

Challenges in Ontology-Guided Data Integration: A Comprehensive Review

Mohamed Hamdy Morse¹, Karam Gouda², and Alaa Eldeen Yassin².

¹ Al-Azhar Information and Decision Support Center

² Department of IS, Faculty of Computers and Artificial Intelligence, Benha University, Banha, 13511, Egypt

ABSTRACT Ontology is a branch of philosophy that studies the nature of reality, and how to classify the things that exist in it. It aims to answer questions such as: "What are the entities that exist?", "What are their properties?", "How related to each other?", and more details about this. Ontology refers to more detailed study of the different knowledge and data to organize them in a way that can be used by different systems to understand and analyze them. One of the most important challenges in the field of ontology is data integration to integrate more than one ontology. Ontology-based data integration provides enhancement of the compatibility between data and knowledge, to improve analyze and understand, and support decisions in complicated situations. This process requires planning to solve some challenges related to how to integrate. Among these challenges is determining the methodology used for integration that is compatible with the nature of the data and the existing work environment. This survey reviews different research on Data integration using ontology. Many methodologies are used in ontology integration, Some used ontology creation to solve the data integration problem and some used ontology matching with NLP techniques, some worked fully automated and some semi-automated. The results varied depending on the methods used and the nature of the data. The paper discusses from previous studies whether ontology is an ideal solution for data integration, methodology used and result of evaluations. The paper proposes a hybrid approach that combines ontology creation and ontology matching as a potential future solution for data integration in scenarios involving multiple data sources, some of which are ontology-based while others are non-ontology-based.

INDEX TERMS Data Integration, Ontology, Ontology Creation, Ontology Matching, NLP techniques.

I. INTRODUCTION

A huge amount of domain knowledge make it difficult for human or even a systematic discovery agent to effectively use it and retrieve the relevant entities. Ontology represents knowledge as a set of concepts within a domain and the relationships between those concepts. Ontology encompasses several data and conceptual models such as sets of terms, classifications and database schemas (Shvaiko and Euzenat, 2008). Growing trend of ontology because ontology probably offers the most interesting opportunity to clarification the details of information. Ontologies have a several operations used for integration such as creation , matching , merging.

Ontology creation is the first operation based on analysis of data to very basic concept /scenarios. Ontology consists of several important concepts, Classes is a groups or collections of objects, Object properties to describe how classes can relate to each other, Data properties like object properties except their domains are typed literals, Individuals (instances) basic data/component of ontology. With the increase of

creation ontology process and the need to integrate this data, some important processes emerged, such as ontology matching.

Ontology matching is required to find semantic correspondences between a pair of entities. Ontology matching divided into four categories: terminological that use lexical similarities between names, comments etc., structural that use the similarities in the structure of the matching ontologies, instance-based that use the classified instance-data in the ontologies, and using background knowledge that rely on external structured resources to find matching entities across different ontologies [12]. Ontology Matching is an effective way to found correspondence and resolve semantic heterogeneity. But it still needs more improvement to overcome the semantic heterogeneity, this improvement by using NLP techniques.

Natural Language Processing (NLP) techniques is a process to understand syntactic, semantic, and sentimental of language. By analyzing the language, the correspondence or difference between words can be identified. Using NLP techniques with ontology matching was a powerful addition to increase the quality

of matching accuracy. The main goal of this work is to strategically advance knowledge in the area of data integration using ontology. This will be further elaborated in the subsequent sections of the survey.

II. Data Integration

Data integration refers to the process of bringing together data from multiple sources across an organization to provide a complete data image to reuse it in another processes. As shown in Figure 1 more than two Data source (DS1 , DS2, DS_n) connected to data integration method to get a unified data. This data is used after integration in many functions such as decision support systems, data analysis and other tasks related to strategic planning for large projects.

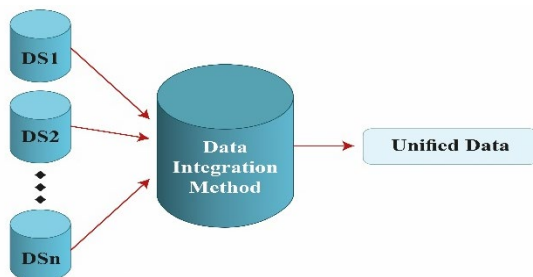


Figure 1. Data Integration Method

Putting data next to each other is not as easy as expected, it needs more study related to:

1- Type of data to be integrated : There are different types of data, Each type requires a method of implementation and a work methodology that may differ from the other type, so a good study and knowledge of all the details are necessary. Study the nature of the data to be integrated in terms of whether it is homogeneous or heterogeneous data and the presence of elements that help and facilitate integration between different data sources.

2- Data volume used : Studying the data volume is very important, as dealing with big data is definitely different from dealing with specific data. The data volume affects the choice of the appropriate methodology to use and the development of a work plan that suits the amount of data used.

3- Methodology used in integration : Determining the appropriate methodology for the data integration process is of utmost importance. The procedures or techniques used to identify, process and analyze information are established. Determine the methodology, fully automated or semi-automated :

- Semi-automated is a process or procedure that is performed by the combined activities of man and machine with both human and machine steps.

- Fully automated is a process or procedure that is performed by only machine.

4- The main purpose of the data integration process : The goal of data integration must be clear in all steps to

choose the most appropriate and suitable from available work environment.

The four points above positively affect the quality of data integration results. Integrating data from different sources, enables us to gain new knowledge and insights that help to make predictive analysis, coordinate complex processes and control systems [9]. Help to Create long time or short time Strategic plans, Which greatly helps in the growth and progress of organizations. Formally, a data integration system Consists of $\langle G; S; M \rangle$, where G is the global schema which provides a reconciled and an integrated schema, S is the source schema which describes the structure of the set of sources participating in the integration process, and M is the mapping between G and S which establishes the connection between the elements of the global schema and those of the sources [10]. With the tremendous technological development over time and the existence of many methods and techniques that provide solutions to the problem of data integration, it has been found that ontology provides a viable solution to the data integration process, with the capability to handle various types, shapes, and large volumes of data. The benefits of data integration are:

1- Data consistency: Data may exist in different sources with different classifications or representations. Integration between these helps to unify concepts and meanings that may be interpreted differently in each source, enhancing consistency between different sources, which will certainly facilitate search and retrieval processes across all different sources.

2- Comprehensiveness of information: Data integration helps provide a comprehensive and accurate picture of complex topics, which will help in collecting and analyzing multiple information in a coordinated manner.

3- Ability of systems to understand and analyze more accurately: Systems that use integrated data and ontologies can be better able to understand relationships between different entities, and draw conclusions or decisions based on common patterns in the integrated data.

4- Knowledge exchange between different systems: Integration helps in exchanging knowledge effectively between different systems. Integration helps in ensuring that all systems “understand” and interpret the same concepts in a consistent manner. This ensures that data is exchanged and analyzed correctly between different systems.

5- Decision Support: Integration helps systems make informed decisions when dealing with multiple types of data and knowledge, helping improve decision making in complex environments to achieve accurate and reliable results.

6- Solution to several problems such as inconsistency and redundancy: Integration can help identify and correct redundancy in data and concepts, reducing duplication of information and increasing its efficiency. Different sources

may sometimes contain contradictory definitions or classifications for the same entities. Integration helps resolve these inconsistencies by developing methods to unify or find agreement between concepts.

7- Supporting innovation: By dealing with multidisciplinary applications, integration can open up new horizons for innovation by allowing the development of applications or solutions that combine knowledge from different fields.

8- AI Models Training: Training AI models to understand relationships between entities in large, complex data, which helps improve the accuracy and quality of the models.

9- Contributing to cultural and social understanding: AI applications that aim to improve communication between different cultures or societies, integration helps ensure accurate representation of different cultural and social knowledge and concepts.

Despite the many advantages of data integration, some of which have been previously, there are some challenges that may face the integration process. For example, will mention some of them:

1- Contrast in concepts and classifications: Sources different in how they classify or represent the same concepts.

2- Inconsistency: Some sources may contain contradictory or conflicting information about the same entities.

3- Complexity: The integration process requires complex technologies and the use of advanced data analysis methods. As the amount of data being integrated increases, it becomes more difficult to verify its validity and quality, especially if the data comes from unreliable or open sources.

4- Privacy and Security.

5- Differences in language and terminology: Determining the exact meaning of data and how to integrate it correctly.

6- Constant changes in data: With data that is constantly changing, such as data collected from the internet or from unverified sources, it can be difficult to continuous integrity of this data.

Despite the existing challenges, data integration remains a subject of study and development due to the increasing volume, types, and sources of existing data. Some people think that linking databases and data integration are the same. The essential differences between the two terms in Table.1.

Table 1: COMPARISON BETWEEN DATA INTEGRATION AND LINKING DATABASE

Feature	linking databases	data integration
Concept	Connect different databases to enable shared access to data through queries.	Combine data from multiple sources into one unified source.
Objective	Enable systems to access data from different databases in real time.	Provide a comprehensive picture of data and analyze it in a unified manner from multiple sources.
Method	Access data via queries that combine data from multiple databases without changing their data.	Extract data from multiple sources, understand it, transform it into a standardized format, and then load it into a central repository.
Data	Data is not stored in unified source, but accessed directly.	Data is stored in only one central repository or unified database.
Benefit	Real-time access to data from different systems and display it in a unified report, supporting decision makers.	Unified data analytics, extract strategic insights across multiple data, support decision makers.
Updates	Access data directly without the need for constant updating.	Requires periodic updates to incorporate new data from different sources.

Any type or form of data can be integrated, such as ontology. The biggest challenge in the field of ontology is the integration between more than one ontology.

III. Ontology

In the field of metaphysics, ontology refers to the systematic inquiry into the nature of existence and reality. It seeks to identify the fundamental kinds of entities that exist, examine how these entities can be categorized, and explore the relationships that structure them. Ontological investigations aim to establish a comprehensive framework of categories that organizes all forms of being into a coherent and logically structured system. The basic question in ontology is "What entities and its relations are needed to describe all truths?", all truth about organization

and all data about this organization. It has vocabulary that describes objects and the relations between them, and has a grammar for using the vocabulary terms to express something meaningful to create Knowledge that help for successful integration. Ontologies are a fundamental part of the Semantic Web concept, allowing knowledge to be represented in a structured, semantic way that systems can understand and interact with automatically. The Semantic Web relies on them to achieve automated understanding of data across the Internet. The basic objective of learning ontology is to extract the right entities and make accurate relationships between this entities from the various data. All types of data can be used, be it unstructured data, semi-structured data or structured data. The only difference is in the methodology of extracting entities and their relationship with each other. It will definitely affect the results as

structured data will of course give higher accuracy unlike unstructured data. There are various variants of Ontology based data Integration [30]:

- **Single- Ontology:** is where the shared vocabulary of all the data sources that need to be integrated is defined in a single global ontology. Data from various data sources are transformed into instances of the global ontology to achieve the data integration. This integration process is typically hard to maintain because it is susceptible to changes in each data source. Any time a change occurs in one of the data sources, a decision has to be made whether to push the change to the global ontology.

- **Multiple — Ontology:** Each data source in a multiple-ontology OBDI is described using its local ontology. It cannot be assumed that these local ontologies share a common vocabulary. Mappings are established between the local ontologies. The integration process consists of three steps: (i) create local ontologies LA, LB, and LC for data sources A, B, and C, respectively, (ii) transform source data of A, B, and C according to their respective local ontologies, and (iii) establish semantic mappings between related ontologies. The drawback of this approach is that semantic mappings among involved ontologies are hard to define and maintain due to varying granularities of the local ontologies. Each inclusion of a new data source requires additional semantic mappings to all existing local ontologies.

- **Hybrid — Ontology :** is characterized by the availability of a shared vocabulary that contains basic terms of a domain that local ontologies should build on via vocabulary/ontology refinement. In this approach, the integration process consists of three steps: (i) define a shared vocabulary V that contains basic terms/concepts of the domain, (ii) create three local ontologies LA, LB and LC by using and/or extending the shared vocabulary V for data sources A, B, and C respectively, and (iii) transform/annotate source data from A, B, and C according to local ontologies LA, LB, and LC.

- **Global as view Ontology:** The central concept of the GAV approach lies in the global ontology definition, which is similar to the hybrid OBDI. GAV OBDI, however, does not require re-development of existing local ontologies due to inter-ontology transformation definitions between local and global ontologies similar to those used in the multiple-ontology .GAV OBDI requires the definition of one local ontology per data source, similar to multiple-ontology and hybrid OBDI. In this method, the integration process consists of four steps (i) Creation of three independent local ontologies LA, LB, and LC (or reuse of existing local ontologies) for data sources A, B, and C respectively. (ii) Transformation of source data in local sources A, B, and C according to local ontologies LA, LB, and LC. (iii) Development of a global ontology G represents a set of common concepts relevant to scenarios, and (iv) Definition of independent mappings between a local repository (i.e., LA, LB, and LC) and the global ontology G to facilitate

data transformation from local ontologies to the global ontology.

The ontology includes a set of basic elements as listed below:

- **Concepts or Classes:** They are classes or types that represent a group of objects or things that share certain properties. Example: In a medical ontology, "patient" and "doctor" might be classes of objects.

- **Relations:** They represent the links or relationships between different objects or concepts within an ontology. These relationships help define how objects interact or relate to each other. Example: the relationship "works for" between a "doctor" object and a "hospital" object, or "treats" between a "doctor" and a "patient".

- **Data Properties:** Properties that describe or provide additional information about objects or concepts. In ontology, properties can be textual, numeric, or date. Example: The property "age" describes the object "sick".

- **Hierarchical Structure:** A structure in which concepts are organized in a hierarchical or tree. Example: The concept "animal" can have sub-concepts such as "mammals" and "birds".

- **Individuals or Instances:** Actual values of existing objects. Example: "Mohamed" could be a value or an object in the class "Patient".

- **Rules:** Rules or laws that define relationships between objects or concepts. For example, "If a person is sick and is receiving treatment, then he or she may need medical care".

- **Constraints:** Conditions or restrictions placed on relationships or properties within an ontology.

OWL is a web ontology language used to represent knowledge in a formal and structured way. Web Ontology Language (OWL) is the latest standard of W3C; it facilitates greater machine interpretability of web content than that supported by XML, RDF or RDF Schema, by providing additional vocabulary along with formal semantics [28]. OWL has three sublanguages [28]:

- 1- OWL Lite is designed for a simple class hierarchy and simple constraints.

- 2- OWL DL is based on description logics and is more expressive than OWL-Lite. Moreover, the reasoning software is able to support complete reasoning for every feature of OWL-DL. This paper focuses on OWL-DL.

- 3- OWL Full is designed for high expressiveness. However, it does not have a computational guarantee.

Various issues or operations, related to ontologies are as follows as shown in Table 2. [26]

Table 2: VARIOUS ISSUES OR OPERATIONS

Operation	Description
Creation	It is the designing and development of an ontology from scratch. In this context, Ontology Engineering is significant, which refers to all things that are needed in the process of design, development and maintenance of an ontology. In other words, it includes methodologies, tools, languages etc.
Merging	It refers to building an ontology by combining knowledge from two or more ontologies into one.
Integration	It refers to the creation of a new ontology by reusing existing ontologies.
Selection	It refers to the selection of an existing ontology for reuse according to its topic coverage, popularity, semantic etc.
Mapping/Matching/Alignment	It refers to mapping classes/entities of various ontologies and matching them for further use. These can be aligned, if needed.
Maintenance	There is a problem of evolution in the existing old ontologies due to complexity of the changes it requires. Therefore, maintenance is required to ensure the consistency of the ontologies.
Import	This refers to the function of including an external ontology in current one.
Export	Its function is to use current ontology into others.
Debugging	This refers to the searching of inconsistencies in the definitions of the concepts.

This survey reviews some research on ontology-based data integration using with operation of creation or matching. A more detailed explanation will be presented in the subsequent parts.

IV. Ontology Creation

Building ontology involves creating attributes and constraints, creating a model, and applying it to domain data. The process of creating an ontology is not easy or simple, but rather complex because it depends mainly on the nature of the data and whether it is structured (data that is organized and easy to manage), unstructured (data that is unorganized and difficult to manage), or semi-structured (part of this data is structured and the other is unstructured). Of course, creating an ontology is considered one of the proposed solutions for data integration, as the ontology is used to unify terms and data structures in a single system or structure that is easy to deal with.

There are many tools for building and working with ontology such as WebOnto, OntoStudio, TopBraid Composer, Protégé and VOWL (Visual Notation for OWL Ontologies). Protégé is the most popular tool used. Using the ontology creation process in data integration is very important as it solves many problems whether the creation methodology is manual, automatic or semi-automatic. It is difficult to evaluate the ontology creation except by analysts of the entire domain or by measuring its fit to the domain systems. A good ontology is impossible to evaluate

because an ontology has different characteristics depending on the applied domain, but it is possible to determine how suitably the ontology fits into the domain [31].

The following are the differences between some of the papers that discussed the ontology creation process in Next. [23] Methodology Type is Automatically, Creation Stages:

1) Read and analyze the schema information of the database and convert it into the base information for the ontology generation.

2) Create a class of ontologies based on the table information of the database.

3) Based on the attribute information of the table, DatatypeProperty of the ontology class and ObjectProperty, which represents

the relation among the classes, are created.

4) Finally, the tuple in the database and the individual data of the OWL ontology are mapped using the generated ontology model.

Implementation of methodology for Smart home database consists of 85 tables.

[24] Methodology Type is Automatically. Creation Stages: 1) The XML document is transformed to XML-Schema using the Trang API for java.

2) The XML-Schema is analyzed using XML-Schema Object Model (XSOM). XSOM is a Java library that allows applications to easily parse XML Schema documents and inspect information in them.

3) The output of XSOM is used as input for the Java Universal Network/Graph framework (JUNG). The JUNG is used for graph-based manipulations. It generates XML-Schema Graph (XSG) that describes the schema in the same way whatever its design style is.

4) The Jena API uses XSG as input to generate OWL entities. Implementation of methodology for XML document represents metadata about scientific publications.

[25] Methodology Type is Automatically. The generator's function modules consist of three parts:

1) Database Analysis Module: Relies on reverse engineering technique to get database metadata from the database (using MySQL).

2) Schema Transformation Module: To convert relational database description into OWL ontology according to predefined rules.

3) OWL Ontology Module: writing the ontology into a document and reading the ontology document to the screen for end user. Implementation of methodology for using MySQL as the original database.

[27] Methodology Type is Semi-Automated. Creation Stages: 1) Recognize and canonicalize table information.

2) Construct mini-ontologies:2 from canonicalized tables.

3) Discover inter-ontology mappings.

4) Merge mini-ontologies into a growing application ontology. Implementation of methodology for Using a Population Database Tables.

[28] Methodology Type is Semi-Automated. Creation Stages: 1) generating the OWL ontology from SQL statements.

2) validating and refining the ontology produced, by a comparison with the E/R diagram.

3) mapping the ontology produced (local ontology) to a global ontology.

4) integrating the global ontology by linking it with the database. Implementation of methodology for Using university database of SQL statements.

[4] Methodology Type is Semi-Automated. This research is the first step of multi-level data collection and integration. Creation Stages:

1) Data Sources Layer: to collect data in one database from a lot of data sources .

2) Ontologies and Mapping Layer: for mapping and transformation, data sources elements for shared vocabulary and global ontology. Implementation of methodology for Summarized information about the population, area type, and financial conditions for all French provinces and cities.

[32] Methodology Type is Automatically. Creation Stages: 1) Generation Phase: the input to the framework is unstructured text corpus and the output is the preliminarily generated KG. Include (Data Cleaning: cleaning irrelevant information is necessary, Knowledge Graph Generator: This generated graph includes the entities and relations with the corresponding confidence score).

2) Refinement Phase: include (Reference Ontologies: to evaluate and verify the generated KGs, Anomalies Exclusion: to exclude irrelevant, illogical, and unrelated nodes (concepts) and relations, Correctness Module: solve problems found, Completion Module: knowledge graph embedding).

3) Mapping Phase: contains (Domain ontologies: the generated KG must be mapped to fit in the domain ontologies to ensure the consistency and Interoperability generated KG with ontologies used in an organization, Consistency check: to solve interoperability issue in the KGs by checking whether all concepts and relations are consistent with the range and object-property in the target domain, The Generated ontology: the KG is trimmed and transformed to a domain ontology). Implementation of methodology for use unstructured text corpus to generate Knowledge graph then generate domain ontology.

The process of creating an ontology is one of the most important processes because it depends on realistic analysis and complete knowledge of the components of the domain. Using creation of ontology as a solution for data integration may be a good solution for data sources that cannot be changed or updated, and this in itself is something that is difficult to find in any area of life. While automatic methods offer higher speed, semi-automatic approaches are generally preferred for their superior accuracy and quality.

V. ONTOLOGY MATCHING

The use of ontology matching in the data integration process is very important as it is used in the presence of more than one ontology and aims to determine the degree of correspondence between them. One type of matching is semantic matching, which relies on measuring the similarity between different entities in an ontology, which relies on the importance of relationships between concepts and meanings rather than on structures. For example, have two ontology source and target. Source ontology have class named "Journal", and target ontology have class named "Periodical". In formal matching, it is difficult to match the words "Journal" and "Periodical" although in semantic matching, the percentage of similarity in meaning may be very large.

Semantic matching can be use techniques such as WordNet or Natural Language Processing (NLP) techniques to identify relationships between words and measure similarity between them. WordNet is a semantic database (Dictionary) that contains links between words based on their meanings. Words are grouped into synonym sets, and this database links words together with relationships to find their synonyms and antonyms. WordNet similarity was employed to obtain the semantic similarities among elements. But the weakness is that the words in WordNet are insufficient. Many elements in ontologies cannot find their correspondences in WordNet. As a result, it is impossible to compute the semantic similarities between these elements with others [44].

Natural language processing (NLP) techniques such as edit distance and word embedding, are used within ontology matching methodologies to increase accuracy of data integration. Ontology matching is an important operation for ontology mapping to find a semantic correspondences between a pair of entities in different ontology. As shown below in Table 3 the summary of ontology matching papers and the different of its methodology.

At the beginning of the table, the type of methodology is shown. Some papers using Fully-automated matching and other using semi-automatically matching. But the papers that have been studied are used Fully-automated matching. In the following "Cardinality", it determine the level of matching. It determines the number of items that are matched between ontology target and ontology source in single matching operation. In our analysis, we determine the techniques applied in each selected research work, including NLP techniques and WordNet. Very few papers do not use NLP techniques and WordNet in the ontology matching process, but most papers use them for their effective role in increasing the quality and accuracy of the matching process.

Table 3: THE SUMMARY OF ONTOLOGY MATCHING PAPERS

Paper No.	Methodology Type	Cardinality	Techniques used
[22]	Automated	1:1	word embedding model, Cosine similarity
[33]	Automated	1:1	TFIDF
[34]	Automated	1:1	String Similarity metrics, WordNet
[35]	Automated	1:1	WordNet, ISUB technique
[36]	Automated	1:1	String Similarity metrics, WordNet
[37]	Automated	1:1	Concept Similarity analysis (CSA2) algorithms
[38]	Automated	1:1	Edit distance, vector distance, WordNet
[39]	Automated	N:M	tokenization, string equality, Levenshtein distance
[40]	Automated	1:1	edit distance
[44]	Automated	1:1	Hybrid (WordNet, Word Embeddings)
[45]	Automated	1:1	WordNet
[46]	Automated	1:1	word embedding (word2vec model), cosine similarity
[47]	Automated	1:1	Vector space model, cosine similarity

The NLP technique is determined by the nature of the data contained within the ontology. However, there is no doubt about the importance of using natural language techniques within the ontology matching methodology, which helped increase the quality and accuracy of ontology-based data integration. Many papers have proven the superiority of using natural language processing techniques over other techniques in ontology matching, due to the effective results they have achieved in achieving matching.

VI. CONCLUSIONS AND FUTURE WORK

In this survey, the importance of ontology-based data integration is presented. The processes of ontology creation and ontology matching are examined, along with the significance of employing Natural Language Processing (NLP) techniques in ontology matching. Ontology creation and matching processes are investigated for their role in facilitating data integration. This study is a good nucleus to identify previous studies in the fields of ontology creation and ontology matching, which is a trend for many current researchers.

We suggested a good study of the nature and form of the data that will be dealt with because it is what will determine the methodology of the work as a whole. With the tremendous development in the size and nature of the existing data, the process of studying and analyzing this data is extremely difficult. In certain data integration scenarios, particularly when some or all of the data sources are non-ontological, challenges may arise that necessitate the combined use of ontology construction and matching as a solution. A semi-automatic approach is often preferred in

such cases, as it offers a balance between precision and efficiency that is typically superior to fully automatic methods.

In future, a hybrid approach will be adopted to formulate a methodology aimed at facilitating ontology-based data integration. This methodology depends on creation and matching processes. NLP techniques are utilized during the matching phase to improve the precision and overall effectiveness of the process. The field of ontology-based data integration is a field that needs a lot of research and study, despite the interest of many in it, but it needs more to reach the highest benefit.

REFERENCE

- [1] Lucilo, Sandro, Ludovic, "A semi-automatic Design methodology for (Big) Data Warehouse Transforming Facts into dimensions", *In The Journal Of IEEE Transactions on Knowledge and Data Engineering*, vol. 14, No. 8, Aug-2015.
- [2] Isam, " Design of Data warehouse model for Decision support at higher education : A case study ", *Article in Information Development*, Vol. 32(5) 1691-1706 (2016).
- [3] Natalya, " Semantic Integration : A survey of ontology-Based Approaches ", *SIGMOD Record*, Vol. 33, No. 4, December 2004.
- [4] Nader, et, " Ontology-based teacher-context data integration ", *SIGMOD*, 2018.
- [5] Gusenkov, et, " On ontology based data integration : problems and solutions ", *IOP Conf. Series: Journal of Physics : Conf. series* 1203(2019) 012059.
- [6] Maxime, et, " Ontology Based RDF integration of heterogeneous data ", *EDBT march-april,2020* , ISBN 978-3-89318-083-7.
- [7] Chuangtao, et, " Ontology learning from relational Database : opportunities for semantic information integration ", *World Scientific (Vietnam Journal of Computer Science)* Vol. 9, No. 1 (2022) 31-57.
- [8] Maryam, et, " Ontology Based Information integration : a survey ", *Archived arxiv:1909.13762v1 [cs.ir]*2019.
- [9] Fritz , et, " Information Integration using the typed graph model ", *Archived arxiv:2107.09592v1 [cs.db]* 2021.
- [10] Ladjel, et, " An automated information integration technique using an ontology-based ", *Hal open science*, ID: hal-03759389, 2022.
- [11] Alberto Salguero, et, " Ontology based framework for data integration ", *Information Science & Applications* (ISSN : 1790-0832), Issue 6, Vol.5 , Jun. 2008.
- [12] Brett Drury, et, " A survey of semantic web technology for agriculture ", *Science Direct*, 2214-3173, 2019.
- [13] Pascal Hitzler, " Semantic Web : A Review of the Field ", *ACM* ISBN 0001-0782/2020/04.
- [14] Zhizun, et, " Knowledge Management Framework For Emerging Infectious Diseases Preparedness and

- Response: Design and Development of Public Health Document Ontology ", *JMIR Res Protocol*, Vol. 6, Iss. 10, e196, 2017.
- [15] WEI Yuan-yuan, et, " From Web Resources to Agricultural Ontology : A Method for Semi-Automatic Construction ", *Science Direct (Journal of Integrative agriculture)* 11(5):775:783,2012.
- [16] Sivadi, et, " A Survey on Semantic Approaches for IOT Data Integration In Smart Cities ", *ICICCT* pp.827-837, 2020.
- [17] Michelle Cheatham and Catia Pesquita, " Semantic Data Integration ", *Archived on Science Direct*, 2016.
- [18] Aamod Khatiwada, et, " Integrating Data Lake Tables ", *In PVLDB*, 16(4): 932 - 945, 2022.
- [19] Karam Gouda, et, " DASS : A Discovery Agent Supporting System ", *In IEEE SMC'99 Conference*, DOI: 10.1109/ICSMC.1999.815672.
- [20] Ines Osman, et, " An Alignment-Based Implementation of a Holistic Ontology Integration Method ", *In Elsevier B.V.* , DOI: 10.1016/j.inffus.2021.01.007.
- [21] Ines Osman, et, " Ontology Integration : Approaches and Challenging Issues ", *In Elsevier B.V.* , doi.org/10.1016/j.inffus.2021.01.007.
- [22] Nesma Mahmoud, et, " Enhanced Ontology Matching For Big Data Integration ", *ICaTAS 2019* , doi:10.1088/1742-6596/1447/1/012028.
- [23] Young B.Park, et, " Methodology for Automatic Ontology Generation Using Database schema Information ", *Hindawi* , doi.org/10.1155/2018/1359174.
- [24] Nora Yahia, et, " Automatic Generation of OWL Ontology from XML Data Source ", *International Journal of Computer Science Issues* 9(2), 2012.
- [25] Shufeng Zhou, et, " Ontology Generator from Relational Database Based on Jena ", *Computer and Information Science* 3(2), 2010 , DOI:10.5539/cis.v3n2p263.
- [26] Swati Negi, et, " An Algorithm for Merging Two Ontologies : A Case Study ", *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 13, Number 12 (2018) pp. 10327-10338.
- [27] YURI A. TIJERINO, et, " Towards Ontology Generation from Tables ", *World Wide Web: Internet and Web Information Systems*, 8, 261–285, 2005 , DOI:10.1007/s11280-005-0360-8.
- [28] Nasser Alalwan, et, " Generating OWL Ontology for Database Integration ", *IEEE* , DOI: 10.1109/SEMAPRO.2009.21.
- [29] Zharko Aleksovski, et, " Using multiple ontologies as background knowledge in ontology matching ", *researchgate* , 242720412, 2008.
- [30] Soumya Mulgund, " Ontology Based Data Integration-Rundown ", *medium.com* , <https://medium.com/@soumyamulgund/ontology-based-data-integration-rundown-cd65c1166cc9>, 2019.
- [31] S. Tartir, I. B. Arpinar, et, " OntoQA: Metric-Based Ontology Quality Analysis ", *Wright State University, Dayton, OH, USA*, 2005.
- [32] Samaa Elnagar, et, " An Automatic Ontology Generation Framework with An Organizational Perspective ", 2020, *Proceedings of the 53rd Hawaii International Conference on System Sciences* , URI: <https://hdl.handle.net/10125/64339> , 978-0-9981331-3-3.
- [33] Doan, et, " Learning to map between ontologies on the semantic web ", 2002, *Proceedings of the 11th international conference on World Wide Web*, ACM, 662-673.
- [34] Lambrix, et, " SAMBO-A System for Aligning and Merging Biomedical Ontologies ", 2011, *Web Semantics: Science, Services and Agents on the World Wide Web*, 4.
- [35] Jiménez-Ruiz, et, " Logmap: Logic-based and scalable ontology matching ", *The Semantic Web–ISWC 2011. Springer*.
- [36] Ichise, " Machine learning approach for ontology mapping using multiple concept similarity measures ", *Computer and Information Science, 2008. ICIS 08. Seventh IEEE/ACIS International Conference on, 2008. IEEE*, 340-346.
- [37] Martinez-Gil, et, " MaF: An ontology matching framework ", 2012, *Journal of Universal Computer Science*, 18, 194-217.
- [38] Li, et, " RiMOM: A dynamic multistrategy ontology alignment framework ", 2009, *Knowledge and Data Engineering, IEEE Transactions on*, 21, 1218-1232.
- [39] Jean-Mary, et, " Ontology matching with semantic verification ", 2009, *Web Semantics: Science, Services and Agents on the World Wide Web*, 7, 235-251.
- [40] Shen, et, " OMReasoner: Combination of Multi-matchers for Ontology Matching: results ", *for OAEI 2014*.
- [41] XIULEI LI, et, " Ontology Matching: State of the Art, Future Challenges, and Thinking Based on Utilized Information ", *IEEE , Digital Object Identifier* 10.1109/ACCESS.2021.30 , VOLUME 9, 2021.
- [42] Cassia Trojahn, et, " Foundational ontologies meet ontology matching: A survey ", *Semantic Web 13* (2022) 685–704 , DOI 10.3233/SW-210447 , Published by IOS Press.
- [43] ZAKARIA HAMANE, et, " ONTOLOGY MATCHING USING DEEP LEARNING ", *Journal of Theoretical and Applied Information Technology* , May 2023. Vol.101. No 10 , ISSN: 1992-8645.
- [44] Yuanzhe Zhang, et, " Ontology Matching with Word Embeddings ", *Springer International Publishing Switzerland 2014* , LNAI 8801, pp. 34–45, 2014.
- [45] Feiyu Lin, et, " A Survey of Exploiting WordNet in Ontology Matching ", 2008, *Springer* , Volume 276, pp. 341–350.
- [46] Xingsi Xue, et, " Matching Transportation Ontologies with Word2Vec and Alignment Extraction Algorithm

- ", *Hindawi, Journal of Advanced Transportation* ,
Volume 2021, Article ID 4439861, 9 pages ,
<https://doi.org/10.1155/2021/4439861>.
- [47] Shu-Chuan Chu, et, " Optimizing Ontology Alignment
in Vector Space ", *Journal of Internet Technology*
Volume 21 (2020) No.1, DOI:
10.3966/160792642020012101002.
- [48] Sarawat Anam, et, " Review of Ontology Matching
Approaches and Challenges ", *International journal of
Computer Science & Network Solutions* , March.2015-
Volume 3.No.3 , ISSN 2345-3397.