Journal of Theoretical and Applied Information Technology
30th November 2020. Vol.98. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ONTOLOGY-BASED SEMANTIC INTEGRATION OF HETEROGENEOUS DATA SOURCES USING ONTOLOGY MAPPING APPROACH

[1]KITTIPHONG SENGLOILUEAN, [2]ROMCHAT KHUNTONG
[1,2]Department of Information Management, Faculty of Humanities and Social Sciences,
Prince of Songkla University, Thailand
E-mail: [1]kittiphong.s@psu.ac.th, [2]romchat.k@psu.ac.th

ABSTRACT

At present, educational institutions have developed their library database systems, or used library package programs. Some educational institutions had many libraries such as the university libraries, the faculty libraries, and the department libraries. However, standards in developing and designing the library systems were differently applied. These differences caused the problems when sharing the link of data from many libraries. That was, it encountered the problem of centralized semantic query because each library stored the same data, but chose to use different vocabulary to describe the data. As there was no common agreement or standard for describing data, it brought about the semantic conflicts of synonyms and homonyms. In this paper, the ontology-based semantic integration of heterogeneous data sources by using the ontology mapping approach was implemented for solving such the problem. The results of the study revealed that the architecture of the semantic integration of heterogeneous data sources using the ontology mapping approach could effectively resolve the semantic conflicts of linking the library data from various sources. Based on the performance evaluation of the proposed architecture, it was found to be highly effective with Precision, Recall, and F-measure, at 90.34%, 91.98%, and 91.15%, respectively.

Keywords: *Ontology Mapping, Semantic Integration, Semantic Conflicts, Library Ontology*

1. INTRODUCTION

Nowadays, educational institutions have severally developed their data systems, especially the library system. It is considered an important system where collecting various kinds of data resources and bodies of knowledge. Some educational institutions have the library systems at the university, the faculty, and the department levels. The data storage is scattered without the centralized standards for the data storage. This leads to the problem of data heterogeneity [1, 2], and it is difficult for sharing and linking the data. In other words, storing the same types of data with different terms brings about the semantic conflicts. The query results do not meet the users' needs and it wastes the time. Unfortunately, the users need to sort the query results once. For example, some library systems use the word "*Resources*" in place of data resources whereas some provide the word "*Materials*" to represent the data resources. Therefore, when data from different sources are shared and the data resources are queried by using the word "*Resources*", the query results will not find "*Materials*". Actually, both "*Resources*" and "*Materials*" store the data resources of the library as considered the library data systems including the university library system, the faculty library system, and the department library system. These data systems are developed within the same organization but different departments. So, the libraries are independently developed without mutual standards or agreements. When there is a need for sharing the data as the centralized data, it results in the semantic conflicts [3, 4], redundancy, and diversity of the data [5-7].

Thus, this research study proposed a technique for solving the problem of semantic data integration, linking heterogeneous data resources, and solving the semantic conflicts by using the ontology mapping as a problem-solving tool. This approach was implemented together with WordNet, DBpedia, and Dublin Core Metadata. The objectives of this study were 1) to design the architecture of the ontology-based semantic integration of heterogeneous data sources using the ontology mapping approach, and 2) to solve the semantic conflicts from the ontology mapping

Journal of Theoretical and Applied Information Technology
30th November 2020. Vol.98. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

process in terms of synonyms and homonyms at the class and property levels.

The goals of this paper, the ontology-based semantic integration of heterogeneous data sources by using the ontology mapping approach was implemented for solving such the problem. Based on the performance evaluation of the proposed architecture, it was found to be highly effective with Precision, Recall, and F-measure, at 90.34%, 91.98%, and 91.15%, respectively (See Table 4).

For contribution and strengths of this paper, it could be summarized as follows.

1) It could effectively convert the data sources from various ontological data sources with the wrapper process.

2) It could effectively resolve the semantic conflicts of synonyms and homonyms from the ontology mapping process by using WordNet, DBpedia, and the Dublin Core Metadata.

3) It could query the semantic data effectively from various kinds of ontological links.

4) The architecture of the ontology-based semantic integration of heterogeneous data sources using the ontology mapping resulted in a high efficiency as evidenced in the experimental results in Table 4.

The remainders of this paper are organized as follows: Section 2 describes the theoretical background and related works; Section 3 describes the proposed architecture; Section 4 presents the evaluation; and Section 5 presents the conclusion and plans for future work.

2. THEORETICAL BACKGROUND AND RELATED WORKS

2.1 Semantic web technologies

Semantic web technologies [8, 9] are technologies initiated from Tim Berners-Lee's concept, and widely known by World Wide Web Consortium (W3C) [10]. The semantic web technology can enhance the ability of the current web technology for computers or the agent software, which humans cannot understand the meanings of the machine-understandable web data in order to meet their understandings, and enable to apply the data for further processing effectively.

2.1.1 Ontology technologies

Ontology technology [11, 12] is one of the key elements used in semantic web technologies. Ontology is a borrowed term from a philosophical study field which studies the nature of things happening in the world by defining and describing what they mean. In general, the ontology is usually used for explicitly describing what interests us in a particular domain without any ambiguity so that humans and computers or the agent software can understand the meanings, eliminate semantic conflicts of the data, and encourage the computer or the agent software to query the word with the same meaning in spite of using different words. The typical ontology components are Class, Properties, Instances, Relationships, and Constraints.

2.1.2 Ontology structure definition language

Ontology definitions are described in RDF, RDFS, or OWL languages developed by the W3C, where the RDF (Resource Description Framework) is a standardized structure for describing the web-based metadata. It is used to describe the data and its relationships in areas of interests based on the fundamental prototype from graphs with the XML language, while the RDFS (RDF Schema) describes the structure of metadata. To describe a resource, the attributes and the property value are defined as nodes to describe the resource whereas the *Predicate* is the attribute defined as a property. And the *Object*, which is the property value, is defined as a class. Then, the relationship between the nodes is established in the form of an RDF Schema [13, 14].

OWL (Ontology Web Language) [15, 16] is a language that can define hierarchical data structures, and describe relational data in a database system. Besides that, it can support the narrative of logical data, data types, and quantity indicators. As a result, the replaced data is more meaningful. The descriptive style is in forms of the class, the property class, and the relationships class to describe the *Entity* and the relationships. Thus, the OWL is considered the language that enables to describe the semantic data and the relationship structure of the system better than other languages.

2.1.3 Inference engine

The inference engine [17, 18] is the software that is created to be able to process data stored in a knowledge base, and find the in-depth query from such a knowledge base. The analogical tool created for the semantic web technology will be used to process the data in the form of the ontology to find the in-depth relationships from the ontology, which can answer the users' semantic query. The examples of well-known analogical tools at this present time are Hermit, Pellet, KAON2, Racer DL, and RDFStore.

2.1.4 Semantic query language

SPARQL (SPARQL Protocol and Query Language) [19-21] is a language format for the data query generated on the RDF standard. Similarly, the SQL (Structured Query Language) is used to query the data from the relational database. The W3C defines SPARQL as the central standardized language for querying the data from documents created by using the RDF document structure.

2.2 WordNet and DBpedia databases

WordNet [22, 23] is an online English vocabulary database designed on a basis of the linguistic theories. It was developed by Princeton University since 1985 to combine a dictionary, a glossary, or a thesaurus. To collect the vocabulary data, it was classified by the word types such as nouns, verbs, adjectives, adverbs, etc., and grouped with the similar meanings, which was called synset. Each synset represents the value of one concept (sense), and links to other sysets with a hierarchical structure in a relationship of synonyms, homonyms, antonyms, hypernyms, hyponym, etc. In this study, it mainly implemented the WordNet as a database in the process of solving the problem of synonym and homonym similarities.

Wikipedia is a large center of bodies of knowledge, where people around the world can improve the contents and it is constantly growing up. In fact, the Wikipedia has a lot of articles with more than 285 languages from all over the world. However, the Wikipedia is stored in the form of HTML, which is understood by humans, but not by the computer. This makes it difficult to further apply the data. As a result of the Wikipedia problem, the DBpedia Project [24-26] was created with the aim of converting the Wikipedia data into the form of the ontology. Moreover, the project had semantic relationships and network, so it is understandable to both humans and computers. All of the data in DBpedia is connected semantically in the form of an open link.

The Dublin Core Metadata [27, 28] is a standard to establish the metadata for web documents that were widely used. The standard by the Dublin Core Metadata Initiative (DCMI) was created by a group of people who arranged the academic conferences consisting of information professionals, computer scientists, database manufacturers, and publishers in Dublin, Ohio, the USA in 1995. The goals of the Dublin Core Metadata are 1) it is a user-friendly structure that can be scaled up or down as needed, 2) it adds details to each core element, or creates a connecting point from the data generated by Dublin Core to another with a heterogeneous structure, and 3) it is an alternative to better use of the digital data, not intended as a replacement for the already-used standard system. The Dublin Core consists of 15 subset elements namely *Title, Creator, Subject, Description, Identifier, Format, Type, Date, Contributor, Publisher, Source, Language, Relation, Coverage, and Rights.*

2.3 Related works

The research on the ontology mapping using description logic and bridging axioms [29] proposed the DL based axiom derivation methodology for solving the problems of synonyms and axiomatic matching in order to link the two ontologies without changing their entity names. The advantages of this research was, it could effectively solve the word similarity of synonyms, and connect the ontology effectively. However, this research was still limited. It could not connect the complicated ontologies. Moreover, it has limitations of the data accuracy of querying from the ontology, and that of using the semantic web rules language (SWRL) to solve problems with complexity levels.

The research on UFOme: An ontology mapping system with strategy prediction capabilities [30] proposed the ontology mapping system through the strategy prediction capabilities to systematically solve the ontology mapping process from both the system users' and users' perspectives. Rules and requirements of the framework arose from a thorough determination of the system accuracy. The highlight of this study was the possibility to suggest the most suitable method to be implemented in the ontology mapping process with consideration of the similarities of the two ontologies. Nevertheless, it was limited in solving the problem of the complicated ontology mapping process, which has not been solved yet. Now the UFOme system has still been on the process of research and experiment. In the future, it will be open for the UFOme users to use and receive feedbacks from the users to make it more effective.

A research study on reusing ontology mappings for query routing in the semantic peer-to-peer environment [31] proposed a solution to the mapping composition with any ontology mapping algorithms, and a query rewriting application on an ontology-based peer-to-peer environment. The strength of this study was, it could solve the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

problem of the word similarity in the ontology mapping process effectively. Additionally, the results of the ontology mapping were accurate. However, this research study also had limitations: 1) it was limited in solving the problem of the complicated ontology mapping process, 2) it was limited in querying the complicated data in terms of accuracy and flexibility, and 3) the efficiency of the data query took a long time due to the large and complicated connection of the data.

3. PROPOSED ARCHITECTURE

3.1 Architectural design of the system

This research study proposed the architecture of the ontology-based semantic integration of heterogeneous data sources using the ontology mapping approach, and solved the problems of the semantic conflicts of the ontology mapping process. It was considered an important research study in the field of the semantic data integration, the ontology mapping, and the semantic query. The architecture design of the system consisted of five layers: 1) Data Source Layer, 2) Wrapper Layer, 3) Ontology Mediation Layer, 4) Semantic Search Layer, and 5) Presentation Layer as shown in Figure 1.

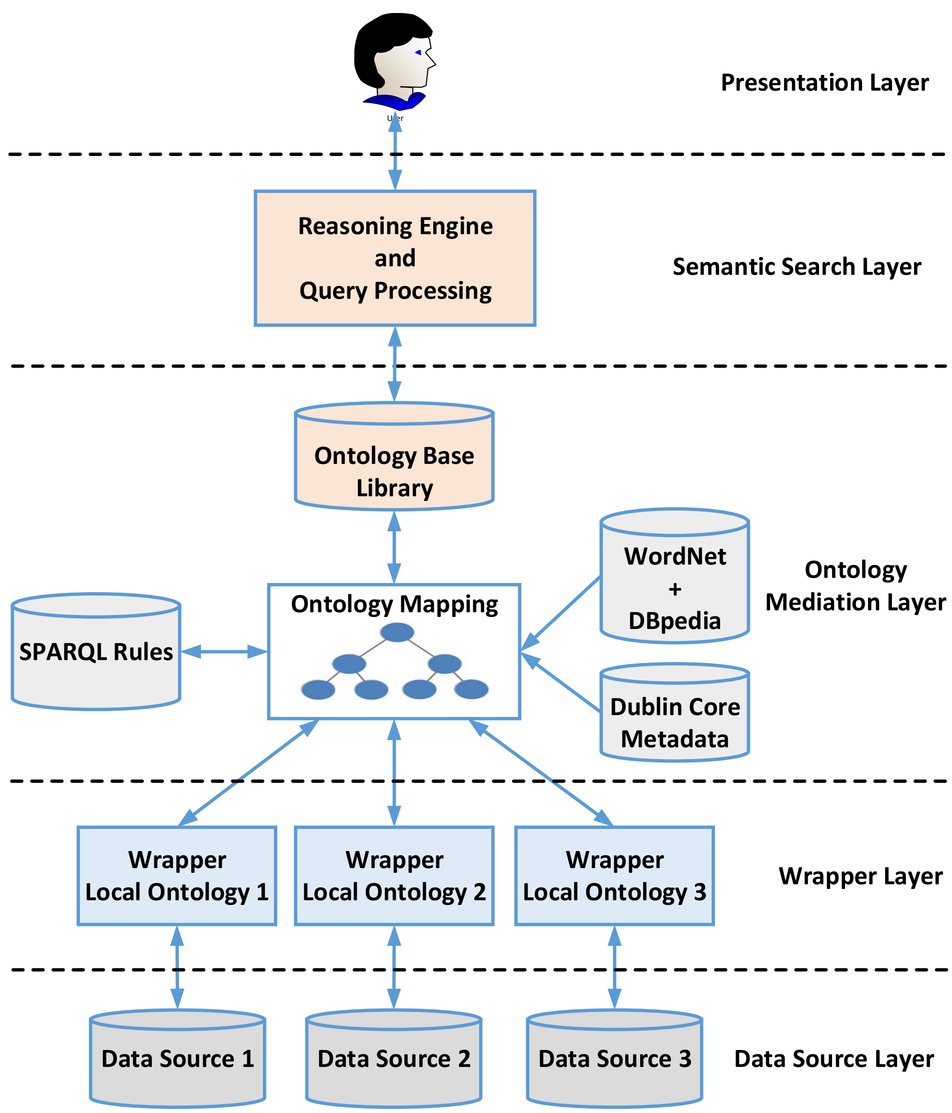

Figure 1: Architecture of Ontology-Based Semantic Integration of Heterogeneous Data Sources Using Ontology Mapping Approach

This section describes the performance of the architecture of the ontology-based semantic integration of heterogeneous data sources using the ontology mapping approach from Figure 1. Each layer was detailed as follows.

3.1.1 Data source layer

Data source layer was a layer that stored the data source and the data structure of the library database system from various data sources. Mostly, it was in the form of a relational database that stored the data of the table structure; the table name, the field name, the data, the primary key, the foreign key, the table relationships, and the field relationships. The problem occurred in this layer was, the same data with different terminology to describe the data led to the semantic conflicts of synonyms and homonyms. Additionally, it brought about the problem of converting the database structure from various sources to the local ontology. The approach for solving such a problem was proposed in the process of *Wrapper Layer*.

3.1.2 Wrapper Layer

Wrapper layer was a layer that converted the database structure from the data source by the wrapper process, which converted the structure of each data source to the form of the ontology, also known as the local ontology, by using the OWL to describe the structure of the local ontology. In the process of converting a relational database to the ontology, tables were generated as classes whereas fields were generated as data properties. The fields with key properties of connection between tables were generated as object properties, while the data were generated as instances of the ontology regarding the OWL structure. Furthermore, the additional relationships and properties were generated to enable the ontology to be completely accurate.

3.1.3 Ontology mediation layer

The ontology mediation layer was very important and considered as the main core of this research study. It served the major function of the semantic data integration and the ontology mapping obtained from the wrapper layer process by using the OWL properties that the World Wide Web Consortium (W3C) had designed and set up the OWL standard to resolve the semantic conflicts. However, in some cases, the properties of the OWL have not been able to resolve all semantic conflicts yet. Thus, SPARQL rules were set up to solve problems that the OWL could not do. In this layer, the processes of the class mapping and the property mapping of each local ontology were combined. It caused the problem when describing the same data but using different terms. Therefore, this study solved the problem of the word similarity of synonym and homonym types of class mapping and property mapping processes using the WordNet, DBpedia, and the Dublin Core Metadata for resolving the semantic conflicts. After completing the ontology mapping with the semantic data integration, it obtained the complete ontology base library. In other words, it was the ontology that obtained from the ontology mapping from different data sources, and that had already solved the semantic conflicts.

As considered the semantic conflicts from the ontology connection and the semantic data integration, it was found that it had the problem of the semantic conflicts of synonyms and homonyms to represent the class level when *Local Ontology 1, Local Ontology 2, and Local Ontology 3* from the wrapper layer were applied to the semantic integration with the ontology mapping process in the ontology mediation layer. For examples, *Local Ontology 1* used "*Resources*" instead of the information resource data, while *Local Ontology 2* used "*Materials*" instead of the information resource data. And *Local Ontology 3* used "*Resources*" instead of the information resource data. To solve the problem of the word similarity, the Wu and Palmer’s equation [32-34] was employed as shown in Equation 1. In which pair of the word similarity obtained the similarity values, its class or property was applied to the ontology mapping.

$$Sim_{wup}(c_1, c_2) = \frac{2 \times depth(lcs(c_1, c_2))}{depth(c_1) + depth(c_2)} \quad (1)$$

From Equation 1, the *depth* represented the distance from the concept to the root of the WordNet hierarchy structure. The Least Common Subsume (*lcs*) value, where *lcs* (*c1*, *c2*) represented the common concept that was the subsume of the *c1* and *c2* concepts, and that was the closest to *c1* and *c2*. The similarity value equaled to 0 <*Simwup* <= 1 because the smallest *depth* did not equal to zero. If the *Simwup value* (*c1*, *c2*) equaled one, it meant that the *c1* and *c2* concepts were in the same synset. It provided the same meaning in spite of using different words.

When considering the semantic conflicts of all three local ontologies, the semantic conflicts between the class, the object property, and the data

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

type property could be summarized as shown in Tables 1, 2, and 3, respectively.

Table 1: Class Semantic Conflicts

Class of Local Ontology 1	Class of Local Ontology 2	Class of Local Ontology 3	Meaning
Resources	Materials	Resources	Collect the data resources
Type	Category	Classification	Collect data of the data type resources
Borrow	Loan	Lend	Collect data of borrowing the data resources
Hold	Reserve	Booking	Collect data of reserving the data resources
Member	Fellow	Partner	Collect the data of the library members
Staff	Staff	Officer	Collect the data of the library staff
Admin	Manager	Staff	Collect the data of the library administrators

Table 2: Semantic Conflicts of Object Property

Object Property of Local Ontology 1	Object Property of Local Ontology 2	Object Property of Local Ontology 3
hasType	hasCategory	hasClassification
hasBorrow	hasLoan	hasLend

Table 3: Semantic Conflicts of Data Type Property

Datatype Property of Local Ontology 1	Datatype Property of Local Ontology 2	Datatype Property of Local Ontology 3
hasTitle	hasMaterialName	hasResourceName
hasBarcode	hasID	hasIdentifier

Methods for solving the semantic conflicts in linking data to obtain the results of semantic query needed to solve the problem of the semantic conflicts before querying the data. To solve such a problem, it needed to apply the linking property of the OWL as follows.

1) For defining a prefix in the OWL, prefixes were used for identifying the data from different data sources, and presenting the differences of the data. For example, *Local Ontology 1: Staff* and *LocalOntology2: Staff* used the word "*Staff*" similarly but using a prefix could distinguish that *LocalOntology1: Staff* was a class of *Local Ontology 1*, with *LocalOntology1* as a prefix, while *LocalOntology2: Staff* was a class of *Local Ontology 2*, with *LocalOntology2* as a prefix. Therefore, when generating the classes and properties, the prefixes were always required to help identify the accurate source of data. In this research study, three ontologies were used for defining the prefixes as follows.

DataSource1 (ont1) = http://www.library1.com/localontology1#
DataSource2 (ont2) = http://www.library2.com/localontology2#
DataSource3 (ont3) = http://www.library3.com/localontology3#

2) The connection of the classes from Table 1 showed the semantic conflicts between the occurred classes. So, *owl:equivalentClass* was employed to connect all three ontologies to indicate the class homogeneity. The portion of the class connection was shown in Figure 2.

```
<owl:Class rdf:about="http://www.library1.com/localontology1#Resources">
   <owl:equivalentClass rdf:resource="http://www.library2.com/localontology2#Materials"/>
   <owl:equivalentClass rdf:resource="http://www.library3.com/localontology3#Resources"/>
</owl:Class>
…
<owl:Class rdf:about="http://www.library1.com/localontology1#Admin">
   <owl:equivalentClass rdf:resource="http://www.library2.com/localontology2#Manager"/>
   <owl:equivalentClass rdf:resource="http://www.library3.com/localontology3#Staff"/>
</owl:Class>
…
<owl:Class rdf:about="http://www.library1.com/localontology1#Staff">
   <owl:equivalentClass rdf:resource="http://www.library2.com/localontology2#Staff"/>
   <owl:equivalentClass rdf:resource="http://www.library3.com/localontology3#Officer"/>
</owl:Class>
```

Figure 2: Class Connection in the Ontology with the OWL

3(The property was connected from the semantic conflicts between the object property and the data type property in Tables 2 and 3, respectively. To connect the property between the ontologies, the property types must be in the same type. In other words, they were in the same object properties or the same data type property as defined with *owl:equivalentProperty*. The results of connection were shown in Figures 3 and 4.

```
<owl:ObjectProperty rdf:about="http://www.library1.com/localontology1#hasHold">
   <owl:equivalentProperty rdf:resource="http://www.library2.com/localontology2#hasReserve"/>
   <owl:equivalentProperty rdf:resource="http://www.library3.com/localontology3#hasBooking"/>
   <rdfs:domain rdf:resource="http://www.library1.com/localontology1#Member"/>
   <rdfs:range rdf:resource="http://www.library1.com/localontology1#Hold"/>
</owl:ObjectProperty>
```

Figure 3: Connection of the Object Property in the Ontology with the OWL

```
<owl:DatatypeProperty rdf:about="http://www.library1.com/localontology1#hasCreator">
   <owl:equivalentProperty rdf:resource="http://www.library2.com/localontology2#hasAuthor"/>
   <owl:equivalentProperty rdf:resource="http://www.library3.com/localontology3#hasInventor"/>
   <rdfs:domain rdf:resource="http://www.library1.com/localontology1#Resources"/>
   <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

Figure 4: Connection of the Data Type Property in the Ontology with the OWL

4) Defining additional features of the OWL

In this research study, the features of the OWL were additionally determined to enable the semantic query to infer the meaning. This also encouraged the system to become more intelligent. Additionally, the SPARQL rules have been created to solve the problem of the ontology mapping process, which the properties of the OWL could not be solved. The SPARQL rules are going to be proposed in detail in the further study. In this paper, it mainly proposed the properties of the OWL.

4.1) Symmetric property was a property that allowed an inference engine to obviously notice the two-way relationship of the data at the instance level between the subject property and the object of the instance property. It defined "*relatedTo*" as *Symmetric Property*, such as *Resources001, relatedTo, Resources002*. Similarly, it could be inferred to *Resources002, relatedTo, Resources001* as well. The symmetric property in the ontology with the OWL was illustrated in Figure 5.

```
<owl:ObjectProperty rdf:about="http://www.library1.com/localontology1#relatedTo">
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#SymmetricProperty"/>
  <rdfs:domain rdf:resource="http://www.library1.com/localontology1#Resources"/>
  <rdfs:range rdf:resource="http://www.library1.com/localontology1#Resources"/>
</owl:ObjectProperty>
```

Figure 5: Symmetric Property in the Ontology in the OWL

4.2) Transitive property was a property that allowed the inference engine to find the relationship of the data in the instance level continuously by defining the property named "*supervisor*" as the transitive property such as *Staff001, supervisor, Staff002* and *Staff002, supervisor, Staff003*. It could be inferred to *Staff001, supervisor, Staff003* as well. The transitive property in the ontology with the OWL was shown in Figure 6.

```
<owl:ObjectProperty rdf:about="http://www.library1.com/localontology1#supervisor">
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#TransitiveProperty"/>
  <rdfs:domain rdf:resource="http://www.library1.com/localontology1#Staff"/>
  <rdfs:range rdf:resource="http://www.library1.com/localontology1#Staff"/>
</owl:ObjectProperty>
```

Figure 6: Transitive Property in Ontology in the OWL

4.3) InverseOf property was a property that allowed two properties to be able to switch between domain and range values by defining the property named "*catalog*" and the property named "*cataloged* by" as *InverseOf* property. So, it could be described that *Librarian001, catalog, Resources002* would allow the system to infer to *Resources002, catalogBy, Librarian001* as well. The *InverseOf* property in the ontology with the OWL was shown in Figure 7.

```
<owl:ObjectProperty rdf:about="http://www.library1.com/localontology1#catalogue">
  <owl:inverseOf rdf:resource="http://www.library1.com/localontology1#catalogueBy"/>
  <rdfs:domain rdf:resource="http://www.library1.com/localontology1#Staff"/>
  <rdfs:range rdf:resource="http://www.library1.com/localontology1#Resources"/>
</owl:ObjectProperty>
```

Figure 7: InverseOf Property in the Ontology with the OWL

3.1.4 Semantic search layer

The semantic search layer was a layer that processed the semantic query commands from the users via using an inference engine to make the semantic query more intelligent, which was the strong point of the research studies on the semantic search. The semantic search prevailed over other search engines with the keywords which could not

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

solve the semantic conflicts of synonyms and homonyms. In addition, the process of query processing was a process of query results from the ontology-based library using the SPARQL. The obtained results of this process were then submitted to the users in the presentation layer, which was described in the process of *Presentation Layer*.

The results of the semantic query using the SPARQL to search the library's data resources from the ontology-based libraries revealed that a total of nine results obtained from all three ontologies: *local ontology1*, *local ontology2*, and *local ontology3* were the results of the inference engine derived from the ontology connection with the properties *owl:equivalentClass*, and *owl:equivalentProperty* as shown in Figure 8.

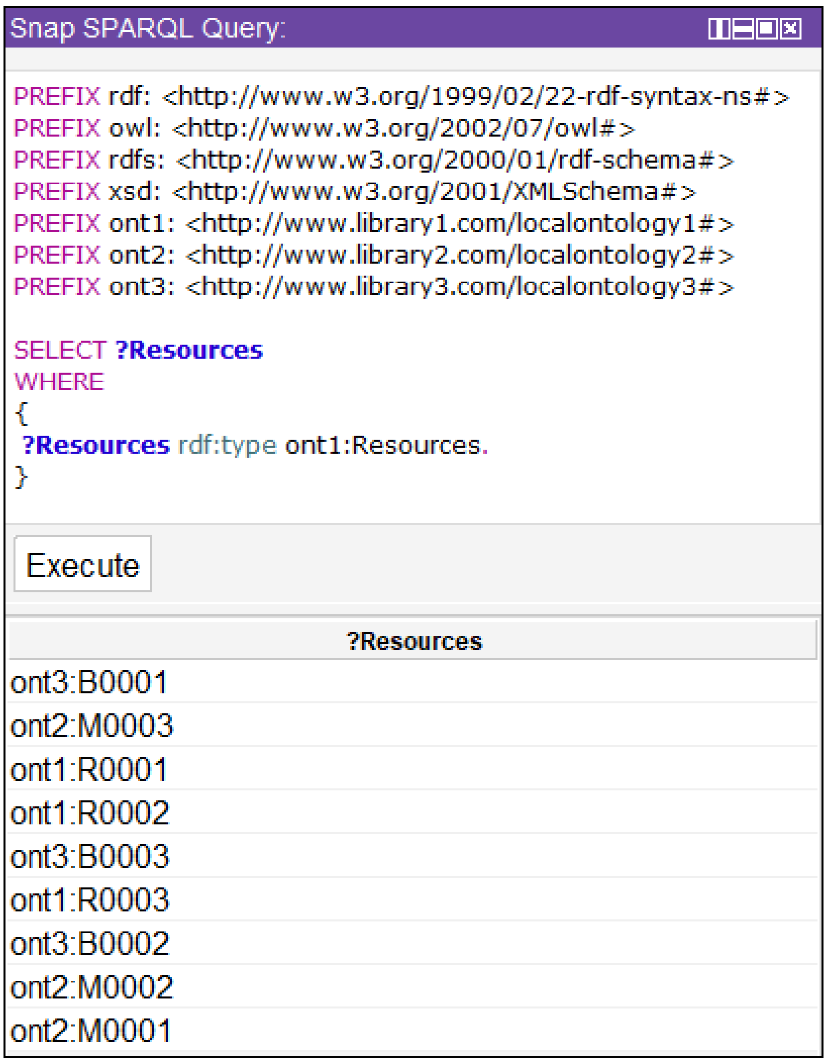

Figure 8: Query Results with Properties owl: equivalent class

The figure below showed the semantic query results with the transitive property by defining a property named "*supervisor*" as a transitive property such as *Staff001, supervisor, Staff002* and *Staff002, supervisor, Staff003*. Thus, it could be inferred to *Staff001, supervisor, Staff003* as well, which followed the principle of the property shown in Figure 9.

Journal of Theoretical and Applied Information Technology
30th November 2020. Vol.98. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

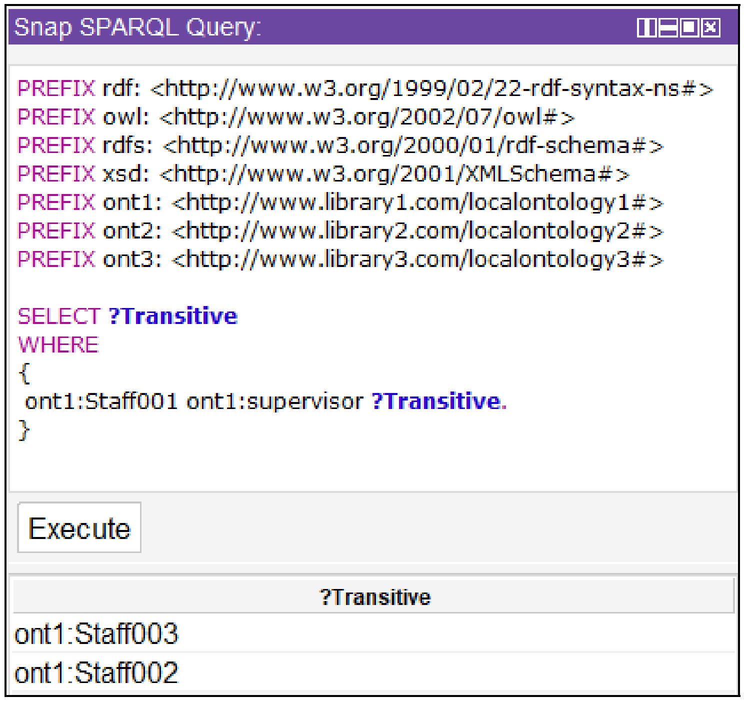

Figure 9: Query Results with the Transitive Property

3.1.5 Presentation layer

The presentation layer was a layer that contacted the users in a semantic query from the ontology-based libraries to search for the required data by using the SPARQL, and to submit the results to the users. According to the experimental results of the architecture of the ontology-based semantic integration of heterogeneous data sources using the ontology mapping approach, it was found that the semantic query results met the users' needs with the short answers. Moreover, the users did not need to rescreen the results. It could resolve the semantic conflicts when using the keyword "*Resources*" instead of data resources. Besides that, the query results could also found "*Materials*" because both "*Resources*" and "*Materials*" stored the library's data resources similarly.

4. EVALUATION

4.1 Experimental design

This section dealt with designing an experiment based on the architecture of the ontology-based semantic integration of heterogeneous data sources using the ontology mapping approach. It was detailed as follows: Three library database systems were tested. These library database systems were designed and developed from different standards and database structures. The architecture in this research study was flexible and able to support different data systems. However, in this test, only three examples were given for solving the problem of the semantic conflicts and the word similarity of synonyms and homonyms in the process of the ontology mapping. In this process, the WordNet and the DBpedia were implemented. These databases were recognized as a large growing knowledge source of data from users around the world to provide the system with the most natural language. Furthermore, the Wu and Palmer's equation, the properties of the OWL, and the SPARQL Rules were used for solving other complicated problems and evaluating the performance to provide the most accurate and flexible system in four areas: 1) ontology mapping: class, 2) ontology mapping: object property, 3) ontology mapping: data type property, and 4) semantic query.

4.2 Measurement and evaluation

F-measure was a value used to measure the fundamental performance of the data, which resulted from the combination of Precision and Recall values in the calculation. The F-measure was considered the representative of Precision and Recall.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$Precision = \left(\frac{ce}{ce + te}\right) 100 \qquad (1)$$

where *ce* referred to the value of the valid data from the ontology mapping, and *te* referred to the value of the invalid data from the ontology mapping.

$$Recall = \left(\frac{ce}{ce + fe}\right) 100 \qquad (2)$$

where *ce* was the value of the valid data from the ontology mapping, and *fe* was the value of the invalid data not from the ontology mapping.

$$F - measure = 2\left(\frac{Precision * Recall}{Precision + Recall}\right) \qquad (3)$$

The evaluation indicated a high accuracy of the architecture of the ontology-based semantic integration of heterogeneous data sources using the ontology mapping approach. In this study, the performance of the system in four areas consisting of 1) ontology mapping: class, 2) ontology mapping: object property, 3) ontology mapping: data type property, and 4) semantic query were measured. The values of Precision, Recall, and F-measure were at 90.34%, 91.98%, and 91.15%, respectively (See Table 4). However, the reason why Precision, Recall, and F-measure values were not definitely at 100% because this present study focused on the ontology mapping to solve the semantic conflicts related to the human's natural language which had a high flexibility and a wide variety of synonyms and homonyms. However, some of which could not be found in the DBpedia and the WordNet.

Table 4: Results of the Proposed Approach within Various Datasets in the Ontology Mapping

	Precision	**Recall**	**F-measure**
Ontology mapping: Class	91.89	94.44	93.15
Ontology mapping: Object property	90.91	92.78	91.84
Ontology mapping: Datatype property	89.34	88.62	88.98
Semantic Query	89.23	92.06	90.63
Total	**90.34**	**91.98**	**91.15**

4.3 Comparison with other approaches

This research study was compared with other related studies presented in Section 2, *Theoretical background and related works*. Each of which was in the same areas. The results of evaluating the performance of each research study could be summarized in Figure 10.

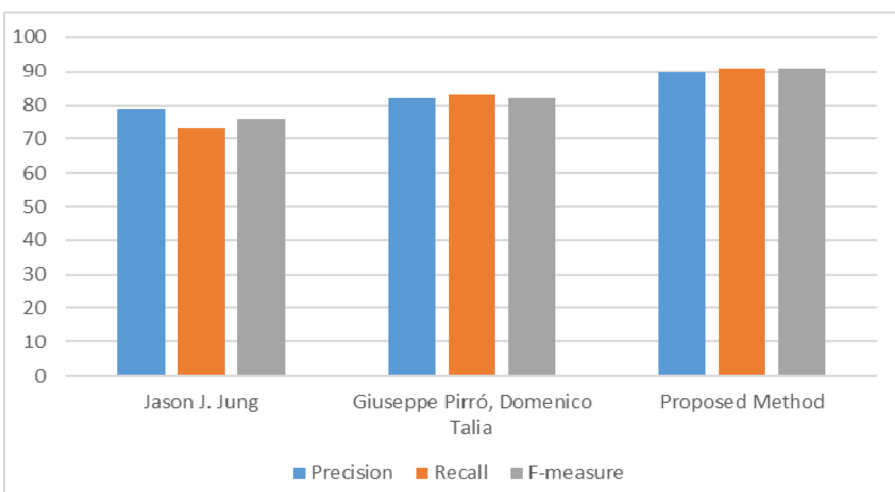

Figure 10: Results of a Comparison with Other Approaches

The results of the performance evaluation of the proposed approach were at a very good level, and different from other research studies presented in Section 2, Theoretical background and related works detailed as follows.

Jason J. Jung conducted the research on UFOme: An ontology mapping system with strategy prediction capabilities [30]. He presented the ontology mapping system by employing strategy prediction capabilities to solve the ontology mapping process systematically from the system users' and the users' perspectives. The framework rules and requirements were arisen from the accurate examination of the system thoroughly. The highlighted point of this research was the possibility to suggest the most suitable method for the ontology mapping process. As considered the similarities of the two ontologies, the performance evaluation of Precision, Recall, and F-measure were at 79%, 73%, and 76%, respectively.

Giuseppe Pirró, Domenico Talia, from the research paper on reusing ontology mappings for query routing in the semantic peer-to-peer environment [31], proposed a solution to the mapping composition with any ontology mapping algorithms, and a query rewriting application on an ontology-based peer-to-peer environment. The highlight of this study was, it could effectively solve the problem of the word similarity in the ontology mapping process, and the results of the ontology mapping were accurate. The results of the performance evaluation of Precision, Recall, and F-measure were at 82%, 83%, and 82%, respectively.

Nevertheless, this present study stood out and differed from other research studies mentioned above in the following areas: 1) it could be converted the heterogeneous data sources to the ontology effectively with the wrapper layer process, 2) it could effectively resolve the semantic conflicts of synonyms and homonyms arisen from the ontology mapping process by implementing the WordNet, the DBpedia, the Dublin Core Metadata, 3) it could effectively query the semantic data from the ontology mapping from heterogeneous data sources, and 4) the architecture of the ontology-based semantic integration of heterogeneous data sources using the ontology mapping approach resulted in a high performance. According to the results shown in Table 4, it could be insisted that this present study was accurate and effective at a very good level.

5. CONCLUSION

This paper presents the architecture of the ontology-based semantic integration of heterogeneous data sources using the ontology mapping approach. It proposes an approach for solving the problem of semantic conflicts of the library system data from heterogeneous data sources through implementing the ontology mapping for the semantic data integration. The results of the study revealed that the ontology-based semantic integration of heterogeneous data sources using the ontology mapping approach could effectively resolve the problem of the semantic conflicts of the ontology mapping from different library sources. Based on the performance evaluation of the proposed architecture, it was found to be highly effective with Precision, Recall, and F-measure at 90.34%, 91.98%, and 91.15%, respectively.

In the future, the approach for solving the problem of the more complicated ontology mapping and the limitations of the OWL will be proposed by generating the SPARQL Rules. Additionally, the problem of the semantic query process will be solved to infer the data more intelligently and effectively.

6. ACKNOWLEDGMENT

This study was granted by the Faculty of Humanities and Social Sciences, Prince of Songkla University, Thailand.

REFRENCES:

[1] V. Jirkovský, M. Obitko, and V. Mařík, "Understanding data heterogeneity in the context of cyber-physical systems integration," IEEE Transactions on Industrial Informatics, vol. 13, pp. 660-667, 2016.

[2] M. Maree and M. Belkhatir, "Addressing semantic heterogeneity through multiple knowledge base assisted merging of domain-specific ontologies," Knowledge-Based Systems, vol. 73, pp. 199-211, 2015.

[3] J. M. Alcaraz Calero, J. M. Marín Pérez, J. Bernal Bernabé, F. J. Garcia Clemente, G. Martínez Pérez, and A. F. Gómez Skarmeta, "Detection of semantic conflicts in ontology and rule-based information systems," Data & Knowledge Engineering, vol. 69, pp. 1117-1137, 2010/11/01/ 2010.

[4] N. Gu, J. Xu, X. Wu, J. Yang, and W. Ye, "Ontology based semantic conflicts resolution in collaborative editing of design documents,"

Advanced Engineering Informatics, vol. 19, pp. 103-111, 2005/04/01/ 2005.

[5] M. d'Aquin and N. F. Noy, "Where to publish and find ontologies? A survey of ontology libraries," Journal of Web Semantics, vol. 11, pp. 96-111, 2012/03/01/ 2012.

[6] J. J. Lastra-Díaz, A. García-Serrano, M. Batet, M. Fernández, and F. Chirigati, "HESML: A scalable ontology-based semantic similarity measures library with a set of reproducible experiments and a replication dataset," Information Systems, vol. 66, pp. 97-118, 2017/06/01/ 2017.

[7] L. C. van der Gaag and H. J. M. Tabachneck-Schijf, "Library-style ontologies to support varying model views," International Journal of Approximate Reasoning, vol. 51, pp. 196-208, 2010/01/01/ 2010.

[8] B. Drury, R. Fernandes, M.-F. Moura, and A. de Andrade Lopes, "A survey of semantic web technology for agriculture," Information Processing in Agriculture, vol. 6, pp. 487-501, 2019/12/01/ 2019.

[9] D. Moussallem, M. Wauer, and A.-C. N. Ngomo, "Machine Translation using Semantic Web Technologies: A Survey," Journal of Web Semantics, vol. 51, pp. 1-19, 2018/08/01/ 2018.

[10] M. Needleman, "The W3C Semantic Web Activity," Serials Review, vol. 29, pp. 63-64, 2003/03/01/ 2003.

[11] M. Mejhed Mkhinini, O. Labbani-Narsis, and C. Nicolle, "Combining UML and ontology: An exploratory survey," Computer Science Review, vol. 35, p. 100223, 2020/02/01/ 2020.

[12] M. Katsumi and M. Fox, "Ontologies for transportation research: A survey," Transportation Research Part C: Emerging Technologies, vol. 89, pp. 53-82, 2018/04/01/ 2018.

[13] V. S. Chooralil and E. Gopinathan, "A Semantic Web Query Optimization Using Resource Description Framework," Procedia Computer Science, vol. 70, pp. 723-732, 2015/01/01/ 2015.

[14] G. Marzano, "Using Resource Description Framework (RDF) for Description and Modeling Place Identity," Procedia Computer Science, vol. 77, pp. 135-140, 2015/01/01/ 2015.

[15] D. Nilavu and R. Sivakumar, "Knowledge Representation Using Type-2 Fuzzy Rough Ontologies in Ontology Web Language," Fuzzy Information and Engineering, vol. 7, pp. 73-99, 2015/03/01/ 2015.

[16] D. Yang, R. Miao, H. Wu, and Y. Zhou, "Product configuration knowledge modeling using ontology web language," Expert Systems with Applications, vol. 36, pp. 4399-4411, 2009/04/01/ 2009.

[17] Z. Wang, G. Tian, and X. Shao, "Home service robot task planning using semantic knowledge and probabilistic inference," Knowledge-Based Systems, vol. 204, p. 106174, 2020/09/27/ 2020.

[18] T. Bloch and R. Sacks, "Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models," Automation in Construction, vol. 91, pp. 256-272, 2018/07/01/ 2018.

[19] P. Ochieng, "PAROT: Translating natural language to SPARQL," Expert Systems with Applications: X, vol. 5, p. 100024, 2020/04/01/ 2020.

[20] D. Wiśniewski, J. Potoniec, A. Ławrynowicz, and C. M. Keet, "Analysis of Ontology Competency Questions and their formalizations in SPARQL-OWL," Journal of Web Semantics, vol. 59, p. 100534, 2019/12/01/ 2019.

[21] J. Potoniec, D. Wiśniewski, A. Ławrynowicz, and C. M. Keet, "Dataset of ontology competency questions to SPARQL-OWL queries translations," Data in Brief, vol. 29, p. 105098, 2020/04/01/ 2020.

[22] F. Leão, K. Revoredo, and F. Baião, "Extending WordNet with UFO foundational ontology," Journal of Web Semantics, vol. 57, p. 100499, 2019/08/01/ 2019.

[23] T. Wei, Y. Lu, H. Chang, Q. Zhou, and X. Bao, "A semantic approach for text clustering using WordNet and lexical chains," Expert Systems with Applications, vol. 42, pp. 2264-2275, 2015/03/01/ 2015

[24] M. A. Paredes-Valverde, M. Á. Rodríguez-García, A. Ruiz-Martínez, R. Valencia-García, and G. Alor-Hernández, "ONLI: An ontology-based system for querying DBpedia using natural language paradigm," Expert Systems with Applications, vol. 42, pp. 5163-5176, 2015/07/15/ 2015.

[25] F. Orlandi and A. Passant, "Modelling provenance of DBpedia resources using Wikipedia contributions," Journal of Web Semantics, vol. 9, pp. 149-164, 2011/07/01/ 2011.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[26] A. Sarkar, U. Marjit, and U. Biswas, "Towards Bengali DBpedia," Procedia Technology, vol. 10, pp. 890-899, 2013/01/01/ 2013.

[27] R. L. Lubas, A. S. Jackson, and I. Schneider, "Using Dublin Core: With information from the Dublin Core Metadata Initiative," in The Metadata Manual, R. L. Lubas, A. S. Jackson, and I. Schneider, Eds., ed: Chandos Publishing, 2013, pp. 45-65.

[28] K.-H. Lee–Smeltzer, "Finding the needle: controlled vocabularies, resource discovery, and Dublin Core," Library Collections, Acquisitions, and Technical Services, vol. 24, pp. 205-215, 2000/06/01/ 2000.

[29] S. K. Kumar and J. A. Harding, "Ontology mapping using description logic and bridging axioms," Computers in Industry, vol. 64, pp. 19-28, 2013/01/01/ 2013.

[30] G. Pirró and D. Talia, "UFOme: An ontology mapping system with strategy prediction capabilities," Data & Knowledge Engineering, vol. 69, pp. 444-471, 2010/05/01/ 2010.

[31] J. J. Jung, "Reusing ontology mappings for query routing in semantic peer-to-peer environment," Information Sciences, vol. 180, pp. 3248-3257, 2010/09/01/ 2010.

[32] A. E. L. Hadi, Y. Madani, R. E. L. Ayachi, and M. Erritali, "A new semantic similarity approach for improving the results of an Arabic search engine," Procedia Computer Science, vol. 151, pp. 1170-1175, 2019/01/01/ 2019.

[33] G. Pirrò, "Semantic Similarity Functions and Measures," in Encyclopedia of Bioinformatics and Computational Biology, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds., ed Oxford: Academic Press, 2019, pp. 877-888.

[34] F. Rahutomo, T. Kitasuka, and M. Aritsugi, "Semantic cosine similarity," in The 7th International Student Conference on Advanced Science and Technology ICAST, 2012.