System Proposal for Integrating Quality Control Data of Components of the Brazilian Oil and Gas Industry

Mario Ricardo Nascimento Marques Junior, Eder Mateus Nunes Gonçalves,
Silvia Silva da Costa Botelho, Emanuel da Silva Diaz Estrada, Danubia Bueno Espindola,
Eduardo Nunes Borges, Werner Luft Botelho, Bruno Machado Lobell and Lucas Silva Marca
Center of Computational Sciences, Federal University of Rio Grande, Rio Grande, Brazil

Keywords: Quality Control, Schema Matching, Ontology, Inspection.

Abstract: Databooks are the main documents for monitoring and validating a work in the oil industry. However, the analysis of these Databooks requires many man-hours of professionals, and will still be subject to failures during these analyzes. An approach to optimize this task would be to use an intelligent search and validation system for this documentation. Once the documents are scanned, algorithms could assist the professional to analyze these Databooks, based on automatic indexing and checks. A problem that arises from this approach is the difficulty to develop a valid conceptual model for all Databooks found and all types of components and structures subject to inspection procedures. As this modeling is done manually by professionals, this task becomes difficult, slow and often inefficient. To mitigate this problem, this work proposes a system to automatically generate a conceptual model from textual descriptions of the inspection elements, making this task simpler. This is done from the development of an ontology that describes the standardized knowledge on inspection of different types of components of the structure of works for oil platforms for the validation of inspection reports and quality certificates regarding criteria of completeness and compliance.

1 INTRODUCTION

In the manufacture and supply of equipment for the Oil Industry, Quality Management is a controlling and standardizing instrument, providing management for a Quality System, covering the organizational structure, procedures, processes and resources (Nascimento, 2010).Inspection is a critical part of the oil industry onshore and offshore due to its influence on maintaining equipment safety, reliability and availability (Rachman and Ratnayake, 2019).

All data related to the inspection of equipment is contained in Databooks. The Databook gathers all the information generated and registered, such as engineering documents and their changes, purchase orders, tests carried out, quality certificates, inspection reports, among others (Duarte, 2010). According to the degree of complexity of the component, a databook can contain thousands of pages.

The analysis of these Databooks is done by trained professionals. These professionals must carefully analyze each page of the Databook looking for abnormalities, such as the lack of an inspection report, an unsigned document or even acts of bad faith. The problem with this analysis is the high time spent, since the professional must manually search the documents for these anomalies. In addition, a Databook must be checked for completeness criteria (if it contains all mandatory documents) and compliance (if all documents are correct as to the procedures adopted). Without an adequate search system, each of these activities can occupy technicians for hours, days or even weeks of work.

One approach to optimize this task would be to use an intelligent search system. Once the documents have been scanned, algorithms could assist the professional in analyzing the Databooks by performing information registration and retrieval functions, in addition to performing completeness and compliance verification functions. However, before these data are stored in a database, they must be pre-processed, as they usually originate from a scanned document or a PDF document (Portable Document Format).

After pre-processing, the extracted data must be stored in a database. However, different companies provide different equipment and services, and con-

Marques Junior, M., Gonçalves, E., Botelho, S., Estrada, E., Espindola, D., Borges, E., Botelho, W., Lobell, B. and Marca, L.
System Proposal for Integrating Quality Control Data of Components of the Brazilian Oil and Gas Industry.
DOI: 10.5220/0010603802110215
In *Proceedings of the 18th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2021)*, pages 211-215
ISBN: 978-989-758-522-7

sequently their reports are delivered in different formats, both in content and in form. Therefore, conceptual modeling, one of the most important steps in a database (Elmasri and Navathe, 2011), should be performed for each item in the Databook, because in this step requirements are analyzed to define the system structure.

This modeling is done manually by a professional, based on his experience, in the domain of the system and certain rules. However, this task becomes difficult when large systems are developed. For the present work, a system is proposed to automatically generate a conceptual model from textual descriptions of the inspection elements, making this task simpler. This is done from the development of an ontology that describes the standardized knowledge on inspection of different types of components of the structure of works for oil platforms for the validation of inspection reports and quality certificates regarding criteria of completeness and compliance. Ontology makes it possible to deal with the diversity of types of components, documents and the correct allocation of data even under different structures. In this way, it is possible to resolve the match between the different data schemes associated with each of the inspected components and structures. Through this approach, data can be inserted into the same database.

This work is part of a larger project that aims to use pre-processing techniques and data science for the automatic verification of completeness, compliance and bad faith in records and documents of purchase, construction and assembly in context *Big Data*. The project involves the development of technologies for registration, retrieval and inference of large databases using technologies of *datalakes* and microservices.

2 FUNDAMENTALS

2.1 Component Inspection for Oil and Gas Industry

Within the scope of the oil industry in Brazil, the document that governs the guidelines related to the inspection of the manufacture of equipment for oil and gas is the ABC of the Inspection of Manufacture (Petróleo-Brasileiro-S.A, 2017).

According to this document, manufacturing inspection ”is the activity carried out for planning and execution purposes aiming to verify, at the premises of the supplier and / or sub-suppliers involved, the conformity of the equipment or materials manufactured with the contractual documents”. An asset is defined as any system, equipment, product or material that the Supplier must deliver to the customer in accordance with the contract.

2.2 Ontology

In the area of Computer Science, ontologies are used for modeling, both in database-based systems and for knowledge representation (Almeida, 2014). An ontology allows the definition of concepts (entities, objects, events, processes), emphasizing their properties, relationships and restrictions, enabling the sharing of knowledge about a given domain, which is represented by a vocabulary (de Freitas, 2017).

According to (Almeida, 2014), two main definitions for ontology in Computer Science are noted in the literature. The first refers to the use of ontological principles to model reality, that is, to provide a description of what exists and to characterize entities (Wand et al., 1999). The other definition refers to the representation of a domain in a logical language for computational purposes. In this case, an ontology consists of a set of statements expressed through a representation language, processable by inference mechanisms (Staab and Studer, 2010).

2.3 Schema Matching

Scheme matching is the problem of generating correspondences between elements of two schemas (Bernstein et al., 2011). A match is a relationship between one or more elements of a scheme and one or more elements of the other.

Figure 1 shows the simplified representation of two schemes and the matches (or correspondences) found in dotted lines. The problem proves to be challenging because the matching task may contain elements described differently (for example an acronym) and still be matched (*PurchaseOrder* and PO), elements whose semantics are partially matched (*Name* representing the full name and *FName* only the first name), elements with similar names but representing different concepts in the schemes (Address representing the billing address and ShipAddress representing the delivery address) and elements without a correspondent in the other scheme (*Product*, *BillTo* and *Customer*).

2.4 Generation Database Schema from an OWL Files

In (Panawong et al., 2016) a method is presented to automatically generate a database schema from an OWL (Ontology Web Language) file. In addition,

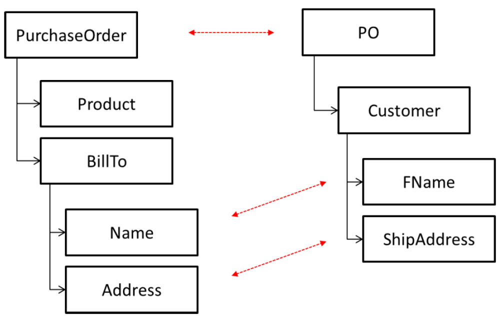

Figure 1: Two schemes and possible marriages (dotted arrows) (Rodrigues, 2013).

a mapping configuration can be created to associate the ontology and the data schema generated by the process. The method is divided into three main processes. First, the inserted OWL file is read and a syntactic analysis of its contents is performed. From this, two types of ontological unit are captured: concept and concept relation. The parsing starts from the root class. Then, the properties of the class are read and recorded. For properties, property types are all important in later processes, so that object ownership (part of) and data ownership (attribute of) are recorded separately along with their restrictions. After scrolling through all properties, all classes and subclasses are scrolled until all classes are read.

From the classes captured in the previous process, the table is generated. For each class, a table is created and its properties are generated as columns of the table. A data property is considered to be a field for entering data, while an object property must be linked to another table using a foreign key following a class constraint. A table name and column tags are generated following a label from an ontology and relationship class. The database table and columns generated together with an ontology class label are written to a file for use in a later process.

In (Afzal et al., 2016) is presented a tool called textit OWLMap, fully automatic to provide a lossless approach to transforming ontology into a relational database format. In the proposed system, initially a user will select an ontology file. Then, information is extracted about ontology constructs using the Jena API. The mapping rules are defined to guarantee a lossless transformation. Based on the mapping rules, ontological constructions are transformed into a relational database.

Mapping rules are used to transform an ontology into a relational database. According to these mapping rules, ontology classes must be transformed into tables, object type properties must be mapped into columns or tables, data type properties must be mapped into columns or tables according to the mapping rules and constraints must be stored in metadata tables.

In (Abo Zidan et al., 2019) a tool is presented that acts as a bridge between a preexisting relational database and the semantic data, which provides a fully automated approach that converts a populated ontology into a relational database without losing the ontological structure, preserving the relations of knowledge between the entities. In the Figure 2 is present the system architecture.

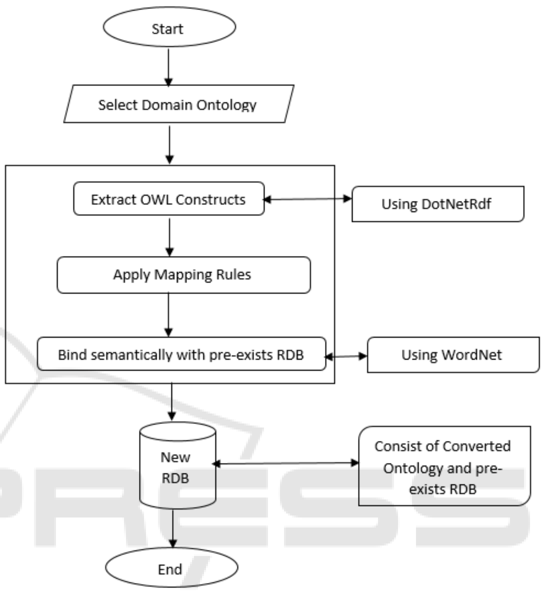

Figure 2: System architecture (Abo Zidan et al., 2019).

As in (Afzal et al., 2016), this tool also uses mapping rules to transform an ontology into a relational database. Mapping rules are necessary for converting an OWL ontology into a relational database. The extraction of classes, properties and the ontological construction to transform the ontology into a relational database is performed through the *DotNetRDF* library.

After identifying the domain ontology and the ontological constructions are extracted using the textit DotNetRDF library extracting the classes, properties with their two object types and data type and other constructions. The proposed conversion mapping rules are applied, then links will be identified between the relational database and the converted ontology using the *WordNet* library. As a result, a relational database is obtained.

2.5 JSON Schema Discovery

In recent years, *JavaScript Object Notation* (JSON) has been gaining popularity, as it provides a

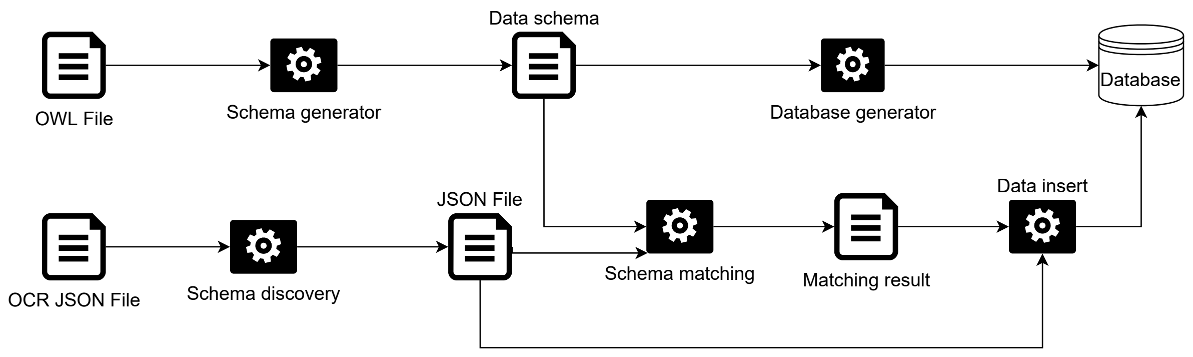

Figure 3: Proposed system.

lightweight data exchange format with great performance (Izquierdo and Cabot, 2013). JSON is a human-readable format that consists of sets of objects (that is, types or concepts) described by name/value pairs (that is, fields or attributes). JSON has no schema, that is, there is no schema that specifies the internal structure of JSON objects, instead, the schema is implicit. Schemaless data is particularly interesting in cases dealing with non-uniform data (for example, non-uniform types or custom fields) or in schema migration (Fowler, 2013), however, it can become a difficulty in data integration scenarios , where it becomes necessary to discover, at least partially, the structure to correctly process the data.

In (Izquierdo and Cabot, 2013) an approach is presented to automatically discover an implicit schema of a set of JSON documents. For this, model-oriented techniques were used to design a process in which initial schema excerpts are discovered from each individual service and then combined to obtain a composite model that describes the underlying domain model of the application, which facilitates understanding of JSON-based services. The approach was implemented in Java and made available on the hosting platform *Github* (Izquierdo, 2014).

3 PROPOSED SYSTEM

As Databooks are usually scanned documents or PDF documents, processing is necessary to extract only the relevant information.Thus, a module is responsible for performing this processing. This module performs this processing through OCR (Optical Character Recognition) algorithms, transforming the Databook into a JSON file. Assuming that the information extracted from a Databook will be contained in a JSON file, the system of Figure 3 has been proposed.

The main objective of the proposed system is to allow the integration of the diversity of types of components, documents and correct data assignment even under different structures. The proposed system has two main functionalities. The first functionality is to generate a database from an ontology. For that, from an OWL file, an algorithm should generate the database schema corresponding to the ontology. From this data schema, the database will be created. For the development of this functionality, the works reported will be maded (Afzal et al., 2016), (Panawong et al., 2016), (Abo Zidan et al., 2019). In addition, the schema generated from the ontology will be used in the schema matching.

The second functionality, on the other hand, consists of, starting from a JSON file of unknown schema, inserting this data in their respective database tables. To do this, from an unknown JSON schema, it is necessary to infer which schema this file corresponds to. To carry out this stage of such functionality, the work presented by (Izquierdo and Cabot, 2013) will be taken as a reference, and libraries that perform this function will also be investigated.

Knowing the schema of the JSON file, the matching schema will be performed, generated from the ontology. To perform the schema matching, the tool proposed by (Aumueller et al., 2005), called *Coma ++* (A system for flexible combination of schema matching approaches), will be used as the basis. This tool was chosen because it allows the matching of different files (for example, OWL and SQL), in addition to being an open source project under the AGPL license. Currently, the tool does not have the ability to work with JSON files, so this capability should be developed. Then, a file containing the matching schema matches is generated. Finally, based on the JSON file and the result of the schema matching, an algorithm inserts the data into the corresponding database tables.

4 CONCLUSION AND FUTURE WORK

This article proposed a system for integrating quality control data from components in the Brazilian oil and gas industry.The system is proposed based on tools already developed, for example *Coma ++* (Aumueller et al., 2005) e *Ontology2RDB*(Abo Zidan et al., 2019). Thus, the main objective of the system is to promote a structure that allows access and management to this data, promoting a better management of equipment quality control. Therefore, in the case of manual data verification, inspectors will perform the task faster due to the ease of access to information. Also, this quality control data can be provided to a system that will infer as to the completeness and compliance of the databooks.

As future work, tests can be cited to validate the functionalities of the tools mentioned in this work. From these tests, it will be possible to evaluate which tool best suits the proposed system.An important step for the implementation of this work is the development of an ontology applied to the inspection of components of the oil industry. First, a search in the literature should be carried out for ontologies applied to this domain and then an ontology developed with the peculiarities of the analyzed domain.Finally, after the development and integration of the proposed system, the effective gains from its use must be evaluated.

REFERENCES

Abo Zidan, R., Almustafa, M., and Tahawi, M. (2019). Ontology2rdb for storing ontology in relational database. *Journal of Theoretical and Applied Information Technology*, 97:2229– 2240.

Afzal, H., Waqas, M., and Naz, T. (2016). Owlmap: fully automatic mapping of ontology into relational database schema. *International journal of advanced computer science and applications*, 7(11):7–15.

Almeida, M. B. (2014). Uma abordagem integrada sobre ontologias: Ciência da informação, ciência da computação e filosofia. *Perspectivas em Ciência da Informação*, 19(3):242–258.

Aumueller, D., Do, H.-H., Massmann, S., and Rahm, E. (2005). Schema and ontology matching with coma++. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 906–908.

Bernstein, P. A., Madhavan, J., and Rahm, E. (2011). Generic schema matching, ten years later. *Proceedings of the VLDB Endowment*, 4(11):695–701.

de Freitas, A. L. P. (2017). Proposta de uma Ontologia para Tratamento de Divergência de Dados na Aplicação de Fusão de Sensores em Plantas Industriais. Master's thesis, Universidade Federal do Rio Grande, Brasil.

Duarte, G. (2010). *O Controle da Qualidade em Processos de Produção Mecânica Não-Seriada*. Escola Politécnica da Universidade de São Paulo.

Elmasri, R. and Navathe, S. B. (2011). *Database systems*, volume 9. Pearson Education Boston, MA.

Fowler, M. (2013). Schemaless data structures. *URL http://martinfowler. com/articles/schemaless*.

Izquierdo, J. L. C. (2014). JSON discoverer tool. https://github.com/SOM-Research/jsonDiscoverer.

Izquierdo, J. L. C. and Cabot, J. (2013). Discovering implicit schemas in json data. In *International Conference on Web Engineering*, pages 68–83. Springer.

Nascimento, R. S. (2010). Qualidade na soldagem em uma empresa fabricante de estruturas metálicas soldadas do setor de óleo e gás. *VI Congresso Nacional de Excelência em Gestão*.

Panawong, J., Ruangrajitpakorn, T., and Buranarach, M. (2016). An automatic database generation and ontology mapping from owl file. In *JIST (Workshops & Posters)*, pages 20–27.

Petróleo-Brasileiro-S.A (2017). ABC da inspeção de fabricação. https://www.petronect.com.br/irj/go/km/docs/pccshrcontent/Site\%20Content\%20\%28Legacy\%29/Portal2018/arquivos/cep/abc_inspecao.pdf.

Rachman, A. and Ratnayake, R. C. (2019). An ontology-based approach for developing offshore and onshore process equipment inspection knowledge base. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 58783, page V003T02A084. American Society of Mechanical Engineers.

Rodrigues, D. d. A. (2013). Casamento de esquemas de banco de dados aplicando aprendizado ativo. Master's thesis, Universidade Federal do Amazonas, Brasil.

Staab, S. and Studer, R. (2010). *Handbook on ontologies*. Springer Science & Business Media.

Wand, Y., Storey, V. C., and Weber, R. (1999). An ontological analysis of the relationship construct in conceptual modeling. *ACM Transactions on Database Systems (TODS)*, 24(4):494–528.