Data and Metadata. 2025; 4:713
doi: 10.56294/dm2025713

ORIGINAL

A Compherence Approach to Collaborative Academic Paper's Ontology Based on Existing Linking Graph Prediction

Un enfoque de coherencia para la ontología colaborativa de artículos académicos basado en la predicción de grafos de enlace existentes

Ahmed Mahdi Abdulkadium[1], Asaad Sabah Hadi[1]

[1]Software Department, College of Information Technology, University of Babylon. Babylon, Iraq.

Cite as: Mahdi Abdulkadium A, Sabah Hadi A. A Compherence Approach to Collaborative Academic Paper's Ontology Based on Existing Linking Graph Prediction. Data and Metadata. 2025; 4:713. https://doi.org/10.56294/dm2025713

Submitted: 27-02-2024 **Revised:** 03-08-2024 **Accepted:** 15-03-2025 **Published:** 16-03-2025

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding Author: Ahmed Mahdi Abdulkadium

ABSTRACT

The current study describes the technological and methodological approach to collaborative ontology development in inter-organizational settings. It depends on formalisation of ontology development cooperation by means of an explicit editorial process, coordinating change proposals between ontology editors in a flexible manner. Added to this is the presence of novel distributed change management of ontologies style, models, and methods. We introduce the Academic Paper Citation Ontology (APCO) as an new layer-style approach to presenting ontologies towards highest independence of the underlying language of the ontologies. We also have attendant manipulation, versioning, capture, storage, and maintenance approaches and methods that exist and which rely on existing notions that are at the cutting-edge. Additionally, we provide a suite of change propagation techniques for supporting the consistency maintenance of distributed replicas of the same ontology. Finally, to increase the domain coverage of FOAF, we have extended it by extracting social interaction facts and relationships from emerging ontology. One specific problem that arises from time to time in enriching and merging ontologies is what this paper is all about: choosing which of the several ontologies available best relates to a particular piece of text associated with an input domain. Artificial Neural Networks (ANNs), more specifically their application in the research field of Natural Language Processing (NLP), are the foundation of the approach proposed. Consider calculating the ontologies' relevance score by combining neural networks and natural language processing.

Keywords: Ontology; Link Prediction; FOAF; GNN and Information Retrieval.

RESUMEN

El presente estudio describe el enfoque tecnológico y metodológico del desarrollo colaborativo de ontologías en entornos interorganizativos. Depende de la formalización de la cooperación en el desarrollo de ontologías mediante un proceso editorial explícito, coordinando las propuestas de cambio entre los editores de ontologías de forma flexible. A esto se añade la presencia de una novedosa gestión de cambios distribuida del estilo, los modelos y los métodos de las ontologías. Introducimos la Ontología de Citas de Artículos Académicos (APCO) como un nuevo enfoque de estilo de capa para presentar ontologías hacia la máxima independencia del lenguaje subyacente de las ontologías. También contamos con enfoques y métodos de manipulación, versionado, captura, almacenamiento y mantenimiento que existen y que se basan en nociones existentes que están a la vanguardia. Además, proporcionamos un conjunto de técnicas de propagación de cambios para apoyar el mantenimiento de la coherencia de réplicas distribuidas de la misma ontología. Por último, para aumentar la cobertura de dominio de FOAF, lo hemos ampliado extrayendo hechos y relaciones de interacción social de la ontología emergente. Un problema específico que surge de vez en cuando al enriquecer y fusionar ontologías es el que se plantea en este artículo: elegir cuál de las diversas ontologías disponibles se relaciona mejor con un fragmento de texto concreto asociado a un dominio de entrada. Las redes neuronales artificiales (RNA), y más concretamente su aplicación en el campo de investigación del Procesamiento del Lenguaje Natural (PLN), son la base del planteamiento propuesto. Considera el cálculo de la puntuación de relevancia de las ontologías mediante la combinación de redes neuronales y procesamiento del lenguaje natural.

Palabras clave: Ontología; Predicción de Enlaces; FOAF; GNN y Recuperación de Información.

INTRODUCTION

There are several similarities between graph networks and ontologies. Although they both use graphics to represent their knowledge, they both do it in their own unique way. Ontologies are digital, representational artifacts of subject matter. It abstracts detailed information about concepts, facts, entities, attributes, etc., and it uses logical relationships between them to represent knowledge and data context. Ontologies also furnish formal, standard, and semantic definitions, allowing data to be normalized so researchers can exchange and use similar data. Data is computable since it exists as an electronic entity, with the implication that machines can reason by executing things like automatically categorizing data and inferring novel relationships. In the social network analysis, a graph-concept network views people as nodes whose relationships (edges) represent a social or behavioral context between them. Apart from qualitative examination of such networks, statistical studies or tests (e.g., centrality, clustering, etc.) may disclose some interesting relationships about the social network of people.(1)

Ontologies are used in both mundane applications, e.g., domain-specific lexicon translation guides, and complex ones, e.g., logically founded property inference settings. In order to model a knowledge domain or conversation, they provide sets of representational primitives. Classes, relations, and their properties—that is, definitions—are the most significant representational primitives. Definition annotations of such primitives provide text information describing their applicability. Restrictions of logical consistency are placed on the application of some extremely ideal ontologies. A class can therefore be defined at least in two fashions: either by its restrictions and interrelations with other classes, or by text annotation definitions. Terms for referring to their primitives are defined on the basis of domains in domain ontologies.

The more developed systems which are based on interoperability of domain ontology should also be reconciled manually among domain ontologies. This is done to incorporate information from texts pertaining to a domain into an ontology. Automatic knowledge conceptualization techniques are less useful since they require human user intervention in order to generate semantics. Growth and merging of the ontology is hence time-consuming and costly.

Ontology selection for a particular task is another troubleshooter's area of expertise. Different ontologies can specialize in different sub-domains or handle different abstraction levels. The choice of the most suitable ontology for a specific document is part of the analysis and enhancement of the text's FAIRness. Having a proper ontology for a specific text database and using it to categorize data presented in the text would make it quicker. It is also easier to link the provided text database with the ontology's already linked data. To compare easily a range of different text documents automatically, it would be helpful to have automatic text classification and associated automatic ontology selection. The articles can further be linked with relevant research data, which also come from other databases, and be compared in an understandable manner.(2)

Traditional IR approaches map the documents based on the words they hold, and not on the concepts they convey. There is linguistic mismatch between consumers and field specialists when they use various vocabulary to express similar topic. The IR systems exclude from the outcome semantically equivalent items that do not share a common lexicon between query and information content. Document representation of an IR system has been able to utilize domain knowledge to represent domain semantics and their relation. Every domain has been using ontologies for explicit domain knowledge modeling and understanding. Thanks to ontology-based document indexing techniques, vocabulary gaps are no longer a problem. These methods are disadvantageous in their heavy dependence upon the quantity and quality of existing data in the input ontology as a starting point. Being prevalent throughout the information retrieval (IR) market, neural networks promise potential resolution of problems with document indexing. Neural network algorithms exploit the semantic relations between the words in settings when they are co-located, although never employing evident semantic relational patterns. Ontologies and semantic lexicons are examples of human-generated knowledge bases specifying such patterns of semantic connections.(3)

Academic paper objective

The usage of various terms to refer to a similar meaning in documents and domain ontology will tackle the term mismatch problem. The second purpose of the proposed study is the display of ideas even when no ideas in a document have its equivalents in domain ontology. This research assumes that the challenge for exhaustive semantic indexing of academic articles into Computer Science Ontology descriptors to the semantic domain of associated ideas in an effort to propel the outcomes to those achieved by experts in an area. This research anticipates improved machine learning and ontology-driven semantic document indexing (MLOntoSDI) information retrieval system to respond to such a challenge.

METHOD

It is our goal with our Friend of a Friend with Benefits Ontology (FOAF) to try to fill in the gap between social network ontologies and the ontologies which would be used in scientific publications.

The Web uses ontology to describe individuals and social relationships. With all of its properties attached to online activity or identity, ontology is primarily interested in people's lives in virtual space: foaf:geekcode, foaf:mbox, foaf:skypeID, foaf:msnID, etc. No information on geographic location or family relationships. It gives the same kind of information about organizations or groups with similar interests in their web presence (e.g., company page). It is especially helpful in defining people on social networking websites like Facebook, Twitter, and Blogspot.

It is a method by which writers can retrieve filtered information on concerts played from other sources through integrating and customizing the academic paper ontology and Friend of A Friend (FOAF) systems for scientific paper recommendation systems in the Facebook network.

Filtering of information is done based on the area of interest specified by the writer's academic panel. In order to integrate our data with a standard ontology, this paper considers GNNAP as a semantic graph-based embedding. According to experimental results, the use of the protege framework in the searching process can speed up the processing of pertinent articles connected by graph embedding according to the concept of link prediction.

Graph Neural Network (GNN)

The link prediction task is one of a variety of machine learning tasks on graphs for which GNNs have been widely applied due to their ability to surmount most of the above limitations.(15)

An Anchor-aware Graph Neural Network combined with Information Entropy (IEA-GNN) learning method for the global position information of graph nodes. After learning the nodes' information entropy, IEA-GNN returns sets of candidate candidates of anchor.

They described how to compute any node's distance to anchor points and gave relative distance information between startup nodes. We use a nonlinear distance-weighted aggregate learning process using candidate set anchor points to retrieve the node feature data. Combining the node representation and global position information with the selected anchor points improves learning. Information entropy-based anchor point selection allows to untangle neighborhood nodes that are homogenous and emphasizes the location differences among nodes by preventing the cluster of the anchor points within the graph. Based on empirical node classification and link prediction on five datasets, IEA-GNN is superior to the baseline model.

IEA-GNN model, which is made up of three modules: node-to-anchor path calculation, anchor point selection, and node embedding calculation.

- The anchor point selection module utilize information entropy to detect anchor points and to estimate the information contribution of nodes in a graph. The graph is a sprinkling of judiciously chosen anchors that interact with other nodes for communicating informative-rich information.
- The node-to-anchor path calculation module, computes the path distance between the specified node and each anchor point within the anchor point sets according to the path calculation function. It also computes the relative position of the node and the number of common nodes that are present in relation to each anchor point.
- Node embedding calculation module, IEA-GNN makes use of a GNN encoder to combine attribute, neighborhood structure, and global position information to obtain the final embedding representation of the node.

Semantic Author and Paper Relationship RDF

Author Paper Publication RDF

A publication of one author will point to a publication of another author, forming the semantic structure of the scientific community, which has some links. Collecting a triple of authors or articles released in the same linking domain, the RDF concept will integrate triple facts.

https://doi.org/10.56294/dm2025713

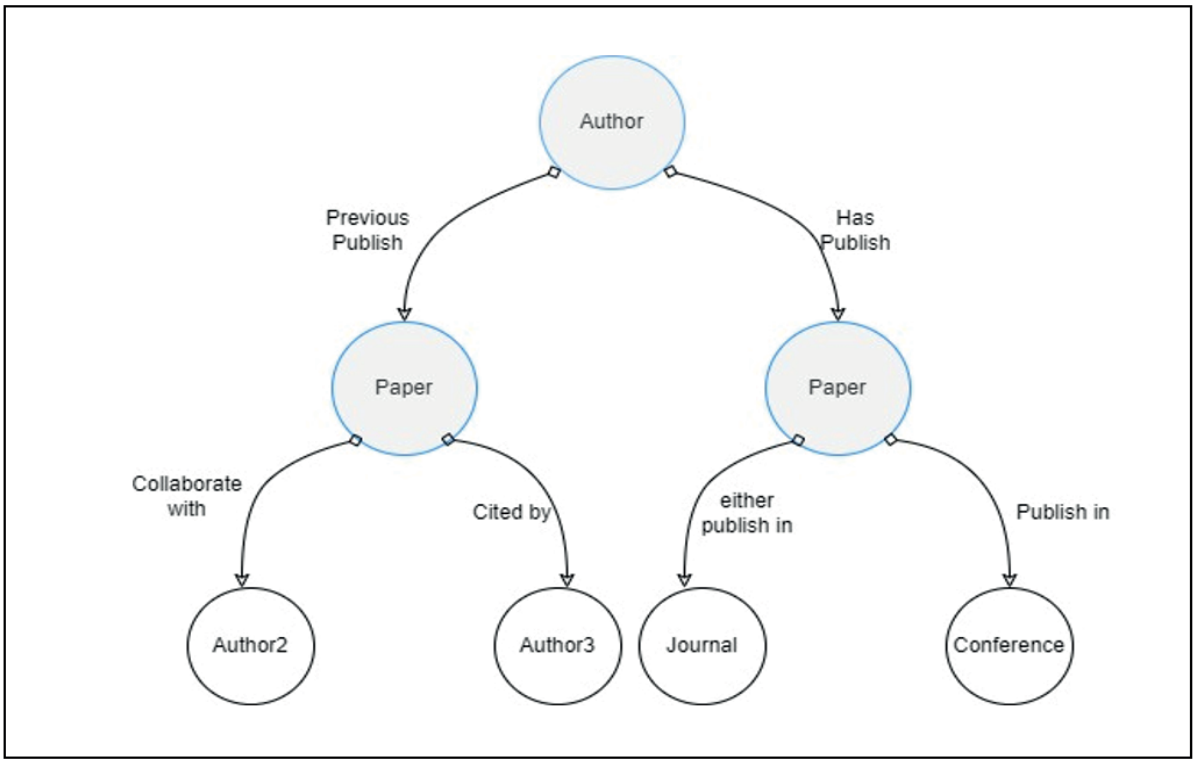

Figure 1. Author RDF

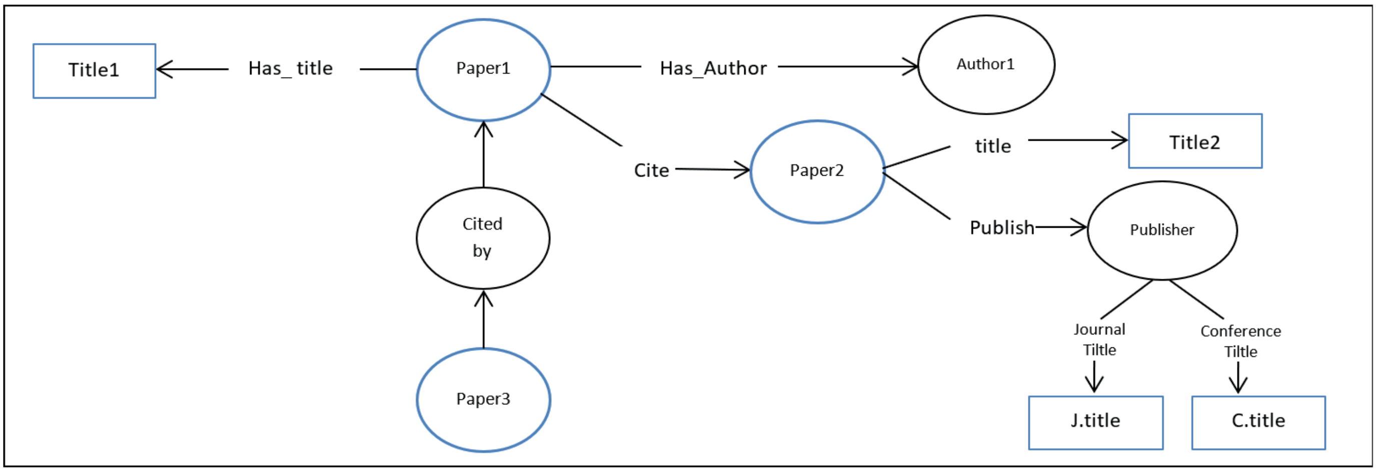

Figure 2. Author Publication RDF

Paper RDF and FOAF Relationship

Another author's piece will cite another author's work, creating numerous connections linking the scientific firm's semantic relationships. When receiving a triple of author or paper publications in the same related field, the RDF notion will include triple data.

Online social networks, or OSNs, are rapidly emerging as the hubs of human emotions. Through the monitoring of depression patients by OSN updates, Improved Depression Recovery Motivation Recommendation System (I-DRMRS) will help speed up life-saving programs. Through increased prediction accuracy, decreased False Positive (FP) rate, and improved detection of suicidal (sybil) thinking people, this work aimed to minimize the number of suicide fatalities. There are three jobs in the I-DRMRS. grouping—location-based grouping and assigning psychiatrists to each cluster is job one. Job 2: classification—using both the TensorFlow (TF) text classifier and also the well-trained TensorFlow (TF) image classifier, to classify suicidal thinking individuals on the basis of the texts and photos which they post on their OSN every day as {suicidal—'s', non-suicidal—'ns'}. Task 3: M-Module and Alarm (motivator module) FOaF—Psychiatrists from each cluster receive the output of the classifier module. Psychiatrists motivate suicidal thoughts for a duration T and observe mood changes. If, even after passing through the M-Module, the progress is not observed by psychiatrists, the alert is given to the suicidal person's Friend of A Friend (FOAF).[17]

https://doi.org/10.56294/dm2025713

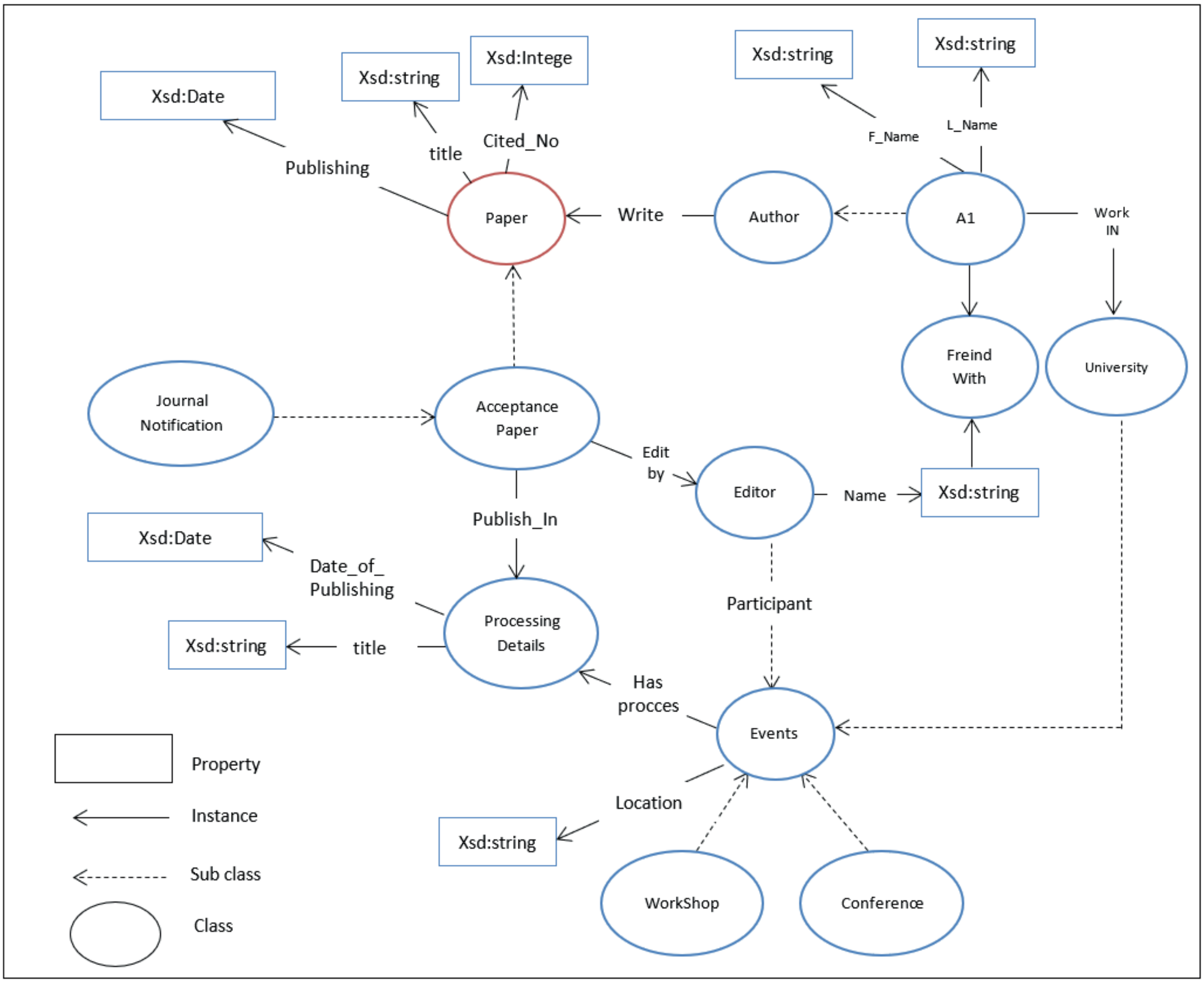

Figure 3. RDF and FOAF Relationship

Table 1. Ontology Construction Steps
Step 1. Determine the domain and scope of the ontology /*Which domain will the ontology will encompass */
Step 2. Define the classes and the class hierarchy /*In knowledge representation, a class is set of individuals or objects. */Classes provide an abstraction mechanism for grouping resources withsimilar characteristics.*/
Step 3. Arranging the classes in a taxonomic (subclass–superclass)
Step 4. Define the properties of classes /*Object attributes establish a connection between people./* /* Properties of datatypes connect people to data values. */
Step 5. Describe the slot's Fact /*Different aspects of slots can describe the kind of value, permitted values, quantity of values (cardinality), and other characteristics of the values the slot can accept/*. /*A slot's maximum number of values is its slot cardinality /*. /*Value kinds that can occupy a slot are described by the slot-value type. String, Number, Boolean, Enumerated */
Step 6. Create instances /*The last step is creating individual instances of classes */
End

Axioms and OWL Ontology by Protégé

Based on the above meta-data, the OWL ontology definition will be built using the protégé framework.

The Protégé framework and ontology editor were developed by the Stanford Center and are an open-source application are utilized to build classes, class hierarchies, variable restrictions, and the property (instance) of class-to-class interactions.

Figure 4 depicts the Academic paper citation APC ontology with classes of many writers, their individual attributes, and characteristics in order to support further data creation, association, and reasoning.

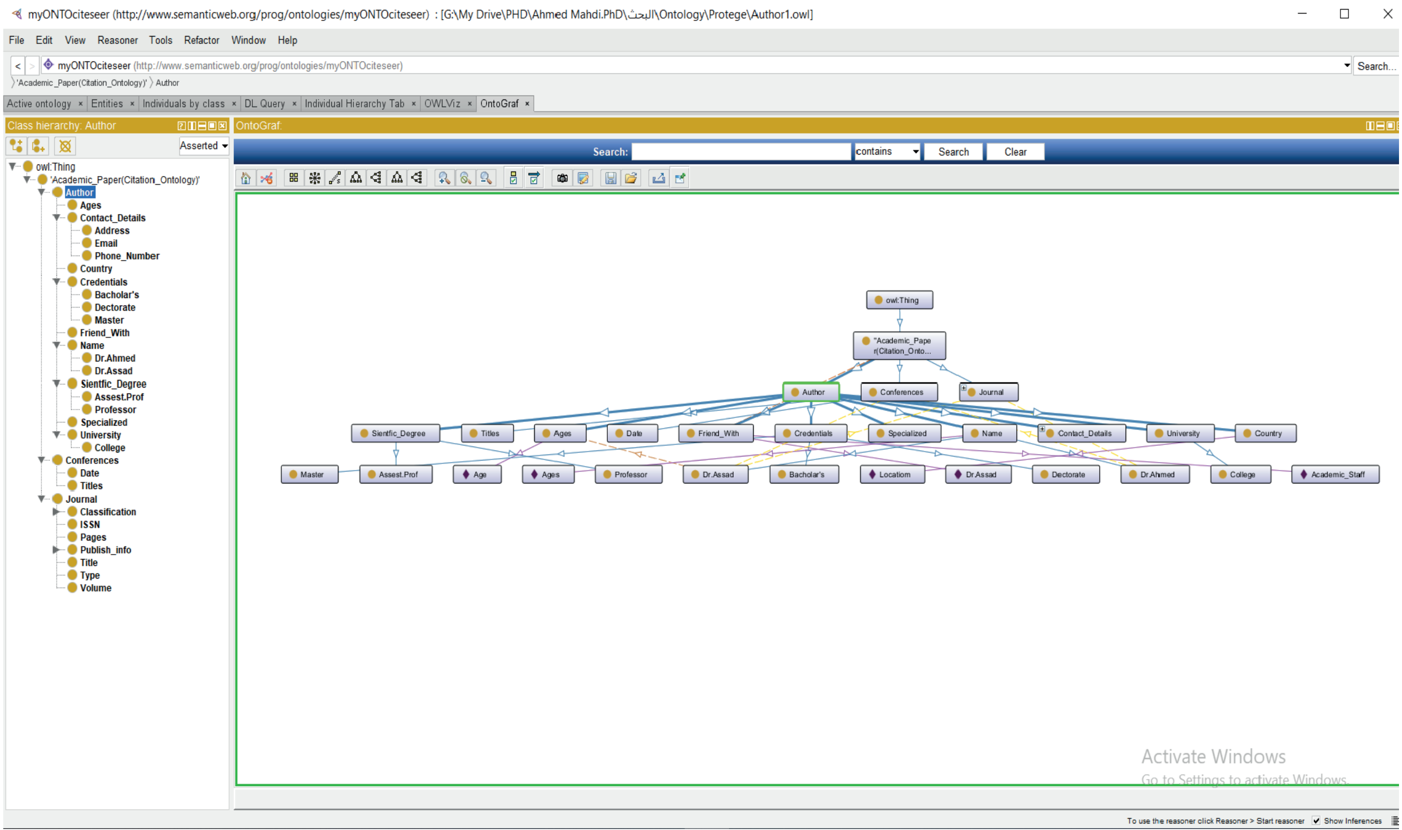

Figure 4. Academic Paper Ontology

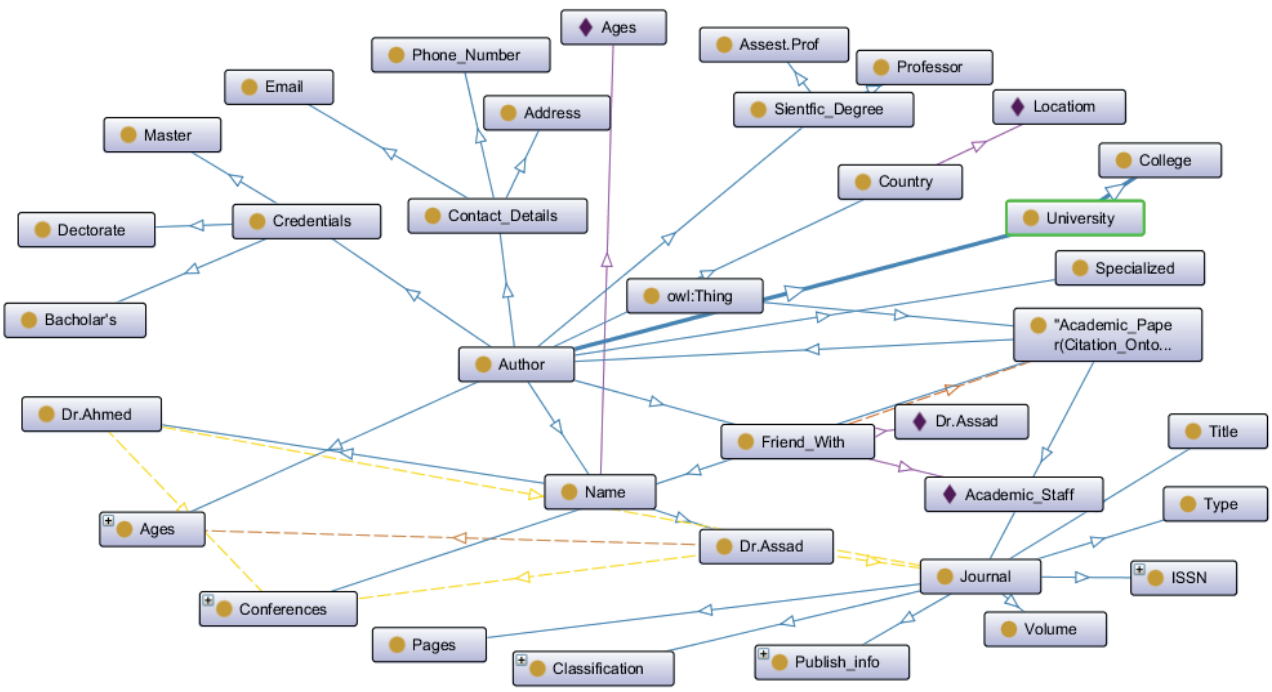

Figure 5. Hierarchy of Academic Papers

Ontology Metric

The quantity of different classes, object properties, data properties, and individuals in our ontology is indicated in the resulting APC Ontology in table 2. Explicitly enumerated in the APC Ontology, the aforementioned quantities are those including entities which are actually built-in by nature, e.g., owl: Thing.

Table 2. APC ontology Metric

Axiom	119
Logical axiom count	62
Declaration axioms count	57
Class count	39
Object property count	6
Data property count	6
Individual count	8
Annotation Property count	1

Experiments of Link Prediction (LP) Model

Table 3 displays the outcomes of the experiment. We highlighted the best baseline approaches with underlining and bolded the top outcomes.

The experimental results reveal that for LP (link prediction) tasks, Negative Link prediction of Author Relationship (NLPAR) surpassed other models. NLPAR is superior with an AUC of 0,866 % in Cora and 0,851 % in Citeeer. Compared with the best baseline method for the AP, NLPAR is 1,30 % superior in Cora and 1,72 % superior in Citeseer.

Table 3. Outcome of Link Prediction on Cora Dataset

The Methods	Info.	Cora Dataset	
		AUC	AP
Spectral-G	G	0,846	0,885
DeepWalk	G	0,831	0,850
GAE	F & G	0,910	0,920
VGAE	F & G	0,914	0,926
ARGA	F & G	0,924	0,932
ARVGA	F & G	0,924	0,926
NLPAR	F & G	0,932	0,940

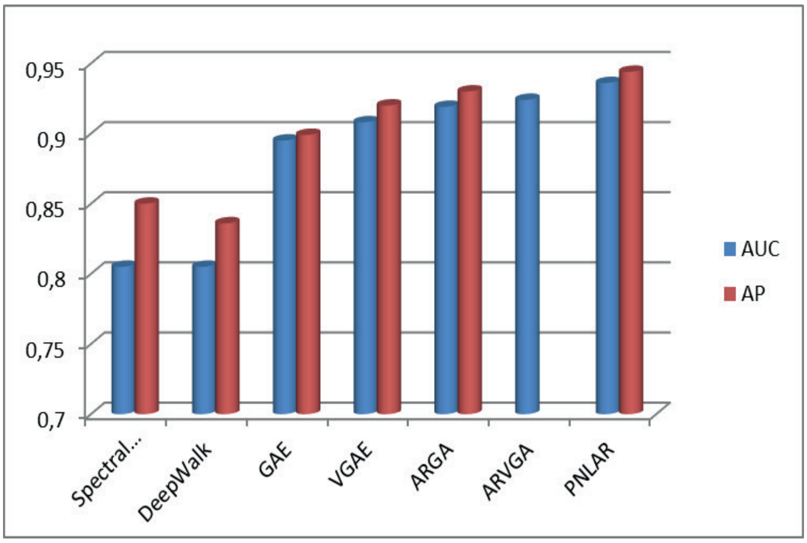

Figure 6. Cora dataset LP visualization

Our proposed model demonstrates the improved expressiveness and performance in two tasks using downstream graphs: link prediction, clustering, and attributed graph embedding based on features the BERT pre-trained model would be able to capture. Based on features the BERT pre-trained model would be able to capture.

Table 4. Connectivity Prediction Outcome on Citeseer Dataset

The methods	Information	Citeseer	
		AUC	AP
Spectral-G	G	0,805	0,850
DeepWalk	G	0,805	0,836
GAE	F & G	0,895	0,899
VGAE	F & G	0,908	0,920
ARGA	F & G	0,919	0,930
ARVGA	F & G	0,924	0,930
NLPAR	F & G	0,936	0,944

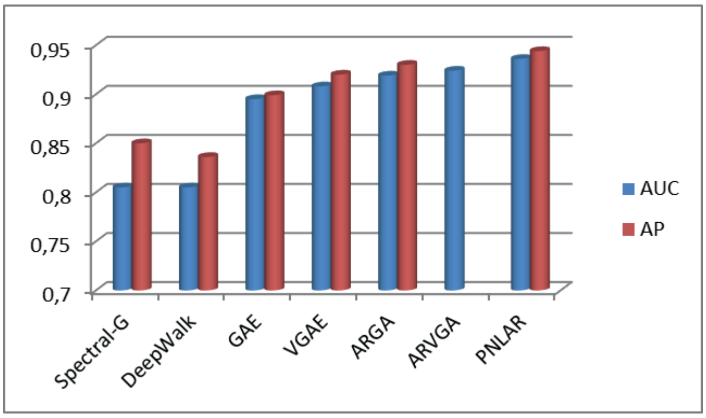

Figure 7. Citeseer dataset LP visualization

CONCLUSIONS

The use of FOAF relation and NLPAR concept can significantly optimize querying results and improve information gathering and retrieval. By utilizing FOAF relation, which is a well-defined approach for describing social relationships, we can extract meaningful information from social networks and use it to enhance the relevance and accuracy of search results. The FOAF relation helps to identify connections and similarities between different individuals, allowing for more targeted and personalized search results.

Furthermore, the application of the Graph Neural Network (GNN) concept can effectively capture the complex relationships and dependencies between different nodes in a graph. GNNs can analyze the local and global features of the nodes, enabling them to learn and identify patterns in the data, which can be used to further optimize the querying results.

Overall, the combination of FOAF relation and a NLPAR concept provides a powerful approach to improve the quality and efficiency of information gathering and retrieval. By leveraging these techniques, we can extract more relevant and accurate information from social networks, which can be beneficial in various domains, including social media, e-commerce, and healthcare, among others.

BIBLIOGRAPHIC REFERENCES

1. Amith, Muhammad "Tuan & Fujimoto, Kayo & Mauldin, Rebecca & Tao, Cui. (2020). Friend of a Friend with Benefits ontology (FOAF+): extending a social network ontology for public health. BMC Medical Informatics and Decision Making. 20. 10.1186/s12911-020-01287-8.

2. Korel, Lukáš, Uladzislau Yorsh, Alexander S. Behr, Norbert Kockmann, and Martin Holeňa. 2023. "Text-to-Ontology Mapping via Natural Language Processing with Application to Search for Relevant Ontologies in Catalysis" Computers 12, no. 1: 14. https://doi.org/10.3390/computers12010014

3. Sharma, A., & Kumar, S. (2023). Machine learning and ontology-based novel semantic document indexing for information retrieval. Computers & Industrial Engineering, 176, 108940.

4. Adrian Mocan and Adrian Mocan, "An Ontology-Based Data Mediation Framework for Semantic Environments."

5. H. Y. Husni, Yeni Kustiyahningsih, Fika Hastarita Rachman, Eka Mala Sari Rochman, "Query expansion using pseudo relevance feedback based on the bahasa version of the wikipedia dataset," AIP Conf. Proc., vol. 2679, no. 1, 2023.

6. Xiaojuan Zhang, "Improving personalised query reformulation with embeddings," SAGE JOURNALS, vol. 48, no. 4, 2021.

7. K. X. Chong Zhang, He Zhu, Xingyu Peng, Junran Wu, "Hierarchical information matters: Text classification via tree based graph neural network." pp. 950–959, 2022.

8. X. F. Yonghao Liu, Renchu Guan, Fausto Giunchiglia, Yanchun Liang, "Deep Attention Diffusion Graph Neural Networks for Text Classificatio." pp. 8142–8152, 2021.

9. K. Blagec, A. Barbosa-Silva, and S. O. and M. Samwald, "a curated, ontology-based, large- scale knowledge graph of artificial intelligence tasks and benchmarks." 2022.

10. S. Jain, S. K.R., and R. Jindal, "A fuzzy ontology framework in information retrieval using semantic query expansion." 2021.

11. Xia, L., Huang, C., Xu, Y., Dai, P., Zhang, X., Yang, H., ... & Bo, L. (2021, May). Knowledge-enhanced hierarchical graph transformer network for multi-behavior recommendation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 5, pp. 4486-4493).

12. H. Bu, J. Xia, and Q. Wu, "A Dynamic Heterogeneous Information Network Embedding Method Based on Meta-Path and Improved Rotate Model," 2022.

13. H. Hong, H. Guo, Y. Lin, X. Yang, Z. Li, and J. Ye, "An Attention-Based Graph Neural Network for Heterogeneous Structural Learning Huiting," 2020.

14. B. Wang, L. Cheng, J. Sheng, Z. Hou, and Y. Chang, "Graph convolutional networks fusing motif - structure information," Sci. Rep., pp. 1–12, 2022, doi: 10.1038/s41598-022-13277-z.

15. X. Liu, X. Li, G. Fiumara, and P. De Meo, "Link prediction approach combined graph neural network with capsule network," Expert Syst. Appl., vol. 212, no. August 2021, p. 118737, 2023, doi: 10.1016/j.eswa.2022.118737.

16. P. Zhang, J. Chen, C. Che, L. Zhang, B. Jin, and Y. Zhu, "IEA-GNN : Anchor-aware graph neural network fused with information entropy for node classification and link prediction," Inf. Sci. (Ny)., vol. 634, no. November 2022, pp. 665–676, 2023, doi: 10.1016/j.ins.2023.03.022.

17. Nedunchezhian, P., Mahalingam, M. The Improved Depression Recovery Motivation Recommendation System (I-DRMRS) in Online Social Networks. SN COMPUT. SCI. 3, 166 (2022). https://doi.org/10.1007/s42979-022-01047-7.

18. Sun, X., & Zhang, L. (2022). Multi-order nearest neighbor prediction for recommendation systems. Digital Signal Processing, 127, 103540.

19. Guo, F., Zhou, W., Wang, Z., Ju, C., Ji, S., & Lu, Q. (2023). A link prediction method based on topological nearest-neighbors similarity in directed networks. Journal of Computational Science, 69, 102002.

FINANCING
No financing.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION
Data curation: Ahmed Mahdi Abdulkadium, Asaad Sabah Hadi.
Methodology: Ahmed Mahdi Abdulkadium, Asaad Sabah Hadi.
Software: Ahmed Mahdi Abdulkadium, Asaad Sabah Hadi.
Drafting - original draft: Ahmed Mahdi Abdulkadium, Asaad Sabah Hadi.
Writing - proofreading and editing: Ahmed Mahdi Abdulkadium, Asaad Sabah Hadi.