A Legacy ERP System Integration Framework based on Ontology Learning

Chuangtao Ma and Bálint Molnár
Faculty of Informatics, Eötvös Loránd University, Budapest 1117, Hungary

Keywords: Integration Framework, Ontology Learning, Legacy ERP Systems, Data Integration.

Abstract: In the past decades, there are various legacy ERP systems that exist in different departments or sub-organizations within the enterprise. The majority of the legacy ERP systems are heterogeneous systems, that may be developed by different software companies under different development framework, which create a big challenge for organizations to develop and implement centralized and integrated management systems based on their existing legacy ERP systems to respond the dynamic business environment with agility. Ontologies are viewed as an effective technology to integrate different data from multiple heterogeneous sources, the ontology learning methods were proposed to achieve (semi-)automated construction of ontologies. This paper proposes a general framework for legacy ERP system integration based on ontology learning to tackle this challenge. Initially, the related literature is reviewed from the perspective of system integration and ontology learning, then an integration framework based on ontology learning is given, and the basic workflow and ontology learning process are analysed and illustrated.

1 INTRODUCTION

The terminology of the legacy information system was proposed by Brodie (Brodie, M.L., 1993) to describe the class of information systems, which are developed a few years ago and their technology isn't the most modern one, but these systems also run normally in organizations to provide the management and decision support currently. Enterprise Resource Planning (ERP) system are developed as a critical assistant tool to provide management decision support and optimal solution for enterprise manager in the past decades. The efficiency of the management and decision was improved by ERP systems significantly that promote the development of the enterprise. Recently, by the growth and expansion of the business, an increasing enterprise plan to, or already have achieved to update and to upgrade their ERP system. Hence, various legacy ERP systems is emerging in different organizations and departments.

The majority of the legacy ERP systems are heterogeneous systems that may be developed by different software companies under different development frameworks, so that fact it creates a big challenge for organizations to implement the integrated business intelligence system based on various legacy ERP information systems, and to achieve collaborative decision making in the process of management. Because of the development and great progress in information systems technology (e.g., service-oriented system architecture, business intelligence, etc), there is a trend that many enterprises plan to develop and implement business intelligence system based on their existing legacy ERP systems to respond to the dynamic business environment quickly and effectively. Therefore, the problems of how to reconcile and integrate these heterogeneous legacy ERP systems efficiently and effectively became an urgent task that should be researched and resolved.

Ontologies are viewed as an effective technology to integrate different data from multiple heterogeneous sources (Das, M., et al, 2015), and have been adopted to solve the problems of data heterogeneity. The results of the integration are determined by the quality of the ontology largely, while the quality of the ontology is limited by the experience of the ontology experts. The process of constructing ontology is a task that requires a lot of time and effort, hence it is practically impossible for ontology experts to construct various domain ontology manually, and implement the integration of various legacy ERP systems on time. Ontology learning framework (OLF) was proposed by Maedche (Maedche, A., et al, 2001) to provide a (semi-)

Ma, C. and Molnár, B.
A Legacy ERP System Integration Framework based on Ontology Learning.
DOI: 10.5220/0007740602310237
In *Proceedings of the 21st International Conference on Enterprise Information Systems (ICEIS 2019)*, pages 231-237
ISBN: 978-989-758-372-8

automated tools and method for ontology modelling, and to achieve the integration of various data for semantic Web.

In this paper, a legacy ERP system integration framework based on ontology learning is proposed to achieve the integration of the various legacy ERP systems effectively and efficiently. The structure of this paper is organized as follows. Section 2 gives a brief description and discussion of the related work, and illustrate the motivation and goals of this paper. Section 3 proposes the general framework for legacy ERP system integration based on ontology learning, and the detailed process of legacy ERP system integration based on ontology learning is illustrated. The conclusion and future direction related to this topic are reported in Section 4.

2 RELATED WORK

2.1 Legacy ERP System Integration

Legacy ERP systems play a significant role in enterprise's daily management, e.g., decision making, production planning, plan executing, cost accounting, and so forth. As we all know, the most prominent elements of ERP systems are supply chain management (SCM) systems and customer relationship management (CRM) systems. However, the large amounts of data are relevant for both enterprise resource planning and supply chain management, data is stored redundantly (Weske, M., 2012). Worse, due to the traits of the legacy ERP systems, including fragile, obsolete, and their interface are unfriendly, the problems of the legacy ERP system integration are becoming more and more serious. Especially, it is crucial for enterprises to develop and integrate a customized ERP system to response the current dynamic business environment (Tommi M.K., et al, 2014). For legacy ERP system integration, there are three levels include, business process integration, data integration, and system integration.

2.1.1 Data Integration

Data integration is the fundamental level in the hierarchical framework for legacy ERP system integration. In generally, ERP system integration could be achieved by the integration framework based on data integration (Huang X.X, et al, 2013). XML is viewed as a useful technology and tool are being adopted to achieve data integration in business information systems (Lampathaki, F., et al, 2009). Meanwhile, technology and tools of data integration based on extract-transform-load (ETL) are emerging, so that a data integration framework was constructed based on ETL tools (Dayal, U., et al, 2009) in order to achieve the data integration by extracting data from distributed database systems. By the development of information technology, a data integration framework based on ontology was proposed (Lv, Y., et al, 2016) to achieve the data integration in wide range. To improve the user friendliness of system interfaces and to integrate the heterogeneous data at semantic level, a lot of linked data were created. However, because of the ontologies are used to standardize the linked-data, and the ontologies underlying linked data are often non-interoperable (He, Y., et al, 2018), which bring a new challenge to integrate the linked-data system.

2.1.2 Business Process Integration

Business process integration is a critical part in integration framework for legacy ERP systems since that is a precondition for ERP system integration. In essence, the business process integration is strongly and intimately connected to workflow integration, hence, a business process framework was proposed based on analysis of the workflow (Kobayashi, T., et al, 2003).

There is a proposal for an integration framework that is workflow oriented, which aims at business integration and automation. The framework is grounded in event process chain approach that applied for ERP systems. This framework was designed to encompass a model that is structure oriented and dedicated to business process applications and workflows (Samaranayake, P., et al, 2006).

In order to clarify the business process integration within the ERP systems, a systematic architecture for business process integration was built to give an overview of the business process integration (Magal, S.R., 2011). Recently, an ontology augmenting XBRL (*eXtensible Business Reporting Language*) extension model and a matching framework based on semantic web technology and augmented ontology were designed to achieve the process integration of financial analysis at semantic level within ERP systems (Bai, L., et al, 2018).

2.1.3 System Integration

System integration is the top-level integration architecture of legacy ERP systems, its objective is to realize centralized management and decision making. System integration provides an approach for effective

integration among different heterogeneous legacy ERP systems. In early research, a prototype of multi-agent enterprise resource planning system that employs software agent theory to achieve ERP system integration (Lea, B.R, et al, 2005). The progress of the cloud computing has created opportunity, a collaborative manufacturing networks was devised for cloud computing platform to achieve the collaboration between of supply chain system and MES (*Manufacturing Excitation System*, *MES*) efficiently (Govindarajan, N., et al, 2017). In practice, an increasing number of tools and platforms were developed to accomplish integration along with modernization of legacy ERP systems. In industrial application fields, Mule ESB (*Enterprise Service Bus*, *ESB*) is a useful system integration component that is integrated by Anypoint platform (Riives, J., et al, 2012). This platform could integrate both the web-based modern ERP systems and legacy systems seamlessly by APIs (*Application Program Interface*, *API*) and graphical data mapping and transformation technology for legacy data. Nevertheless, the efficiency and quality of the integration are determined by the experience of system administrator.

2.2 Ontology Learning

Ontology learning was proposed and considered as a (semi-) automated tool to acquire and to extract knowledge, to identify ontologies, to construct ontologies, to recognize relationships among ontologies, and so forth. This approach made it possible to introduce the methods and algorithms of machine learning into the discipline of ontology engineering (Maedche, A., et al, 2001). The previous researches on ontology learning were focused on knowledge extraction automatically from text database (Lehmann, J., et al, 2014). Initially, ontology learning was applied on areas where *Natural Language Processing* (*NLP*) plays an important role. The fields of ontology learning are expanding gradually from knowledge extraction to semantic matching and ontology mapping accompanied the progress of machine learning.

Ontology learning frameworks were built to provide a useful tool for the purpose of ontology learning. A flexible ontology learning framework was proposed to generate and evaluate an ontology (semi-)automatically through an integrated tool-suite that exploiting ontology learning process (Gacitua, R., et al, 2008). Through advancements of ontology learning technology, frameworks have been enriched greatly so that a framework can include functions for learning, of elements – that could be either lexical or pieces of ontological knowledge –, of methods, and of evaluation of results (Shamsfard, M., et al, 2003).

Nowadays, in big data era, there are various representations of knowledge and dynamic relationships within different domain ontologies, which increase the complexity of domain ontology and bring a new challenge for ontology learning. Furthermore, the research topics of ontology learning are focused on the improvement of learning methods and algorithms. For instance, an approach for optimizing ontology learning frameworks is founded on the method of seeking for near-optimal input weights was proposed to integrate multiple and heterogeneous evidence sources (Wohlgenannt, G., et al, 2015). In addition, a domain ontology learning method based on LDA (*Latent Dirichlet Allocation*, *LDA*) model was proposed to improve the capability of knowledge representation for ontology content (Hong, W., et al, 2017). Similarly, a partial multi-dividing ontology algorithm was proposed to optimize the ontology learning model for improving the efficiency of ontology learning (Gao, W., et al, 2018). However, limited number of studies regards to ontology learning are available on *NLP*, moreover rigorous studies for system integration based on ontology learning have not been attempted to. It can be seen from the aforementioned studies that research about legacy ERP system integration and ontology learning frameworks mainly focuses on data integration, business process integration, and system integration respectively. However, the results of integration are restricted to descriptions of common properties at simple syntactic level, thereby it can't meet the requirement of heterogeneous business information system integration within the dynamic business environment. While the research about ontology learning framework mainly focuses on the improvement of learning algorithms and methods, the fields of application for ontology learning are constrained on *NLP*, e.g., semantic analysis, knowledge acquisition, content recommendation, and so forth. To the best of our knowledge, there was no attempt to explore the potential of Information System Integration based on ontology learning.

The motivation of this paper is to solve the problems of various legacy ERP system integration in the dynamic business environment from the perspective of ontology learning, and the objective is to build a general integration framework for the legacy ERP systems based on ontology learning.

3 INTEGRATION FRAMEWORKS BASED ON ONTOLOGY LEARNING

3.1 Architecture for Legacy ERP System Integration

For legacy ERP systems, data heterogeneity issues will occur if a logical data item address is stored multiple times in different sub-systems. Worse, the majority heterogeneity of legacy ERP systems usually caused by the various semantics of the attributes, thereby these semantic differences need to be eliminated in the process of ERP system integration. Information systems integration based on ontology learning does not only realize the system integration at syntax level to meet the integrated requirement for complex heterogeneous system in dynamic business environment, but it also improve automatically the quality and efficiency of the information system integration at semantic level by ontology learning and mapping.

In order to construct integration frameworks based on ontology learning, the architecture for legacy ERP system integration should be analysed and considered. Therefore, an architecture for legacy ERP system integration was illustrated in Figure 1.

As shown in Figure 1, the legacy ERP system integration within different sub-organizations was analysed. In this integration architecture for legacy ERP systems, data integration was emphasized, because it bridges business process integration and information system integration. Meanwhile, there are many-to-many mapping of relationships between business processes and database entities. Thereby, a conclusion can be drawn that data integration play a vital role in legacy ERP system integration. Thus, data integration was selected to illustrate the integration framework based on ontology learning.

3.2 Integration Framework based on Ontology Learning

For legacy ERP systems, there are various data collections in heterogeneous databases that need to be integrated. These data collections, namely, database tables in relational database management systems, were created under different kinds of database management systems that exist various data organization philosophies, so this fact causes a huge obstacle for integrating. To solve this problem, Extensible Mark-up Language (XML) was developed to provide an opportunity for a unified description of heterogenous data in a semi-structured way, and then

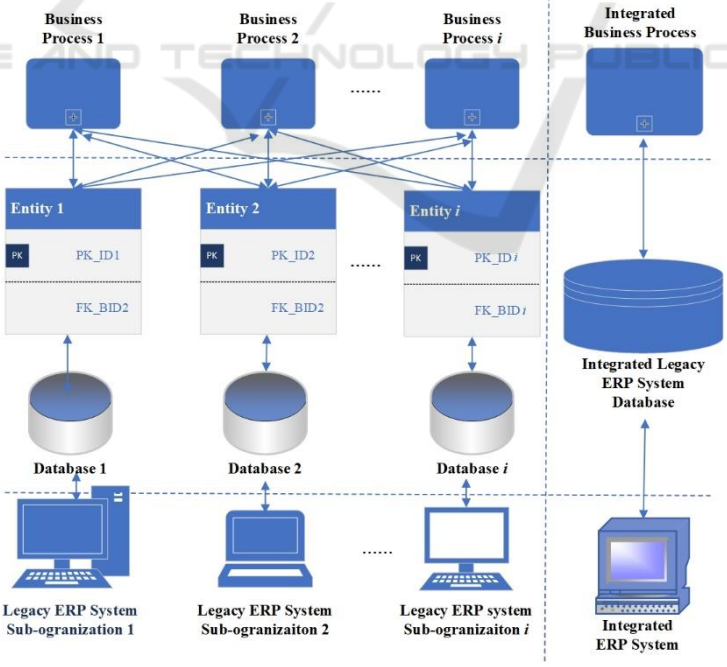

Figure 1: Architecture for legacy ERP system integration.

data can be transformed and loaded into heterogeneous database structures. Similarly, there are several heterogeneous data collections are represented of the same entities, while the syntactic description differs widely. The typical example of this phenomenon is that there are various database field names and types in different legacy ERP systems for the similar data items of the same entities.

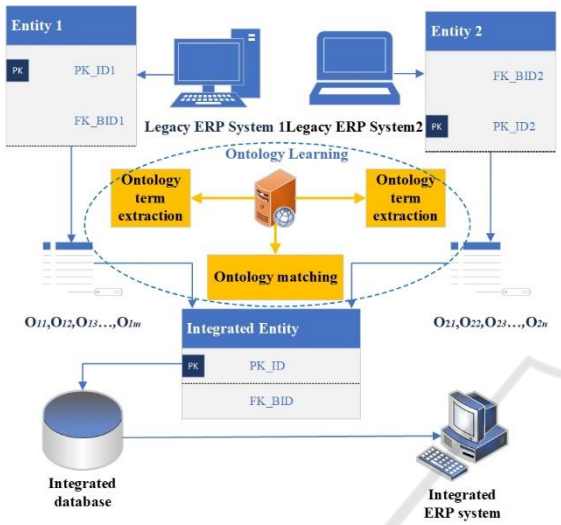

Figure 2: Legacy ERP system integration framework based on ontology learning.

In the above described complex situation, extensible mark-up languages are incapable of accomplishing database integration effectively. To tackle this problem, a legacy ERP system integration framework based on ontology learning is proposed, and then the schematic diagram of legacy ERP system integration framework is given in Figure 2.

In Figure 2, the basic workflows of the legacy ERP system integration based on ontology learning are illustrated. The workflows of the legacy ERP system integration based on ontology learning could be summarized as follows: (i) Identify the entities from legacy ERP systems database tables; (ii) Extract the ontology terms from database scripts documents; (iii) Match the different ontology terms of different entities; (iv) Map onto an integrated entity based on the matching results of the ontology terms; (v) Create and develop the integrated database schema and its implementation.

3.3 Ontology Learning for Legacy ERP System Integration

To depict the roles of ontology learning in integration framework for legacy ERP systems proposed in Section 3.2, the detailed process of ontology learning for legacy ERP system integration is analysed and illustrated in Figure 3.

It is visible from Figure 3 that the input of ontology learning is text document sets, e.g., database scripts, which contain system entities, and the output of the ontology learning is an integrated data table. The key steps of ontology learning for legacy ERP system integration is described as follows.

Step 1 Pre-process of Entities Text Document: There are various formats and naming conventions among text documents for different entities, e.g. date formats, fields name, and so forth, which produces a huge obstacle for ontology learning. Hence, the text document sets and text corpora should be pre-processed (semi-) automatically at initial phase of ontology learning for achieving the ontology con-

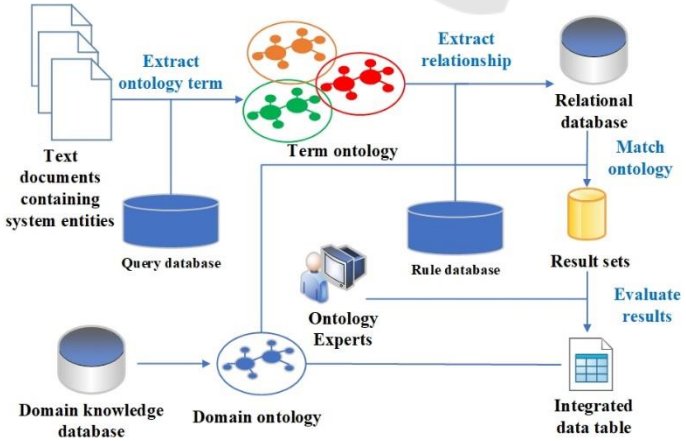

Figure 3: Ontology learning process for integration of the legacy ERP systems.

struction and mapping based on linguistic methods. The methods for text document pre-processing include clustering, dimension reduction, and linguistic processing. Due to the clustering method is mainly used to extract the relations among data, thus text notation, parsing, and lemmatization could be adopted to pre-process the text document.

Step 2 Extract Ontology Terms from the Text Document: It is a critical step to obtain ontology terms for ontology learning, thereby, ontology terms for representing knowledge, concepts within ontology should be extracted after pre-processing text documents. In this step, the natural language processing techniques, such as semantic analysis, co-occurrence analysis, and hyponymy detection, could be applied to carry out term extraction from the text documents and database scripts. Due to the results of the ontology learning are largely determined by the quality of the ontology terms. Thus, it is necessary for us to consider how to improve the term quality and the degree of automation in the process of ontology term extraction.

Step 3 Recognize and Extract the Relationships Among Ontology Terms: Relationships among ontology terms are essential for ontology mapping and matching, in this stage, all ontology terms will be analysed with the help of association rule-base and the domain ontology to recognize and extract the relationships among ontology terms. The method for recognition and extraction of relationships among ontology terms could be classified in natural language processing approaches and statistical approaches. The method based on *NLP* techniques include dependency analysis, lexicon for syntactic pattern, while method based on statistical techniques include hierarchical clustering, and association rule mining (Asim, M.N., et al, 2018). The typical relationship between different ontology terms includes: X is similar with Y, X is equal with Y, X is irrelevant considering Y, and so forth.

Step 4 Matching Ontology Terms based on its Relationships: Matching the different ontology terms according to its relationships is conducive to integrate the different legacy ERP systems at data integration level. On performing the ontology terms matching based on their relationships, the axioms for relationship among the ontology terms should be firstly extracted by applying inductive logic programming. Then sets of relationships for ontology terms will be obtained, these relationships will provide a constructive advices and basis to support the integrate process of the different database entities within the different legacy ERP systems.

Step 5 Evaluate the Results of Matching Ontology Terms: It is necessary to check the consistency of ontology terms and then to evaluate the results of matching ontology terms before the process of integrating the different database entities within heterogeneous legacy ERP systems. The method for evaluating the results of matching ontology terms could be classified into four types: gold standard based, application based, data-driven based and manual evaluation method. Meanwhile, ontology learning is a multi-level process, which increases the difficulty in the process of results evaluation. Therefore, the domain ontology and knowledge of ontology experts should be introduced into the process of results evaluation.

4 CONCLUSIONS

In this paper, ontology learning technology was employed to solve the integration problems of legacy ERP systems. And then, a general integration framework based on ontology learning is presented to integrate various legacy ERP systems effectively and efficiently. Data integration was selected to demonstrate the integration process of the legacy ERP systems based on ontology learning, and the key steps of the legacy ERP system integration based on ontology learning were given.

The presented study proposed a general system integration framework based on ontology learning. However, due to current study at its preliminary stage, there are several issues that should be answered in further investigation and study. According to the current study, the process of ontology construction is a precondition of the system integration based on ontology learning. Therefore, the method of extract and construct ontology terms from legacy ERP systems database scripts should be investigated and studied further. Moreover, the problem of how to extract and learn the relationships between different ontology terms from legacy ERP systems should be considered and studied. The last but not least, the problem of how to evaluate the efficiency and check the consistency of the ontology terms in the process of legacy ERP system integration based on ontology learning also should be studied, because the consistency between different ontology terms will influence the quality of the mapping relationships, eventually, it will determine the accuracy and effectiveness of integration results of legacy ERP systems.

ACKNOWLEDGEMENTS

This work was supported by grants of the European Union co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00002) and the China Scholarship Council (201808610145).

REFERENCES

Brodie, M.L., 1993. The promise of distributed computing and the challenges of legacy information systems. *Interoperable database systems*, pp. 1-31.

Das, M., Cheng, J.C. and Law, K.H., 2015. An ontology-based web service framework for construction supply chain collaboration and management. *Engineering, Construction and Architectural Management*, 22(5), pp.551-572.

Maedche, A. and Staab, S., 2001. Ontology learning for the semantic web. *IEEE Intelligent systems*, 16(2), pp.72-79.

Weske, M., 2012. Business process management architectures. In *Business Process Management*, pp. 25-39. Springer, Berlin.

Tommi M.K., Maglyas, A., & Smolander, K., 2014. ERP System Integration: An Inter-organizational Challenge in the Dynamic Business Environment. *International Conference on Enterprise Information Systems*, pp.39-56. Springer: Lisbon.

Huang, X. X., Zhu, W., 2013. An enterprise data integration ERP system conversion system design and implementation. *Applied Mechanics and Materials*, 433-435 (5), pp. 1765-1769.

Lampathaki, F., Mouzakitis, S., Gionis, G., Charalabidis, Y. and Askounis, D., 2009. Business to business interoperability: A current review of XML data integration standards. *Computer Standards & Interfaces*, 31(6), pp.1045-1055.

Dayal, U., Castellanos, M., Simitsis, A. and Wilkinson, K., 2009. Data integration flows for business intelligence. *12th International Conference on Extending Database Technology: Advances in Database Technology*, pp. 1-11. ACM: Saint Petersburg.

Lv, Y., Ni, Y., Zhou, H. and Chen, L., 2016. Multi-level ontology integration model for business collaboration. *The International Journal of Advanced Manufacturing Technology*, 84(1-4), pp.445-451.

He, Y., Xiang, Z., Zheng, J., Lin, Y., Overton, J.A. and Ong, E., 2018. The eXtensible ontology development (XOD) principles and tool implementation to support ontology interoperability. *Journal of biomedical semantics*, 9(3), p.1-10.

Kobayashi, T., Tamaki, M. and Komoda, N., 2003. Business process integration as a solution to the implementation of supply chain management systems. *Information & Management*, 40(8), pp.769-780.

Samaranayake, P., & Chan, F. T. S., 2006. Business process integration and automation in ERP system environment: integration of applications and workflows. *The 2nd International Conference on Information Management and Business*, pp.155-162. University of Western Sydney: Sydney.

Magal, S.R. and Word, J., 2011. Integrated business processes with ERP systems. pp.315-348. Wiley Publishing, Hoboken.

Bai, L., Koveos, P. and Liu, M., 2018. Applying an ontology-augmenting XBRL model to accounting information system for business integration. *Asia-Pacific Journal of Accounting & Economics*, 25(1-2), pp.75-97.

Lea, B.R., Gupta, M.C. and Yu, W.B., 2005. A prototype multi-agent ERP system: an integrated architecture and a conceptual framework. *Technovation*, 25(4), pp.433-441.

Govindarajan, N., Ferrer, B. R., Xu, X., Nieto, A., & Lastra, J. L. M. ,2017. An approach for integrating legacy systems in the manufacturing industry. The *14th International Conference on Industrial Informatics*, pp.683–688. IEEE: Poitiers.

Riives, J., Karjust, K., Küttner, R., Lemmik, R., Koov, K. and Lavin, J., 2012. Software development platform for integrated manufacturing engineering system. T*he 8th International DAAAM Baltic Conference "Industrial Engineering*, pp. 555-560. Tallinn University of Technology: Tallinn.

Maedche, A. and Staab, S., 2001. Ontology learning for the semantic web. *IEEE Intelligent systems*, 16(2), pp.72-79.

Lehmann, J. and Voelker, J., 2014. An introduction to ontology learning. Perspectives on Ontology Learning. pp.7-14. IOS Press: Amsterdam.

Gacitua, R., Sawyer, P. and Rayson, P., 2008. A flexible framework to experiment with ontology learning techniques. *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pp. 153-166.Springer: London.

Shamsfard, M. and Barforoush, A.A., 2003. The state of the art in ontology learning: a framework for comparison. *The Knowledge Engineering Review*, 18(4), pp.293-316.

Wohlgenannt, G., Belk, S. and Rohrer, K., 2015. Optimizing ontology learning systems that use heterogeneous sources of evidence. *International Workshop on Multi-disciplinary Trends in Artificial Intelligence*, pp.137-148. Springer: Hanoi.

Hong, W., Hao, Z. and Shi, J., 2017. Research and Application on Domain Ontology Learning Method Based on LDA. *Journal of Software*, 12(4), pp.265-273.

Gao, W., Guirao, J.L., Basavanagoud, B. and Wu, J., 2018. Partial multi-dividing ontology learning algorithm. *Information Sciences*, 467, pp.35-58.

Asim,M.N., Wasim,M. and Khan,M.U.G., 2018. A survey of ontology learning techniques and applications. *Database*, 2018, pp.1-24.