FUOYE Journal of Engineering and Technology, Volume 7, Issue 3, September 2022 ISSN: 2579-0617 (Paper), 2579-0625 (Online)

A Novel Model for the Semantic Enrichment of an Ontology Alignment System

*[1,2]Muhammad D. Abdullahi, [1]Aliyu Salisu and [1]Donfack A. F. Kana
[1]Department of Computer Science, Ahmadu Bello University, Zaria, Nigeria
[2]Iya Abubakar Institute of ICT, Ahmadu Bello University, Zaria, Nigeria
mdabdullahi|aliyusalisu|donfackkana@abu.edu.ng

Received: 17-MAR-2022; Reviewed: 12-APR-2022; Accepted: 21-JUN-2022
https://doi.org/10.46792/fuoyejet.v7i3.819

ORIGINAL RESEARCH

Abstract- In many applications, ontology alignment is a difficult challenge, and it is a major concern for interoperability and domain specialists in information systems. Although there are various methods for ontology alignment, most of them ignore ontological links such as subsumption and aggregation. Furthermore, traditional alignment methods provide little or no information about the underlying structure of the correspondences between ideas that they identify, limiting them to basic links between matched concepts. However, many actions, such as ontology mergers, ontology evolution, or data conversion, require more detailed information, such as the actual relationship type of correspondences matches or information about the cardinality of a correspondence (one-to-one or one-to-many). We employed an enrichment technique to build an upgrade to the present ontology alignment tool (Falcon AO++) that recognizes and adds more semantic information to a created ontology mapping in this study. The improved mapping now supports equal, is-a, inverse is-a, part-of, has-a, and related semantic connection types. To enable semi-automated mapping to correct and detect more sorts of correspondences, the enrichment technique leverages a variety of linguistic, structural, and background knowledge. As an outcome, more expressive mappings may be created. In terms of precision, recall, and f-measure, Falcon-AO++ and eFalcon-AO perform better by 18.1 percent, 20.2 percent, and 18.6 percent, respectively.

Keywords- Falcon AO++, Ontology, Matching, Mapping, Merging, Semantic, Relationship, Cardinality

1 Introduction

Ontology matching plays an integral part in data integration and semantic web. The result of matching two or more ontologies of the same domain is a list of correspondences between the two ontologies, which is often called mapping or alignment. Popular alignment techniques often produce undefined matching results, signalling that two concepts or ontology entities are related in some manner but not revealing what type of relationship exists (Kittiphong et al, 2019) (Jérôme et al. 2017). Unfortunately, such a specification appears to be impossible based on the generally used exploratory and limited approaches for determining correspondences (Kimmig et al., 2011).

Some matching methods focus solely on returning equal and disjoint relations (Alhassan, 2015), which is a connection between two concepts that implies that they express the same thing. True equivalency relationships appear unusual in broader align scenarios, but more particular ones like subsumption (is-a) or aggregation (part-of) may arise more frequently (Le, 2020). Other ways may be able to accommodate a variety of relation types, but they do not specify the types in the schema. True equivalency relations are uncommon in broad alignment contexts, but more particular ones like subsumption (is-a) or aggregation (part-of) are more common (Le, 2020).

*Corresponding Author

Section B- ELECTRICAL/ COMPUTER ENGINEERING & RELATED SCIENCES
Can be cited as:
Abdullahi M.D., Salisu A. and Kana D.A.F (2022): A Novel Model for the Semantic Enrichment of an Ontology Alignment System, *FUOYE Journal of Engineering and Technology* (FUOYEJET), 7(3), 299-304. http://doi.org/10.46792/fuoyejet.v7i3.819

Other techniques may enable a wide range of relation types, but they do not specify the kinds in the mapping. This is primarily owing to the fact that establishing the sort of semantic correspondence between two objects is more difficult than determining their synthetic relatedness (Elena, 2018). A matching technique can determine the relationship between two terms: city hall and city, based on lexicographic similarities, but it is more difficult to characterize the connection type (part-of) (Chantal, et al. 2017).

Knowing the correspondence type link between two mapped concepts has a number of advantages (Fausto, et al. 2014). If the two ontologies must be integrated, a typical matching technique has discovered the two correspondences (Beverages, Beverages) and (Red Wines, Wines), but no correspondence type has been given. On the other hand, the OT (Target Ontology) that will be created from the two input ontologies is expected to have a leading concept (Beverages) and four sub-concepts (Juices Liquors, Beer & Champagne).

Regardless, a crucial topic in the Target Ontology OT is which of the two equivalent concepts (Red Wines and Wines) should be picked. If the relationship is judged to be of the equal relation type, just one of the ideas be included in the OT (which is the assumption if no relation type of connection is specified). If the sort of relationship is presumed to be equivalent, a random concept, such as Wine or Red Wines, must be selected. As a result of the ontology merger, the first entry, Red Wines, may be added to the target ontology OT. The concept of Wine is removed from the combined ontology in this situation. As a solution to this problem, many lexicographic resources built by SemRep (Pavel & Jérôme, 2018) were leveraged to offer expressiveness to the correspondences produced by Falcon AO++ (Jauro, 2014). In this paper, we illustrate how to use different approaches to apply relation types to an existing mapping, as well as how each idea relates to the others. The improvement of semantic mapping, also known as determining the different sorts of correspondences in a given alignment set, is the focus of this research.

We introduce and demonstrate the SemRep mapping enriching tool, which determines a type of semantic relationship for each relationship using a pre-calculated ontology mapping (Elena et al., 2018). In comparison to prior techniques, SemRep focuses primarily on language laws and linguistic perspectives. Linguistics is, by and large, the key to the appropriate type of relationship matching and determination (Steffen et al., 2020). It demonstrates how to use several approaches to compute the semantic relation types between two mapping ideas using these linguistic principles. This research's findings and experiences can be used to match schema and ontology in general.

The following is how the rest of the paper is organized: The second section of this paper discusses related works. A brief overview of relation kinds is offered in section 3. The pattern of the six relations types is described in Section 4. Section 5 on evaluation and results describes the suggested eFalcon-AO framework architecture. Section 6 provides discussion and analysis. Finally, conclusion is provided in section 7.

2 Related Works

Most matching approaches, for example, (Le, 2020), Sambo (Manvi & Sanjay, 2021), Falcon-AO (Alhassan, 2015), do not empathize with the types of relationships between entities in mapping ontologies (Chantal & Brigitte, 2017). Some matching tools employ WordNet to determine the types of correspondences; (Alhassan, 2015) which indicates that it would be possible to provide a relationship type (Manjula, K., & Dinesh, 2012), but most matching tools do not use this semantic expertise Most techniques only find simple correspondences; for instance, Complicated correspondences, according to (Matthias, 2018), imply the presence of a relation type other than equal, yet most techniques do not consider relation types to be such. So far, only a few approaches to semantic matching have been developed (Pavel & Jérôme, 2018). 2018 (Pavel & Jérôme).

Several other publications dealt with the detection of semantic correspondence types, but only find a handful and, in most cases, do not provide a thorough explanation. The approach (Matthias, 2018) finds correspondences and relationship forms between concepts from a variety of crawling ontologies using the Swoogle search engine. The strategy supports relationships that are equal, subset, or mismatched. The tough problem of recognizing complicated correspondence, unlike relationship type identification, has yet to be satisfactorily overcome.

The Falcon AO++ alignment tool was extended to SemRep to allow for the determination of relation types using its lexical background information, transforming it into a semantic enrichment tool; the enrichment tool eFalcon-AO could discover six different connection kinds (equal, is-a, inverse is-a, part-of, has-a and related). It appears to be the only open-source tool that distinguishes between the types is-a (subsumption) and part-of (aggregation), which are rarely separated in other related techniques. Because all that is required is a basic list of concept correspondences, eFalcon-AO is a very versatile tool that can technically handle any mapping.

3 Relations Types

Ontology matching relationship types are equivalent to linguistic relations (Sachi, *et al.* 2021), (Pavel & Jérôme, 2018); nevertheless, scientific literature uses distinct terminologies, and relation types are often not fully defined. Synonym relations are known as equivalence, (Sachi, et al, 2021), is equal (Ankur, 2020), or same-as, whereas hyponym relations are known as is-a, (Steffen, 2020), kind-of, (Yongfang, *et al.*, 2018), subsumption (Zhuoyu, 2018) (Association, 2021). SubClassOf and sameAs are terms used in OWL to describe hyponym or synonym relationships between concepts. The authors (Elena, 2018) and (Chantal *et al.*, 2017) use description logic notation to represent semantic relations in part. Equivalence (=), more-general (($\geq$), overlapping (($\sqsupset$), and mismatch ($\perp$) are all mentioned.

4 Proposed eFalcon-AO Relations Pattern

4.1 Part-of and Has-a Format

The patterns for part-of and has-a relations are less flexible, thus they're used less often in description sentences. As in "A CMOS is the hardware within a computer," the prepositions in and of, as well as the adverb within, all imply part-of relationships. The part-of and has-a trends are listed in Table 1.

Table 1. Type Styles

Part-of Forms	"in" "of" "as part of"
Has-a Forms	"having" "with" "consisting of"

4.2 Is-a Relation Format

The most flexible and important Relation Type-Pattern is the is-a form. Simple examples include "X is a Y" or "X is Y" (plural), while more complicated ones include "X is any variety of a Y" or "X is generally any form of Y." With a second (temporal) adverb, such as usually, traditionally, or generally, such patterns can be found.

http://doi.org/10.46792/fuoyejet.v7i3.819 http://journal.engineering.fuoye.edu.ng/

Table 2 shows some common 'is-a pattern' examples as found in SemRep resource definition statements.

Table 2. Is-a pertter

Is-a Forms	"is a" "is typically a" "is any form of" "is a class of"

4.3 Equal-Relation Format

Equal-relations forms are regularly used in itemizations and are frequently represented by basic words like "and" or "or," for example, a specific vehicle is a bike or a bicycle. There are a couple of more complex forms that were first introduced in Table 3. Outside of itemizations, binary patterns can be used in definition phrases, such as A stands for B (in abbreviations), A is either a B synonym or B is an A synonym. However, such patterns are quite exceptional, as most Wikipedia articles have only one definition. " As a result, instead of a descriptive sentence like "A car is an automotive synonym," papers like "An automobile, auto car, motor car, or car is a wheeled motor vehicle.

Table 3. Equal Format

Equal Forms	''A, B, and C" "A also called B" "A, also known as B or C"

4.4 Proposed Framework

Falcon-AO++ was extended with an external background resource, Semantic Repository, in our system, the Semantic Enrichment of Falcon-AO++ Correspondences (eFalcon AO), with Falcon-AO++ as the base system (SemRep). This research work's major focus includes, among other things, discovering and labelling ontology correspondences with semantic relation types. This process is known as mapping enhancement since it results in a semantically richer mapping. The approach establishes the semantic relation type of the associated concepts, by first calculating an initial mapping using Falcon AO++ ontology mapping outputs and then using SemRep lexical background knowledge resources for the enrichment. As a result, the initial mapping is the input to the mapping enrichment approach (eFalcon-AO), and the output is an enriched mapping, which gives each Correspondence a specific semantic relation type. Background knowledge resources in SemRep, which also include the right use of dictionaries and thesauri, provide a more comprehensive set of resources that help Falcon AO++ produce enhanced output alignments. Six different relation types are determined using these SemRep resources (WordNet, UML, Thesauri etc.): equality, is-a (subsumption), inverse is-a, part-of (aggregation), and inverse part-of (Has-a) relations.

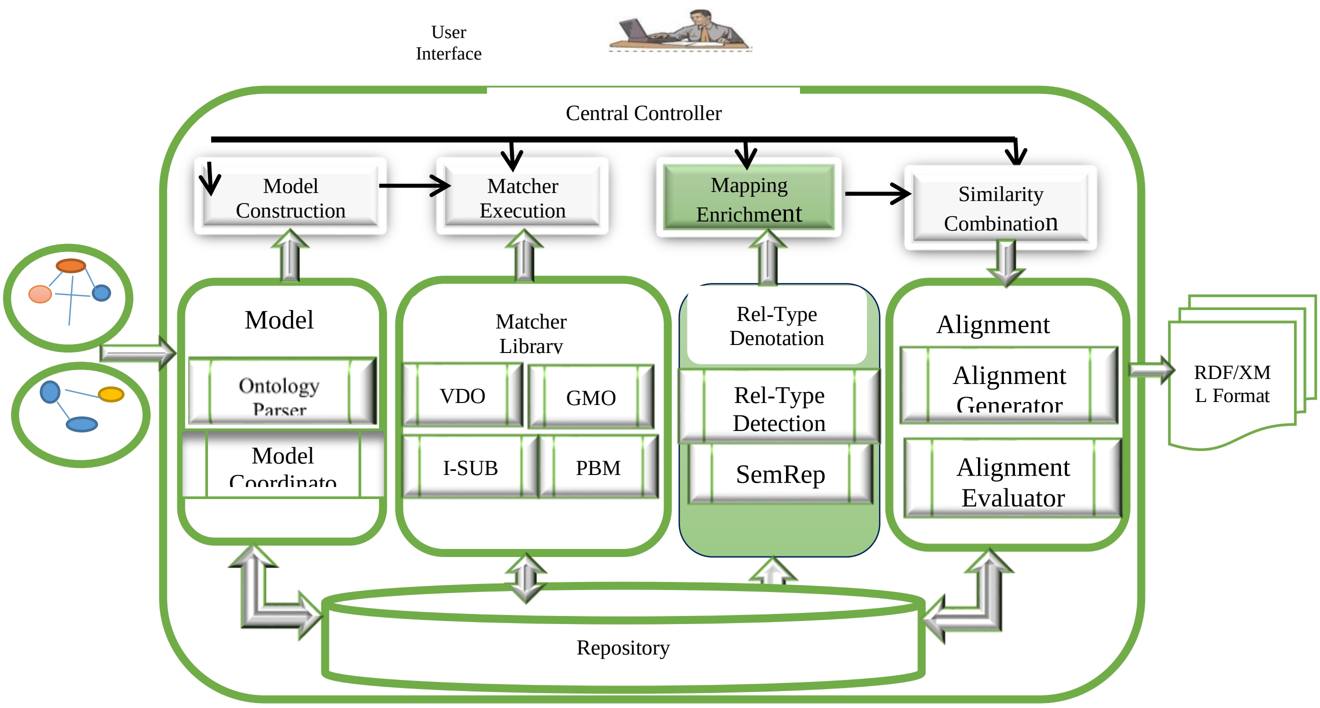

Fig. 1: Proposed eFalcon AO Architecture

http://doi.org/10.46792/fuoyejet.v7i3.819 http://journal.engineering.fuoye.edu.ng/

5 Evaluation and Results

The influence of the relation types recognized and denoted by SemRep on the accuracy of the alignment outcome was tested using six pairings of ontologies from the OAEI conference track dataset and reference alignments. The system's alignment results were compared to the alignments in the reference alignment. The usual measurements of Precision, Recall, and F-measure for evaluating ontology alignment systems was employed for evaluation.

$$\text{Precision} = \frac{\text{No. of correct found mappings}}{\text{Total No. of found mappings}} \quad (1)$$

$$\text{Recall} = \frac{\text{No. of correct found mappings}}{\text{Total No. of existing mappings}} \quad (2)$$

$$\text{F-measure} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

6 Analysis and Discussion

As test data, the OAEI conference track dataset was employed. The tests were run on a system with a 2.20GHz Intel (R) Core (TM) i5 CPU, 6 GB RAM, and Windows 8.1. Tables IV and V compare the results of Falcon-AO++ and eFalcon-AO. Table 4 lists the number of existing alignments in the reference alignment, the number of alignments found by each system, and the number of correct alignments found (i.e., the number of found alignments that exist in the reference alignment) for each pair of input ontologies. The input ontologies must be matched. The system may automatically return six different relationship types for the eFalcon-AO using the embedded SemRep resources.

The enhanced system can return (equal, is-a, inverse is-a, part-of, has-a, and related) in the input (Table 4), which increased the number of correct found alignments from 4 (in Falcon-AO++) to 6 (in eFalcon-AO), and the equivalent and is-a (subsumption) input increase the number of found alignments from 9 (in Falcon-AO++) to 11 (in eFalcon-AO) (in eFalcon-AO). The findings of the study clearly demonstrated that recognizing and annotating the various semantic relation kinds of corresponding concepts might improve alignment outcomes for ontology merging and evolution.

The comparative result of Falcon-AO++ and eFalcon-AO demonstrates an improved performance of 18.1 percent, 20.2 percent, and 18.6 percent in terms of; average precision, recall, and f-measure correspondingly, as indicated in Table 5, and illustrated in Figures 2, 3 and 4.

7 Conclusion

The general objective of this paper is to extend an Ontology Alignment System Falcon AO++ with SemRep which is a collection of lexicographic resources to enrich the mapping correspondences of the system. The system tries to discover a set of matches between the concepts (classes, properties, etc.) of two ontologies and their corresponding relationships. In this paper, the equivalence, is-a, inverse is-a, part-of, has-a, and related relations were considered using a degree of confidence in the range [0,1]. Detecting the semantic relation types of the mapped concepts are believed to be highly reliable in the field of ontology merging and evolution in the Semantic Enrichment system (eFalcon-AO) developed. The major requirement for the alignment system is to accept input ontologies and discover the semantic relation types that exist between the corresponding ontologies using the Semantic Repository (SemRep) resources.

For the future work; Machine Learning techniques for mining semantic mappings can be incorporated, V-doc can be extended to consider further neighbours rather than only one-step neighbours and the turning time of the frame work can be improved.

Table 4. Alignment Result

Inputs			Falcon AO++		eFalcon_AO							
S/N	Input Ontologies	No. of Available Alignments in reference Alignment	No. of Return alignment	No. of Precise found Alignments	Relation Types: Is-a	Inv. Is-a	Part-of	Has-a	Equal	related	No. of found alignments	No. of Correct found alignments
1	Confof Iasted	11	9	4	1			1			11	6
2	Confof Sigkdd	5	7	4			1			1	8	6
3	Edas Ekaw	20	23	13	1		2				26	16
4	Edas Iasted	11	19	7	2	1				1	23	11
5	Ekan Isated	13	10	7		2		1			12	10
6	Ekan Sigkdd	9	11	7		1				1	13	9
7	Iasted Sigkdd	22	15	13			1	2			18	16
Average		**13**	**13.42857**	**7.857143**							**13.42857**	**10.28571**

http://doi.org/10.46792/fuoyejet.v7i3.819 http://journal.engineering.fuoye.edu.ng/

Table 5. Alignment results

		Falcon AO++			eFalcon_AO		
S/N	Input Ontologies	Precision	Recall	F-Measure	Precision	Recall	F-Measure
1	Confof Iasted	0.444444	0.363636	0.4	0.666667	0.545455	0.6
2	Confof Sigkdd	0.571429	0.8	0.666667	0.714286	1	0.833333
3	Edas Ekaw	0.565217	0.65	0.604651	0.73913	0.85	0.790698
4	Edas Iasted	0.368421	0.636364	0.466667	0.578947	1	0.733333
5	Ekan Isated	0.7	0.538462	0.608696	0.9	0.692308	0.782609
6.	Ekan Sigkdd	0.636364	0.777778	0.7	0.818182	1	0.9
7.	Iasted Sigkdd	0.866667	0.590909	0.702703	1	0.681818	0.810811
Average		**0.59322**	**0.62245**	**0.592769**	**0.773887**	**0.824226**	**0.778683**

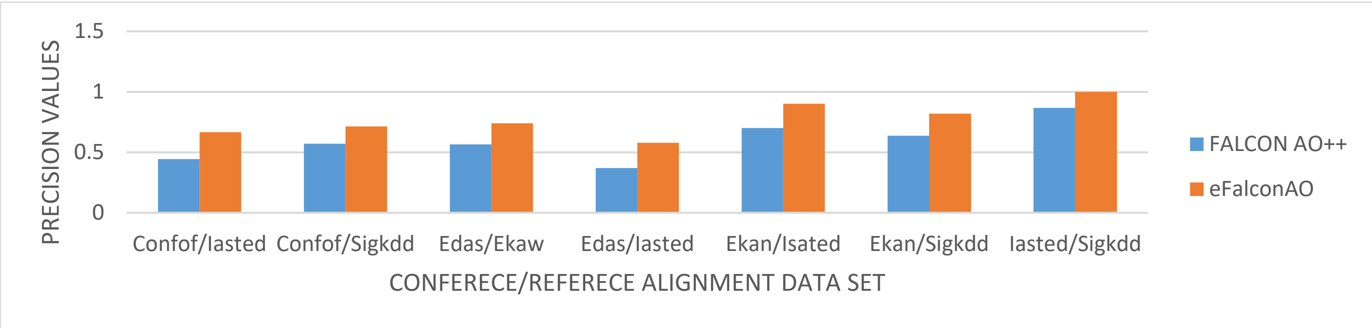

Fig. 2: Precision Comparative Result of Falcon AO++ and eFalconAO

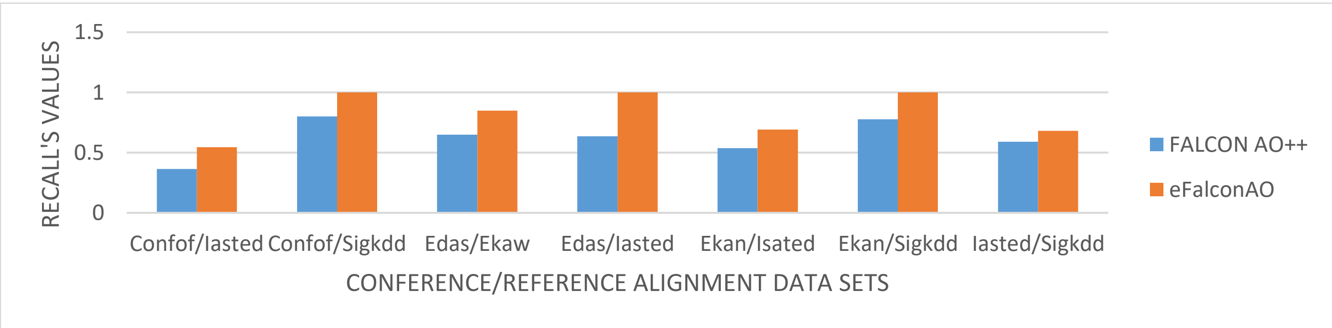

Fig. 3: Recall Comparative Result of Falcon AO++ and eFalconAO

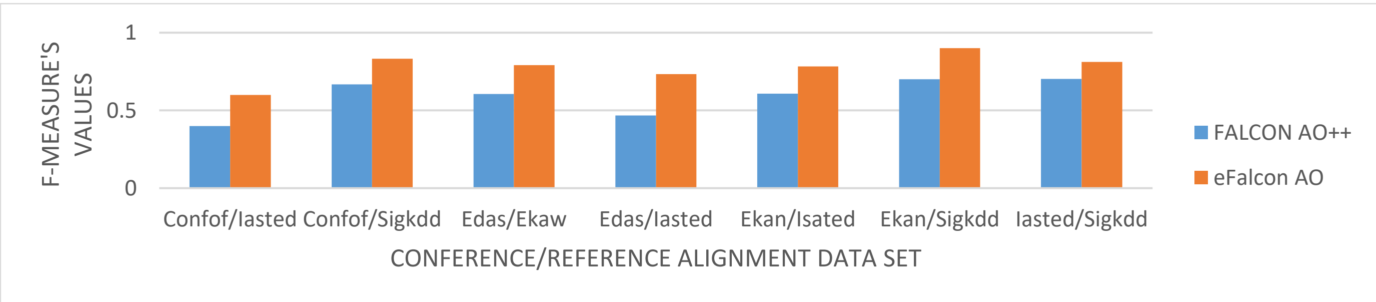

Fig. 4: F-measures Comparative Result of Falcon AO++ and eFalconAO

http://doi.org/10.46792/fuoyejet.v7i3.819 http://journal.engineering.fuoye.edu.ng/

REFERENCES

Alhassan, B. B. (2015). *Extending an Ontology Alignment System with a Lexical Database.* Kaduna: ABU Press Ltd.

Alibuhtto, M. C., & Mahat, N. I. (2020). Distance based k-means clustering algorithm for determining number of clusters for high dimensional data. *Decision Science Letters*, 51-58.

Ankur, A. (2020). Quality assurance and enrichment of biological and biomedical ontologies. *BMC Medical Informatics and Decision Making*, 1-4.

Association, G. W. (2021, 07 29). *WordNets in the World*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/WordNet

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 1-27.

Chantal, R., & Brigitte, S. (2017). Exploiting WordNet as Background Knowledge. *Proceedings of the 2nd International Conference on Ontology Matching* (pp. 291–295). Aachen -Germany: CEUR-WS.

Chuncheng, X., Tingsong, J., & Baobao, C. (2018). ERSOM: A Structural Ontology Matching Approach Using Automatically Learned Entity Representation. *Key Laboratory of Computational Linguistics, Ministry of Education, China*, 1-11.

Elena, B. (2018). Exploiting Relation Extraction for Ontology Alignment. *Jena University Language and Information Engineering (JULIE),* 1-8.

Fausto, G., Mikalai, Y., & Pavel, S. (2014). S-match: an Algorithm and an Implementation of Semantic Matching. *In Proceedings of the First European Semantic Web Symposium* (pp. 61-75). Germany

Geng, Z., Chengchang, Z., & Huayu, Z. (2018). Improved K-means Algorithm Based on Density Canopy. *Knowledge-Based Systems*.

Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in RFM analysis based on k-means. *Indonesian Journal of Electrical Engineering and Computer Science*, 470-477.

Huanyu, L., Zlatan, D., Daniel, F., & Valentina, I. (2019). User validation in ontology alignment: functional assessment and impact. *The Knowledge Engineering Review*, 33.

Jauro, F. (2014). *Design and Implementation of Falcon AO++.* Kaduna: ABU Press ltd.

Jérôme, D., Fabrice, G., & Henri, B. (2017). Association Rule Ontology Matching Approach. *International Journal on Semantic Web & Information Systems*, (pp. 27-49).

Kimmig, M., Monperus, M., & Mezini, M. (2011). Querying Source Code with Natural language. *26th IEEE/ACM International conference on Automated Software Engineering* (pp. 376-379). USA.

Kittiphong, S., & Romchat, K. (2019). Ontology-Based Semantic Integration of Heterogeneous Data Sources Using Ontology. *Journal of Theoretical and Applied Information Technology*, 1-14.

Kumar, S., Ujjal, M., & Utpal, B. (July 2015). Automatically Converting Tabular Data to RDF: An Ontological Approach. *International Journal of Web & Semantic Technology (IJWesT)*, 6(3)

Le, D.-H. (2020). UFO: A tool for unifying biomedical ontology-based semantic similarity calculation, enrichment analysis and visualization. *IBM Research Almaden-USA*, 1-8.

Manjula, S. K., K., C. S., & Dinesh, A. U. (2012). Semantic Plagiarism Detection System Using Ontology Mapping. *Advanced Computing: An International Journal*, 3.

Manvi, B., & Sanjay, K. J. (2021). Semantic Enrichment for Non-factoid Question Answering. *Springer*, 1-15.

Marutho, D., Handaka, S. H., Wijaya, E., & Muljono. (2018). The Determination of Cluster Number at k-mean using Elbow Method and Purity Evaluation on Headline News. *2018 International Seminar on Application for Technology of Information and Communication (iSemantic)* (pp. 533-538). IEEE.

Matthias Pfaff, H. (2018). A web-based system architecture for ontology based data integration in the domain of IT. *Enterprise Information Systems*, 236-258.

Mccallum, A., Nigam, K., & Ungar, L. (2000). Efficient clustering of high-dimensional data sets with application to reference matching. *In Proceedings of the Sixth ACM SIUKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 169-178). Boston, MA, USA.

Merlin, F. J., & Ravi, L. (2017). Ontology mediation method for building multilingual ontologies. *Int. j. inf. tecnol.*, 1-9.

Muningsih, E., & Yogyakarta, A. B. (2017). Optimasi jumlah cluster k-means dengan metode elbow untuk pemetaan pelanggan. *Pros. Semin. Nas. ELINVO*, 105–114.

Patil, C., & Baidari, I. (2019). Estimating the Optimal Number of Clusters k in a Dataset Using Data Depth. *Data Science and Engineering*, 132-140.

Patric, A., & Erhard, R. (2014). Enriching ontology mappings with semantic relations. *Elsevier*, 1-18.

Pavel, S., & Jérôme, E. (2018). A Survey of Schema-Based Matching Approaches. *Springer*, 33.

Pavelt, S. E. (2020). A Survey of Schema-Based Matching Approaches. *Journal on Data Semantic*, 1-18.

Pilar, E., María, d. R.-G., Jesús, P., & Gustavo, C. (2020). An Ontology-Based Framework for Publishing and Exploiting Linked Open Data: A Use Case on Water Resources Management. *Applied Science MPDI*, 1-17.

Ronald, B., Alain, A., & Philippe, N. (May 2017). A Semantic Metadata Enrichment Software Ecosystem Based on Topic Metadata Enrichments. *International Journal of Data Mining & Knowledge Management Process (IJDKP),* Vol.7, No.3.

Sachi, N., Mohanty, G., & Nalinipriya, O. P. (2021). *Semantic and NLP-Based Retrieval from Covid-19 Ontology.* USA: Journal of Semantic Web.

Sebastian, R., & Maria, M. (2019). The Semantic Asset Administration Shell. *Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS,* 1-16.

Shqiponja, A., Vasilis, E., & Ronald, F. (2020). Ontology-Enriched Query Answering on Relational Databases. *IBM Research*, 18.

Steffen, S. (2020). *Product Taxonomy Matching in E-Commerce Environments.* Mannheim, University of Mannheim.

Syakur, M. A., Khotimah, B. K., Rochman, E. M., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. *IOP Conf. Ser. Mater. Sci. Eng*.

Tomašev, N., & Radovanović, M. (2016). Clustering evaluation in high-dimensional data. *Unsupervised Learning Algorithms*, 71–107.

Uma, M. D., & Meera, G. G. (2020). Scalable information retrieval system in semantic web by query expansion and ontological-based LSA ranking similarity measurement. *Int. J. Advanced Intelligence Paradigms*, 1-23.

Wulandari, S. (2020). Analyze K-Value Selected Method of K-Means Clustering Algorithm to Clustering Province Based on Disease Case. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 121-124.

Yongfang, W., Yangfang, T., & Yongfang, Y. (2018). Determination of Semantic Types of Tags in Social Tagging Systems. *Knowledge organization*, 1-14.

Yu, C., & Zhang, R. (2014). Research of FCM algorithm based on canopy clustering algorithm under cloud environment. *Comput. Sci.*, 316–319.

Zhang, D., Shou, Y., & Xu, J. (2017). An improved parallel K-means algorithm based on MapReduce. *Int. J. Embedded Systems*, 275–282.

Zhuoyu, H., Jowers, A., & Dehghan-Manshadi, M. (2020). Smart manufacturing and DVSM based on an Ontological approach. *Elsevier*, 1-15.

http://doi.org/10.46792/fuoyejet.v7i3.819 http://journal.engineering.fuoye.edu.ng/