Knowing the Unknown: Unshielding the Mysteries of Semantic Web in Health Care Domain

Pallavi Nagpal, Deepika Chaudhary and Jaiteg Singh

Chitkara University Institute of Engineering &Technology, Punjab, India.

Abstract

Health care is one of the popular segments wherein a huge amount of data is being generated every second and in a heterogeneous format. It is only the use of semantic technologies that can act as a guard for data integration and enriching it by adding proper semantics. The Semantic Web gives us a sensible and versatile answer to perceive proficiency across the congregation of healthcare data. Eventually the ultimate goal of improving the health care practices depends on how the healthcare data is linked together. The major challenges to chase the goal are not only limited to enable the integration of heterogeneous data sources but also requires the development of certain tools for efficient search, ontology management, and data analytics. The major objective of this study is to understand the technologies behind the Semantic Web and to unshield the mysteries of application of Semantic Web in the healthcare segment. The data for such an extensive survey was gathered from repositories of Pub Med, Scopus, Web of Science, and Google Scholar along with few medical science libraries.

Keywords 1

Semantic Web, Healthcare, Ontology, Knowledge base, SNOMED-CT

ACI’21: Workshop on Advances in Computational Intelligence at ISIC 2021, February 25-27, 2021, Delhi, India
EMAIL: pallavihptu@gmail.com (P.Nagpal); deepika.chaudhary@chitkara.edu.in (D.Chaudhary.); jaiteg.singh@chitkara.edu.in (J.Singh).

1. Introduction

The better definition of Semantic Web (SW) always begins with the word Semantic. In simple words semantics signifies meaning. Meaning allows effective use of data. Meaning is mostly missing in many of the information sources and is the job of the user or programming instruction to deliver the same. For example, if we simply look at the <H1> tag it is used to highlight major headings but as technical people, it is understood that the text which is surrounded by <H1> tag is of more importance to the user when compared to other. Basic semantics are attached for the search engines with <META> tags; however, they are just solitary keywords and lacks when linked. Semantics accord keywords with effective meaning through relationships. The Semantic Web in simple terms is a web of data paraphrased and linked together in ways to build context that follow a grammar & language construct. Although these semantics can also be added with a help of programming language in that case also formal standards are missing and also these semantics cannot be shared, aggregated, and validated. One can also attach these semantics through programming languages but in that case, too, there are no formal standards to be followed, and sharing, aggregation, and validation among statements becomes complex if not impossible. The Semantic Web all together supports the evolutionary nature of WWW. The Semantic Web architecture is based on the Layered Approach. These layers have some strong key dependencies between them. The Semantic Web layer cake presented in Figure 1 highlights these dependencies.

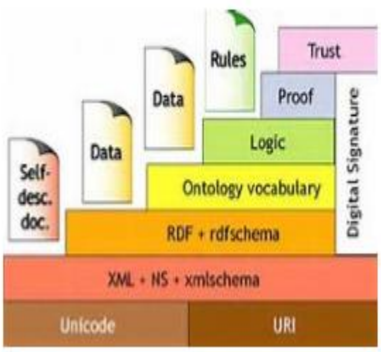

Figure 1 : The Semantic Web Stack

Layer 1 Uniform Resource Identifier (URI) – This layer uniquely identifies the available resources on the Internet. This layer ensures the uniqueness of Objects.

Layer 2 Extensible Mark-up Language (XML) – A language that lets it underlying user to create web pages with a defined vocabulary. This layer also supports sharing documents across the web.

Layer 3 & Layer 4 Resource Description Framework (RDF) - RDF acts like an Entity-Relationship(ER) model which is drafted for writing a set of statements to describe objects (resources) available on the web. RDF Schema on the other hand is used to write ontologies to represent relationships between web objects [26]. For integrating richness to the interoperability of data among various domains Ontology Web Language (OWL) was developed.

Layer 5 Proof & Trust layer - This layer is all about the representation of logics to deduce one document from another and there validations. The trust layer ensures the faith through the use of digital signatures.

To conclude we can state that the Semantic Web can be better understood as an extension of the WWW, it extends itself through the usable, systematized semantics that is based on research in knowledge/facts representation and the logic to approach the goal of ubiquitous information sharing. When compared to WWW it's the Semantic Web which primarily consists of statements for application consumption. These statements are not just plain statements for human interpretations but also include logic and act as meaningful links which can directly be interpreted by machines.

The outcome of this paper is to provide insights on the recent advancements done in this field. The more specific objectives can be summarized as follows:

a) To create a theoretical foundation in the area of data standards and interoperability, ontologies and its visualization, semantic data repositories and Semantic Web user Interfaces.

b) To reveal the current state of art of application of Semantic Web in Health Care Domain.

c) To un-shield the successful use cases of new approaches, techniques and Semantic Web applications in the field of health care data management.

The next section presents the present day challenges which require attention of the researcher community.

2. Healthcare Challenges in Current Scenario

The use of information systems and technologically advanced medical devices in various health care hospitals is producing and will continue to produce vast amount of data. This huge collection needs to be explored and transformed in such a way that it can be converted into valuable information which can thus improve the healthcare processes. Which is not easy and can be daunting because of various challenges? The following challenges needs to be addressed:

a) Interoperability of health and medical data.
b) Personalization of Ontologies and its visualization.
c) Explosion of health data and Semantic Data Repositories
d) Development of new user friendly interfaces

Semantic Web is attaining approval in addressing these challenges and therefore World Wide Web consortium (WWWC) established the Health Care and Life Science Interest Group in the year 2015 (HCLSIG,2015) with the objective to defend and support the use of Semantic Web technologies in health care segment. The rest part of the paper is organized as follows section 3 provides a background and short outline of prevailing literature for the mentioned challenges, section 4 highlights the new techniques and approaches relevant to health care segment, section 5 concludes the study.

3. Literature Review on Semantic Web Technologies in Health Care Domain

This section presents a short outline of a few of the prevailing literature towards Semantic Web technologies in the health care segment. In this phase, a detailed study of literature has been carried on to improve the quality and safety of web applications in the health care domain.

3.1 Data Standards and Interoperability

3.1.1 Theoretical Foundation

The automation of various health care services; usage of various medical information systems and other technological instruments have put together and still is generating volumes of medical data and that too in heterogeneous form. The huge volume of data has to be processed in a meaningful form to produce some valuable knowledge. This knowledge can lead to sound health care practices and thus can be of utmost use for humans[26]. When the data is coming from multiple sources it's the interoperability standards that act as a bridge to integrate and exchange the data across systems and services. To achieve this requires Schema matching techniques that can transform the data from human-understandable to machine-ready format. This technique when implemented can produce good healthcare services thereby reducing the cost by eliminating duplicate operations. The authors in their work have highlighted the meaning of interoperability for Semantic Web applications which can be defined as the vocabulary, organization, and structure of data required to integrate the data from multiple institutions [28]. Figure 2 gives a glimpse of various standards dealing with interoperability and the changes concerning time.

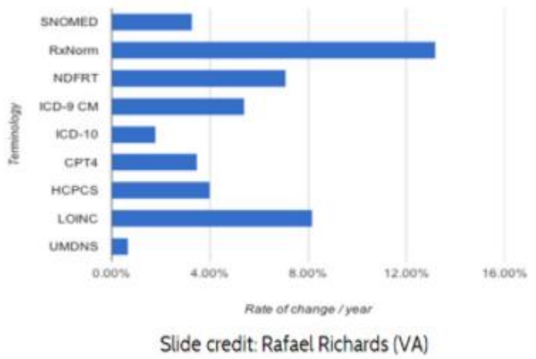

Figure 2: Rate of Change/Year (https://yosemiteproject.org/interoperability-roadmap/)

The interoperability can be categorized into two parts a) functional and b) semantic where functional deals with the common procedures and semantic deals with the framing of a common language which machine can understand in the end to end communication. To make applications fully interoperable requires standards. As per the literature mentioned the standards can be divided into a) Vocabulary or ontology Standards b) Data interchange and Integration standards c) Health record maintenance standards. These standards are established through four methods a) Adhoc b) Defacto b) Government-mandated) Consensus.

3.1.2 Present State of Art

Below we present the literature survey done in this direction:

The authors worked on schema matching techniques and presented a method for automating the process for matching schema at the field level; they achieved 71.8% when mapping the four staged process which includes string matching and substring matching [25]. The authors in their paper have discussed the standards required in Clinical data management [15]. Clinical data management deals with signs, operations, medicines, and lab values for a particular patient. The data here can only be interoperable when in a structured form. This transformation is a challenge and requires a lot of research. Further data analysis also requires the data to be converted in a structured form. The authors have given the FAIR guiding principle as one of the possible solutions to this problem. The acronym of FAIR is "Findability", "Accessibility", "Interoperability", and "Reusability". The authors in the study have mentioned Fast Healthcare Interoperability Resources (FHIR) as a newbie in the area of biomedical informatics and healthcare [26]. These guidelines when followed can provide a technological edge over health level seven (HL7). It was also discussed in the Yosemitea manifesto (https://www.dataversity.net/ semantic-interoperability-future-healthcare-data/) that for data exchange the Resource Description Framework (RDF) to be used. Existing data standards to be mapped with RDF's, Government agencies should mandate RDF as a Universal Healthcare exchange language. These standards can be adopted in the area of legacy systems, modern medical entities, and healthcare information systems. This standard can also leverage RESTFUL web services. The given section categories the literature review underwent by various researchers.

1. Private and public sector organization to align their systems with Fast Healthcare Interoperability Resources is the upcoming standard for data exchange. (https://www.dataversity.net/ semantic-interoperability-future-healthcare-data/).s [14]
2. Development of various standards for providing data interoperability; XML, GPS, Web Services, and Security, TCP/IP, 802.11, GPS [12]
3. Development of Reference Information Model (RIM) (HL7 Version 3 (V3) - a suite of specifications based on HL7's Reference Information Model (RIM)), its data elements, terminology, clinical statements, templates, document architecture. (http://www.hl7.org/implement/standards/rim.cfm).
4. Development of Collaborative Standard Hubs for quality improvement in the healthcare segment across rural and urban states (Srivastava et al., 2020).

3.2 Ontologies & its Visualization

3.2.1 Theoretical Foundation

Ontologies play a very important role in Semantic Web applications. They act as a common vocabulary for a specific domain consisting of a terminological collection of terms together with the rules to combine these terms and form relations. They also act as a basis for interoperability.

3.2.2 Present State of Art

Here, we first survey the work done in this direction. The authors in this paper did an extensive survey on the current state of ontology development in the area of health care [20]. They have studied various papers in the time frame of 2009 and 2018. They classified the work done by various researchers and proved that at present also there is a dire need for the development of new ontologies for delivering effective healthcare services. The authors in this paper have highlighted the role of ontology in decision making and how ontology that defines concepts such as disease, location, and environment and the interrelationships can influence the process of decision making in the public sector domain [26]. They also focused on the various mapping methods to work on interoperability issues. Further, they have developed a centralized knowledge base for healthcare systems specifically for the Tamil Nadu Region of India. The author in this paper has integrated the data which is coming from different sources and thus created an ontology [19]. They have followed four-way steps which include identification of data sources, the formalization of concepts, performed audit, and thus formulated ontology. In the table given below we summarize the work done in the direction of creating new ontologies into past, Work in progress, and Future Scope. The authors stated that today also lot of data which is published by big organization like WHO is published in proprietary format and not in accordance with Semantic Web standards and therefore it is still very difficult to integrate and further process those chunks [28]. Although a lot of research work (WHOGO, 2015), (PubMed, 2015), (NIH, 2015) has been carried in this direction still there is a long way ahead. The table below summarizes the literature.

1. Development of a Centralized knowledge base. Personalization of ontologies for describing the class hierarchies among chronically ill patients to form a decision support system for chronically ill patients (Riaño et al., 2012)
2. Introduction of Electronic Health care system to improve the various parameters of health care services [19].
3. Improvement was done in various organizations dealing in the service delivery segment, availability of reliable health data for healthcare providers, and improves upon in the public health system. A lot of new ontologies were also developed. The process of ontology validation was also improved [27].
4. Development of comprehensive monitoring frameworks in the field of maternal health informatics that would be created with the consensus of people practicing it and also on the ontology-based data integration approach. These frameworks will facilitate the research and evidence-based decision support systems [11].
5. To work upon improving the quality of health services using semantic web technologies. Majorly the privacy and security, trust, risks, and social implications and the quality of information are important and play an important role in the semantic web areas [15].
6. HCLS Knowledge Base is designed as a knowledge base where data from multiple sources (PubMed, Clinical Trials) have been stored. WHOsGHO – From WHO statistics about 3 million of data has been converted to RDF [23]. Other projects similar to the one mentioned above like DailyMed (2015) would continue to generate data from the health domain based on the principles of linked data [23]
7. The data hub is currently indexing hundred of datasets which are tagged under the healthcare category. The requirement of a lot of new projects in this direction which can convert the data stored in a proprietary format to linked data [23]

3.3 Semantic Websites user Interface

3.3.1 Theoretical Foundation

There are multiple projects such as CardioShare and Bio2RDF which have certain capabilities for navigating and querying the underlying Semantic Web data. However they lag intuition and can be more improvised.

3.3.2 Present State of Art

The projects which are designed in this direction should have the capabilities for searching and navigating through the Semantic Web data [28]. There are certain projects which have these capabilities like Bio2RDF and CardioShare, but these projects are limited and are not that intuitive which means a novice user will find it very difficult to explore and visualize the RDF triples. Therefore, there is a strong need to develop a good interface while developing the Semantic Web Applications. The architecture is so simple that even a beginner can explore it without much difficulty. The authors in this paper have designed and Semantic Web Portal (SWP) which is a light weight portal to browse and visualize the data generated and that too in meaningful and friendly way [9]. This system was deployed in Indiana University Health

Care Center to store and visualize the semantic information from one place and was used by multiple users i.e the patients the doctors, the practitioners to look all semantic information in one place.

1. Design and development of few interfaces to explore and navigate the Semantic Web data. The portals can only read the JSON files [18]
2. To make this process more interactive the researchers are working on the strategies which can also read the dynamically generated JSON objects. For novice users, the researchers are working to make the interface user friendly and easy to explore [18]
3. Semantic Graph mining to identify and rank the nodes and relationships. Development of user friendly and more intuitive interface and graph visualisation methods to navigate through various ontologies (https://arxiv.org/abs/2008.03053).
4. Representation of Semantic data in proprietary formats- A lot of big organizations share the data in PDF or spreadsheet format which makes it difficult to integrate. CardioSHARE (Vandervalk, McCarthy& Wilkinson, 2008) represents a decentralized web service framework that provides a SPARQL endpoint that enables querying transparently resources in the "deep web" from distributed and independent source.
5. Bio2RDF in this project 11 billion of data that comes from various heterogeneous sources have been ported to RDF formats (https://bio2rdf.org/).Bio2RDF in this project 11 billion of data that comes from various heterogeneous sources have been ported to RDF formats (https://bio2rdf.org/).

3.4Semantic Web and Reasoning

3.4.1Theoretical Foundation

The success behind any Semantic Web application lies in the inference process. Semantic Analysis is the mapping between syntax and the meaning of the sentences. Semantic analysis is used to check whether inserted sentence is accurate or not. The Semantic Web application designers are extremely benefitted if they can select a suitable reasoner as per the design of the application. A reasoner can therefore be defined as a piece of code which is able to infer logical consequences among a set of declared facts or axioms [8]. Correctness, efficiency, soundness and completeness of the new inferences drawn are some of the important attributes of a good reasoner. FaCT++, Pellet, HermiT, Kaon2, Hoolet are the few examples of Semantic Web reasoners. The reasoners can be categorized into various categories as per the various OWL profiles. In the study done by [9] the authors have a clearly described the various types of reasoners.

4. Conclusions

Although the technologies in the underlying domain have contributed a lot in health care segment still this study reveals that there still certain challenges that needs to be resolved. This survey reveals the present day challenges which require an ultimate attention of the researcher .The end users can be benefitted in case we get an optimal solution to the challenges mentioned herein.

5. Future Work and advancements in Semantic Health Care Segment

Healthcare segment is generating a lot of data continuously. Semantic interoperability of data can help the human community a lot. The patients, the doctors, various organizations can get the maximum benefits from this. The figure below represents the future issues that need to be addressed to make this technology a huge success.

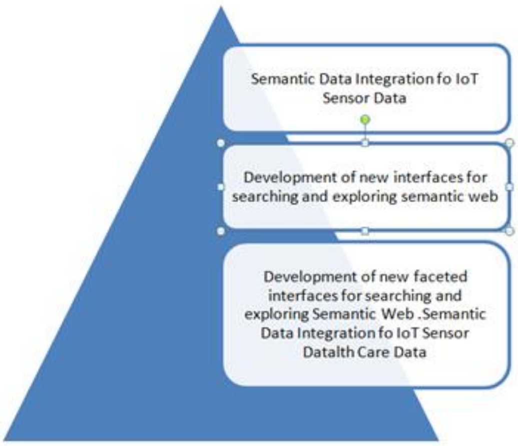

Figure 3: Advancements in Semantic HealthCare

a) Semantic Data Integration for IoT Sensor Data

In present scenario, about 35 billion of IoT devices are connected and it is predicted that this number would grow around 120 billion in 2025, which would be generating around 180 trillion gigabytes of data. This data comes from various heterogeneous devices thus making the formats incompatible to integrate. Which creates a significant problem for IoT application developers? Semantic Annotations and Clustering can be used as a method to integrate this data which is a challenge and can be considered as a future scope [3]

b) Development of new faceted interfaces for searching and exploring semantic web health care data

Irrespective of the complexity and size of data interfaces act as central linkage between human computer interactions. This becomes more complex when data is coming from various heterogeneous systems. Therefore, there is a need of designing new faceted and interactive interfaces [24]

c) Integration of Machine learning algorithms into Semantic web reasoners

Certain new reasoners are required in the area of the health care domain to generate inferences that can learn from themselves and are based on ANN or deep learning techniques [13].

References

[1] Andročec, D., Novak, M., & Oreški, D. (2018). Using Semantic Web for Internet of Things Interoperability: A Systematic Review. International Journal on Semantic Web and Information Systems (IJSWIS), 14(4), 147-171.

[2] Allemang, D. and Hendler, J., (2011). 'Semantic web for the working ontologist: effective modeling in RDFS and OWL'. Elsevier.

[3] Balakrishna S., Thirumaran M., Solanki V.K. (2020) IoT Sensor Data Integration in Healthcare using Semantics and Machine Learning Approaches. In: Balas V., Solanki V., Kumar R., Ahad M. (eds) A Handbook of Internet of Things in Biomedical and Cyber Physical System. Intelligent Systems Reference Library, vol 165. Springer, Cham. https://doi.org/10.1007/978-3-030-23983-1_11

[4] Berners-Lee, T., Hall, W., Hendler, J.A., O'Hara, K., Shadbolt, N. and Weitzner, D.J., 2006. A framework for web science. Foundations and Trends® in Web Science, 1(1), pp.1-130.

[5] Bergman M.K. (2018) Platforms and Knowledge Management. In: A Knowledge Representation Practionary. Springer, Cham. https://doi.org/10.1007/978-3-319-98092-8_12

[6] Bernstein, A., Hendler, J., & Noy, N. (2016). A new look of the Semantic Web.

[7] Brooks, P., & Avera, H. (2010). Standards and interoperability in healthcare information systems: Current status, problems, and research issues. In Fifth MWAIS Conference.

[8] Jain, S., Jain, N. K., & Goel, C. K. (2009). Reasoning in EHCPRs system. Int. J. Open Problems Compt. Math, 2(2).

[9] Ding, Y., Sun, Y., Chen, B., Borner, K., Ding, L., Wild, D., & Toma, I. (2010). Semantic web portal: a platform for better browsing and visualizing semantic data. In Active Media Technology (pp. 448-460). Springer Berlin Heidelberg

[10] El Asikri, M. *et al.* (2016) 'Contribution to ontologies building using the semantic web and web mining', *Proceedings - 2016 International Conference on Engineering and MIS, ICEMIS 2016*, (February 2018). doi: 10.1109/ICEMIS.2016.7745329.

[11] Henao, J., Quintana, Y. and Safran, C. (2019) 'An informatics framework for maternal and child health (MCH) monitoring', *Studies in Health Technology and Informatics*, 257(February 2020), pp. 157–162. doi: 10.3233/978-1-61499-951-5-157.

[12] Kubben, P. (2019) *2019_Book_FundamentalsOfClinicalData Scie*. Available at: papers3://publication/uuid/5CA1CB66-E0F7-4439-866E-935EFEAF8720.

[13] Kulmanov, M., Smaili, F. Z., Gao, X., & Hoehndorf, R. (2020). Semantic similarity and machine learning with ontologies. Briefings in Bioinformatics.

[14] Mandel, J. C. *et al.* (2016) 'SMART on FHIR: A standards-based, interoperable apps platform for electronic health records', *Journal of the American Medical*

Informatics Association, 23(5), pp. 899–908. doi: 10.1093/jamia/ocv189.

[15] Ngo, Q. T. and Tao, N. B. (2020) 'Determinants of semantic web technology adoption from IT professionals' perspective: Industry competition, organization innovativeness, and data management capability', *International Journal of Data and Network Science*, 4(3), pp. 271–288. doi: 10.5267/j.ijdns.2020.6.005.

[16] Liyanage H, Krause P, de Lusignan S Using ontologies to improve semantic interoperability in health data BMJ Health & Care Informatics 2015; 22: doi: 10.14236/jhi.v22i2.159

[17] Okikiola, F. M., Ikotun, A. M., Adelokun, A. P., & Ishola, P. E. (2020). A Systematic Review of Health Care Ontology. Asian Journal of Research in Computer Science, 5(1), 15-28.

[18] Oren, E., Heitmann, B., & Decker, S. (2008). ActiveRDF: Embedding Semantic Web data into object-oriented languages. Journal of Web Semantics, 6(3), 191-202.

[19] Sachdeva, S., & Bhalla, S. (2012). Semantic interoperability in standardized electronic health record databases. Journal of Data and Information Quality (JDIQ), 3(1), 1-37.

[20] Satti, F. A., Ali Khan, W., Ali, T., Hussain, J., Yu, H. W., Kim, S., & Lee, S. (2020). Semantic Bridge for Resolving Healthcare Data Interoperability. In *34th International Conference on Information Networking, ICOIN 2020* (pp. 86-91). [9016461] (International Conference on Information Networking; Vol. 2020-January). IEEE Computer Society. https://doi.org/10.1109/ICOIN48656.2020.9016461

[21] Saripalle, R. K. (2019). Fast Health Interoperability Resources (FHIR): Current status in the healthcare system. International Journal of E-Health and Medical Communications (IJEHMC), 10(1), 76-93.

[22] Schulz, S., Stegwee, R., & Chronaki, C. (2019). Standards in healthcare data. In Fundamentals of Clinical Data Science (pp. 19-36). Springer, Cham.

[23] Zenuni, Xhemal & Raufi, Bujar & Ismaili, Florie & Ajdari, Jaumin. (2015). State of the Art of Semantic Web for Healthcare. Procedia - Social and Behavioral Sciences. 195. 1990-1998. 10.1016/j.sbspro.2015.06.213.

[24] Zhang, Guo-Qiang et al. 'Ontologies as Nested Facet Systems for Human–data Interaction'. 1 Jan. 2020: 79 – 86.Informatics 2015; 22: doi: 10.14236/jhi.v 22i2.159.

[25] Patel, A., & Jain, S. (2018). Formalisms of representing knowledge. Procedia Computer Science, 125, 542-549.

[26] Jain, S. (2018). Intelligent decision support for unconventional emergencies. In Exploring Intelligent Decision Support Systems (pp. 199-219). Springer, Cham.

[27] Jain, S., & Meyer, V. (2018). Evaluation and refinement of emergency situation ontology. International Journal of Information and Education Technology, 8(10), 713-719.

[28] Devi, R., Mehrotra, D., & Baazaoui-Zghal, H. (2020, June). An r2rml-based Approach to Map Dengue Patient Database to Ontology. In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 790-795). IEEE.