ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/esingbemi-princewill-ebietomere/122

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science

https://urfpublishers.com/journal/artificial-intelligence

Vol: 2 & Iss: 2

Research Article

Semantic Web Technologies and Machine Learning: A Conjoined Silver Bullet for Big Data

Esingbemi Princewill Ebietomere[1,2]* and Ganiyu Olabode Ajisegiri[3]

[1]Department of Computer Science, University of Benin, Nigeria

[2]School of Engineering and Computing, Regent College London, UK

[3]School of Built Environment, Engineering and Computing , Leeds Beckett University, UK

Citation: Ebietomere EP, Ajisegiri GO. Semantic Web Technologies and Machine Learning: A Conjoined Silver Bullet for Big Data. *J Artif Intell Mach Learn & Data Sci 2024*, 2(2), 438-444. DOI: doi.org/10.51219/JAIMLD/esingbemi-princewill-ebietomere/122

Received: 31 May, 2024; **Accepted:** 14 June, 2024; **Published:** 15 June, 2024

***Corresponding author:** Esingbemi Princewill Ebietomere, Department of Computer Science, University of Benin, Nigeria; School of Engineering and Computing, Regent College London, UK. E-mail: princewill.ebietomere@uniben.edu

ABSTRACT

Big data is a concept used to describe data/knowledge characterised by massive size (volume), rapid growth (velocity), and heterogeneity (variety). For efficient and effective decision making in every organisation intrinsically and intricately driven by data/knowledge, it is expedient to guarantee the veracity of this data/knowledge. Unfortunately, the physiognomies of this data impose a grave tailback which has made it impossible for traditional data processing and management techniques to thrive. Critical among these snags are interoperability and integration. Grippingly, semantic web technologies-a collection of standard, methods, and tools for the realisation of semantic web vision-have been touted as a panacea to ameliorating these problems. At the heart of semantic web technologies is machine learning-a computing methodology, and a subdomain of artificial intelligence (AI). This report explicates the intricacies of big data, its benefits, challenges, and as well, the semantic web technologies and machine learning. Furthermore, how these two disruptive technologies interoperate to conjointly address the problems (interoperability, and integration) acting as clog in the wheel to seamless interpretation, processing and management of big data for efficient and effective decision making by machine and human was exposed. At the end of this report, readers are expected to have insight into some critical problems associated with big data, the need to address these problems, and very importantly, how the combination of semantic web technologies and machine learning conjointly help address these problems.

Keywords: Big data, Semantic web, Semantic web technologies, Machine learning, Interoperability, Integration

1. Introduction

The indispensability of data/knowledge to the survivability of every organisation and society, particularly, in a digital economy (data driven economy) cannot be overemphasised[1,2]. A fact further bolstered with the crusade to make machines intuitive, autonomous, and as well perform repetitive task on behalf of human. These data are often big data whose veracity should essentially be guaranteed for it to be valuable for efficient and effective decision making[3,4] by human and machine. However, ensuring reliability and obtaining value from big data is undoubtedly herculean[5] as this data is usually characterised by massive size (volume), swiftness in its creation (velocity), and heterogeneous nature (variety)-all interplaying to create tailbacks in seamless organisation, management, and processing of such data.

These physiognomies of big data on the one hand can be greatly beneficial as the analyses of such massive data can give great insight for excellent decision making[6–8]-as obvious in many fields or domains including biological sciences, health, law, agriculture, and manufacturing-particularly, with the emergence of industry 4.0 compelling machine to machine interaction[2,3,7,8]. On the other hand, these features birthed several difficulties, pressing among which are; interoperability, and integration[4-6,8,9] which have not only rendered the traditional data processing mechanisms impotent but have as well led to the conception, birth, proliferation, and exploration of several disruptive stratagems and technologies. In this report, some of the important technologies revolutionising the acquisition, organisation, analysis, and management of big data are exposed.

The web-an exceptional mechanism transforming the internet-is a concept often misinterpreted as the internet by many. It is a compounded system of interconnected apparatuses that has become a global repository[10] and by nature, an undisputable instance of big data pool. It has over the years transmuted from read only (web 1.0), through read and write (web 2.0), to read, write, and execute (web 3.0, also known as the semantic web). The semantic web has several definitions and appellations ascribed to it. These varying descriptions are as a result of the several perspectives from which the concept is viewed; a situation complicated by the seemingly use of the word "semantic", and interconnectedness with other terminologies such as knowledge graph, linked data, web of data, and linked open data, which has seen it misconstrued for these terminologies[10–13]. Following from this, a consensus definition has remained evasive for the semantic web. While the semantic web can agreeably be referred to web of data, there is a fine line between the semantic web and terminologies like knowledge graph, linked data, and linked open data[10,11,14-17].

It is apposite to mention that semantic web at provenance began as a vision, but arguably has become a field[10,11,13,18]. Furthermore, the core technologies responsible for the realisation of the semantic web are collectively referred to as the semantic web technologies. Interestingly, the success of these technologies is also dependent on other technologies (methods/methodologies/tools) and notable among these is machine learning[19] -a computing methodology, and a sub-domain of artificial intelligence (AI)[20].

The crux of this report is to espouse how semantic web technologies and machine learning are working conjointly towards ameliorating problems associated with big data-in particular interoperability, and integration-acting as clog in the wheel to the seamless organisation, management, and processing of big data for efficient and effective decision making in a data driven economy. This report, will help researchers particularly, new entrants in this domain gain insight into some critical problems associated with big data, the need to address these problems, and very importantly, how the combination of semantic web technologies and machine learning conjoinedly help address these problems. This report is organised as follows: Section 2, expounds the concept of big data, its importance and challenges. Section 3 explicates the semantic web technologies, some of its notable categorisations, and its role in handling big data. Section 4 elucidates machine learning, some of its notable categorisations, and it role in big data. Section 5 describes how semantic web technologies and machine learning interoperate to achieve seamless interpretation, and processing of big data. Section 6 holds conclusion.

2. Big Data

There are varying descriptions of the concept "Big data" in literature. However, there seem to be a consensus on the characteristics of the concept. As earlier hinted, big data is characterised by several "Vs ", with the core ones being volume, velocity, and variety. In recent times, others like value, veracity, and variability have also become popular. A notable definition reflective of these characteristics is that of NIST[21] which states "Big Data consists of extensive datasets-primarily in the characteristics of volume, velocity, variety, and/or variability-that require a scalable architecture for efficient storage, manipulation, and analysis". This section describes these characteristics, and as well discusses the several challenges imposed by these features on its seamless management, and processing.

2.1. Characteristics of big data

i. **Volume:** for data to be described as big data, it should be such with massive volume. These data can be structured, unstructured, or semi structured[5]. Often times, the volume of this data makes it grim for it to be stored in conventional relational databases thus, obfuscating its organisation and processing with traditional techniques. A classical scenario is that of the web. The advent of web 2.0 brought about information overload on the web and the failure of most tools to cope with the volume of data on the web birthed the semantic web-which proposes an innovative way of organising data for ease of consumption and interpretation by human, and machine alike.

ii. **Velocity:** This defines the frequency at which data is generated/created. If the velocity of data is high, then such data is termed "big data". With the advent of several disruptive technologies like the internet of things (IOTs), and cyber-physical systems as evident in industry 4.0-an advanced manufacturing model used to refer to modifications linked to automation fields fused with information technology, and driven by knowledge, experimentation, and innovation[8,22]-this characteristic of big data has become more ubiquitous[8,23]. Additionally, this feature of big data imposes a strain on computational resources as a result of the fast pace at which the data is generated, thus requiring real time analysis of data for veritable decision making.

iii. **Variety:** This refers to the generation of data from diverse sources and varying forms (text, audio, and video) and formats (txt, json, xml, and rdf). This nature of data has led to the creation of tailback that makes it gruelling for continuous data organisation, sharing, and processing. This bottleneck obviously, necessitated the need for integration of such data for ease of interoperation of the interplaying components.

iv. **Veracity:** Though this is not often listed as a core characteristic of big data, but it has become one of the most desired feature of big data in recent time owing to several failures traced to lack of proper attention to it. The essence of data acquisition particularly in data driven economy is to generate insight for efficient decision making by human and machine for increased productivity; a process which is either marred or boosted by the veracity (quality-accuracy, exactness) of data deployed for analyses[24]. Consequently, it is critical to guarantee the veracity of data[3,4,8].

v. **Value:** This is one of the features of data that came with the advent of data mining-also referred to as knowledge discovery on database (KDD)-whose kernel and end product according to Ebietomere and Ekuobase[10], is "knowledge"-often operationally encapsulated as data[8]. The value of data is the wealth (economic and social) derived from data usage[21], and it is a function of data veracity-that is, the veracity of data is a determinant of its value and soundness of ensuing decision.

vi. **Variability:** This characteristic is closely related to variety. But the underlining difference being that, the same set of data may have different meaning and formatted differently from different data sources[21]; thus, resulting in data inconsistency-a condition that further complicates the management and sharing of such data among interplaying components.

2.2. Importance of big data

The importance of big data is evident in its deployment and observable in many fields including business, law, health, biological sciences, and manufacturing. A typical case in manufacturing is evident with the emergence of industry 4.0 as aforementioned, where efficient and effective monitoring, prediction, diagnosis, coordination, discovery, mining, recommendation, and question answering, are strongly reliant on big data analytics for human, and machine interaction (processability, and interpretation)[2,7,8,23].

2.3. Challenges of big data

Ensuing from the descriptions of big data physiognomies, problems such as that associated with storage, management, interoperability, and integration are intuitively evident[6]. In this report, emphasis will be on interoperability, and integration due to their criticality on the performance of organisations leveraging big data[6] and the fact that addressing these two problems proffer solution to many others. These problems are briefly described in the following sub-sections.

2.3.1. Interoperability: The term “interoperability” has several definitions adduced to it, with varying modification of the term to suit the domain of discourse[25]. It is not unusual to see definitions or descriptions from structural, and semantic perspectives, with the later perspective birthed with the advent of the semantic web. Some of the notable definitions of interoperability include: (i) that, captured in the compilation of IEEE standard computer glossaries which according to Tolk[26], sees interoperability “as the ability of two or more systems to exchange information and use the information thus exchanged”, and (ii) that captured in the work of Sachdeva and Bhalla[3], which stated “IEEE Standard 1073 defines semantic interoperability as shared data types, shared terminologies, and shared coding”. From the definitions, the following facts are evident; (a) for seamless exchange of information between two systems there is need for consistency in format. Ensuring this though grueling with multiplicity of data sources and formats, but can be realised via standardisation[27], (b) the information exchanged must be interpretable and understandable by the systems involved for it to be useful. This is possible if a uniform vocabulary is adopted across the systems exchanging and sharing information.

2.3.2. Integration: Data integration is simply the process of consolidating disparate data from diverse sources into a single pool (e.g. data warehouse, data lake, data lakehouse) in order to provide users with a unified view[28], for business intelligence or analytics. The complications in integrating data has seen it been described by some experts as a big, dirty, and hairy problem[5,29]. Attempts have been made in literature to also, formally define data integration for ease of comprehension. This is apparent in the work of Lenzerini[28] where data integration was defined as a triple; I = (G,S,M). G in the triple, denotes a global schema, expressed in a language L_G over an alphabet A_G; S denotes a source schema, expressed in a language L_S over an alphabet A_S; and M represents a mapping between G and S, established by a set of assertions of the forms; $qS \sim q_G$, and $q_G \sim q_S$, where q_S and q_G are two queries of the same arity, over S, and G respectively[28]. It is worthwhile to state that, while the problem of integration is a product of data variety, that of interoperability is a product of data variety and variability. Consequently, addressing the problem of integration can help ease the problem of interoperability as it is perceived that integration heralds interoperability[6]. How to address these issues is the focus of subsequent sections

3. Semantic Web Technologies

The description of semantic web technologies is evident in the semantic web layer cake-which has undergone several modifications over the years to depict the current realities in the realisation of the goal of semantic web[10]. An emblematic standard layer cake is shown in **(Figure 1).**

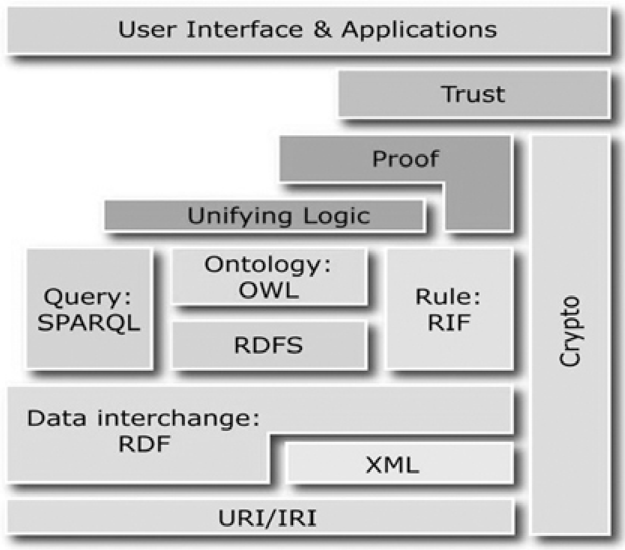

Figure 1: A semantic web layer cake[30].

From **(Figure 1)**, the slice in the architecture that is bounded below by XML portion, and above by the Unifying Logic is what is essentially referred to as semantic web technologies. Several categorisations of these technologies abound in literature[10,13,31-33]. Noteworthy among these categorisations is the interesting perspective taken by Ebietomere and Ekuobase[10], which conveys semantic web technologies along three primary dimensions; (i) standards (e.g. Extensible Markup Language, Resource Description Framework/Resource Description Framework Schema, Web Ontology Language, and SPARQL), (ii) methods (e.g. contextual analysis, natural language understanding, knowledge graph, linked data, and ontology) and (iii) tools (e.g. knowledge annotation tools, knowledge acquisition tools, and knowledge representation tools). It is from this perspective semantic web technologies are viewed in this report. Consequently, semantic web technologies can simply be seen as a collection of standards, methods, and tools interoperating towards the actualisation of the vision of machines’ seamless understanding of data or information and performance of cognitive tasks as surrogates. This report discusses important and popular semantic web technologies that have been deployed in mitigating the problems of interoperability, and integration.

3.1. RDF/RDFS

RDF and RDFS are often used interchangeably in many literature, however, there is a distinct dichotomy between them. While RDF is a data model represented as a triple in the form of subject, predicate, and object, RDFS is a representational language consisting of vocabularies for describing resources[34,35] with several variations including RDFS++, and RDFa[30,36]. A formal description of RDF can help one further appreciate its nitty-gritty. Thus, assuming I is the set of all IRIs, B is a set of blank nodes, L is a set of literals, and t denotes a triple; then, the RDF triple may be defined as $t = <s, p, o>$, where $s \in I \cup B$, $p \in I$, and $o \in I \cup B \cup L$, where s, p, and o are subject, predicate, and object respectively[29]. Besides, a `collection of RDF triples may be viewed as a labelled multi-graph where the subjects and the objects are the nodes in the graph and the properties posing as connectors between the nodes to form an edge shown as [37]. A classical application of RDF/RDFS is the Friend of a Friend (FOAF) project which describes people and the relations between them.

3.2. Web Ontology Language (OWL)

OWL is the most formalised language for knowledge representation-deeply rooted in classical logic. Its high degree of expressivity occasioned by its formalism has seen it found application across numerous domains-e,g. law, agriculture, and biological sciences-and thus, has become a de-facto language for knowledge representation (ontology). There are basically three sub-languages of OWL; (i) OWL lite (ii) OWL DL, and (iii) OWL full[38]. The degree of expressivity and formality increases form OWL lite, through OWL DL, to OWL full. Very importantly, which of the sub-languages to employ at a given instance is a function of the application's demand. The most recent version of OWL is the OWL 2, which has RDF/XML as its primary exchange syntax[10].

3.3. Ontology

The term ontology has its root in Philosophy and has over the years diffused to the field of Artificial Intelligence (AI) and Computer Science particularly with the advent of the semantic web. This diffusion across several fields has undoubtedly seen its definition metamorphosed. Ontology at provenance, according to Smith[39], was described as "the science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality". In recent times, a notable definition that seems realistic, relatable, and resonates with the common perception of the concept is that of Noy and McGuiness[40], which labels the concept thus; an ontology is a formal explicit description of concepts in a domain of discourse (classes-often referred to as concepts), properties of individual concept describing several features and attributes of the concept (slots-often referred to as roles or properties), and restrictions on slots (facets-often referred to as role restrictions). Its capability to help enforce seamless sharing of common understanding, knowledge reusability, domain knowledge disambiguation, and separation of domain knowledge from operational knowledge has seen it gain usage in many applications across several domains-for example, knowledge representation, semantic search and retrieval systems, and very importantly, interoperability[10,27].

3.4. Linked data

This is another method that has found usage, particularly, in data intensive applications. Many of its definitions in literature suggests the concept is synonymous to open data as can be seen in Rocha and Prazeres[41]. A more pragmatic description is that where linked data is understood as a set of best practices or guiding principles deployed for publishing and associating possibly unstructured data or a mix on the web (repository) for ease of readability, comprehension, and consumption by machine[10,14,29,42]. The essence of linked data is to foster proper linking, integration, and reuse of vast and heterogeneous data/knowledge in a repository[29,42,43]. Its potency in achieving tripod responsibility of linking, integration, and reuse has also, seen it gain popularity in many applications-for instance, digital libraries, and security. It is pertinent to mention that several techniques and standards exist for publishing linked data, and popular among these are RDF-for data/knowledge representation, and SPARQL-for querying data/knowledge from repository.

4. Machine Learning

The goal of AI is to make machine perform cognitive task on behalf of human in a domain of discourse by leveraging big data[10]. Achieving this goal undoubtedly requires proper coordination and management of complex interactions of several processes; a demand that has led to the proliferation of several competing algorithms and methods. Interestingly, machine learning is one of such computational methods-driven by statistical algorithms. Basically, machine learning can be classified as supervised, unsupervised, and reinforcement contingent on the approach or method deployed. These approaches are highlighted subsequently.

4.1. Supervised learning

It is used to refer to a machine learning model trained on labelled (with predefined tag) data. Here the training data includes the desired output. The model can then learn with the labels and ultimately ensures correct identification. It must be noted that labelling data is expensive (time, and cost), but its efficacy has seen supervised learning become arguably, the most rampant machine learning model used today. Typical supervised learning tasks include; classification (e.g. support vector machine, random forest), and regression (e.g. simple linear regression, and lasso regression)

4.2. Unsupervised learning

This refers to a scenario where a model is trained with data that are unlabelled (without predefined tag). The model finds a pattern in the data and tries to ensure correct identification[20]. This approach is often employed when; (i) there are no standard labels, (ii) labelling is ridiculously expensive (time, and cost), and (iii) there exists massive data to be explored. Typical unsupervised learning tasks include; clustering (using k-means, and principal component analysis), and association (association market basket analysis). It is pertinent to note that many machine learning models have explored a combination of supervised and unsupervised learning-known as semi-supervised learning[20]-which leverages the strengths of both.

4.3. Reinforcement learning

This involves a trial and error scenario[20], where the model interacts with a dynamic environment to achieve a certain goal. The crux of this approach is in learning from experience to better improve performance. This type of machine learning approach is particularly evident in autonomous driving vehicle, and game playing.

Generally, machine learning systems can be (i) descriptive-in which case data is leveraged to proffer explanation to happenings, (ii) predictive-in which case data is used to predict happenings, and (iii) prescriptive-where data is leveraged to make suggestions on what actions should be taken. The use of machine learning is evident in many applications including; recommendation systems, spam email filtering, self-driving vehicles, and internet of things. Again, it must be mentioned that what has been described in this section is just a brief overview of machine learning sufficient to help appreciate the discourse in the subsequent section.

5. Interoperation of Semantic Web Technologies and Machine Learning

Basically, systems that rely on semantic web resources and machine learning components are referred to as Semantic Web Machine Learning (SWeML) systems[44]. The interoperation of semantic web technologies, and machine learning particularly, vis-à-vis how both are helping each other accomplish tasks that somewhat would have been impossible or arduous to accomplish by either; is evident in many spheres including knowledge acquisition, and knowledge representation[45-48]. To avoid deviating from the goal of this report, this section only details some of these coactions that border on addressing the problem of interoperability and integration.

5.1. Machine learning techniques in construction of ontology for interoperability

As earlier mentioned, ontology is the corner stone of semantic web, and being a cutting-edge technology that guarantees sharing of common knowledge, and domain knowledge disambiguation among others, ontology has been described as a panacea to the problem of interoperability[3,27] and as such has been used extensively, as evident in Sachdeva and Bhalla (3), Azarm, and Peyton[27,49,50]. Crafting ontology for interoperability is a delicate task that requires high level of expertise (technical knowhow) and huge domain knowledge in any area of application. Unfortunately, many domain experts lack the technical knowhow in creating ontologies, particularly with the crudeness of most of the early tools for its creation as they were often manually driven and unable to cope with big data. Attempts to eradicating these bottlenecks, necessitated the drifting from manual (though the gold standard but tedious and time consuming) to (semi) automatic (less tedious, fast, but less accurate) ontology construction.

The automatic construction of ontology is largely data driven and relies heavily on machine learning for discovery/extraction of keywords/concepts, association, and clustering from massive data, hence rapidly expediting the entire process of construction. This is reflective in some powerful ontology construction editors like OntoEdit-Text-To-Ontology[45], and OntoGen[46]. For instance, OntoGen allows ontology to be bootstrapped from a corpus by employing several machine learning techniques. It employs latent semantic indexing (LSI) for extracting background knowledge from text documents, singular value decomposition (SVD) and bag of word representation of text documents for extracting words with similar meaning, k-means clustering for partitioning words into clusters, and either support vector machine (SVM) or centroid vectors for extraction of keywords[46].

After extracting the necessary concepts with machine learning techniques, one can then make these concepts machine interpretable and processible using semantic web technologies such as RDF and OWL (RDF/XML). The working together of these two disruptive technologies evidently helps to simplify the entire process of ontology construction from scratch to finish in the face of big data.

5.2. Leveraging machine learning techniques and linked data for integration

Leveraging machine learning for data mapping, autonomous learning, and data processing has undisputedly helped ease the handling of big data, as many of these activities were often performed with traditional tools which requires a lot of effort and time to set up, thus making it inapt for big data. Grippingly, the emergence of linked data-a semantic web technologies method driven by RDF/OWL, and SPARQL for big data representation and query processing respectively-the process of integration has become more seamless[29]. A typical ecosystem for linked data is as depicted in **(Figure 2)**.

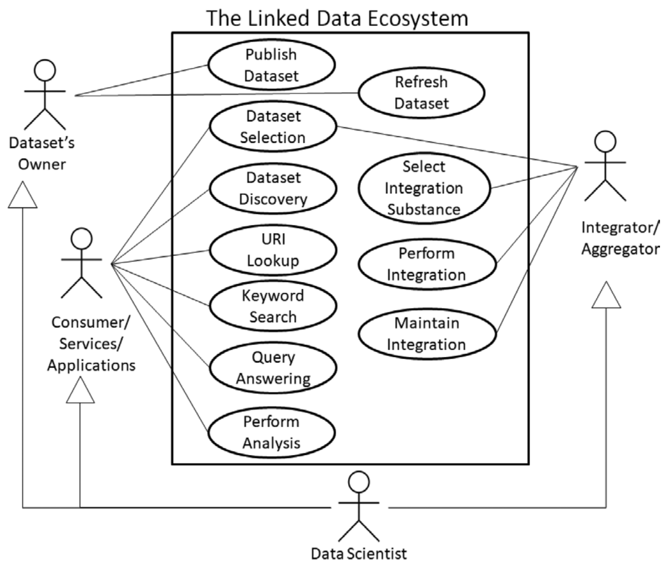

Figure 2: Linked Data Ecosystem use case diagram[29].

(Figure 2) shows four actors;

(i) Dataset's Owner-which could be organisations (private, and public) or individuals that are responsible for publishing and updating datasets

(ii) Consumer/Services/Applications-which refers to entities, services, and applications that consume data for diverse reasons, for example, data selection, data discovery, and analysis

(iii) Integrator/Aggregator-this could be organisations or individuals responsible for integrated access services provision by integrating data and as well maintaining data from different sources

(iv) Data Scientist-an instance of an integrator/aggregator. Observably, many of the processes in the ecosystem, are driven by machine learning techniques.

Overall, the deployment of machine learning techniques has enhanced rapid development of interoperable ontologies, knowledge graphs, and linked data. It is germane to state that, while ontology and knowledge graph have been extensively deployed in handling the problem of interoperability, linked data has been employed in addressing the problem of integration.

6. Conclusion

The report started with the description of big data, its characteristics, importance, and some of its challenges. It was made clear that the physiognomies of big data birthed several challenges which has made its processing and management gruelling, and critical among these challenges are interoperability and integration. Next, the technologies that can help address these grave challenges were exposed. This began with the description of some key semantic web technologies-standards (e.g. RDF, and OWL), methods (e.g. ontology, and linked data), and tools (e.g. knowledge Acquisition/representation tool-OntoGen). It was made clear that the potency of these technologies in addressing the problem of interoperability and integration is profound in its proliferation for handling such problems across several domains. Subsequently, the concept of machine learning, its approaches, and areas of applications were exposed. Also, it became obvious what machine learning has been used to achieve in the manipulation and processing of big data. Finally, the interoperations of semantic web technologies and machine learning were exposed. It was made clear that the mechanisms behind the workings of most of the tools for creating ontology. Knowledge graph, and linked data are driven by machine learning techniques. Consequently, one can conclude that semantic web technologies and machine learning are acting as conjoined silver bullet in addressing the critical challenges of big data-interoperability and integration; a trend that will undoubtedly continue.

7. References

1. Gokalp MO, Kayabay K, Akyol MA, Eren PE, Kocyigit A. Big data for industry 4.0: A conceptual framework. 2016 International Conference on Computational Science and Computational Intelligence (CSCI). 2016.
2. Berges I, Ramirez-Durán VJ. Illarramendi A. A semantic approach for big data exploration in industry 4.0. Big Data Research 2021;25: 100222.
3. Sachdeva S. Bhalla S. Semantic interoperability in standardized electronic health record databases. J Data Informat Quality 2012;3: 1-37.
4. Demidova E, Dsouza A, Gottschalk S, Tempelmeier N, Yu R. Creating knowledge graphs for geographic data on the web. ACM SIGWEB Newsletter 2022; 1-8.
5. Golshan B, Halevy A, Mihaila G, Tan W-C. Data Integration. Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems. 2017; 101-106.
6. Kadadi A, Agrawal R, Nyamful C, Atiq R. Challenges of data integration and interoperability in big data. IEEE Xplore 2014;38-40.
7. Krishnan R, Samaranayake VA, Jagannathan S. A Hierarchical dimension reduction approach for big data with application to fault diagnostics. Big Data Research 2019;18: 100121.
8. Ebietomere EP, Ekuobase GO. Semantic Search Engine in Industry 4.0. In: Patel A, Debnath NC, (edn.,) Semantic Technologies for Intelligent Industry 4.0 Applications Rivers Publishers 2023; 1-23.
9. Ostrowski D, Kim MA Semantic based framework for the purpose of big data integration. 2017 IEEE 11th International Conference on Semantic Computing (ICSC) 2017.
10. Ebietomere EP, Ekuobase GO. Semantic Web Technologies. In: Patel A, Debnath NC and Bhushan B. (1st edn.,)Semantic Web Technologies: Research and Applications. CRC Press 2022; 1-24.
11. Machado LM, Borges MM, Souza RR. The Evolution of the concept of semantic web in the context of wikipedia: an exploratory approach to study the collective conceptualization in a digital collaborative environment. Publications 2018;6: 44.
12. Machado LMO, Souza RR, Graça Simões M. 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. JASIST 2019;70: 701-714.
13. Hitzler P. A review of the semantic web field. Communications of the ACM 2021;64: 76-83.
14. Bizer C, Heath T, Berners-Lee T. Linked Data-The Story So Far. International Journal on Semantic Web and Information Systems 2009;5: 1-22.
15. Ehrlinger L, Woß W. Towards a Definition of Knowledge Graphs. Sematics 2016.
16. Bellomarini L, Sallinger E. Vahdati S. Chapter 2 Knowledge Graphs: The layered perspective. Knowledge Graphs and Big Data Processing 2020; 20-34.
17. Hao X, Ji Z, Li X, et al. Construction and application of a knowledge graph. Remote Sensing 2021;13(13): 2511.
18. Berners-Lee T. Semantic Web roadmap. 1998.
19. Tresp V, Bundschus M, Rettinger A. Huang Y. Towards Machine Learning on the Semantic Web. In: da Costa PCG, et al. (edn.,) Uncertainty Reasoning for the Semantic Web 2008; 282-314.
20. IBM What is machine learning? IBM 2023.
21. NIST. NIST Big Data Interoperability Framework: Volume 1 Definitions. NIST Special Publication 2019;1: 53.
22. Carvalho NGP, Cazarini EW. Industry 4.0-What Is It? In Ortiz H. Industry 4.0, Current Status and Future Trend. (edn.), Intechopen 2020; 3-11.
23. Ekuobase GO, Ebietomere EP. Semantic Web Services: The Interoperability Middleware Technology for Industry 4.0. In: Patel A, Debnath NC. (edn.,) Semantic technologies for intelligent industry 4.0 applications. River Publishers 2023; 25-43
24. Ortiz JH, Marroquin WG, Cifuentes LZ. Industry 4.0: Current status and future trends. Industry 4.0 - Current Status and future trends 2020.
25. Jochem R, Knothe T. Interoperability requirements derived from interoperability dimensions. Enterprise Interoperability II 2007; 39-50.
26. Tolk A. Interoperability, Composability, and their implications for distributed simulation: towards mathematical foundations of simulation interoperability. In: 2013 IEEE/ACM 17th International Symposium on Distributed Simulation and Real Time Applications 2013; 3-9.
27. Ebietomere E, Nse U, Ekuobase B, Ekuobase G. Crafting electronic medical record ontology for interoperability. The African Journal of Information Systems 2021;13.
28. Lenzerini M. Data integration. Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems 2002.
29. Mountantonakis M, Tzitzikas Y. Large-scale Semantic Integration of Linked Data. ACM Computing Surveys 2019;52: 1-40.
30. Sheth A. Semantic Web: Technologies and Applications for the Real-World. Kno.e.sis Publications 2007.
31. Matthews B. Semantic Web Technologies. JISC Technology and Standards Watch 2005; 20.
32. Hogan A. Linked Data and the Semantic Web Standards. In: Harth A, Hose K, Schenkel R. (edn.,) Linked Data Management 2013; 1-53.

33. Patel A, Jain S. Present and future of semantic web technologies: a research statement. IJCA 2019;43: 413-422.

34. Ciobanu G, Horne R, Sassone V. A descriptive type foundation for RDF Schema. J Logic Algeb Methods Programming 2016;85: 681-706.

35. Xue X, Wu X, Chen J. Optimizing ontology alignment through an interactive compact genetic algorithm. ACM Transactions on Management Information Systems 2021;12: 1-17.

36. W3C. Resource Description Framework (RDF): Concepts and Abstract Syntax. 2024.

37. Wylot M, Hauswirth M, Cudré-Mauroux P, Sakr S. RDF Data Storage and Query Processing Schemes. ACM Computing Surveys 2018;51: 1-36.

38. Horridge M, Jupp S, Moulton G, Rector A, Stevens R, Wroe C. A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools Edition 1.1. 2007.

39. Smith B. Ontology. In: Floridi L. (edn.,) Blackwell Guide to the Philosophy of Computing and Information. Blackwell 2003; 155-166.

40. Noy, N. and McGuinness, D. Ontology Development 101: A guide to creating your first ontology. 2001.

41. Rocha A, Prazeres C. LDoW-PaN: Linked Data on the Web-Presentation and Navigation. ACM Transactions on the Web 2017;11: 1-42.

42. Haller A, Fernández JD, Kamdar MR, Polleres A. What Are Links in Linked Open Data? A Characterization and Evaluation of Links between Knowledge Graphs on the Web. Journal of Data and Information Quality, 2020;12: 1-34.

43. Mountantonakis M, Tzitzikas Y. Scalable methods for measuring the connectivity and quality of large numbers of linked datasets. Journal of Data and Information Quality, 2018;9: 1-49.

44. Breit A, Waltersdorfer L, Ekaputra FJ, et al. Combining machine learning and semantic web: A systematic mapping study. ACM Computing Surveys 2023;55: 1-41.

45. Maedche A, Staab S. Ontology learning for the Semantic Web. IEEE Intelligent Systems 2001;16: 72-79.

46. Fortuna, B, Grobelnik M, Mladenic D. OntoGen: Semi-automatic Ontology Editor. Lecture Notes in Computer Science 2007;309-318.

47. Elnagar S, Yoon V, Thomas M. An Automatic Ontology Generation Framework with an Organizational Perspective. Proceedings of the 53rd Hawaii International Conference on System Sciences 2020.

48. Zulkipli ZZ, Maskat R, Teo NHI. A systematic literature review of automatic ontology construction. IJEECS 2022;28: 878-889.

49. Azarm M, Peyton L. An ontology for a patient-centric healthcare interoperability framework. Proceedings of the International Workshop on Software Engineering in Healthcare Systems 2018.

50. Adel E, El-Sappagh S, Barakat S, Hu J-W, Elmogy M. An extended semantic interoperability model for distributed electronic health record based on fuzzy ontology semantics. Electronics 2021;10: 1733.