The Semantic Web Status: A Literature Review

Haochen Wang
13500198 2698251

Universiteit van Amsterdam
Vrije Universiteit Amsterdam
haochen.wang@student.uva.nl

Abstract. Semantic Web has been proposed around 30 years and the semantic web technologies evolve quickly in the meantime. Although there is not a wide use of Semantic Web yet, part of applications and industrial software have adopted and deployed a small scale of the Semantic Web. Over these 20 years' development of semantics, technologies were carried out to deal with challenges in knowledge representation, data semantic integration, and structural technology stack. This paper reviews the major development of the Semantic Web during the last decade and the future trend with challenges by analyzing relevant literature. The paper proposes research questions, reviews literature according to the research questions, concludes and summarizes the answers.

Keywords: Semantic Web · Semantic web technology · Web services · Linked Data · Data integration · Literature review.

1 Introduction

People are familiar with the Web and more than a trillion web pages make up the existing web, the World Wide Web(WWW). Up to January 2022, there are around 1.18 billion websites around the world. The speed of new websites created is fast as well, which is over 250 thousand every working day.[1] The Web contains a huge amount of information.

The Web allows users to acquire their interested topics practically. Accessing documents seems efficient via the Web. Users are able to search and obtain the information by keywords. However, it is not the case that users deal with non-text-based content, such as PDF, image, or video. From this aspect, the limitations and problems of the existing web can be perceived. Therefore, the Web needs to be extended so that information is able to be accessed with semantics. Semantic Web consists primarily of documents that humans read and information that computers manipulate.[2]

Here, the term "Semantic Web" is a W3C's vision of the web. It can be understood by being splitted in"semantic" and "web". From the word "semantic", semantic can be regarded as a study of meaning.[3] The purpose is to recognize the characteristics employed to distinguish an entity from plain contents.[4] From the other word "web", the Semantic Web aims to build a version of Web that is readable, interpretable, and processable by machines. The semantic web technology is used to generalize the techniques adopted in the Semantic Web. Subsequently, computational machines have abilities to define meaningful interpretations appearing in a similar manner that people deal with information to complete their objectives.

In general, the term"Semantic Web" usually suggests both the Semantic Web elementally as a huge public repository of datasets which can be comprehended by computers and the semantic web technologies as a group of techniques for knowledge representation and data processing.[7] There are two most significant features of the Semantic Web. The one is that it establishes based on the fundamentals and techniques of the existing Web, for example, indexing and naming schema. The other is that its concepts' meaning should be shared with a machine-readable representation, which can be known and handled by reasoners in some software.

The semantic web technologies facilitate humans and machines to generate Linked Data. The operations with Linked Data include assembling vocabularies and defining rules. Since the 2000s, some communities have begun to promote paradigms of specifications that can help knowledge bases disposed on the Web for biology[1], medicine, genomics, and related fields.[5] Today, more organizations adopt the mechanism of structured knowledge bases to publish prime data privately. E-government and eScience are two typical application paradigms that deploy Semantic Web and technologies using independent standards to provide services internally.

P. Hitzler mentioned in his review that 'Semantic Web may or may not come into existence someday, and indeed some members of the research field may argue that part of it has already been built.'[18] It may sound contradictory about the Semantic Web and its technologies. To have a better understanding of Semantic Web, the main objective of this paper is to conduct a literature review of the Semantic Web and its technologies status, how they developed in a decade from the 2000s to 2020s, what triumphs and challenges of their research and applications are, and how will the Semantic Web develop tomorrow. Therefore, this paper manages to grep and conclude the concepts by analyzing and comparing the existing literature from 2000 to 2021. Also, some important topics close to the Semantic Web are discussed, such as Linked Data and Web Ontology Language(OWL) which is a group of typical knowledge representation languages for ontologies.

The review is organized into five sections. The background knowledge and basic terminologies have been illustrated in this section as an introduction part. The next section defines three research questions and describes how they motivate the literature study. Then, four subsections discuss the Semantic Web in the 2000s, in the 2020s, Linked Data semantic integration, and the future trend respectively. They can provide solid literal evidence of the research questions. In the section of Conclusion, there are the answers to the research questions. The

[1] for example, Open Biological and Biomedical Ontology `https://obofoundry.org/`

last section summarizes the review and has a discussion on the next step of the semantic web technologies.

2 Research Questions

This literature review is conducted by following the procedure that is raising research questions, analyzing and discussing the certain period literature works through the group of accurate research questions, and reporting the analysis. The idea is to deeply grasp how the Semantic Web develops recently and identify to what extent it will deploy in the future. Three research questions are defined to make the objectives more concrete and describe their motivations.

- **RQ1**: What is the current status of the Semantic Web in applications?
  - The aim is to investigate the literature that presents surveys and evaluations on the using and developing of Semantic Web in applications from the 2000s to 2020s and identify the state-of-art technologies, but not to provide a concrete implementation.
- **RQ2**: How Linked Data organizes information on the Semantic Web?
  - Linked Data is a collection of interrelated datasets which can show relationships among data and can be accessed on the Semantic Web. The aim is to investigate the effort that has been made in Linked Data integration and identify some of the main processed and tools.
- **RQ3**: What is the challenges in widely using the Semantic Web in enterprise?
  - The aim is to investigate the existing and potential obstacles in general applications, enterprise from technological and social aspects. The other aim is to identify the expected developments in semantic web technologies.

The literature study aims at these research questions and emphasizes the main topic that is the current status of the Semantic Web and the semantic web technologies. The next section will survey and review the literature on the main topic from four aspects.

3 Literature Study and Review

3.1 The Semantic Web in the 2000s

In this subsection, some of works published between 2000 and 2009 on the Semantic Web is reviewed, most of which are surveys. They provided the adoption of the Semantic Web and the development of the semantic web technologies in the 2000s.

Like the development of many other technologies and applications, the Semantic Web was born under inspiring ideas, some exposed problems, industry practical needs, and attainable conditions. By the mid-1990s, several research groups have come up with an idea that if the Web markup contained some notations which can be understood by computers, then improved Web-based tasks could be done by users like searching, querying, and faceted browsing.[8] Some researchers realized the limitations of the existing Web at that time.

The Semantic Web was introduced to settle two explicit issues with the WWW, as known as Web 2.0, which are from accessing data and enabling delegation.[4] For accessing data, documents on the WWW are indexed by simple texts and given access by links. Due to this character and searching algorithm, ambiguous concepts of one word cannot be distinguished and it is not feasible to conduct complicated matching associated with inference. The result is the "best fit" for one search. The final usable result needs a combination of content from multiple sources. The integration work has to be done by humans. Underlying data is not available. Many websites are generated from databases, which are not easy to search and use. For enabling delegation, information from data increases fast but the browsing devices are not used to infer or compute meaningful tasks for users. Delegation can assign tasks to machines including the integration and analysis of information as well as sense-making.

Shadbol et al.[5] thought the Semantic Web was attainable at that time drawing on some technologies from Artificial Intelligence(AI) research in the last 50 years. They considered that "the Semantic Web is a Web of actionable information derived from data through semantic theories." It is from an aspect to demonstrate the principle of the Semantic Web dealing data in a machine-comprehensible and processable manner. This cognition of the Semantic Web occurred after the initial article from *Scientific American* about the Semantic Web in 2001.[2] The Semantic Web would engage appropriate data to a particular context environment. They emphasized that AI would be one of the contributing disciplines in knowledge representation and ontology engineering. However, in 2006, it had a limited ability to collaborate with heterogeneous data. Especially in a large-scale situation, some researchers argued that "the Semantic Web has failed to delivery large-scaled data or information as an agent."

Moreover, there was a constant need for the integration of knowledge. Although some use of ontologies occurred, it was necessary to come up with a standard of Web and data. A great increasing need for shared semantics and a web of data is to integrate data components as information. It referred to ontologies to adopt common conceptualization. Ontologies first served biology, medicine. There were language standards that could be deployed on the Web for such fields. Therefore, the need to understand systems attended to a demanding requirement for data integration. In spite of some applications like eScience and e-government, the Semantic Web was not yet widespread adoption. The situation is similar or even the same as today.

Furthermore, conventional relational databases are not very effective or efficient to store, manipulate, and index a growing amount of documents in natural language. The area of information retrieval and extraction with natural language processing emerged. Thanks to these improvements, knowledge engineering can help to represent knowledge in a machine-understandable method. Semantic technologies began to provide descriptions of data, languages, and infrastructures facilitating computers to operate such things. Adding tags to semi-structured information as annotations allows computers to understand and process the value in the field. Employing semantic annotations, computational resources can recognize and manipulate the information. Some logical languages were employed in the context of the Semantic Web to express the sense of data. One of them is the Description Logic.[9] It is defined by many subordinate logic languages and works for many practical applications although some behavior is not ideally performed.[10] This could be solved by OWL discussed later.

Some language standards come to provide a foundation for semantic interoperability. There are lots of methods to communicate the semantics of data from HTML, for example, META, which is the anchor tag to encode semantic information. Universal Resource Identifiers(URI) can identify resources and provide the global network effects that drive the benefits of the Web.[12] Anyone can link, refer to, and retrieve from URIs. It showed the idea of shared information. It provides objects and relations that allow machines to obtain and process data or information directly for the Semantic Web.

The most important specification standard is Resource Description Framework(RDF). It provides a straightforward and powerful representation language in triples for URI.[5] RDF as defined in 1997 and became a W3C standard in 2004, which drew attention to the specification and promoted widespread deployment to enhance the Semantic Web. RDF semantically describes resources on the Web and forms a directed graph. This is a flexible data model that suits the context of the public Web. Furthermore, based on RDF, many derivative standards and languages occurred. For example, SPARQL is a query language of RDF families. It uses statements in RDF and defines a basic language for ontologies.[13]

With the improvement of greater expressivity in object and relation descriptions, Web Ontology Language(OWL) was born. OWL extends RDF to a full-complete Description Logic. This reflects and overcomes the problem in the previous paragraph about Description Logic. OWL defines properties of objects and relations to be inverse, transitive, symmetric, or functional.[14] It also defines the membership of instances for classed or hierarchies. All these new features provide more flexibility and expressivity for entities in semantic. Thanks to much extensive improvement, OWL has many variants and sub-languages, such as OWL Lite. Further, OWL has its successive and upgraded version, OWL2.[15] OWL enables efficient representations of ontologies and decisions that fit the requirement of greater expression in object and relation descriptions. In 2006, the authors saw the increasing adoption of OWL and continuous needs for tools to support their productions and applications.

To complete OWL with rules as a language framework, Rule Interchange Format(RIF) focuses more on rules and inference supporting and interoperating across various rule-based formats. It addresses the difficulty in specifying a formality to capture all knowledge within a certain domain. Inference combines AI methods to extend various logic to capture causal, temporal, and probabilistic knowledge.[16] This extends the Semantic Wen to include reasoning. However, it was just a start in 2006.

By then, the semantic web technologies have reached a stage where they could help design the architecture of networks. Following Open Systems Interconnection(OSI) model, Tim Berners-Lee initially come up with a similar conceptual structure of the Semantic Web, shown in Figure 1.[17] This OSI-like model first proposes an encapsulation method to integrate data and describe the relation. The data with information is encoded with Unicode and URI refers to resources. XML, NameSpace, and Schema are used as syntactic descriptions. The upper five layers of semantics are familiar which are discussed before. The layering has two main duties. One is to prevent a higher layer duplication of implementing functionalities provided by a lower layer. The other is to allow applications in each layer only to understand an interface provided by the lower layer and to give interpretations of definitions to the higher layer.

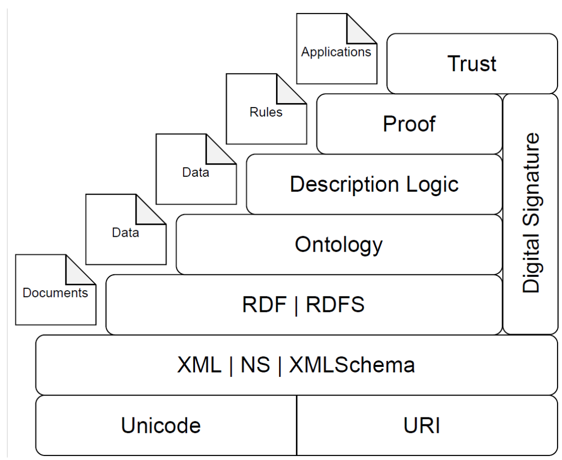

Fig. 1. OSI-like model of the Semantic Web

Unfortunately, this type of modelling has been turned out to be incorrect in the next several years' research[4]. The updated model is reviewed in the next sub-section.

The Semantic Web in the 2000s faced many challenges. Recalling the great need for data integration, there are also increasing needs for people and organizations to make their ontologies available. Industries wanted to make substantial reuse of existing ontologies and to enrich linked information space. Challenges in large-scale decentralized information repositories and construction of a Semantic Web browser together with requirements for new rules of integration and peer-to-peer protocols were factors that let many researchers look down and under-estimate the Semantic Web at that time.

3.2 Web 3.0 and Technologies in the 2020s

After 20 years' development. Semantic Web and the semantic web technologies grow fast among research, applications, even industries. The era of the Web turns to Web 3.0, which is specifically the Semantic Web. The Semantic Web field become more diverse. It has turned to an enduring goal to widely use the Semantic Web. All potential methods and tools to create and maintain the Semantic Web are also taken into account.[18] Some new and improved technologies occur such as Linked Data and knowledge graphs. More and more investigations and applications use these with the W3C standards. This sub-section will review some research and technology improvement around 20 years and the point of view to the Semantic Web field nowadays.

Firstly, as mentioned before, researchers started to realize that the OSI-like model for the Semantic Web1 did not properly describe the structure model of the Semantic Web. Recalling the model in the 2000s, there were two major downwards in the description and implementation. One is downward of compatibility. Application at any layer in the model should be able to use and interpret information from the lower layers. This is not the case in the OSI-like model. The other is downward of internal OWL. Ontology vocabulary is obtained by OWL. The implementation in OWL would interfere with the higher layer. For example, one type of OWL, OWL Lite, contains a Description Logic. This is another conflict with the OSI-like model.

The problem was reflected into an updated model of the Semantic Web, shown in Figure 2.[4] The updated model has a more accurate to describe a realistic implementation of the Semantic Web in accordance with developing and deploying. Compared to the formal OSI-like model, SPARQL as a query language is regarded as a substitute to some formal logic patterns. The layers are able to interact and some of them cooperate to fulfil the interface requirement. OWL and RIF become the branchings for ontology, logic, and rules.

The alternative model of the Semantic Web is better to illustrate the components and the structure. This is motivated by the development and improvement of research on semantic web technologies. Similarly, the updated model will someday be updated in the future. Not surprisingly, it already has shortages discovered by later studies.[4] Actually up to now, there might not be a complete theory that can be applied to select one integral language or structural model that syntactically restricts and semantically extends first-order logic for description.

Next, researchers have reached an agreement that metadata is a basic substance to interconnect via the Semantic Web. In order to represent machine-processable metadata, it requires a set of operations to define associations to the content on the Semantic Web. Every item needs registering meta statements.[4] One accustomed procedure follows tagging, taxonomy, and ontology.

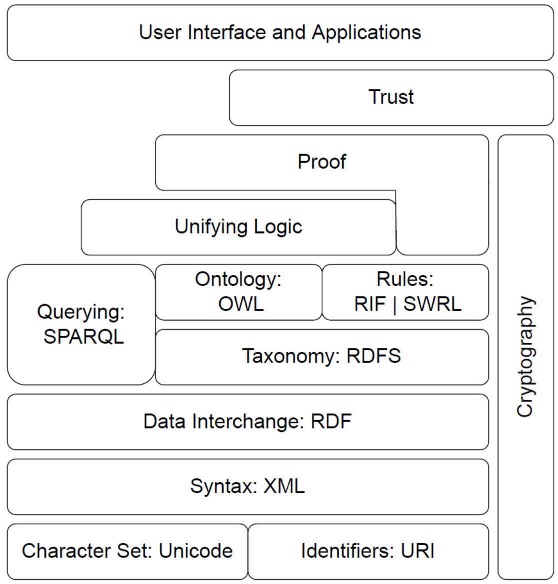

Fig. 2. Updated model of the Semantic Web

Tagging is a simple technique that supports lists of keywords extracted from natural languages. It is freely predefined by a controlled vocabulary, namely subject indexing.[19] The subject indexing is based on a lexical database, for example, WordNet. However, tagging may not be necessarily useful for unfamiliar concepts. Therefore, it is not enough only to control vocabularies; also to manage their usage. With this requirement, the procedure moves to the next step, taxonomy.

Taxonomy is a schema of describing classification organized in a hierarchical structure.[20] One straightforward taxonomy can be specified in RDFS which supports a hierarchy of classes and properties.[4] The classes and properties must be fulfilled with formal definitions, OWL or RIF, when attaching usages with vocabularies. Here, the last step is ontology, which is to conduct an explicit and formal specification of shared conceptualization.[21][22]

Formal ontologies mostly develop constantly during these 20 years. Based on ad-hoc modeling, the methodology for the development of ontologies has been under research but has not yet led to any breakthrough output.[18] It is turned out to be tough to maintain, share, and reuse. On the contrary, ontologies are now one of the major means of attaining data integration and sharing in the context of the Semantic Web. A propulsive idea is that existing ontologies ought to be reused by other users and applications. To deal with such a dilemma, Linked Data would become the next generation design principle of organizing information on the Semantic Web for both research and applications.

Then, Linked Data was focused on organizing and integrating data on the Semantic Web. It is made up of a large set of RDF graphs. Each RDF graph is linked with the sense of data. The amount of such public linked RDF graphs has been growing significantly in recent years. On May 1st, 2007, only 12 RDF graphs made up the Linked Open Data Cloud(LOD Cloud). As of May 5th, 2021, 1301 RDF graphs consist of a larger one and are continuously enriched. In particular, there is a huge increment during these ten years for no more than 100 graphs in 2009 but over 1300 nowadays.[23] L Rietveld et al.[24] reports that over 37 billion RDF triples are available extracted from more than 650,000 documents, which is only a selective component of all RDF graphs that can be accessed in 2015. Besides the large scale in amount, a rich diversity of topics are covered in LOD Cloud. The topics include social networking, scientific publications, life sciences, media, linguistics, media, government, and geography.[18]

With the boom of Linked Data, ontologies gradually do not play a leading prominent role in data integration. They turn to be often used as a schema that organizes the internal structure of graph-based datasets. Compared to the overpromise and depth of research on ontology, Linked Data, for example, Linked Data Cloud, is rather simplistic and only a sort of shallow representation of data. Recalling the unreliable reusability of ontologies, it is a simple approach based on triples and links between RDF datasets that carries more realistic promises for data integration and management.

Following this idea, RDF-based and graph-based data thesaurus was developed and recommended, such as occurring early SKOS.[2] It was later this time, from 2009 when SKOS was assigned as a standard. In 2011, Schema.org[3] appeared on the Web with its evolutionary and structural data management.[25]. It suggests website providers annotate entities with links on Schema.org vocabularies as metadata. Another important effort is Wikidata[4], launched in 2012.[26] It by January 2022 contains over 96 million data items, has had more than 1.5 billion edits since Wikidata was set up, and over 23,000 active users.

While there are many outstanding applications of Linked Data, shallow and non-expressive schema utilized in Linked Data appears to be a dominant challenge. In the meanwhile, Knowledge Graphs by Google emerged and drew related research and applications.[27]

People may have used an application of Knowledge Graphs when searching on Google. Google deploys Knowledge Graphs to navigate from one node to others in a graph structure by directing the active hyperlinks. Other than Google, the technology of knowledge graph has taken an outstanding place in leading information technology companies, such as Apple, Microsoft, and Meta Platforms(known as Facebook).[18] The biggest difference is apparently the changeover between academic research and uses in industry directly. Therefore, recent improvement around Knowledge Graphs is supported and pushed by powerful industrial use cases.

[2] https://www.w3.org/TR/2009/REC-skos-reference-20090818/
[3] https://schema.org/docs/schemas.html
[4] https://www.wikidata.org/wiki/Wikidata:Main_Page

Looking back at the history of the Semantic Web, especially of ontology and Linked Data discussed before, Knowledge Graphs is mostly a new idea and field out of the Semantic Web. However, many issues and obstacles for Knowledge Graphs continue to exist the same as they are for Linked Data. According to the current challenges proposed in Noy et al.[28], research with concrete substance to deal with those questions has been carried out.

3.3 Linked Data Semantic Integration

As discussed briefly in the previous, the Semantic Web was limited in domains in the 2000s. Shadbol et al.[5] thought that it required an increasing amount of data exposed in RDF. Information and knowledge should be uptaken in reusing RDF data for one's own and others. In 2006, still many applications either imported legacy data or redeposited it into a single large repository. This was a public and common problem hindering wide use of the Semantic Web. For the reason that research and applications operated in particular domains, it was not trivial to make RDF and URI public on the Web. The key appeared in the 2010s, the management of ontologies and knowledge has come into research and deployment for public use, like Linked Open Data.

The main target of Linked Data is linking and integration to ease the process of data discovery, analysis, and to offer integrated querying.[29] The major principles of Linked Data were officially proposed in the recommendation of W3C[5]: URI as the namespace, HTTP as the access method, stored in RDF, queried by SPARQL, and interacted URI links.[30] More specifically, data described by different schemas can be first transformed and later integrated in this flexible way.

To have an overview of the semantic integration, a landscape illustrates the integration process in a structured way. The multidimensional spaces are defined by the Cartesian product shown in Figure 3.[29] Each dimension is briefly described in the following:

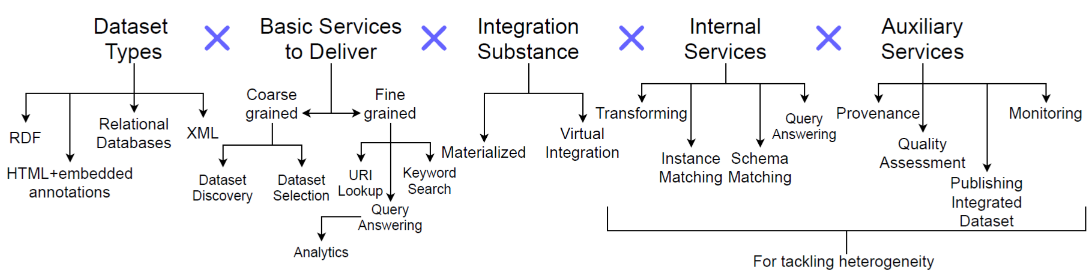

Fig. 3. The dimensions of data integration landscape

[5] https://www.w3.org/DesignIssues/LinkedData.html

– *DatasetTypes* Different types of dataset can be used as an input with embedded annotations. The dataset's owner is responsible for this step.
– *BasicServicesToDeliver* The main purpose of semantic integration is to deliver external services. The services will be offered by integrating several datasets.
– *IntegrationSubstance* The different integration substances include physical (materialized) process, virtual integration and their specialization.
– *InternalServices* The services are used during the integration process to join the data pieces.
– *AuxiliaryServices* The services can be optionally exploited either before or after the integration process, concerning provenance, accessing the quality, publishing, monitoring, and more.

Among the five dimensions, *InternalServices* contains important processes to integrate triples in RDF. Some different processes could be followed according to the survey.[29] One of the main processes is analyzed. There are three steps utilized for specifying and matching in semantic integration.

Step 1: Top-level Ontology-based or Competency Query-based Integration One typical process can be observed in Figure 4.[31] The integrated datasets should have the specification of providing either ontology schemas or/and competency queries. Then, the data of the individual dataset should be transformed in agreement with the integrated schema. Information integration is commonly completed within the context of databases by reconciling data from different sources under a general standard of schema.

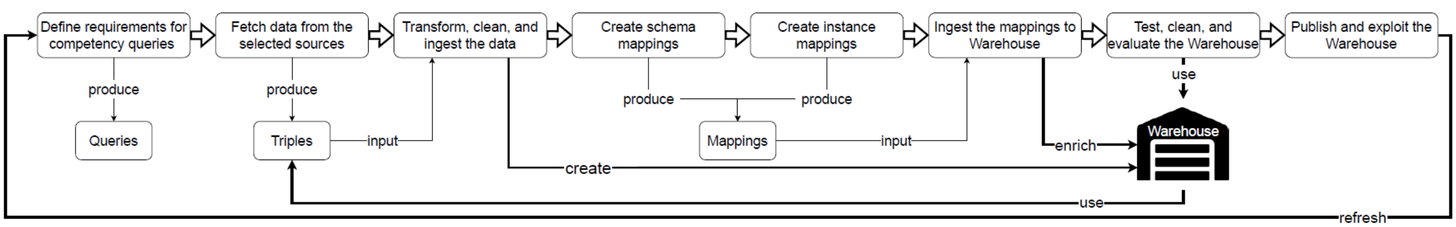

Fig. 4. Top Level Ontology-based Integration

Step 2: Automatic General Purpose Integration An integrated dataset can be produced as illustrated by a UML state diagram in Figure 5, which is a suggestive lifecycle model.[32] The aim is to build a general-purpose integration of a collection of datasets.[29] Moreover, it is a best-effort methodology and pretty burdensome to guarantee the integration satisfies the criteria of accuracy, validity, and completeness. It begins with collecting hundreds of datasets from public catalogs[6] and then it works out a transitive and symmetric closure in OWL to discover all equivalent entities among the datasets. Eventually, it creates

[6] for example, DataHub.io `https://datahub.io/`

indexes and performs measurements to offer services such as object co-reference discovery and selection. Each process for semantic integrating can be constructed into a procedure from serial transitions among the states in Figure 5.

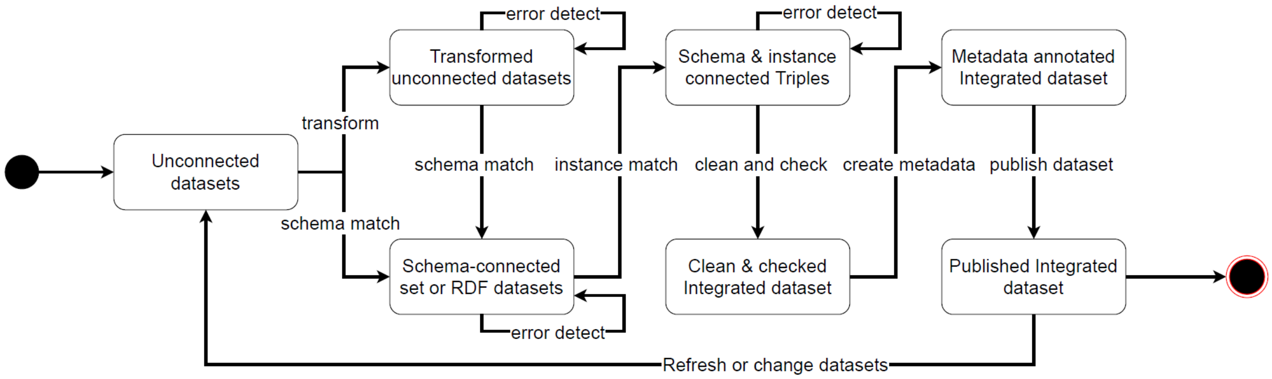

Fig. 5. Lifecyle of an Integrated Dataset

Step 3: Composite Processes The integration process in this step comprises two sub-process: the first process aims at discovering and selecting several most related datasets as the information needs, the second process aims at sub-level integration of the previously selected datasets. One of the composite processes can be regarded as a point in the five-dimension space mentioned before. The number of candidate datasets can be various depending on the abundance of integrated datasets.

The existing integration tools, like MatWare[31] and LODsyndesis[32] illustrated before, have only been employed on a small scale of datasets. And they cannot trivially expand to a large scale. As a result, there are many services for providing RDF datasets but each of them only presents a limited field of semantic integration. The Semantic Web requires large-scale data integration to hold up Linked Data datasets. The only best sharing attempt is Linked Open Data Cloud.

3.4 Interdisciplinary and Future Trend

As discussed before, the Semantic Web is not only driven by certain fields or methodologies. It is a multi-discipline topic. This characteristic is named interdisciplinary. For example, AI, like machine learning, shares many techniques and solutions to the Semantic Web field. In this sub-section, firstly, the interdisciplinary from some close topics is reviewed. Then, the outlook of future development and obstacles suffered in applications and enterprise are analyzed.

The field of database is another closely related to the Semantic Web. Graph-based structured data has a warehouse with data and metadata management.

In particular, researchers and developers pay much attention to the integration of heterogeneous sources. In the field of Big Data, the Semantic Web emphasizes mainly the various sources of data.[33]

For natural language query answering and automated constructing knowledge graphs from texts, Natural Language Processing(NLP) plays an important role in data integration. Machine learning, especially deep learning, has the capability to improve challenging tasks appearing in the context of the Semantic Web. Meanwhile, the semantic web technologies are contributing their potential to promote explainable AI.[34] Many other fields like cyber systems and the Internet of Things are researched and considered in the adoption of the semantic web technologies as well as smart manufacturing, smart energy grids, and building management.[35]

Through many discussions in literature, some of the most pressing challenges can be found in industrial knowledge graphs, ontology alignment, information extraction, question answering, ontology design patterns, and more. The Semantic Web needs consolidation most at this stage. The consolidation will come into being across the sub-fields, leading to application-oriented processes.[18] Surprisingly, this consolidation has been already taking place in industry as start-ups and multinationals adopt many of the semantic web technologies. However, many resources, like industrial knowledge graphs[36], are not shared, probably intending to protect their competitiveness. It still needs time and effort to have the corresponding software solutions becoming more widely used.

The Semantic Web expects a long-lasting and steady development of semantic technologies in the mainstream and infrastructures. Estimated by Gartner[7], by 2024, wide applications of semantic technologies will increasingly come into enterprise as a standard approach. And from 2025 to 2027, semantics will likely become a ubiquitous reality in an omnipresent Web, namely “semantic environment”.

Especially, two particular technology topics are proposed for growing in significance in the following several years: Internet of Things(IoT) and Ubiquitous Data Streams(UDS).[4] IoT raises challenges that everything is consuming and providing data on the Internet. IoT reflects the abilities of the Semantic Web to organize and mediate data in a large scale. Followed by the increment of IoT devices, UDS raises challenges by creating streaming data that IoT will generate, and how information in the streaming is extracted semantically and shared in an automated method considering privacy and trust.

Over these years, researchers and developers make efforts to tackle some critical gaps for industrial uptake including the scalability of reasoning, lightweight ontologies, and their maintenance. And then, the semantic web technologies will be ready for the mainstream so that the Semantic Web will be prosperous. Now, more needs for computing resources combined with the semantic technologies are suffering long-term challenges. Some of the prominent challenges appear in reasoning for a large scale of the Semantic Web, automated annotation, and semantic data mining for real-time data streams.

[7] `https://www.gartner.com/en`

Specifically, the following aspects can be primary challenges and obstacles at:

- *Security and Privacy* Security and trust models of the semantic context are important for publicly shared data and service usages.[4]. The challenges will arise from the Web ubiquity and heterogeneity of data streams. With an increment of participants, provenance issues will come from social aspects, like dependence on the social web and Collective Intelligence.
- *Sensor Networks* As mentioned before, IoT is growing and has been used in sensors in infrastructures. Ongoing effort on sensors with semantics can give rise to the uptake of the semantic web technologies. Such semantic sensors will face challenges from object description, real-time reasoning, service execution environments, and scalability.[4] Scalability and real-time reasoning are mainly open challenges. The ideal performance at an acceptable scale will perhaps remain inadequate for ten years.
- *Virtual Worlds* A persistent and decentralized online 3D virtual environment is heated nowadays, like Metaverse. The Semantic Web can provide the key requirement of it which contains extensive knowledge management to maintain semantic data in a virtual world like Metaverse and in the real world. Although it has been achieved for a limited scale only with a centralized approach, the goals of world-scale and distributed solutions are still huge challenges for further study.
- *Social Technology* The combination of social technologies and the Semantic Web to knowledge extraction and reasoning is the key of Collective Intelligence[37]. It enables decentralized workflows through the involvement of the Web.[4] Challenges lie in annotations and ontology development to capture and describe the involvement of the communities and their semantic meanings. Lightweight ontology and reasoning should be the main promoter of Collective Intelligence solutions.

In terms of industries, the Semantic Web has potential to be adopted in the following fields: energy industry, data integration production, multimedia annotation, health industry, smart life, and urban computing. In the following 20 years, the Semantic Web is believed to be as ubiquitous as the Web (WWW) is today. Furthermore, it will be improved with the capabilities of scalability, real-time reasoning, and world-scale knowledge management. The Semantic Web and the semantic web technologies are going to meet industrial adoptions as a mainstream. Certainly, there are both cheering prospects ahead as well as adventurous challenges.

4 Conclusion

The Semantic Web develops with time and has become not only a version of the Web, Web 3.0, but also brings about a set of technologies, namely the semantic web technologies. The field of the Semantic Web is a comprehensive discipline similar to Artificial Intelligence. It is given a character of interdisciplinary which has connections academically with AI, Big Data, Knowledge Engineering, NLP, Computer Vision, and more as well as relations industrially with Urban Computing, Metaverse, Collaborative Intelligence, IoT, and more. At the early stage of the Semantic Web in the 2000s, research focused on the representation of information, which settled down and emphasized Ontology with Description Logic and Rule-based Taxonomy. With Linked Data improving and the need for sharing, the methodology of Linked Data semantic integration has been developed and completed. Researchers have carried out some reliable manners to organize and manage Linked Data exploited on the Semantic Web. Nowadays, the Semantic Web is becoming the mainstream of data sharing, discovery, integration, and reusing. More and more applications and enterprises will show up to provide better solutions facing the large-scale data era. This section will summarize the literature review and answer the research questions proposed in Section 2.

- **RQ1**: What is the current status of the Semantic Web in applications?

  An abundance of knowledge and techniques have been produced in the Semantic Web field respecting efficient information management for data sharing, discovery, integration, and reusing. The accumulated contributions and improvement of the semantic web technologies are presented into applications, including public schema repositories, knowledge graphs in industry, ontology modelling applications, Wikidata, and Linked Open Data Cloud. The Semantic Web is welcoming its mainstream in IT for many organizations and companies show their preference to the adoption of the Semantic Web. Although some challenges still exist in knowledge extraction and large-scale data integration, there have been a few famous adoptions in applications like voice assistants Apple Siri and Microsoft Cortana, searching service Google Knowledge Graph, and Facebook Open Graph It is going to be ubiquitous as people understand and use WWW today.

- **RQ2**: How Linked Data organizes information on the Semantic Web?

  There are five dimensions to make up the semantic integration of Linked Data: the type of datasets, basic service to deliver, integration substance, internal services, and auxiliary services. The key processes lie in the integration substance. There are normally three steps of the process: elementary integration based on top-level ontology or competency query, automatic general-purpose integration, and composite integration. The data on the Semantic Web can be successfully organized using a group of integration tools in a small scale of datasets. Insufficiently, the integrated RDF dataset services today only present on a single specific domain not abundant enough. Linked Open Data Cloud is recommended since it contains a great number of links to datasets for simplifying the integration process with other sources.

- **RQ3**: What is the challenges in widely using the Semantic Web in enterprise?

  The challenges mainly from three aspects: issues from interdisciplinary, large-scale integration, and manual operation. The Semantic Web has a characteristic of interdisciplinary. Therefore, pursuing progress in the Semantic Web requires dedications from many subfields. And one urgent requirement is to find out an efficient method to piece together contributions or modifications of these sub-fields. To provide better and applicable solutions, the Semantic Web is facing challenges from security and privacy, sensor networks, virtual worlds, and social technology. A large-scale, even world-scale, data integration, and knowledge management is quite a huge obstacle in widely deploying and exploiting the Semantic Web. With the using areas of the Semantic Web expanding, the information is getting richer and richer so that no one existing repository can hold up all the required knowledge. It is not reliable merely to depend on rule-based integration and lots of manual interventions and operations in constructing datasets.

5 Summary and Discussion

The growing number of the Web users and websites has elicited the improvement of the Web technologies. The Semantic Web as a "new" type of the Web was born in the mid-1990s, later became a recommended version by W3C, and developed steadily more than 20 years. There have been many attempts in research, applications, industries, and enterprise. With artificial intelligence succeeding in many fields and computing power increasing, scientists and engineers start to discover more probabilities that can make the Semantic Web attainable on a larger scale. In particular, still many triples and RDF graphs are not ideally built as the literature illustrated. Many of the knowledge extraction and annotations are done by hand.

The next step in our opinion is to possibly implement an automated pipeline concerning AI techniques to explore and integrate data from the real world. This can motivate the large-scale data integration so as to push the Semantic Web into ubiquitous use cases. A pioneer of the AI-based semantic knowledge graph project is Semantic Artificial Intelligence (Semantic AI). In some way, it provides an AI strategy of data management based on machine learning to get implemented along the whole Linked Data lifecycle.

Thanks to more extensive and deeper research, people are looking forward to the prosperous prospects of the Semantic Web.

References

1. Siteefy: How Many Websites Are There in the World? (2022) https://siteefy.com/how-many-websites-are-there/. Accessed 12 January 2022
2. Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." Scientific American 284.5 (2001): 34-43.
3. Wikipedia: Semantics. http://en.wikipedia.org/wiki/Semantics (2022). Accessed 12 January 2022
4. Domingue, John, Dieter Fensel, and James A. Hendler, eds. Handbook of semantic web technologies. Springer Science & Business Media, 2011.
5. Shadbolt, Nigel, Tim Berners-Lee, and Wendy Hall. "The semantic web revisited." IEEE intelligent systems 21.3 (2006): 96-101.
6. Hitzler, Pascal. "A review of the semantic web field." Communications of the ACM 64.2 (2021): 76-83.
7. Dadkhah, Mahboubeh, Saeed Araban, and Samad Paydar. "A systematic literature review on semantic web enabled software testing." Journal of Systems and Software 162 (2020): 110485.
8. Patel-Schneider, Peter F. "OWL web ontology language semantics and abstract syntax, W3C Recommendation." http://www.w3.org/TR/2004/REC-owl-semantics-20040210/ (2004).
9. Baader, Franz, et al., eds. The description logic handbook: Theory, implementation and applications. Cambridge university press, 2003.
10. Horrocks, Ian. "Using an expressive description logic: FaCT or fiction?." KR 98 (1998): 636-645.
11. World Wide Web Consortium: The global structure of an HTML document, W3C Recommendation. https://www.w3.org/TR/html401/struct/global.html#edef-META. Accessed 13 January 2022
12. Masinter, Larry, Tim Berners-Lee, and Roy T. Fielding. "Uniform resource identifier (URI): Generic syntax." Network Working Group: Fremont, CA, USA (2005).
13. Brickley, Dan. "RDF vocabulary description language 1.0: RDF schema." https://www.w3.org/TR/rdf-schema/ (2004).
14. Bechhofer, Sean, et al. "OWL web ontology language reference." W3C recommendation 10.2 (2004): 1-53.
15. Horrocks, Ian, et al. "OWL 2 web ontology language profiles." W3C recommendation, W3C, Oct (2009).
16. Stevens, Robert, et al. "Ontologies in bioinformatics." Handbook on ontologies. Springer, Berlin, Heidelberg, 2004. 635-657.
17. Berners-Lee, Tim, and Dan Connolly. "Delta: an ontology for the distribution of differences between RDF graphs." World Wide Web, https://www.w3.org/DesignIssues/Diff (2004): 4-3.
18. Hitzler, Pascal. "A review of the semantic web field." Communications of the ACM 64.2 (2021): 76-83.
19. Wikipedia: Controlled vocabulary. https://en.wikipedia.org/wiki/Controlled_vocabulary (2022). Accessed 14 January 2022
20. Wikipedia: Taxonomies. https://en.wikipedia.org/wiki/Taxonomy (2022). Accessed 14 January 2022
21. Grimm, Stephan. "A unifying formal ontology model." Proceedings of the International Conference on Knowledge Engineering and Ontology Design (KEOD 2009), Funchal. 2009.

22. Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquis. 5(2), 199–220 (1993)
23. The Linked Open Data Cloud: History. https://lod-cloud.net/ (2021). Accessed 16 January 2022
24. Rietveld, Laurens, Wouter Beek, and Stefan Schlobach. "LOD lab: Experiments at LOD scale." International semantic web conference. Springer, Cham, 2015.
25. Guha, Ramanathan V., Dan Brickley, and Steve Macbeth. "Schema. org: evolution of structured data on the web." Communications of the ACM 59.2 (2016): 44-51.
26. Vrandečić, Denny, and Markus Krötzsch. "Wikidata: a free collaborative knowledgebase." Communications of the ACM 57.10 (2014): 78-85.
27. Google The Keyword: Introducing the Knowledge Graph: things, not strings https://blog.google/products/search/introducing-knowledge-graph-things-not/ (2012) Accessed 16 January 2022
28. Noy, Natasha, et al. "Industry-scale knowledge graphs: lessons and challenges." Communications of the ACM 62.8 (2019): 36-43.
29. Mountantonakis, Michalis, and Yannis Tzitzikas. "Large-scale semantic integration of linked data: A survey." ACM Computing Surveys (CSUR) 52.5 (2019): 1-40.
30. Linked Data- Design isseus: https://www.w3.org/DesignIssues/LinkedData.html (2009). Last accessed: 21 January 2022
31. Tzitzikas, Yannis, et al. "Matware: Constructing and exploiting domain specific warehouses by aggregating semantic data." European Semantic Web Conference. Springer, Cham, 2014.
32. Mountantonakis, Michalis, and Yannis Tzitzikas. "Scalable methods for measuring the connectivity and quality of large numbers of linked datasets." Journal of Data and Information Quality (JDIQ) 9.3 (2018): 1-49.
33. Janowicz, Krzysztof, et al. "Why the data train needs semantic rails." AI Magazine 36.1 (2015): 5-14.
34. d'Amato, Claudia. "Machine learning for the semantic web: Lessons learnt and next research directions." Semantic Web 11.1 (2020): 195-203.
35. Sabou, Marta, et al. "Semantics for cyber-physical systems: A cross-domain perspective." Semantic Web 11.1 (2020): 115-124.
36. Noy, Natasha, et al. "Industry-scale knowledge graphs: lessons and challenges." Communications of the ACM 62.8 (2019): 36-43.
37. Peeters, Marieke MM, et al. "Hybrid collective intelligence in a human–AI society." AI SOCIETY 36.1 (2021): 217-238.