly included with Internet infrastructure and services, instead of the particular way in which they are associated to the IoT. Dynamicity, diversity, networks, data, and the heterogeneity of devices are key issues of IoT technologies. Semantic web technologies have been proved fruitful in different areas in dealing with the heterogeneity issue in (i) interconnecting such data (ii) inferring novel knowledge in developing intelligent applications (iii) providing interoperability in data management. However, one of the challenges with existing IoT applications is that the devices are not (or little) compatible with each other because their data are dependent on proprietary formats and they do not employ common vocabulary to explain compatible IoT data. SWTs have already shown their benefits in domains other than IoT. Semantic web technologies are used in IoT to reduce the challenge in dealing with interoperability of data produced by devices already employed in real life. The Semantic Web deals with IoT and WoT. Upon realizing the true potential of semantic technologies, various IoT frameworks have been proposed which address the data interoperability issues using SWTs and standards.

For the development of the IoT, many semantic technologies such as ontologies, semantic annotation, RDF, linked data, semantic web service, and so on can be used as a principal solution. The use of ontology with a semantic description for data will make it interoperable for clients who share and utilize the same ontology. Through semantic technologies in IoT, we can handle interoperability, efficient data processing, resource discovery, integration, reasoning, and querying. Likewise, semantic technologies have demonstrated their usefulness in different areas, and a few among various problems that SWTs are addressing are to (i) reduce heterogeneity by using semantic interoperability, (ii) provide simple integration of data application, (iii) infer and mine new knowledge to develop applications and provide smart solutions, and (iv) provide interoperability between different data processes with representation, management, and storage of data.

5.5. *Mining and analytics*

Nowadays, as a huge amount of information is available on the web, more than one billion pages are indexed via search engines, and therefore, searching for the desired information is an extremely tricky task. This richness of resources has encouraged the requirement for developing automatic mining techniques on the World Wide Web, thus giving rise to the term 'Web Mining.' SWTs intend to address the issue of extracting information from the web by offering machine-understandable semantics to give better machine support for the client. An integrated approach of SWTs and web mining provides a better method for the mining of related and semantic information from the web, thus giving rise to the term 'Semantic Web Mining.' The combination of SWTs and web mining enhances the consequences of web mining through utilizing the semantic structure in the web. Ontology has the capacity to mine data from a large pool of data. It is used during preprocessing for improving clustering results. SWTs make Web mining simpler to accomplish and can also enhance the efficiency of the process. By using semantic web technologies with linked open data, we support knowledge discovery.

The endeavor behind the Semantic Web is to include machine-interpretable, semantic annotation (comment) in web documents in order to get knowledge rather than unorganized data. Various kinds of research work have been done; however, the integration of the semantic web with linked open data in the area of mining is still to be closed. In data mining, selection of the data according to user need is considered as a challenging task because there is no relationship between the data sets. We use an ontology approach for mapping the data over the web. An ontological framework for mining searches the data and derives only those results which are related to the search parameter. The data existing on the web are interlinked to each other via ontology which provides successful searching. There are two key difficulties in semantic web mining: first, the nature of data is unstructured and second, extraction of data from the web. The difficulty in the present search engine is the vague behavior of words. For this cause, a search engine is enabled to differentiate the meaning and therefore it returns all the websites which contain given search parameter. We can resolve this problem by using semantic technologies which have the capacity to differentiate the meanings of words. The Semantic web enables knowledge mining instead of data mining because it incorporates the domain knowledge and semantic annotation property which provides the meaning of the data. Therefore, SWTs help people explore the facts from the specifics to answer specific questions.

6. How other domains go hand in hand with semantic web

Machine learning (ML) and Natural Language Processing (NLP) utilize the services of semantic web technologies and semantic web technologies use these domains for achieving better results. Therefore, it can be said that these two domains and the semantic web cannot live separately. In this section, we elaborate the working of these two domains and SWTs with respect to each other.

6.1. *Machine learning*

Machine learning means how to formulate the problem and how to represent the data. Semantic web technologies provide ontological background knowledge and a standardized model for representation of data. Semantic web principles explain the concept of Meta data and it also has large calibre similar to the data model for data communication and integration. Ontology matching is a vital phase of the semantic web; therefore, many methods are used for this purpose, such as Probabilistic approaches, theorem proving, heuristic and Rule-based methods, and Machine-learning Reasoning and Graph analysis in which machine-learning algorithms are very popular because they provide good similarities matching between concepts. Machine learning solutions have been developed to support ontology learning (regardless of the possibility of not being fully automatic), deep annotation (reconciling of databases and ontologies) and annotation via information mining. Machine learning will progressively be utilized to examine distributed data sources and maintain reasoning and querying over the semantic web. Machine-learning approaches have been used for the semiautomatic annotation of unstructured or unorganized data, organization of ontologies, and to combine semantic information into web mining. Machine learning helps the semantic web via ontology evaluation, construction, management and evolution, refinement, and additionally the merging, mapping, and alignment of ontologies. It compares numerous similarity measures and algorithms used to map or merge two ontologies with machine-learning algorithms. The objective of the Semantic Web is to permit machines to recognize this data. With regard to the Semantic Web, Meta data assist the machine to process the data according to their meaning. The Semantic web accomplishes various tasks for machine learning such as using various tools to explain and exchange data for later use by machine-learning methods in a canonical model, utilizing ontological structures to enhance the machine-learning task, and offering background facts to guide machine learning.

For instance, a person will have the capacity to give a computer an instruction, for example, 'search all terrorist attacks which have been targeted by Al-Qaida perpetrator groups in India.' Nonsemantically, a computer may have the capacity to do a search for Al-Qaida perpetrator groups and return the consequences for the client to sort through. This search would be done on word-matching without the machine understanding. With the help of SWTs, the computer would have the capacity to access ontologies so that it could recognize (or appear to understand) the relationship between terrorist attacks, Al-Qaida, and India.

6.2. Natural language processing

The Semantic Web is a challenge for natural language generation techniques to scale up the huge amount of data. On the contrary, there is a need in the semantic web technologies that provide access to the machine-oriented Web of data. The natural language generation provides ways for displaying semantic data in an organized, coherent, and accessible manner. Natural language generation methods address the generation of information from huge multiple data sets. These methods can be additionally developed on semantic web data to produce summary data that communicate the most vital content in a dataset. Methods for ontology summarization contribute to the generation of summary data from huge datasets by using topographical measures over RDF in order to judge what content is most vital. But the methods of content selection in natural language generation depend on the context that was communicated in the target text. The methods of ontology summarization are used in natural language generation and enhance the fitness of text-based summaries as an intelligible manner of introducing data to humans. The Semantic web requires machine-interpretable semantics for handling textual information. Natural language techniques (information extraction, term recognition) give meaning to unstructured or unorganized data or text. When merging the meaning in ontology, it becomes useable across the entire web and allows many processes like reasoning and querying to be carried out. The state-of-the-art natural language systems are able to provide users sensitive access to the riches of the textual data via ordinary languages.

The combination of NLP and SWTs deals with structured and unstructured data that are not feasible by traditional relational methods; for example, a machine reader is a tool that converts natural language text into proper structured knowledge and the latter, according to shared semantics, can be interpreted by machines. We used this combination in various applications where we are conflicting with a huge volume of unstructured or unorganized information. We extract structured data from a text document by using NLP and then link this extracted data via semantic technologies. Combining both technologies, we can fetch related information and relate that information to rigorous scientific data (e.g. on organic compounds, pathways, etc.) that have previously been declared and are being placed in an organized database. The mixture of NLP and SWTs provides an environment by which we can put the complex query and fetch actual logical answer. We can draw an appropriate structure for the textual data with the help of the RDF (it is a data model for the semantic web) and natural language processing technologies. This combination permits consumption to grow along with NLP strengths without redesign of any applications that utilize those data. Therefore, the semantic web can enhance the technical level of natural language technologies and NLP can also help in delivering and using a better semantic web. The real challenges for the researcher are to fetch the NLP techniques nearer to ontological engineering and to expand existing NLP techniques to ontology-based applications. NLP provides a better interface to the client by which users easily access the document over the web.

7. Discussion and conclusion

The paper is structured into five sections, namely Introduction, Review of related works, Semantic web and Technologies, Toward the growth of Semantic Web (Computational Intelligence, Evolutionary and Swarm Computing, Knowledge Representation of Smarter Data), Domains related to Semantic Web Technologies (Sensor network, cloud computing, Big data, Internet of Things, Mining and Analytics), and How other domains go hand in hand with Semantic Web (Machine Learning, Natural Language Processing). Art of review differs from other reviews in three key areas: first, we present those domains that enhance the growth of semantic web, secondly we explain all domains that utilize the semantic web technologies, and third, we mention those domains that go hand in hand with semantic web technologies. This paper provides answers to various questions like Can SWTs be applied in different fields?, How SWTs are helpful for other domains?, How other domains enhance the progress of semantic web?, What are the research directions for the semantic web and its related domains?, Why and how SWTs are important?, Which type of Knowledge Representation languages are used by the Semantic Web?, Which type of problems require immediate attention of the researchers?, How can we reduce the existing issues of the Semantic web with the help of other domains?, What is the current progress of the Semantic Web? and so on. To the best of our knowledge, it is a first attempt to bring all domains related to SWTs into one platform.

However, we acknowledge that 'SWTs,' as described in this survey, can be helpful in all the 10 domains. This review paper interprets and presents a comprehensive overview about semantic web technologies as well as its related domains and research directions of that particular related domain that provides future directions and guidelines for researchers. We have concluded that semantic technologies are very helpful in people's future lives. So both industry and academia give their best efforts to tackle all those challenging issues of the semantic web, and they also work on promoting the progress of the semantic web. This survey provides pointers for further reading at many places.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is the part of the Defence Research and Development Organization funded project titled 'Intelligent Real Time Situation Awareness and Decision Support System for Indian Defence' with grant number ERIP/ER/1506048/M/01/1611.

Notes on contributors

Archana Patel received the MCA degree from National Institute of Technology (NIT) Kurukshetra, India in 2016. She is pursuing PhD from NIT Kurukshetra. She has two years experience as JRF in Defence Research and Development Organisation (DRDO) funded project. Her research interests include Semantic Web, Big Data, Ontological Engineering. and Knowledge Warehouse.

Sarika Jain graduated from Kurukshetra University (India) in 1998 and did her master from Jawaharlal Nehru University (India) in 2001. Her doctorate is in the field of Knowledge Representation in Artificial Intelligence. which was awarded in 2011. She has served in the field of education for 17 years and is currently serving at National Institute of Technology, Kurukshetra. Her current research interests are in the area of Intelligent Systems, including Knowledge Representation, Ontological Engineering and Semantic Web Technologies. She is a member of IEEE and ACM and a Life Member of Computer Society of India.

ORCID

Archana Patel http://orcid.org/0000-0002-7505-7226
Sarika Jain http://orcid.org/0000-0002-7432-8506

References

[1] Bittencourt II, Isotani S, Costa E, et al. Research directions on semantic web and education. Interdiscipl Stud Comput Sci. 2008;19(1):60–67.
[2] Markellou P, Mousourouli I, Spiros S, et al. Using semantic web mining technologies for personalized e-learning experiences. In: Proceedings of the web-based education; 2005. p. 461–826.
[3] Kenekayoro PT. Semantic web-the future of the web. Afr J Math Comput Sci Res. 2011;4(3):113–116.
[4] Abaalkhail R, Guthier B, Alharthi R, et al. Survey on ontologies for affective states and their influences. Semant Web. 2018;9(4): 441–458.
[5] Bizer C, Heath T, Berners-Lee T. Linked data: the story so far. In: Semantic services, interoperability and web applications: emerging concepts. Hershey (PA): IGI Global; 2011. p. 205–227.
[6] Harth A, Hose K, Schenkel R. Linked data & the semantic web standards. In: Linked data management. Chapman and Hall; 2016. p. 28–73.
[7] Kumar VK. Semantic web approach towards interoperability and privacy issues in social networks. Int J Web Serv Comput. 2014;5(3): 13–17.
[8] Dhenakaran SS, Yasodha S. Semantic web mining: a critical review. Int J Comput Sci Inf Technol. 2011;2(5):2258–2261.
[9] Stumme G, Hotho A, Berendt B. Semantic web mining: state of the art and future directions. J Web Semant. 2006;4(2):124–143.
[10] Sridevi K, UmaRani DR. A survey of semantic based solutions to web mining. Int J Emerg Trend Technol Comput Sci. 2012;1:50–57.
[11] Berendt B, Hotho A, Mladenic D, et al. A roadmap for web mining: from web to semantic web. In: Web mining: from web to semantic web. Berlin: Springer; 2004. p. 1–22.
[12] Huang J, Dang J, Huhns MN, et al. Use artificial neural network to align biological ontologies. BMC Genomics. 2008;9(2):S16.
[13] Caliusco ML, Stegmayer G. Semantic web technologies and artificial neural networks for intelligent web knowledge source discovery. In: Emergent web intelligence: advanced semantic technologies. London: Springer; 2010. p. 17–36.
[14] Chen H, Wu Z, Cudre-Mauroux P. Semantic web meets computational intelligence: state of the art and perspectives. IEEE Comput Intell M. 2012;7(2):67–74.
[15] Yu W, Chen J. The state-of-the-art in web-scale semantic information processing for cloud computing. arXiv preprint arXiv:1305.4228. 2013.
[16] Sheth A, Ranabahu A. Semantic modeling for cloud computing, part 2. IEEE Internet Comput. 2010;14(4):81–84.
[17] Androcec D, Vrcek N, Seva J. Cloud computing ontologies: a systematic review. In: Proc. of MOPAS; 2012.
[18] Kotis K, Katasonov A. Semantic interoperability on the internet of things: the semantic smart gateway framework. Int J Distr Syst Technol. 2013;4(3):47–69.
[19] Zeng D, Guo S, Cheng Z. The web of things: a survey (invited paper). J Commun. 2011;6(6):424–438.
[20] Kotis K, Katasonov A. Semantic interoperability on the web of things: The semantic smart gateway framework. In: 2012 Sixth international conference on complex, intelligent and software intensive systems (CISIS); 2012. p. 630–635.
[21] Zafeiropoulos A, Spanos DE, Arkoulis S, et al. Data management in sensor networks using semantic web technologies. Data management in semantic web; 2009.
[22] Khan I, Rifat J, Fatima ZE, et al. A data annotation architecture for semantic applications in virtualized wireless sensor networks. arXiv preprint arXiv:1501.07139. 2015.
[23] Compton M, Henson CA, Lefort L, et al. A survey of the semantic specification of sensors; 2009.
[24] Zhang S, Wang H, Huang W, et al. Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG. Optik-Int J Light Electron Opt. 2018;157:866–872.
[25] Patel A, Jain S. Formalisms of representing knowledge. Procedia Comput Sci. 2018;125:542–549.