Semantic Web and Linked Data - Summary Paper

Sai Vardhan Reddy Ainavolu
Department of Software Engineering, *Arizona State University*, Tempe, United States of America
sainavol@asu.edu

***Abstract*—This is a summary of the inception, evolution, application and future of semantic web engineering.**

***Index Terms*—Semantic Web Engineering, Linked Data, Data Integration, Future of the web**

I. Introduction

At the very first World Wide Web conference in the year 1994 Tim Berners-Lee envisioned a concept where web agents talk to each other by understanding the underlying semantics of the content; enabling a host of complicated actions which were considered far-fetched back then very much possible [1]. This idea took the name Semantic Web.

II. Evolution of the Web

Initially a collection of documents, the web evolved to what it is today. A massive and sophisticated collection of data that is routinely processed to extract information for inter and intra-operation of the web services. A major contributor to this progress was the establishment of the standards for the web. This allowed for a consistent understanding of the meaning between the services, which in turn made the websites that would otherwise be static to dynamic. The transactions of data that occurred in these dynamic web pages meant that the web was not just a wiki but a collection of data. Advancements in the field of Artificial Intelligence helped us to extract meaningful information out of this data which is being put to good use.

III. Technologies for the Semantic Web

Documents and underlying data were hard to process as it was primarily geared toward human interpreters with no information on the semantics of the content. A Layered architecture was developed to go along with the web infrastructure to make the semantics of the content more widely available over the web. The first layer is built on top of XML which allows for XML-based generic software and document languages to directly run on the semantic web. The second layer is made up of the Resource Definition Framework(RDF). RDF allows for simple expression of data in the form of triplets. The data types and properties for RDF are defined by RDF Schema(RDF-S). The last layer is the ontology layer defined by DAMN+OIL. DAMN+OIL makes up for some of the missing features in RDF-S for the AI community, like description logic. It also allows for more efficient and advanced querying [5].

IV. Linked Data

The foundations and the technologies developed over the years allowed us to link the data together moving us one step closer to realizing the vision of the Semantic Web. A crucial step in establishing an agent that communicates smoothly with all the web services is to link all the data together. RDF and ontology of the content help us to consistently link the together appropriately [2].

A. Linked open data cloud

People around the world have contributed to making existing data sets with open licenses are available throughout the web by publishing the data connected with RDF.

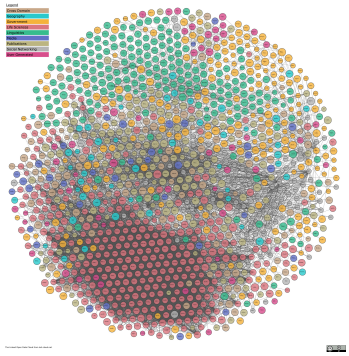

Fig. 1. Linked data open cloud [7]

V. Need for Data Integration

The data used by web services are heterogeneous and ever-changing. To keep up with the ever-changing data we need to automate or have a dynamic data source. Here we can leverage linked data. Since the linked data is from a global source it is always updated. All the services contributing to the linked data and using it will allow for better interoperability between the services [1].

VI. Applications of Linked Data

Linked data has applications in several domains. Web of data can be exploited in the following ways: Linked Data Browsers, Linked Search Engines, and some domain-specific applications [2].

A. Linked Data Browser

A linked data browser allows us to traverse the data links described by RDFs. We can jump from one source to another by clicking on the links. This provides a good human interface to traverse the web of data.

B. Linked Search Engines

1) Human Oriented Search Engine: This is similar to that of Google search where the engine crawls the web of data based on keywords or some description. Falcon and SWSE are some of the implementations of this Engine.

2) Application Oriented Indexes: This allows other linked data applications to query the web of data by exposing APIs. The primary purpose of these engines is to avoid every application to implement their own crawler of the web of data.

C. Domain specific Applications

Not all applications can take advantage of generic search engines like Application Oriented Indexes. To meet their domain-specific needs some applications integrate linked data into their applications. For example Revyu a film reviewing site uses linked data to retrieve more information about a film that is being reviewed.

VII. Research Challenges

There are many challenges we are yet to overcome before using the web as a global database that can be accessed by agents to interact with various services.

A. User Interfaces

Eventually relying on the web for data implies that it should be compatible with user interactions. When a user navigates to a page linked data browser should allow the same navigation. For instance, we can go front and back in a wiki but this is hard to recreate in the linked data as we would have dynamically recreate the page with a different focal point.

B. Link Maintenance

Depending on contributors alone to maintain the links is not feasible and scalable. Links have to be regularly updated for the web of data to be a consistent source for the applications.

C. Privacy

Protecting the privacy of users in a distributed system is a big challenge. Every contributor has to be verified along with the content they post and based on the laws of different countries the data must be subject to different policies.

VIII. Future of semantic web

The vision of the semantic web can be broken into two parts: Acting on Human decisions or Suggesting Humans.

A. Acting on Human decisions

The web needs to be completely machine-readable to achieve this is for all the use cases but for some use cases this is already achieved by manually setup workflows. For example, we can book flight, car, and hotel for a trip together based on our previous bookings.

B. Suggesting Humans

This is where there is some conflict between the semantic web and artificial intelligence. Recommending the user's next action is a key part of the vision for the semantic web which is already being achieved by AI as we speak. It is easier to add AI workflows to the services as you just need to append some necessary infrastructure but not make significant changes to infrastructure like in the case of the semantic web.

IX. Conclusion

While the semantic web has come a long way from a pipe dream progress in the fields of AI has raised some questions about its scope and relevance in the current world. Regardless of what entails in the future for the semantic web, it has certainly contributed to the progress of interoperability between the services. Universal interoperability between the services is something that not every service owner or user subscribes to but for the fields that it's desired like research and public domain services, there seems to be no other alternative. Thus making it relevant for those use cases.

References

[1] N. Shadbolt, T. Berners-Lee, and W. Hall, "The Semantic Web Revisited," IEEE Intelligent Systems, vol. 21, no. 3, pp. 96–101, 2006.

[2] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data - the story so far," International Journal on Semantic Web and Information Systems, vol. 5, no. 3, pp. 1–22, 2009.

[3] L. M. Machado, R. R. Souza, and M. Graça Simões, "Semantic web or web of data? A diachronic study (1999 to 2017) of the publications of Tim Berners-Lee and the World Wide Web Consortium," Journal of the Association for Information Science and Technology, 2019.

[4] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," Scientific American, vol. 284, no. 5, pp. 34–43, 2001.

[5] Ossenbruggen, Jacco van, Lynda Hardman and Lloyd Rutledge. "Hypermedia and the Semantic Web: A Research Agenda." J. Digit. Inf. 3 (2002).

[6] An, Yuan. "Data Semantics : Data Integration and the Semantic Web." .

[7] J. P. McCrae, "The linked open data cloud," The Linked Open Data Cloud. [Online]. Available: https://lod-cloud.net/. [Accessed: 30-Aug-2022].