

Article

Orchestrating Distributed Computing and Web Technology with Semantic Web and Big Data

Sandy Victor Amanoul¹, Lozan M. Abdulrahman², Rozin Majeed Abdullah¹, Riyadh Qashi³

¹ Engineering Department, Technical College of Engineering, Duhok Polytechnic University, Duhok, Iraq,

² Information Technology Department, Technical College of Administration, Duhok Polytechnic University, Duhok, Iraq,

³ Vocational School Center 7, Electrical Engineering of the City of Leipzig, Leipzig, Germany,

*Corresponding author: Riyadh Qashi (email: r.qashi@bsz7-Leipzig.de)

Received 17 September 2023; Accepted 16 October 2023; Published 15 December 2023

Abstract: Complex data systems are incapable of processing large data volumes, rendering the task of retrieving pertinent information unattainable. The advent of the Internet has amplified the significance of accessible and readily available information. Additionally, it receives support from the World Wide Web Consortium (W3C) and global organizations responsible for establishing web standards, such as Web Ontology, Inc. It expands the functionality of the website to facilitate the retrieval, integration, and transmission of information. In recent years, several major organizations have shown a strong inclination towards using semantic technologies for the purpose of collecting Big Data. Undoubtedly, there are other advantages of integrating this into the Creative. It enhances the ability of end-users to manage data from many repositories, focuses on changing the corporate environment and the user experience, and incorporates individual definitions and integrates several data sources. Furthermore, the market's evolving expectations and contemporary organizational practices require an adaptable but all-encompassing information strategy. Integration of data warehousing may be achieved by the use of scattered corporate ontologies. This study explores the impact of the Semantic Web on enhancing the intelligence of Big Data. It analyses the obstacles and opportunities associated with the integration of Big Data with the Semantic Web.

Keywords: Distributed Computing, Cloud Computing, Web Technology, Semantic Web, Big Data.

Copyright © 2023 Journal of Smart Internet of Things (JSIoT) published by Future Science for Digital Publishing and Sciendo . This is an open access article license CC BY (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

Any company, such as retail distribution and orders, has undergone significant transformations to become more fully computerized because of the synergy between data processing and systems analysis. Data procedures and company programs that often can handle the commodity details are therefore held as data. Nevertheless, the increasing complexity of business and electronic data

processes is challenging companies by the complexity. In the context of extensive data, this issue is often discussed [1]. Due to its comprehensive studies and application opportunities, big data has become more aware and appreciated. Generally, More significant amounts of complex knowledge and various variations in categories are referred to as "big data." When you are dealing with big data, you should expect to gather all three data types: Organized, semi-structured, and unstructured data. Approximately 85% of the data consists of hierarchical and semi-structured data. Because large-data systems are more complex, they usually have to work with many sources of data. Data's structural heterogeneity refers to different data that cannot be directly mapped into one another stored in the data model. Incoherent data cannot reflect and cannot understand each other, cannot remember the link among data sets, and are called heterogeneous semantic data. Thus, with the increased complexity of goods and consumer demands, businesses collect more and processing details.

Many businesses have to juggle a vast amount of complicated (big) data through several information networks as they seek to leverage and incorporate big data. Due to these factors, most organizations find that today's IT framework cannot fit within their existing structures. To increase uncertainty and redundancy in the data means that products have no universal standards for comparisons, which may negatively affect the ability to convey details. Therefore, additional costs can arise for product information presentation, search, and maintenance. Furthermore, there may be more false orders, mis deliveries, and requests from customers. Therefore, it is essential to implement both syntactic and semantical restrictions. It avoids these confusing or misleading meanings of broad data sets [2].

The semantic Web is an extended form of the current Web that makes searching, re-use, combine and share data more accessible. Semantic Web is therefore then viewed as an integration device across different content, data applications, and frameworks. Big data today usually consists of 3 V: volume, variety, and speed. Various data are discussed for different data formats and many different data sources. The challenges of the sort of extensive data are thus crucial in solving many real-world problems. As an integrator, semantic Web incorporates data from different sources, such as web services, relational databases, spreadsheets, and other formats.

2. Semantic Web

The semantic Web refers to make web pages understandable. The reason for this additional importance is that machines can use these things. Ontology, by contrast, is the key driver of the semantic Web and defines how data can be integrated between heterogeneous sources of information and how the links between them can also be explained. This standard describes the name, definitions, relationships, and features of various entities within a specific field. Through creating common standards for data representation, the semantic Web encourages inclusion. By enabling semantic content to be considered in internet-open data, Semi-structured and unstructured data must become the standard on the Network Extensions (a protocol and query language for semantic web data). Though massive amounts of extensive data are still at present a problem, NoSQL databases have found a solution in the Resource Description Framework (RDF), resource to do not from sources of limitless data[3].

Instead of a technique or tool, NoSQL databases are a more critical way of philosophy. It shows a set of methods that can be experts in big data management. Compared to traditional relational databases, The extensible markup database approach, storage using NoSQL web ontology (OWL), is considerably different (XML). Our goal is to improve the site by connecting data and enriching the linkage of data so that users can grasp it. It promotes new-level service possible on the web with this. Using RDF as the data model for the sharing and representation of new data, the Semantic Web group made a choice vital for realizing the full potential of the technology, which is aimed at

addressing its difficulties and at achieving its goals. Despite the reputation and maturity of the relational models, XML was chosen for RDF support both for research and business systems. Several core formats were developed to help model, characterize and query data in a more prosperous way[4].

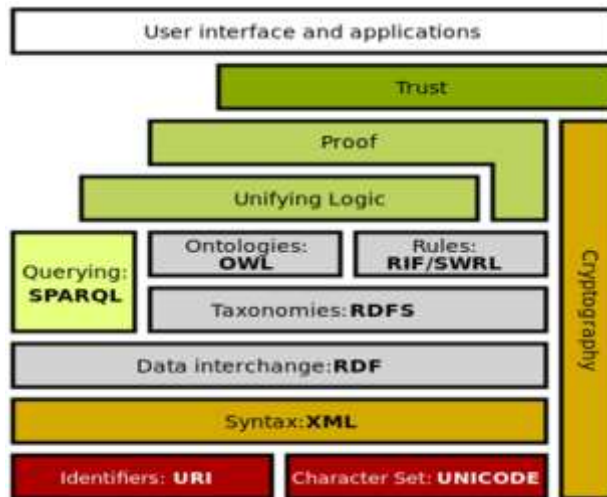


Figure 1: Semantic web.

2.1. Knowledge representation

A topological view of the semantic web is not practical for desktop use. RDF (resource description framework) provides new tools to construct semantics and entice and application authors. RDF is not fit for the desktop. It is common to use graph models to store data related to subjects, the relations between objects, and data in which the subjects and objects are involved. As such, RDF is the universal representation language for semantic details. When storing information in a diagram model, you should use a connection between the topic and the subject. In contrast, the object and subject get along by themselves and while storing for each other alone.

The Relational Database to RDF Mapping Language, or R2RML for short, is a language that streamlines the process of translating data from a relational database model to RDF format. It is most often referred to as R2RML. Consequently, it is possible to create a mapping between relational data sources and RDF that is not only flexible but also expandable. This is because of the fact that this is possible. This language makes it possible to convert massive datasets into RDF format in a short period of time and with a high degree of efficiency. This is made possible via the use of this language. It is necessary to offer meaning and interpretation in order to keep data in a linked form, which is made possible by conceptual RDF [5]. This is a need. Data storage requires this as a requirement in order to proceed. When it comes to accomplishing this purpose to the fullest extent that is practically possible, the development of ontologies is an essential must. defines a relative vocabulary as a collection of ideas and the meanings of those concepts that are generated through the use of an ontology that is fully explained. This is the definition of a relative vocabulary.

Through the use of this strategy, it is possible to create reusable terms that are capable of being included into a broad range of terminological systems. It is possible to make use of these words in these systems. To take use of the expressive features that this language provides, it is usual practice to create ontologies using the Web Ontology Language (OWL). This is done in order to make good use of the language. Additionally, there is the possibility that they may be expressed in RDF ontology (S), which is an additional factor to take into account. When it comes to the breadth and longevity of its language, OWL is better than RDF, despite the fact that both OWL and RDF are

vocabularies that are equal to one another. In terms of its vocabulary, RDF is equivalent to Open Web Language (OWL). Each and every one of the things that are described by RDF(S) consists of classes, properties, as well as the scope and range of topics and objects that are connected by a property. An abbreviation for the Resource Description Framework is denoted by the letter RDF(S). In contrast, it is possible to make use of OWL in order to construct semantic linkages. This is something that can be done. With the help of either the domain or the range of this property, it is possible to restrict the classes that are allowed to be connected together. The use of any of the two qualities is enough for accomplishing this goal. The descriptive capabilities of RDF(S) have been expanded as a result of the formalization of the most recent version of OWL, which is known as OWL2. In addition, OWL has just undergone a modification that has resulted in expanded capabilities for the provision of explanations that are logically descriptive . It is possible for classes and object attributes to have a wide variety of defined characteristics, and these characteristics may be expressed in a variety of different ways. The qualified cardinality limitations and the limits in the data range are examples of the kinds of characteristics that might be taken into consideration for inclusion in this category. It is possible for ontologies to make use of other languages, such as SWRL (Semantic Web Rule Language), for the purpose of rule representation [[6]. This is accomplished via the use of rules. Furthermore, this is in addition to the capacity to articulate reasoning. The formation of links between different sorts is a result of these features, which are the element responsible for the establishment of connectivity.

2.2.Semantic Annotation

RDFa is a notion that was first conceived by the World Wide Web Consortium (W3C), which is the entity that is responsible for its own creation. The purpose of this action was to provide HTML, XHTML, and certain XML document types with extra extensions. It is possible for web pages to include certain bits of information. This is made possible with the assistance of these extensions. There are two other lightweight forms that are utilized and exploited within the area of semantic metadata. These formats include RDFa and one other lightweight format. The formats in question are employed and exploited. In this particular setting, the formats that are being taken into consideration are RDFa and JSON-LD, which is an abbreviation that stands for JavaScript Object Notation for Linked Data. The definition of HTML characteristics that are included in HTML pages has been included into the HTML extension standard, which has been changed to include this new feature as an additional characteristic. This new characteristic is an additional characteristic. In the process of doing this, the goal is to integrate data from machines that may be used in subsequent operations. This is done with the objective of integrating the data. The fact that this method is even more lightweight and less expressive than RDFa or JSON-LD is something that must be kept in mind at all times. Regarding this particular matter, it is important to take it into mind. It is essential to have this specific piece of information in mind with regard to this problem. Consumers are able to acquire search results that are more accurate and complete than those they would obtain if they did not utilize search engines . This is because consumers who use search engines like Google to seek answers to their discovery questions are able to acquire search results that are more accurate and thorough. The HTML websites that are able to include semantic metadata into their discovery information are the ones that are able to make this a reality [7].

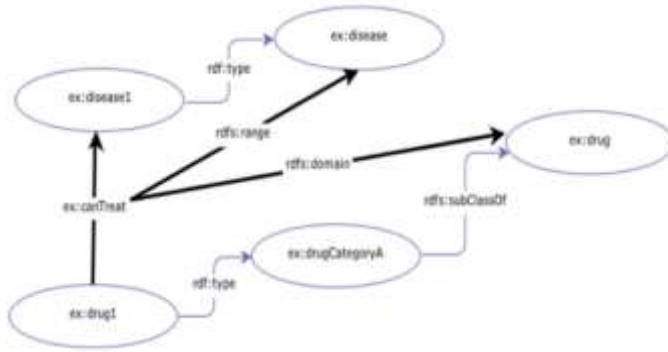


Figure 2: Semantic annotation

2.3.Semantic Queries For Search

Simple Protocol and RDF Query Language, often known as SPARQL, is a query language that was designed specifically for RDF, which at the time was an acronym that stood for Resource Description Framework. Simple Protocol and RDF Query Language is the name of the query language. RDF was the inspiration for the creation of SPARQL. Instead of retrieving single data sets that are not connected to one another, it is possible to get linked data by using queries that are more intricate than the ones that are used to retrieve single data sets. This is in contrast to the retrieval of single data sets that are not related to one another. The Restricted Document Format (RDF) was the inspiration for the development of the query language known as SPARQL, which was designed expressly for that purpose. It is now able to conduct complex searches on connected datasets as a result of the use of SPARQL, which in turn makes it feasible to recognize relationships that were not before known. There is a situation in which nobody loses out. When domain-specific ontologies are used, it is possible to convert the data that were obtained during the discovery phase into a format that is compatible with linked data. This is made possible via the utilization of domain-specific ontologies. This action is taken in order to achieve the objective that was stated earlier about. Efforts are now being made in order to develop a greater number of semantic knowledge bases that are able to be searched using SPARQL. These efforts are currently being performed. Furthermore, in order to get the desired results, these efforts involve combining SPARQL with other technologies, such as text mining and similarity tagging, in order to enhance semantic searches. This is done in order to guarantee the desired outcomes. This action is taken in order to accomplish the goals that have been set [8].

2.4.Rules, Reasoning & Inferencing

It is possible to make advantage of the inherent logical description capacity of ontology, or one of the rules-based languages that have been established in the semantic web, in order to increase the degree of complexity of the relations that are supplied by ontology. This is done for the aim of enhancing the amount of complexity presented by ontology. You have the choice between any of these two alternatives, which will increase the level of intricacy. You have the flexibility to choose any of these two options depending on your preferences. Both of these choices are up for consideration, and they are both accessible for selection. The web toolkit is able to build rules, reason with them, and make conclusions, in addition to the skills it already has. For the web toolkit, this is an extra component that is included. This is yet another component that is a part of the web toolkit that is provided all together[9]. The administration of complicated rule-based constraints inside the semantic web-based toolbox has been made easier by the development of a number of rule languages . These rule languages have been established in order to allow the simplification of

mathematical operations. The management of these limits was made easier as a result of this action being taken. In order to make the process of simplifying mathematical operations simpler to carry out, this was one of the most important things! Semantic SWRL is a system that improves the axioms of Open Web Language (OWL) by combining Horn-like notions with the Open Web Language (OWL) versions of OWL DL and OWL Lite, as well as the ML version of the Markup Language with the Unary/Binary Rule. This enhancement is accomplished by combining the aforementioned principles. Through the use of the Markup Language in its ML form, this objective is successfully realized. The Unary/Binary Rule is also included into this system, which is an additional characteristic. In order to make it feasible to carry out this activity, the OWL versions have been modified to contain rules that are comparable to those found in Horn. The construction of SHACL, which replaced SPIN, was accomplished via the process of creating inferencing rules in RDF syntax. In addition, SPIN was successfully manufactured using this method. For the purpose of providing a genuine picture of the advanced SPARQL restrictions that were in existence at the time, this was done according to [10].

2.5.Increasing Use Of Semantic Web Technologies

It is of the utmost requirement to develop these technologies and acknowledge the significance of semantic Web technologies in order to encourage breakthroughs in a wide variety of businesses. This is a necessity of the highest order. This action is taken in order to guarantee that these technological breakthroughs can be accomplished via the use of these specific technologies. During the same period of time when there has been a growth in the utilization of semantic technologies, the market is seeing the development of new technologies that are oriented in a way that is semantic[11] . An example of a knowledge graph is shown here for the purpose of giving an illustration. The following is an example of a knowledge graph. In spite of this, it took some time for this concept to accomplish the goal of achieving universal acceptance, just as it took some time for central technology to reach a point where it was available to the whole public. The robustness of these designs has been improved, and linked entity charts have been used in order to provide a description of these designs. The goal of this endeavour is to accomplish the design description. Using approaches that are based on the Semantic Web platform, it is possible to develop one-of-a-kind apps that make use of network connection techniques for the goal of information analysis. This may be performed via the implementation of methods. This is one of the many options that are open to you. This particular scenario is one of the numerous possibilities that may occur. However, in order for these approaches to be used in an efficient manner, it is of the utmost importance that the ontologies and the large datasets be in a format that can be accessed by a computer [12]. This is an essential must. This is a requirement that has to be provided for. For the purpose of making effective use of these technologies, it is necessary to fulfil this criteria, which is an important prerequisite. A further point to consider is that in order for the algorithms that are utilized for machine learning to be utilized for this purpose, it will first be necessary for them to go through an upgrade .

3. Big Data

Big data may be described as data that conventional methods cannot handle. Huge volumes, different speeds, and different varieties cause numerous other management and process challenges and problems. Big data is growing exponentially, and new technologies are therefore important to develop. The leading big data technologies are Hadoop, Map Reduce, and No SQL. These technologies are beneficial in the management of big data.

It is the most significant difficulty. Extensive data exceed structured data. Because of its amorphous form, it is tough to manipulate big data. Big data, volume, speed, and variety are three fundamental properties[13].

- Volume

The enormous amount is the essential property of big data. Today, big data are growing exponentially. The data is all over and is produced very quickly. Social media and server logs generate enormous amounts of data every day[14].

- Velocity

Speed is another essential big data property. That is how quickly data are generated. Now, because of social media and the digital world, big data is generated quickly. Thousands of GB of data are produced daily across digital world.

- Variety

Regarding big data, variety is the prime directive. If big data is to succeed, it must deal with the complexities of identity. Buckets and Bocks Semi-structured data is also provided. The information which has not yet been designed for relational databases is difficult to deal with[15].

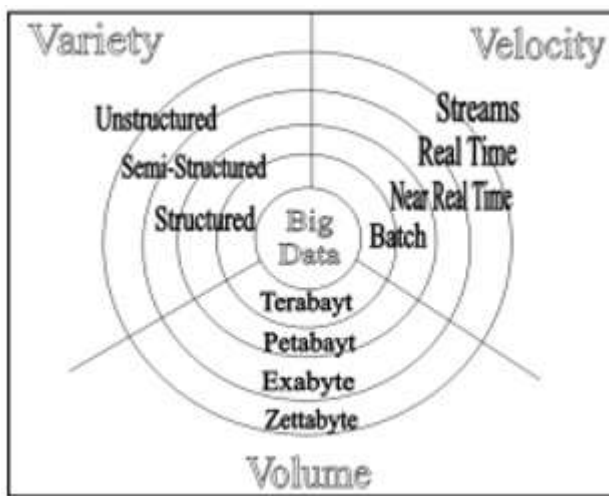


Figure 3: Big data

There is no data structure itself as a significant challenge. It is tough to deal with big data in traditional technologies and databases because it is entirely unstructured. Traditional databases for structured data are mainly developed.

3.1 Challenges of integrating big data

Big data is a concept that encompasses large and unstructured data sets with scarce use or utilization applications. it's a time-consuming task to synthesize this impressive index and, for example, this mixing, growth, knowledge, quest, data acquisition, information extraction, storage, and representation, and representation may be complex[16].

- a. Data Management

One of the most critical tasks to integrate big data is data management. When the data is huge, it becomes difficult to manage this data. NoSQL databases can be used to avoid these problems. Contemporary database design offers a wide variety of functionality in terms of data access and manipulation. So that a large amount of data can be stored, these databases use the key-value pair concept[17].

- b. Bad data

Integration of data is the most critical challenge of any database project. Heritage data must be purified before the changes and membership, or organizations will most likely face goodness data issues honestly later. The pollution of legacy information has an increased impact; in nature, it focuses on the large volume of information customers. The choices will also be made in this regard if this information degenerates. During cleaning data used for use within the integrated framework, issues of unknown information quality cannot be developed irregularly [18].

c. Lack of Skills

New data management systems, such as NOSQL databases and data warehouse architectures, appear in the marketplace regularly, and consumers switch from relational databases to more immediate plans. Innovations of big data are absent because of the scarcity of people with expertise [19].

d. Transmission of Data into Big Data Structure

Few citizens have called for various clients to report on separate data platforms for varying reasons. Moreover, they cannot consider the adjustable information consistency, access, and mobility of data instead of storing it comprehensively. Both the fun of trial and error was associated with data processing and replication.

e. Extracting Information

In the most convenient use data, data is available, existing data accumulation is expanded, and end-customer access is facilitated through corporate knowledge instruments to the actual purpose of data disclosure. It turns into a test in a vast coordinated approach to ensuring the right time for data customers [20].

f. Synchronization Data Sources

When large amounts of knowledge from several places will enter a single information stage quickly and in parallel, we can learn about the problem of replication. This gathers the information from a single location if it no longer refers to another. It also helps with meanings, analogies, thoughts, and metadata workflows. Typically, the task of data management, data retrieval, and synchronization, and data dissemination all lead to a data state of flux[21].

g. Other challenge

Other challenges that might arise during data integration include the vast amount of data, cost of setting, reliability, correct information, and data change rate. It is only a test to practice the enormous amount of data with the practical speed with the critical purpose to provide the necessary customers with the data at the moment of their request. Additionally, data collection support is obtained when the trading data is supported by one source or by buyers as well as[22].

4. Related work

It was hypothesized by Banane and Belangour [23] that the NoSQL Big Data technology known as MongoDB might be used as a viable technique for the construction of a database system that could be deployed in the year 2020. One of the most significant benefits is that MongoDB was developed with the express purpose of storing RDF data in the JSON format. This served as the fundamental impetus for the creation of it. RDF, which is an acronym for Resource Description Framework or Resource Data Framework, is stored in the NoSQL database in the form of JSON, which is an abbreviation for JavaScript Object Notation. JSON is an acronym for the Resource Data Framework. It is possible to store RDF in the form of JSON.

It is in the year 2020 when Saravana Kumar and Santhosh [24] come up with an alternative method of addressing the examination of data. In order to construct this technique, the semantic structure of the data that is discovered on the internet acts as the basis. In addition to the fact that it is equipped with features for the extraction and selection of data for the purpose of data processing

and retrieval, it also contains standard properties that are essential for the successful mapping of text. Utilizing this strategy leads to a decrease in the dimensionality of the vector, which in turn leads to an increase in the efficiency with which information can be accessed with more ease. The results of the simulation show that the recommended reduction in dimensionality technique is able to obtain a high degree of accuracy when the T-Order dimension is removed from the Bayes and Decision Trees categories. This is shown by the fact that the T-Order dimension is removed. This is something that is shown by the outcomes of the simulation that was carried out. When contrasted with the conventional methods of selection that are used for classification algorithms, this strategy offers a substantial distinct benefit.

By demonstrating in [25] how the RDF/XML ontology framework may be used to represent a broad variety of knowledge operators, Patel et al. published their findings in 2019. Significant repercussions were discovered over the course of the case study that used the field ontology for the purpose of analyzing emergency events. The system of communication is improved, and broad adoption of the World Wide Web is ensured, which is an essential component for all applications. In the case that information is shown as a single unit, the system of communication is increased. One of the most significant benefits is the capability to efficiently handle a wide variety of semantic difficulties that varies in terms of the degree of complexity with which they are presented. Some of the problems that need to be taken into consideration are insufficient and confusing information, massive data integration, duplication, ambiguity, and imprecise word usage. In addition, the process of mapping and acquiring new information are two additional problems that should be taken into consideration.

An extensive data environment was investigated as part of the research that Ostrowski and his colleagues (2016) carried out, which was referred to as [26]. This was one of the components of the investigation that they carried out. A significant portion of their effort was focused on ensuring that the quality of our data sources was sufficiently handled, in addition to building a data strategy and structure. This was the primary focus of their work. By using this framework, which makes use of a common ontology field, it is possible to do semi-automated mapping of a broad variety of data sources via the usage of this framework. Within the framework of this method, the many advantages that may be acquired via the use of extra data sources are shown. Supply Chain Risk Detection was shown to be an efficient approach via the use of a test as well as a case scenario. Manual labour has historically been the method of choice for carrying out the process of ontological federation in the past. The combination of semantics with annotations, on the other hand, has the potential to provide a major improvement in the efficiency of ontology federation.

In accordance with the assertion that was stated by Barba-González et al. (2019) in reference [27], the BIGOWL solution makes it possible to carry out studies on huge datasets. At the present, a model for an RDF database setup will include an RDF/programmatic semantic query engine as a component of its overall structure. The first experiment was focused on optimizing traffic in real-world circumstances, while the second experiment was concerned with analyzing public data that was kept in the cloud. Due to the fact that both of these studies were carried out in the field, they offered proof that the fundamental concept was correctly conceived.

In the paper that they published in 2019, Das et al. [28] presented a complete method that incorporates the usage of a cloud search engine. This technique was included in the publication. This particular search engine makes use of semantic technology in order to get information that is relevant to the user's search. This information is obtained on the basis of the information that is provided by the user. Consumers are able to get exact data and results that are related with their search when they make use of a Semantic Search Engine that is stored in the cloud. This eliminates the need for them to go through a number of different websites. The search may be carried out with the assistance of modern big data technologies, which make the process of discovering and acquiring data far less difficult than it was in the past. This makes it feasible to carry out the search

process.

For the purpose of achieving asymmetrical and transverse equivalence closure, Mountainantonakis and Tzitzikas[29], (2018) provide a set of approaches and algorithms that are comparable to one another and that complement one another. This action is taken in order to accomplish the objective of reaching a state of closure. Through the use of indexing, it is possible to guarantee that all data sets are accessible to all individuals, and connection is utilized to ascertain the extent of the organization's potential to contact people all over the globe. A huge number of data sets are produced in a methodical manner on a consistent basis in order to get a complete global view. This is done in order to achieve the best possible results. In conclusion, various data sets are linked to one another in order to improve the overall knowledge of the database and the information that it includes. This is done in order to increase the overall knowledge of the database. Approximately two billion consecutive comparisons are used in order to evaluate whether or not an accelerated method offers any additional benefits. This is done in order to assess whether or not an expedited method offers any additional benefits. In order to ascertain whether or if a faster procedure confers any additional advantages, this is carried out.

Through the use of Linked Semantic Web Data in their professional endeavours, Hyvonen [30]conducted an investigation of the efforts that Finland has made to establish national ontological and semantic cultural heritage portals. The presence of these portals is an illustration of the work that Finland has conducted. When it comes to publishing content for the Semantic Web and Cultural Heritage, the "Sampo model" is a method that is used with the intention of achieving the aim of performing this function. There are three fundamental elements that make up this system: 1) a "business model" that is designed to integrate and govern the process of content gathering and distribution via the use of a common ontology; 2) a common ontology; and 3) a common lexicon. 2) In order to put this strategy into action, you will need to design an interface that allows data to be reused and accessible by other applications in a manner that is distinct from how it is accessed by other applications. Only a single SPARQL endpoint is used to store the information, and this endpoint is the only one of its kind. Three) In order to receive and analyze knowledge, a paradigm that consists of two stages is used. One of the initial steps in the paradigm is the use of a filter, which is done in order to limit the range of things that are of interest. Even though the somaticized faceted search is the major filter, the tools that are available on the Digital Humanities website are meant to either be visualized or studied. This is the case despite the fact that the search is the key component.

An architecture that was developed by Michel et al. in 2018 [31]aims to combine structured tools with those that are not represented as RDF. This is the purpose of the architecture. The use of the concept of connected data is what allows this to be done. The 'Micro-SPARQL' Endpoint will be included into an application model that is built on microservices, and the microservices architecture will be used in order to build a connection with the Web APIs. In conclusion, this will be the case. SPARQL is expected to exhibit a wide variety of practical uses of SPARQL Micro-based services within the realm of biodiversity. This is something that is projected to happen. Constructing a core platform that would enable independent suppliers to connect data from a variety of resources that are both scalable and trustworthy might be accomplished by the establishment of a SPARQL service. This would be a significant step towards achieving the goal of constructing a core platform.

In the research that Jabbar et al. [32]conducted and published in 2018, they proposed the use of a combination of three distinct data models all on their own. Relationship-based, semantic, and big-data systems are the types of data and information that are included into these models. These two elements were brought together with the purpose of resolving issues and improving functioning in order to achieve the desired results. An empirical investigation is used to demonstrate the construction of a realistic implementation of merging data from a relational database (RDB) into a

resource description framework (RDF). This is shown via the examination of the data.

According to the documentation that can be found in [33], Rai et al., 2021 presented a method for the synthesis of semantic web services that makes use of keywords. You can find this information in the paperwork mentioned above. Throughout the course of the development process, the use of safeguarded methodologies is utilized in order to complete the creation of the eBook Search ontology, which is founded on OWL-DL itself. In this specific case, the ontology of the Search for eBooks was built by making use of a wide variety of criteria, and SQWRL and SWRL were used in order to accomplish this objective. A consistency check of the ontology was carried out with the assistance of a pellet reasoner. A check of this kind was performed. When it is used, ontology is significant because it makes it possible to create content for a specific digital library by making use of certain keywords. This is the reason why ontology is so important.

The semantic knowledge data quality architecture assessment that is supplied by Liu et al., 2020 [34] enables users to grasp the semantic value of image big data in connection to requests and the origin of ideas. This is made possible by the fact that users are able to understand the information. The effective completion of this assessment was made possible by the availability of a dataset containing semantic information. It is conceivable for our Integrated Data Science and Technology Hub (IDSTH) to effectively solve the problem of sample analysis in order to ensure that synanthropic elimination operations are carried out in an accurate way. This is something that is achievable. When developing the SHR that we have developed, we have taken into consideration not only the amount of data that is relevant but also the degree of semantic similarity that exists between the two entities. As a result of this, we are now in a position to draw judgements on the evaluation from a completely different point of view. Furthermore, in order to ensure that the assessment process is carried out in real time, our groundbreaking framework integrates a wide range of educational and therapeutic technologies. This is done in order to guarantee that the evaluation method is carried out. Among the technologies that fall under this category are offline data processing and online assessment capabilities.

Within the scope of their study that was published in [35], Dividino and colleagues presented a framework for the integration of data. This framework was built with the goal of managing the many data streams that are produced by the several types of Internet of Things infrastructure, namely maritime and marine sensors. This framework was developed with the intention of handling that data. The system enables the processing of data as well as the querying of data in real time. Both of these activities are completed simultaneously. For the purpose of enhancing the knowledge that is already accessible in the maritime domain, the current technique makes use of the ideas and technology that are linked with the Semantic Web. Additionally, a significant amount of attention is devoted to streaming models, in addition to the engine that is responsible for the processing of data streams on a massive scale. This attention is paid in addition to the engine. There are three different RMSA prediction scenarios that are detailed in this article in order to illustrate that the technique is applicable in real-world situations. (1) It is possible to accomplish the detection of vessels that are approaching regions with thick ice, (2) the detection of vessels that are approaching places with severe weather conditions, and (3) the detection of vessels that are approaching regions with speed limitations that are not being followed. All of these detections can be accomplished. Each and every one of these detections is in fact attainable.

Everyone involved was Mouad and his coworkers. More specifically, the authors of reference [36] recommend reformatting a SPARQL query in their paper, which is scheduled to be released in the year 2020. This is mentioned in the context of the article. A model-driven approach is being used in the project with the purpose of achieving these objectives in order to ensure the greatest possible level of success for the project. Pig, Hive, and Spark are the three approaches that have had metamodels constructed for them. A mapping has been established for the first query language for Big Data, and that mapping has been established. It is true that both of these changes have taken

place. Through the use of the Atlas Language, which is a dataset that estimates a large number of data sets, we were able to evaluate the efficacy of our newly developed method by performing RDF validation on a substantial amount of RDF information. With this information, we were able to establish whether or not our approach was successful. For the purpose of providing reliable findings, the verification was carried out on a server that can be depended upon.

In the paper that they produced in the year 2020, Qiu and his colleagues [37], presented the SGCCE concept extraction strategy as a potential method for accomplishing their goals. This method makes use of semantic information in order to extract concepts from large amounts of data. Its goal is to do this. Additionally, it provides a one-of-a-kind method of assessment that is known as SBE, which is its own unique approach. When doing an analysis of the idea extraction strategy, a semantic technique known as SBE is used. This technique is used to analyze the process. In light of the outcomes of the investigation into idea extraction, it has been discovered that SGCCE is far more effective than expected. Having finished a significant number of demonstrations, this conclusion was reached after the conclusion was reached.

A research study was provided by Ramay et al. (2021), in which the authors offered data models for the purpose of comparing and mapping NoSQL, RDB, and Semantic Web for the purpose of transformation [38]. This study was presented as part of a research study. Through the use of the Semantic Web, the purpose of this project is to improve the efficacy and efficiency of data gathering and storage systems for the purpose of performing annotations on massive amounts of data. Through the use of the Semantic Web, this objective will be fulfilled. It is the goal of this project to reduce the amount of time that it takes to answer to inquiries. This is in addition to ensuring that the data format is compatible with the web and that it has a bigger quantity of semantic information. The usage of a dataset that is composed of medicines and is then modified is done in order to offer aid in determining the location of big data. This is done in order to provide assistance. In order to graphically display the data that has been modified, the RDF graph and the bar graph are used immediately after the completion of the processing of the model that was recommended. The ability to get discoveries in a timely and accurate manner is made possible by big data, which is a very advantageous feature.

In 2018, Bellini and Nesi [39] conducted an evaluation that included both a study of the RDF-Shops that are now operational and an examination of the restrictions that are associated with the utilization of RDF shops for the provision of intelligent city services. Both of these aspects were included in the evaluation. For the purpose of providing a full understanding of the use of an RDF repository as the basis for Smart City models and applications, the evaluation model is offered here. Support for the requirements is provided by the RDF evaluation model, which also provides support for a set of criteria that enhances the standard that is presently in place for RDF storage. An examination of the similarities and differences across a number of well-known RDF companies, such as Virtuoso, GraphDB, Oracle, and StarDog, amongst others, was carried out at the same time. The documentation, which is constructed on the basis of the Florence Smart City models, includes the implementation of the proposed Smart City RDF benchmark, datasets, and user-friendly tools. All of these components are included in the documentation.

Taelman et al. [40] were able to improve the storage of RDF data (Resource Description Framework) by including the bidirectional delta chain into the Hybrid storage strategy (OSTRICH). This allowed them to achieve their goal. All of this made it possible for the data to be stored in a more effective manner. Based on the findings of the studies, it can be concluded that the adoption of this adjustment was successful in addressing the fundamental problem of scalability in the unidirectional delta chain, particularly with respect to the time of intakes. Throughout the course of the procedure, this was shown on a number of instances. The implementation of our ground-breaking strategy, which is referred to as COBRA, was the method that made this accomplishment possible. The successful completion of this task was made possible by the use of this strategy. As

a direct result of this, the speed was enhanced by forty-one percent. As a result of this alteration, the total amount of storage space for two of the three datasets has been reduced by thirteen percent in comparison to the amount that was initially available. With the exception of the BEAR-A dataset for VQ, the performance of all query types that involve versioning has been increased by 21%. This improvement is not applicable to the BEAR-A dataset. As a result, the performance of each and every query type is enhanced. The COBRA approach, which makes use of a bi-directional delta chain and has request times of one millisecond or less, is an excellent choice for RDF archives when it comes to apps that are used for online searches. This is due to the fact that the latency of the network might occur at periods that are extremely slow.

In the paper that Kraus et al. released in 2018, [41], they not only provide a detailed analysis, but they also issue a call to action to the scientific, clinical, and administrative stakeholders that are involved in tackling the challenges that are often associated with the utilization of big healthcare data. In order to accomplish the objective of establishing entire data integration, we present a realistic strategy to accomplishing this aim. The use of iterative processes, concurrent upgrades, and extensive data analytics are the means by which this objective is attained.

In the piece of writing that bears the title "Alkhamisi and Saleh, 2020" [42], the authors discuss the relevance of ontology and explain how it is connected to the human experience. Furthermore, it illustrates the practical uses of the customer's holistic viewpoint within the framework of semantic integration inside situations that occur in the real world. This is a significant contribution. This is the conclusion that I would want to reach, which is that ontology-based models have the potential to provide considerable benefits in the management of huge amounts of data across a broad variety of applications.

In the course of their inquiry, Mehmood et al. [43], (2019) suggested the utilisation of a data lake methodology as a way for the purpose of collecting all of the data for the goal of analysis. Using this approach, a wide range of data technologies are used in the capacities that are most appropriate for them. Collecting data, storing it, integrating it, analyzing it, and finally displaying the findings of the analysis once it has been carried out are all made easier by the platform. Within the context of the CUTLER project, this solution is the first attempt to incorporate a broad variety of data sources from four pilot locations.

As an extra point of interest, in the year 2019, Hyun Jeong Ban and 김학선 [44], suggested the notion of undertaking a semantic network analysis by gathering terms associated to hotel packages from Google. In order to acquire the data that was obtained between February 1, 2016, and January 31, 2019, SCTM, which is an acronym that stands for "Smart Crawling and Text Mining," was utilized. This was done in order to acquire the specific information. The software application known as Ucinet6 was utilized in order to graphically portray the data and attract attention to the semantic linkages that exist between the data. The analysis was the major concentrate of that specific programming. According to the conclusions of the research, the terms "travel," "book," "news," "room," and "years" were the ones that had the highest links with the hotel Paquet. The purpose of this research is to get a deeper comprehension of the interrelationships between words by examining the magnitude of the own vector and the significant role it plays in the process. It offers support to the hotel business in the process of producing price packages that are competitive within the framework of hotel packaging trends within the hotel industry.

According to the findings of their investigation, Maniou [45], (2021) said that there is a relationship between social networking and the dissemination of cultural material that is made possible by the media and news sources. An exhaustive investigation into all of the relevant historical and contextual information is carried out within the framework of a qualitative study. The purpose of this investigation is to ascertain the significance of the material for semantic research.

In order to overcome the problem of volume, Vieira et al., 2020 [46] developed a test to analyze

the data. This exercise was carried out. The results of this test provided a representation of the manner in which the majority of the data would be used into the conclusion. The basis for the data models that are built by organizations is comprised of the fundamental ideas and technologies that are involved with handling enormous amounts of data. To put it into perspective, the only way they are able to mimic supply chains is by using basic mathematical concepts and simulations. The problem of recognizing and overcoming these difficulties may be approached via the use of simulation, which is an option that might be considered.

Elzein et al. conducted a comprehensive analysis of the difficulties that are now being faced, the prospects for growth, and the ongoing enhancements that are being made in the cloud administration of huge RDF data in their paper that was published in 2018 [47]. In addition to this, they presented insightful insights and significant takeaways from research that were conducted on the handling of vast volumes of data. Through the completion of this course, students will be able to acquire a grasp of the basic concepts of RDF data manipulation, which will enable them to comprehend the most recent innovations and accomplishments in RDF graph creation. The study also includes a comparative analysis of the effectiveness of the different query processing algorithms and storage systems that are now in use. This is an extra aspect of the research. The purpose of this document is to provide an overview of the prospective possibilities and problems that now exist in the area of RDF in order to guide extensive research endeavours.

In their work that was published in 2017, Ullah and colleagues [48], proposed the concept of semantic interoperability across heterogeneous Internet of Things (IoT) devices in the health care industry for applications that include massive amounts of data. They referred to this idea as SIMB IoT. The Semantic Interoperability Model proposes a framework for suggesting appropriate medications for various ailments based on data collected from a variety of IoT sensors, while also taking into consideration the potential adverse effects of such a recommendation. This framework is based on both the data collected and the potential adverse effects of such a recommendation. Two datasets have been selected with the intention of carrying out an analysis consisting of large amounts of data. One of the databases has a plethora of information on diseases that are brought on by medicines, while the other dataset gives information on the bad effects that certain drugs could have. Both datasets are available online. Each and every piece of information that is sent back and forth between the patient and the physician is properly recorded and delivered in a manner that is both efficient and relevant. The topic of a proposal is a lightweight model that is capable of delivering annotations for vast amounts of data across a number of devices that are a part of the Internet of Things (IoT) and have varied characteristics. This model is intended to be used in conjunction with one another.

Jia and Li (2020) [49] presented an improved approach for assessing the intensity of emotions in the emotional lexicon. This method was available to the authors. The application of this strategy was made to a wide range of distinct types of feelings. It is recommended that you make use of word2vec in order to train the emotive word vectors that are present in the microblog entries from the very beginning. It is necessary to choose benchmark words from each of the categories of an emotional expression in order to ascertain the degree of similarity that exists between the expression vector and the word vector. Consequently, the emotional intensity of a wide range of various sorts of emotive phrases will be presented in this manner. after this, we analyse the impact that negative words and degree adverbs have on the emotional predisposition of different categories of inspiring words, and we make semantic suggestions for the purpose of computational analysis after this investigation. When all is said and done, we integrate the emotion dictionary with the semantic pattern rules in order to compute the emotional strength and categorization of the many emotional categories that can be found in the microblogs. This allows us to determine how the microblogs are expressed emotionally.

An overview of the different methodologies that are presently being utilized for the

construction of triplestores was published by Iancu and Georgescu [50], (2018) within the context of cloud settings. The authors placed a particular focus on triplestores that are used as Database-as-a-Service (DBaaS). A comparison and analysis of the different options that were discovered is carried out based on the findings. It offers useful ways for picking the online data storage solution that is the most fit for your requirements, regardless of whether it is on-premises or as a Database-as-a-Service (DBaaS).

Table 1: summary of reviewed papers about semantic web and big data

REF.	AUTHOR	YEAR	PROBLEM	SEMANTIC WEB STANDARDS	OBJECTIVE
[23]	Banane and Belangour,	2020	Big Data performs the storage and retrieval of big data but does not support RDF data.	RDF, Ontology, SPARQL	we make use of mobile learning as a reference point of comparison to show a new meaning representation of Big Data
[24]	Saravana Kumar and Santhosh	2020	The massive flow of Internet data poses problems inefficient knowledge recovery	Direct mapping	Centered on the semantic Web data system, a new approach is proposed. It integrates both function extraction and data mapping feature collection techniques with standard features for efficient text mapping.
[25]	Patel et al.	2019	The basis of information on the intelligent Web should be as general as possible, given the complex existence of data.	RDF XML	Representing information as a unit seeks to reduce classical ontological problems to allow AI applications to conduct human-like actions
[26]	Ostrowski et al	2016	Big Data Collection Extract meaning	SPARQL	to use the Semantic Web opportunities to incorporate and scale large data technology projects
[27]	Barba-González et al	2019	Integrating the information field with a particular analytical challenge	BIGOWL RFD	The method known as BIGOWL proposes a philosophical basis for Big Data Analytics.
[28]	Das et al	2019	Searching the precise details is very difficult to navigate through the web page. This means choices are made based on organizations' imperfections and lack of knowledge.	RDF	Implements a cloud search engine in a semantic technology database that extracts the relevant user data
[29]	Mountantonakis and Tzitzikas	2018	Due to the vast number of LOD datasets, it is not simple to locate such links.	URIs	methods and algorithms for the semantic resolution of cross and hierarchical data (multidimensional) indices
[30]	Hyvonen	2020	The semantic challenge in the publication and use of data on the Web on cultural Heritage is anticipating the Cultural Heritage's heterogeneous content (CH).	SPARQL	Establish a national ontology heritage system and semantic portals focused on linked web data
[31]	Michel et al	2018	the difficulty of handling vast amounts of high-speed data	SPARQL URIs	It offers access to a limited resources-focused virtual graph—and uses micro-services design concepts to describe architecture to query the web APIs by using SPARQL
[31]	Jabbar et al.	2018	Raw information provided by IoT formation Big Data can generate extremely unstructured and heterogeneous data forms	RDF SPARQL XQuery SQL	Propose a combination of three main data structures, including relational, semantic, and big data and metadata involving problems and improved capabilities
[32]	Rai et al.	2021	E-book Ontological quest for the chapters and their required paragraphs with the Rules	OWL, SWRL	Get keyword-based semantic web service composition

[33]	Liu et al.	2020	their lack of evaluation of DQ at the relevance dimension, in particular for unstructured image data focused on semantic similarity	IDSTH	offer a semantic awareness data quality appraisal architecture that will help to help users to understand the semantic importance of picture significant data source and requirements
[34]	Dividino et al.	2018	Info drift is an ongoing problem in the current military world, creating scenarios in which data are either underused or not used for decision-making in more severe cases.	RDF, Ontology, SPARQL	Data aggregation architecture for heterogeneous data sources of marine and marine sensors of the various types of IoT networks and real-time interrogation and data management
[35]	Mouad et. al.,	2020	Hive, Pig, and Spark allow for the querying and interpreting of data without being suited for semantic data.	SPARQL RDF	Turn the SPARQL question into the Hive, Pig, or Spark software if the consumer chooses.
[36]	Qiu et al.,	2020	Find efficient insight and procedure information from large data volumes.	Direct mapping, SGCCE	Extract more valuable big data insights and expertise gathered in smart cities.
[37]	Ramay et al.,	2021	When a transition occurs, data and metadata must be used or sent together for mapping with no information loss.	NoSQL, RDB, RDF	Comparing and mapping data models between NoSQL, RDB, and Semantic Web for transformation aims to minimize query response time and ensure online and semantically enhanced information formats consistency.
[38]	Bellini and Nesi	2017	RDF stores can be used with knowledge base integration for online and smartphone apps heterogeneous information	RDF	The restrictions on RDF shops usage in smart cities were assessed in conjunction with existing RDF stores.
[39]	Taelman et al.	2020	must be able to store and query those emerging RDF datasets through your releases	RDF	Improved recovery of RDF archives by using a bi-directional delta chains under the hybrid storage technique (OSTRICH).
[40]	Ullah et al	2017	IoT devices are extremely heterogeneous regarding basic communication protocols, data formats, and technologies	RDF SPARQL	annotated patients' data and made it semantically interoperable. SPARQL query is used to extract records from RDF graph
[41]	Iancu and Georgescu	2018	storing large volumes of data in triplestore databases, which are currently the matter of choice for storing semantic web data	RDF, RDFS, SPARQL, OWL, GRDDL, RDFa, RDB2RDF, R2RML, Direct Mapping	choosing the most appropriate database solution for semantic web data representation, both on premises or as DBaaS

5. Discussion

The above cases reflect the latest semi-organized, unstructured, and hierarchical ontological study to integrate large-scale data sources. To boost efficiency, different sources and data methods are used. And more works have used minimum and fundamental semantic ontology and have not been entirely contrasted with big data. In addition, it is an early phase in the growth and still a more excellent job for science, particularly in the fields of health, business, and finance, to incorporate Big Data into ontology, semantical web technologies to resolve the diversity of challenges. Both Big Data and Semantic Web Technologies are linked to real challenges. Two separate tools of research. In addition, knowledge is the focus of all technology that can be further implemented or more comprehensive. The emphasis is also on It can thus be said that both regions must continue to make more progress and that huge, fast, unstructured, and unsafe data around us should be given priority.

6. Conclusion

Nowadays, big data is very complex and comes from many sources, including hierarchical, semi-structured, and unstructured sources. Many companies want to combine this variety of data and execute essential organizational and predictive functions on this data. Still, it might be necessary to note that this integration is challenged in this phase. The complexities of big data don't stop with the sheer volume of data, but a single point of failure doesn't handle the variety of formats and where it is stored, and the power over distribution. Properties of large data applications, such as workload and length, affect the overall efficiency of big data architectures. E.g., In batch processing, Hadoop is excellent but works well with vast datasets but not with interactives. An ontology helps integrate heterogeneous and schema-less files. Established knowledge graphs will display social data as RDF triples. A SPARQL query will answer an RDF graph's data needs. The way to extract value from significant data innovations is to apply them to parallelized queries. Large scale RDF management is reliant on distributed RDF triple technology.. This study paper highlights some difficulties and addresses the role of the semantic web and big data in solving these problems by semantic data integration.

References

- [1] Sadeeq, M. M., Abdulkareem, N. M., Zeebaree, S. R., Ahmed, D. M., Sami, A. S., & Zebari, R. R. (2021). IoT and Cloud computing issues, challenges and opportunities: A review. *Qubahan Academic Journal*, 1(2), 1-7.
- [2] Zeebaree, S. R., Shukur, H. M., Haji, L. M., Zebari, R. R., Jacksi, K., & Abas, S. M. (2020). Characteristics and analysis of Hadoop distributed systems. *Technology Reports of Kansai University*, 62(4), 1555-1564.
- [3] Abdullah, P. Y., Zeebaree, S., Jacksi, K., & Zeabri, R. R. (2020). An HRM system for small and medium enterprises (SMEs) based on cloud computing technology. *International Journal of Research-GRANTHAALAYAH*, 8(8), 56-64.
- [4] Saeed, J., & Zeebaree, S. (2021). Skin lesion classification based on deep convolutional neural networks architectures. *Journal of Applied Science and Technology Trends*, 2(01), 41-51.
- [5] Zeebaree, S., Zebari, R. R., Jacksi, K., & Hasan, D. A. (2019). Security approaches for integrated enterprise systems performance: A Review. *International Journal of Science and Technology Research*, 8(12), 2485-2489.
- [6] Abdullah, P. Y., Zeebaree, S., Shukur, H. M., & Jacksi, K. (2020). HRM system using cloud computing for Small and Medium Enterprises (SMEs). *Technology Reports of Kansai University*, 62(04), 04.
- [7] Salim, N. O., Zeebaree, S. R., Sadeeq, M. A., Radie, A., Shukur, H. M., & Rashid, Z. N. (2021). Study for food recognition system using deep learning. In *Journal of Physics: Conference Series*, 1963(1). IOP Publishing, p. 012014.
- [8] Jacksi, K., Ibrahim, R. K., Zeebaree, S. R., Zebari, R. R., & Sadeeq, M. A. (2020). Clustering documents based on semantic similarity using HAC and K-mean algorithms. In *2020 International Conference on Advanced Science and Engineering (ICOASE)*. IEEE, pp. 205-210.
- [9] Malallah, H. S., Qashi, R., Abdulrahman, L. M., Omer, M. A., & Yazdeen, A. A. (2023). Performance Analysis of Enterprise Cloud Computing: A Review. *Journal of Applied Science and Technology Trends*, 4(01), 01-12.
- [10] Zebari, R. R., Zeebaree, S. R., Rashid, Z. N., Shukur, H. M., Alkhayyat, A., & Sadeeq, M. A. (2021). A Review on Automation Artificial Neural Networks based on Evolutionary Algorithms. In *2021 14th International Conference on Developments in eSystems Engineering (DeSE)*. IEEE, pp. 235-240.
- [11] Yazdeen, A. A., Qashi, R., Malallah, H. S., Abdulrahman, L. M., & Omer, M. A. (2023). Internet of Things Impact on Web Technology and Enterprise Systems. *Journal of Applied Science and Technology Trends*, 4(01), 19-33.
- [12] Mostafa, S. A., et al. (2019). Applying Trajectory Tracking and Positioning Techniques for Real-time Autonomous Flight Performance Assessment of UAV Systems. *Journal of Southwest Jiaotong University*, 54(3).
- [13] Hammed, Z. S., Ameen, S. Y., & Zeebaree, S. R. (2023). Investigation of 5G wireless communication with dust and sand storms. *Journal of Communications*, 18(1).
- [14] Abdulrahman, L. M., Ahmed, S. H., Rashid, Z. N., Jghef, Y. S., Ghazi, T. M., & Jader, U. H. (2023). Web Phishing Detection Using Web Crawling, Cloud Infrastructure and Deep Learning Framework. *Journal of Applied Science and Technology Trends*, 4(01), 54-71.
- [15] Abdulrahman, L. M., Zeebaree, S. R., & Omar, N. (2022). State of Art Survey for Designing and

- Implementing Regional Tourism Web-based Systems. *Academic Journal of Nawroz University*, 11(3), 100-112.
- [16] Mohsin, S., Salim, B. W., Mohamedsaeed, A. K., Ibrahim, B. F., & Zeebaree, S. R. (2024). American Sign Language Recognition Based on Transfer Learning Algorithms. *International Journal of Intelligent Systems and Applications in Engineering*, 12(5s), 390-399.
 - [17] Ibrahim, R. K., et al. (2022). Clustering Document based on Semantic Similarity Using Graph Base Spectral Algorithm. In *2022 5th International Conference on Engineering Technology and its Applications (IICETA)*. IEEE, pp. 254-259.
 - [18] Zangana, H. M., & Zeebaree, S. R. (2024). Distributed Systems for Artificial Intelligence in Cloud Computing: A Review of AI-Powered Applications and Services. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, 5(1), 1-20.
 - [19] Omer, M. A., Yazdeen, A. A., Malallah, H. S., & Abdulrahman, L. M. (2022). A Survey on Cloud Security: Concepts, Types, Limitations, and Challenges. *Journal of Applied Science and Technology Trends*, 3(02), 47-57.
 - [20] Abdulkareem, N. M., & Zeebaree, S. R. (2022). Optimization of Load Balancing Algorithms to Deal with DDoS Attacks Using Whale Optimization Algorithm. *Journal of Duhok University*, 25(2), 65-85.
 - [21] Ibrahim, R. K., Zeebaree, S. R., Jacksi, K., Sadeeq, M. A., Shukur, H. M., & Alkhayyat, A. (2021). Clustering document based semantic similarity system using TFIDF and k-mean. In *2021 International Conference on Advanced Computer Applications (ACA)*. IEEE, pp. 28-33.
 - [22] Salim, B. W., & Zeebaree, S. R. (2023). Kurdish Sign Language Recognition Based on Transfer Learning. *International Journal of Intelligent Systems and Applications in Engineering*, 11(6s), 232-245.
 - [23] Banane, M., & Belangour, A. (2020). Towards a New Scalable Big Data System Semantic Web Applied on Mobile Learning. *International Journal of Interactive Mobile Technologies*, 14(1).
 - [24] Kumar, C. S., & Santhosh, R. (2020). Effective information retrieval and feature minimization technique for semantic web data. *Computers & Electrical Engineering*, 81, 106518.
 - [25] Patel, A., Jain, S., & Shandilya, S. K. (2018). Data of semantic web as a unit of knowledge. **Journal of Web Engineering*.
 - [26] Ostrowski, D., Rychtyckyj, N., MacNeille, P., & Kim, M. (2016). Integration of big data using semantic web technologies. In *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*. IEEE, pp. 382-385.
 - [27] Barba-González, C., García-Nieto, J., Roldán-García, M. M., Navas-Delgado, I., Nebro, A. J., & Aldana-Montes, J. F. (2019). BIGOWL: Knowledge-centered big data analytics. *Expert Systems with Applications*, 115, 543-556.
 - [28] Das, N. N., Chowdhary, M., Luthra, R., & Garg, S. (2019). Semantic Big Data Searching in Cloud Storage. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*. IEEE, pp. 351-355.
 - [29] Mountantonakis, M., & Tzitzikas, Y. (2018). High-performance methods for linked open data connectivity analytics. *Information*, 9(6), 134.
 - [30] Hyvönen, E. (2020). "Sampo" model and semantic portals for digital humanities on the Semantic Web. In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020)*. CEUR-WS.org.
 - [31] Michel, F., Faron Zucker, C., Gargominy, O., & Gandon, F. (2018). Integration of web APIs and linked data using SPARQL micro-services—Application to biodiversity use cases. *Information*, 9(12), 310.
 - [32] Jabbar, S., et al. (2018). A methodology of real-time data fusion for localized big data analytics. *IEEE Access*, 6, 24510-24520.
 - [33] Rai, S., Chaurasiya, A., & Bind, S. (n.d.). Semantic Web-based Service Composition for E-learning.
 - [34] Liu, Y., Wang, Y., Zhou, K., Yang, Y., & Liu, Y. (2020). Semantic-aware data quality assessment for image big data. *Future Generation Computer Systems*, 102, 53-65.
 - [35] Dividino, R., Soares, A., Matwin, S., Isenor, A. W., Webb, S., & Brousseau, M. (2018). Semantic integration of real-time heterogeneous data streams for ocean-related decision making. *Defence Research and Development Canada= Recherche et développement pour la...*
 - [36] Banane, M., & Belangour, A. (2020). A new system for massive RDF data management using Big Data query languages Pig, Hive, and Spark. *International Journal of Computing and Digital Systems*, 9(2), 259-270.
 - [37] Qiu, J., Chai, Y., Tian, Z., Du, X., & Guizani, M. (2019). Automatic concept extraction based on semantic graphs from big data in a smart city. *IEEE Transactions on Computational Social Systems*, 7(1), 225-233.
 - [38] Ramay, W. Y., Cheng-Yin, X., Rahman, S. u., & Habib, M. A. (2021). Hybrid approach for big data localization and semantic annotation. *Concurrency and Computation: Practice and Experience*, 33(4), e4955.

- [39] Bellini, P., & Nesi, P. (2018). Performance assessment of RDF graph databases for smart city services. *Journal of Visual Languages & Computing*, 45, 24-38.
- [40] Taelman, R., Mahieu, T., Vanbrabant, M., & Verborgh, R. (2022). Optimizing storage of RDF archives using bidirectional delta chains. *Semantic Web*, 13(4), 705-734.
- [41] Kraus, J. M., et al. (2018). Big data and precision medicine: Challenges and strategies with healthcare data. *International Journal of Data Science and Analytics*, 6, 241-249.
- [42] Alkhamisi, A. O., & Saleh, M. (2020). Ontology opportunities and challenges: Discussions from semantic data integration perspectives. In *2020 6th Conference on data science and machine learning applications (CDMA)*. IEEE, pp. 134-140.
- [43] Mehmood, H., et al. (2019). Implementing big data lake for heterogeneous data sources. In *2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)*. IEEE, pp. 37-44.
- [44] Ban, H.-J., & Kim, H.-S. (2019). Semantic network analysis of hotel package through big data. *Culinary Science & Hospitality Research*, 25(2), 110-119.
- [45] Maniou, T. A. (2021). Semantic analysis of cultural heritage news propagation in social media: Assessing the role of media and journalists in the era of big data. *Sustainability*, 13(1), 341.
- [46] Vieira, A. A., Dias, L. M., Santos, M. Y., Pereira, G. A., & Oliveira, J. A. (2020). On the use of simulation as a Big Data semantic validator for supply chain management. *Simulation Modelling Practice and Theory*, 98, 101985.
- [47] Elzein, N. M., Majid, M. A., Hashem, I. A. T., Yaqoob, I., Alaba, F. A., & Imran, M. (2018). Managing big RDF data in clouds: Challenges, opportunities, and solutions. *Sustainable Cities and Society*, 39, 375-386.
- [48] Ullah, F., Habib, M. A., Farhan, M., Khalid, S., Durrani, M. Y., & Jabbar, S. (2017). Semantic interoperability for big-data in heterogeneous IoT infrastructure for healthcare. *Sustainable Cities and Society*, 34, 90-96.
- [49] Jia, K., & Li, Z. (2020). Chinese micro-blog sentiment classification based on emotion dictionary and semantic rules. In *2020 International Conference on Computer Information and Big Data Applications (CIBDA)*. IEEE, pp. 309-312.
- [50] Vidal, M.-E., Endris, K. M., Jozashoori, S., Karim, F., & Palma, G. (2019). Semantic data integration of big biomedical data for supporting personalized medicine. *Current Trends in Semantic Web Technologies: Theory and Practice*, 25-56.