2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering | 978-1-6654-1569-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/ECTIDAMTNCON51128.2021.9425757

Ontology Model for Automatic Duplication Reducing for Industrial Standard Assessment

Roungsan Chaisricharoen[1,2], Wanus Srimaharaj[1], Phakphoom Boonyanant[1], Teeravisit Laohapensaeng[1,2], Kanchit Pamanee[3]

[1]*School of Information Technology, Mae Fah Luang University*, Chiang Rai, Thailand

[2]*Computer and Communication Engineering for Capacity Building Research Center, School of Information Technology, Mae Fah Luang University*, Chiang Rai, Thailand

[3]*Epsilon Solution & Consultant Co., Ltd.*, Bangkok, Thailand

roungsan.cha@mfu.ac.th, wanus.sri10@lamduan.mfu.ac.th, teeravisit.lao@mfu.ac.th, phakphoom.boo@mfu.ac.th, kanchit@epsilonsolution.com

***Abstract*—The entrepreneurs found that data science was redundant and wasteful of management resources in several industry-standard assessments which are duplicated. Due to the lack of advanced technology professionals, the application of highly complex technology is a barrier to implementation. Ontology models and artificial intelligence are effective means of solving such problems. Therefore, this study aims to address the ontological problems that can interpret and rule out complex artificial intelligence issues and ambiguous interpretations. It allows quick and repeatable actions to be performed automatically without an administrator. The proposed ontology model reflects the requirements in the industry-standard while K-Nearest Neighbor (K-NN) algorithm separates the data groups and integrated assessment suite that minimizes the redundancy of requirements from standards.**

Keywords—Industrial Standard, Ontology, Artificial Intelligence, N-Nearest Neighbor, K-NN

I. INTRODUCTION

At present, the production of products has to be certified by the involved certification process with a more detailed and complex audit trail. The industry sector needs data acquisition in assistance programs for standard audits such as ISO, PSM AIA, HAZOP, etc. [1], which is obtainable by collecting information from various industries. However, it found that the requirements in each standard were redundant and interchangeable in key areas [2] that could be greater than 50-60%. Without reducing these redundancies when incorporating audit issues of all standards, it could be plenty of requirements since one product has to be certified by multiple standards. This is the concept of developing a redundancy check system to identify interoperable issues. It is helpful to reduce the development time of the package for customers who have requirements to support different standards for the specific industry product regarding the benefit of reducing the development cost in software package [3]. Industrial customers also benefit from significantly reducing the burden of managing standards [4]. In this regard, it found that they could apply manual data analysis to find duplicate requirements. Therefore, there is a need for researchers to help solve the problem of complex artificial intelligence and the ambiguous interpretation for a small software company whose development team consists of a small number of intermediate and advanced programmers.

II. LITERATURE REVIEW

The main disadvantage of today's standards relies on qualitative classification techniques or statistical methods. Previously, a spectral quantitative method was presented for the current industry classification along with introducing the financial markets implicit industry classification standards [5]. The result was found that the implicit classification of the market provided better results. However, it turned out that different grouping techniques offered similar classifications as a cluster.

Furthermore, an important part of the global benchmarking ability is the degree to which data is classified. The standard has been developed to classify business establishments for the collection, analysis, and dissemination of statistical data relevant to the business economy [6]. Although the national system has been around in recent years, business classification is across national borders. The United Nations Department of Statistics works with the International Monetary Fund and the World Bank as well as the International Standard Industrial Classification (ISIC) which is one of the challenging statistical efforts being undertaken by an international organization, to make national accounting data interchangeable through the standard system.

Moreover, the industry-standard classification system is the cornerstone of quantitative research in many fields. Information collected and encoded based on such systems will often provide the most complete and accessible material for use in investigating, quantifying, and mapping the industrial landscape and understanding the whole economy. The previous research used examples of the biotech industry to continually test the suitability of industrial classification systems, such as the Swedish Industrial Classification System [7]. However, due to the relatively stable classification, it indicates that researchers try to capture some new emerging industries and troubled growth. Most understanding of the dynamics of economic development comes from aggregated company data, which is often categorized and identified using the interrelated system of national and international industry classification standards.

Some research developed the concept of a national cluster template and presented a systematic method for identifying industrial standards [8]. The results were based on the obsolete Standard Industrial Classification (SIC) system for Industry classification. The results were updated by using the 1997 Benchmark Input-Output account for the United States, based on the North American Industrial Classification System (NAICS). The advantage of the developed concept is to provide more comprehensive blended templates. A given cluster template can be the foundation for a regional economic development strategy.

It shows that previous research works relied on the quantitative and statistical methods that related to some standardized organization. Though those methods provide enough efficiency for general cases of industrial standard classification, it might spend several times for operating all requirement checklists. Consequently, this study aims to design the automatic mode based on ontology and machine learning methods for reducing the redundancy of industrial standard requirements.

III. Theories and Methods

A. System Design

The conceptual framework of the system lies in the process of reducing redundant requirements (reduce repeated checklist) as part of programs to support industry standard audits defined as an integrated assessment system. The redundant requirement reduction system compiles requirements from various standards. The customer has to be reviewed and analyzed to identify the central set of requirements for all required standards by eliminating the duplicate requirements of each standard. This allows industrial customers to easily implement standardized quality controls through various devices, shown in Figure 1.

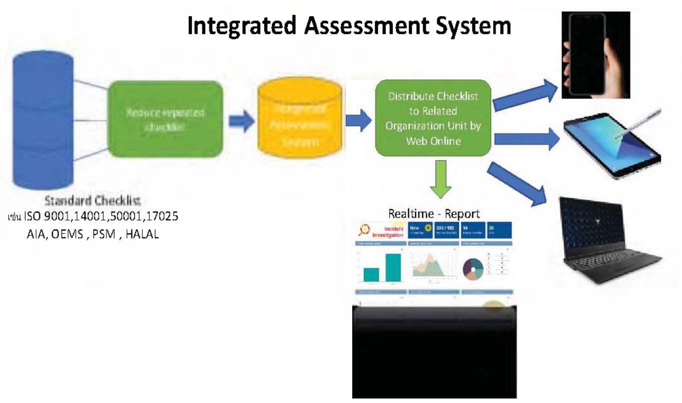

Figure1 Overview of Framework

To reduce and combine these requirements initially, this study develops a data code for various standard requirements. According to the ontology principle [9], which is the creation of a knowledge base structure or region which has a common concept and understanding, it is used to define a model within the knowledge domain to describe the interested domain. The retrieval message and the right are used for making an ontology model where these requirements can be stored in a relational database appropriately when the database has been searched. In the next step, these requirements will be classified and grouped through a data classification [10] to combine similar requirements into a single group for creating a central specification as sample data.

B. Operational Design

1. Industrial Standard Requirements

This process describes the details of the requirements of various industrial standards. It is important to summarize the points that are similar in each standard and other elements.

2. Preparation of Ontology Model Requirements

This process will determine the model of ontology for analyzing requirements in various industrial standards. It is divided into three levels as follows.

2.1 Meta-Level defines a group of data types addressing in various specifications.

2.2 Intentional-Level identifies the working principle of various elements addressing in the specification.

2.3 Extensional-Level specifies the details of the principles specified at the Intentional-Level stage.

3. Preparation of Ontological Database

In this process, it will implement the ontology model requirements to develop an ontological database, shown in Figure 2 [9]. The information can be accessed through simple sequencing and sorting.

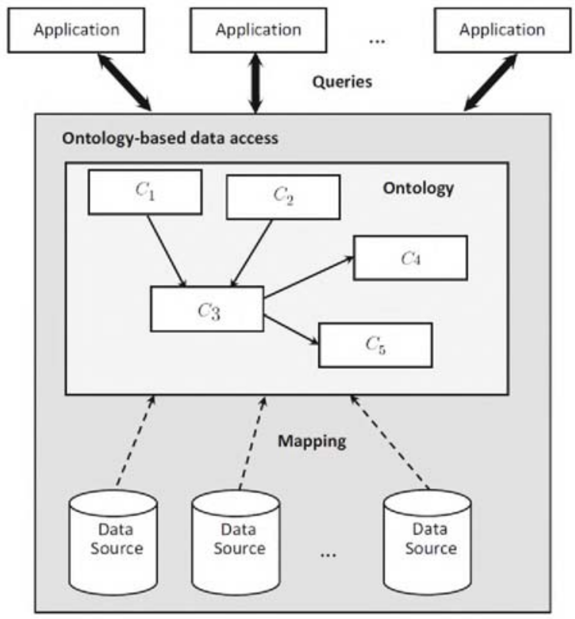

Figure 2 Ontological Database [9]

4. Design of a Separation System for Sorting of Industrial Standard Template

In this process, K-Nearest Neighbor (K-NN) algorithms, data mining techniques are used to classify the linked label that shows as a composed-of relationship from the ontology model [10]. This will result in groups of terms with similar factors distributed into precise groups of template results. It combines all data into central requirements that can be responsive to smaller requirements.

5. Performance Testing and Analysis

In this process, it is a test and correction of the overall accuracy was encountering any problems in the proposed system.

IV. Experimental Results

A. *Extraction of Necessary Information*

9,978 samples have been used in this study. It consists of variables including Template Type, Topic ID, Question Type, Active Status, Related Standard, and Result Template (Standard Template). Several variables are shown in Figure 3-6.

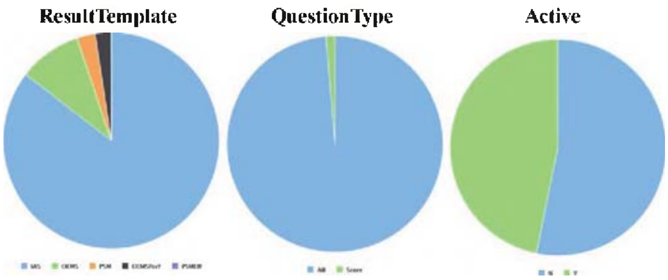

Figure 3 Variables of Result Template, Question Type, and Active Status

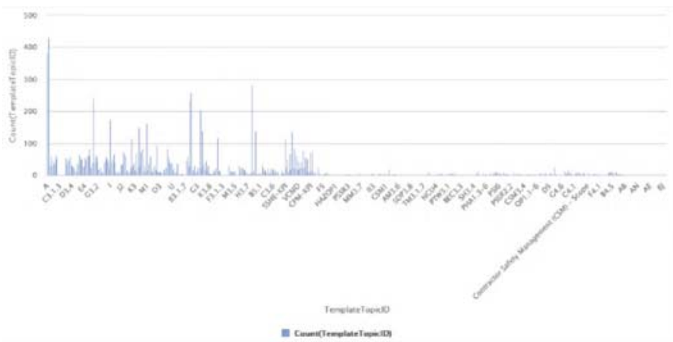

Figure 4 Variables of Template Topic ID

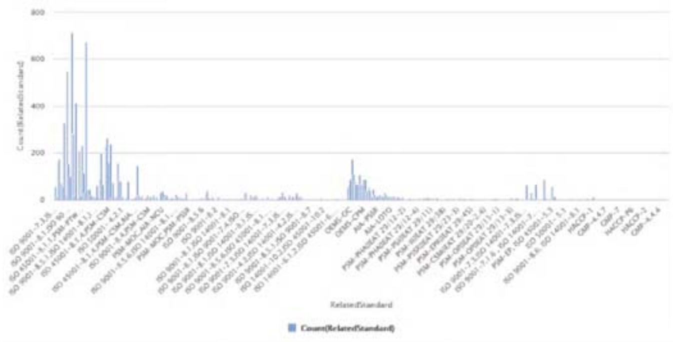

Figure 5 Variables of Related Standard

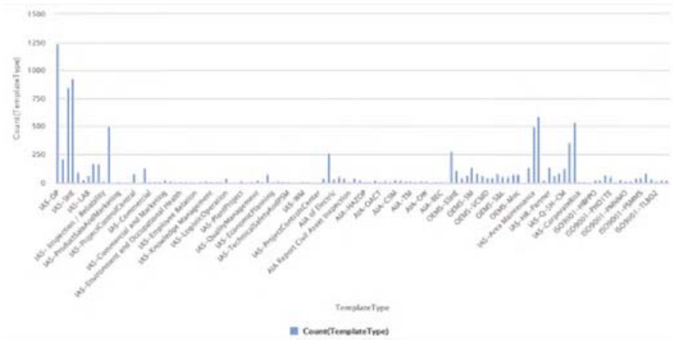

Figure 6 Variables of Template Type

B. *Ontology Model*

Ontology is the study of the types of things, which is used to summarize problems with common attributes and relationships of existing entities. The ontology includes putting attributes into categories and links all categories with each other. Moreover, it represents the types of relationships between categories in this study. The ontological model was created to reduce automatic repetition to assess industry standards based on conceptual and class assignments, shown in Figure 7.

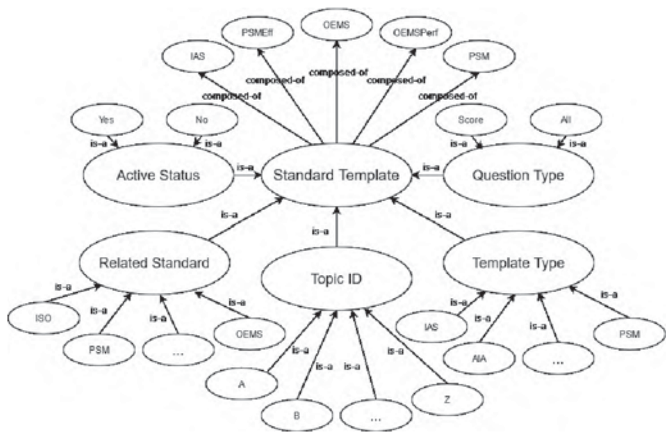

Figure 7 Spatial Ontology

The spatial configuration is linked between the five roles including Active Status, Question Type, Related Standard, Topic ID, and Template Type. There are labeled as a spatial ontology relation. These nodes are presented as linked labels where the spatial relation is a linked label that shows a composed-of relationship with the composing labels of spatial roles including IAS, PSMEff, OEMS, OEMSPerf, and PSM. Each class has the necessary characteristics to separate industry-standard requirements in the step of data classification.

C. Data Classification

K-NN algorithm is applied for classifying the linked labels and separating industrial standard requirements relying on the composed-of relationship. The model for data classification was tested for accuracy utilizing 10-fold cross-validation which divides information into 10 parts. The same amount of information is used to test the performance of a model of data classification results by the K-NN algorithm as shown in Table 1. The lowest accuracy is at a percentage of 99.93 while the lowest recall at a percentage of 97.82 for the PSM standard.

Table 1 Data Classification Results

Prediction Value	Actual Values					Accuracy
	IAS	PSMEff	OEMS	OEMSPerf	PSM	
IAS	8,525	0	0	0	6	99.93%
PSMEff	0	5	0	0	0	100%
OEMS	0	0	938	0	0	100%
OEMSPerf	0	0	0	235	0	100%
PSM	0	0	0	0	269	100%
Class Recall	100%	100%	100%	100%	97.82%	

Besides, the implementation of data classification by K-NN algorithms has percentage accuracy at 99.94 with only a 0.10 error rate. It can be concluded that the classification of data by the algorithm K-NN is highly reliable and can be applied appropriately.

V. Conclusion

Ontology model and data classification system have been used for separating industrial standard requirements in this study. The proposed ontology model deals with problems in the redundant checklist in the industry-standard where the K-NN algorithm has been used as a classification method for developing the system for each customer. As a result, it shows that the proposed model yields high-performance results that can be applied in further cases of industrial standards.

In future work, a new study of the industrial standards required by the customer will be conducted. It will regard a waste of time and the cost of work. However, it will focus on the industrial standard redundancy system which can reduce the development time efficiently.

Acknowledgments

This work is supported by the Capacity building and ExchaNge towards attaining Technological Research and modernizing Academic Learning-CENTRAL, which is funded by the Erasmus+ program (Project ID: 598914, Reference: 598914-EPP-1-2018-1-DK-EPPKA2-CBHE-JP).

References

[1] AIChE Academy (2019). Chemical Engineering Training + Education. 2019 Course Catalog.

[2] ISO (International Organization for Standardization). (2015). Standardization and Innovation: ISO-CERN Conference Proceedings, 13–14 November 2014.

[3] Kirrmann, H., & Dzung, D. (2006, June). Selecting a standard redundancy method for highly available industrial networks. In 2006 IEEE International Workshop on Factory Communication Systems (pp. 386-390). IEEE.

[4] Blind, K., Jungmittag, A., & Mangelsdorf, A. (2011). The economic benefits of standardization. DIN German Institute for Standardization. http://www. din. de/sixcms_upload/media/2896/DIN_GNN_2011_engl_akt_neu. pdf and Summary of Results: http://www. din. de/sixcms_upload/media/2896/economic_benefits_ standardization. pdf.

[5] Nagy, L., & Ormos, M. (2018). Review Of Global Industry Classification. In ECMS (pp. 66-73).

[6] Picker, A. D. (2011). International economic indicators and central banks (Vol. 392). John Wiley & Sons.

[7] Waxell, A. (2009). To capture an emerging industry: using industry standards to identify the biotech industry. International Journal of Social Research Methodology, 12(5), 441-456.

[8] Kelton, C. M., Pasquale, M. K., & Rebelein, R. P. (2008). Using the North American Industry Classification System (NAICS) to identify national industry cluster templates for applied regional analysis. Regional Studies, 42(3), 305-321.

[9] Simari GI et al., Ontology-Based Data Access Leveraging Subjective Reports, Springer; (2019)

[10] Nilsson, NJ (1965). Learning machines McGraw-Hill. New York, 19652.