A Systematic Mapping Study of Database Resources to Ontology via Reverse Engineering

Aftab Ahmed Abbasi and Narayan K. Kulathuramaiyer
Faculty of Computer Science and Information Technology (FCSIT),
University Malaysia Sarawak (UNIMAS), Kuching, Malaysia

Abstract: This study aims at proposing transformation technique to build OWL ontology from relational database resources by following SQL-DDL (Structured Query Language - Data Definition Language) written version. Though, the databases has proposed the best form of techniques in order to store in memory for the purpose of managing data and retrieving data or for the purpose of recovering it for further functions. However, databases has been observed semantically absence in achieving world-wide objectives based on both web and data integration semantically. Semantic web is supported by ontology that has ability to represent data in meaningful, rich and readable form for both human and machine. This issue can be overcome through ontology would help in improving databases by semantic functions. This approach can capture meta-data from tuples by analysis of database relations. The meta-data can assist to take out several features of semantic and would not infer from Structured Query Language (SQL). As conceptual model is richer in semantics, so this study gets conceptual (EER) model by applying reverse engineering technique. Finally, the generated ontology is validated and enhanced through comparison with that of database conceptual model which is also known as EER diagram, for achieving the highest ontology.

Key words: Deep web, transformation technique, databases, semantic web, ontology, meta-data

INTRODUCTION

Presently, the research in the world of semantic web has drawn great focus for solving complex issues in the integration of data. Semantic web is an extension in current web providing a standard way for exchanging information between machine to machine and people. The ontology is known as the focal point of the semantic web. Ontology is an unequivocal subtle element in a typical conceptualized concept (Gruber, 1993). Moreover, this is described as a strictly unequivocal course of actions encircled in continuously sorted out course to depict conceptions based on certain treatise basis intended for learning (Swartout et al., 1997). Ontology expect a basic part in comprehension the considered database interoperability because of gigantic qualities (Noy and McGuinness, 2001), for instance sharing information structure, reusing and including rich and machine understandable semantic to data in it. Using ontology means, it is not projected as the replacement for certain developments of database. Truly, database is considered as more risky when compared to ontology for securing broad scale of these data sets. Hence, this study finds a structure partner for both the database techniques as well as ontology methods. Whereas, databases techniques are broadly used and addressable. However, this separating of rationality is non-use-able for the most part of the approaches. Additionally, adding to a ontology beginning with no outside help is dismal, uninteresting, both slanted and work-expensive at the season of building one by hand exhibits the same detriments (Li et al., 2005; Stojanovic et al., 2002). This suggested course of action used a couple of guidelines for changing an offered database to a ontology which can be used either for manual change or for an automated change of integrity processes (Stojanovic et al., 2002). In this regard, the terms mapping and transformation should be differentiated clearly. The change from database to the ontology infers for making ontology by following the perceptive model of databases by using a couple of procedures either by hand or through a structure applied by logical form databases and ontology which are remained existed (Astrova et al., 2007a).

Literature review: In the research of (Astrova et al., 2007b), Slight the levels of relational tenet or the break standard at the time of mapping of tables by focussing on destinations diagraming. The fundamental approaches for

Corresponding Author: Aftab Ahmed Abbasi, Faculty of Computer Science and Information Technology (FCSIT),
University Malaysia Sarawak (UNIMAS), Kuching, Malaysia

730
change to OWL are described well in (Xu et al., 2006; Barrasa et al., 2003; Upadhyaya and Kumar, 2005; Li et al., 2005). Regardless, a couple of statutes in these technique are not the world over as they discharge to cover certain sorts of component or relationship, for occasion, unary or twofold relationship with extra attributes or ternary and higher associations.

Particular procedures can be depended upon a blend of sources including schematic relations, ER plots, deferred ER, SQL-DDL, database tuples, examination of HTML pages and enquiry of client request to add to the tenet for the change of a relational database into ontology. A gigantic piece of these frameworks utilize a blend of sources. The considered change was at initially connected with that from a relational model to an object-oriented model, then from a social to (RDF) model which is known as an ontological model. Stojanovic et al. (2002) added to the models for changing databases to edge methods of reasoning. Regardless, their standards are manual and don’t make OWL ontologies. In like way, they don’t cover in their measures the issues of composite attribute or multi-referred attribute. Buccella et al. (2004) showed general measures which aren’t possible. Similarly, an unambiguous framework as discussed in Li et al. (2005), Benslimane et al. (2006), Buccella et al. (2004), Barrasa et al. (2003) and Tirmizi et al. (2008) has misdirected the prerequisites if there should be the occasions of discontinuities and ISA relation (class dynamic structures). There are differing overviews in their gages, for instance, they can’t be connected with all circumstances which might be appeared in changed databases. The framework (Sonia and Khan, 2008; Upadhyaya and Kumar, 2005) removes a computed model (EER Model) from the source and a white later structures the change rule. The downside of such techniques is that they add to the theory without focal metadata from relations. Benslimane et al. (2006), the system depends on upon the likelihood that the semantics cleared by inspecting HTML structures will be utilized to conform and overhaul the social example. This research merges an enormous measure of human encouraged attempts. In like way, ontologies can confine after any change to the structure of the HTML pages on which they are based upon. This strategy by multifaceted nature bargains satisfactorily with jumbled issues, for occurrence, seeing discontinuity and ISA-A relationships. This can be coordinated unary and parallel relationship with extra characteristics, ternary associations, higher associations and multi-regarded properties in this strategy, however most past technique have neglect to do in that purpose of restriction. In this, three frameworks are used including ER model, relational model and the databases tuples analysis.

Overview of database resources and owl ontology:
Resource models (conceptual model and logical models) are playing important role for creation of databases. For databases creation, the relational databases approach widely accepted approach for storing, managing and retrieving data than other approaches such as the network and hierarchical or object-oriented. Further, relational databases are moreover by and large recognized by databases models or programming architects, customers and sellers. More than 70% data set is kept away in these relational databases (Hu and Qiu, 2007). However, the field experts comprehensively can open relational databases to include relational object databases or object directed databases. Moreover, databases have various positive benefits that include data need securing just one times, data can be different ranges, data unaltering and data free. Database can easily back up applying certain security coatings which are flexible.

These reasons of databases are negated by some broad prevention for event, not supportive to the extent attribute names, table, weak semantics, disagreeable layout and weak process. Incidentally, the most obstacles lay in the framework of database in light of two stages: Construction of the sensible model as entity relationship (ER) diagram and change it into relational model for the most part work in SQL-DDL. The issue lies in the liberal semantic challenges in the midst of the arrangement of changing over the model ER to SQL as mentioned by Benslimane et al. (2006).

Consequently, the databases response in terms of necessities and the relational model’s inadequacy has been relied on the ontology model to determine by having one of the ontology’s languages. Similarly, the OWL (Web ontology languages) has alluded to as a modish systematized as W3C alleviating world-class device decipher limit within web substance that kept up by XML and RDF arrangement giving included substance vocabulary based on routine semantics.

Comparitive study of database models and ontology capabilities: This part starts with the brief introduction of different database source models characteristics in an ontology context. As seen in Table 1, there are two database sources which include the conceptual model of EERD and the logical model of RM. The table shows the limitations of each source. Since, the two database models are considered as being the sources to this approach which can capture any ontology context as long as it is available in one of the sources.
Table 1: Shows the comparison between EER, RM (SQL) and OWL models

<table>
<thead>
<tr>
<th>Specification-context</th>
<th>Conceptual model (EERD)</th>
<th>Logical model relational model</th>
<th>Ontology model OWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Possible</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Relationship</td>
<td>Possible</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>(has child) relationship</td>
<td>relationship</td>
<td>relationship</td>
<td>object</td>
</tr>
<tr>
<td>Cardinalities</td>
<td>Possible</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Value restriction</td>
<td>Not possible</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Has value</td>
<td>Not possible</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Keys</td>
<td>Possible</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Ternary and higher relationship</td>
<td></td>
<td>Possible</td>
<td>Not</td>
</tr>
<tr>
<td>Inheritance</td>
<td>Possible</td>
<td>Not possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Multiple valued</td>
<td>Possible</td>
<td>Not directly</td>
<td>Possible</td>
</tr>
<tr>
<td>Equivalent</td>
<td>Not possible</td>
<td>Not possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Transitive</td>
<td>Not possible</td>
<td>Not possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Symmetric</td>
<td>Not possible</td>
<td>Not possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Inverse property</td>
<td>Possible</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Functional</td>
<td>Possible</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Inverse functional</td>
<td>Not possible</td>
<td>Not possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Disjoint class</td>
<td>Possible</td>
<td>Not possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Sub property</td>
<td>Possible</td>
<td>Not possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Enumerated</td>
<td>Not possible</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Individual</td>
<td>Not possible</td>
<td>Possible</td>
<td>Possible</td>
</tr>
</tbody>
</table>

The context available in the RM model will be captured automatically whereas the context available in the EER model can be obtained with the assistance of database expert through inverse technique of engineering. Similarly, the use of the relational model (SQL) can make transformation system automatically. Conversely, using the conceptual database model (EER) with RM can add more semantics to the ontology.

From Table 1, it can be observed that if any one of the database sources (EERD or RM) is capable of representing an ontological context, this approach can acquire these semantics. Generally, this kind of approach can extract the semantic from the RM source if it is accessible, otherwise the semantic will be obtained from the EER source; this ensures that there is no conflict between the results obtained by the two sources.

Table 1 also shows some limitations. For example, there are many different modeling contexts which can not be obtained from both sources, e.g., transitive property, symmetric property and equivalent (class/property). This clearly shows that the automatic transformation from database models to ontological models has some limitations as there are many disparities between the two models.

Finally, with all the advantages of a database, there are still some general limitations which might affect the extraction of semantics from database sources such as the meaningless names of some tables or attributes, poor semantics and bad design intended to improve the performance. This approach tries to overcome these limitations by using both models including logical model and conceptual model of databases. This study of the study divided further A to E subsections.

Translating extended entity relationship to OWL ontology: Both EER and ontological model are considered as the conceptual modelling (Sugumaran and Storey, 2002) but this study focuses on EER as the ER model is called as pure relational model. This means that the ER does not include advanced topics such as inheritance (specialization/generalization). Database design goes through two types of modelling. The first is the conceptual model (EER) and the second the RM model. Therefore, OWL ontology can be generated either from an Extended Entity Relationship model or a Relational model. This study has shown a way for producing OWL ontology from the conceptual model of the database EER. In fact, it is suggested that using EER as an ontology source immediately after finishing the database conception stages, i.e., the new-born EER model is the best source from which to obtain the semantic since the actual database structure does not suffer from any modification. Both database models are used to enhance the target ontology. The general rule for translating an EER model to an OWL model consisted of three steps:

- Entities are represented by OWL classes
- Relationships can be translated into a pair of object properties. It also can be translated into two pairs of inverse object properties with a class according to the needs
- Attributes can form data type properties or a class with data-types properties as may apply

Indeed, these steps were also applicable to an ER model since EER contained all the ER constructs. The difference in defining names between the database and the ontology is considered to be the important point to be addressed. In the database, any ER components are identified by their unique names whereas for ontology, OWL components are identified by URIs.

MATERIALS AND METHODS

The conversion process: This study explains the proposed procedure to generate ontology from a relational database. This procedure consists of seven steps:

- Using reverse engineering methodologies on a logical model written in SQL in order to capture syntax and schema structure (relations, attributes, inclusion dependencies, etc.) and the hidden semantics
• Analyse the structures obtained and semantics in order to infer the best corresponding match constructed from the ontology side
• Apply the transformation technique to generate a complete ontology schema which includes classes, sub-assumption relationships, object properties and data-type properties
• Checking the ontology’s consistency through an ontology reason in order to detect the superfluous relationships and to remove redundant data
• Validate and refine the produced ontology by the information obtained from the conceptual database model (EER)
• Evaluate the ontology
• Migrate data in order to produce a knowledge base

This approach co-exists in semantic annotation in semantic web community and is known by the database community as reverse engineering (Stojanovic et al., 2002).

Creation of classes: The process consisted of many steps:

• All entities (strong or weak) will be translated into an OWL: class
• All composite attributes will be transformed into an OWL: class
• All n-ray relationships where n>2 will be transformed into an OWL: class
• All (many-to-many) relationships with supplementary attributes will form an OWL: class

Data-types creation: The process of producing data-types obtained from entity attributes or relationship attributes is outlined below. All attributes of an entity will be converted to data-type properties. Moreover, their domain will be the class corresponding to their owner entity. The range is their equivalent XML types. No constraints for the attributes in EER seem presented including NOT NULL or UNIQUE; therefore we treat attributes as outlined. The OWL: cardinality, equal to one for primary key, the alternative key and the discriminator attribute of a weak entity. In addition, this restriction is applicable for all the attribute parts of a composite attribute.

All other attributes will be assigned to the function or owl: functional, except the multivalued attribute part of an entity or a complex attribute. Complex attribute is an arbitrary nested composite and multivalued attribute. This would not represent each internal composite attribute of the complex attribute here by a class, instead it represents its leaf node with functional restriction and for the multivalued attribute case, this approach can represent it without restriction. For instance, an individual has many than one residency and each one residency has a simple address and multiple phones. The composite attributes were shown between parentheses ( ), their components separated by commas and the multivalued between braces { }. The following example demonstrates a complex attribute: {Address-phone ((Phone (Area-code, Phone-number)), Address (House-number, Street-name, City, Postcode)}.

For relationship attributes, if the relationship is converted to a class, then each attribute belonging to this relationship will be assigned to the class corresponding to it as its domain and the equivalent XML types as its range. However, there were two cases of a relationship that did not form a class: for (1-M) relationship, the attributes belonging to this type of relationship will be allocated to the class corresponding to the entity of the M-side as their domain. For (1-1) relationship, the attributes belonging to this type of relationship can be allocated to either class involved in the relationship.

Object-properties creation: Each relationship is signified through a couple of objective properties. It implicitly means they will be inverse properties of each other except the special cases. There are three aspects of relationships.

Representing binary relationship: Each relationship will have pair of object properties each the inverse of the other, the domain and the range represented by the classes corresponds to the entities participating in this relationship. For (1-1) relationship both object properties will be Functional. For (1-M) relationship the 1-side is Functional.

Representing self-relationship: Self-relationship is a case of a binary relationship in which both the domain and range refer to the same entity. Object properties representing this relationship could use the class corresponding to the entity holding self-relationship as domain and range.

Binary relationships between weak and strong entity: In this case, the existence of an instance of a weak entity was based on the existence of its instance record in the strong entity. Therefore, this relationship can be created by an object with cardinality of one. The domain was the
class corresponding to the weak entity and the range was the class corresponding to the strong entity.

**Binary relationships (M-N) cardinality:** This relationship implied two relationships. This approach constructs only a single object property for each relationship without its inverse and there would be only two object properties instead of four object properties. The first one would be assigned to the class corresponding to the M-side entity as its domain and the range is the class of N-side. The second object property would have the class corresponding to the N-side entity as its domain and the range as M-side class. Finally, the two objects would possess the inverse properties of each other.

**General object properties characteristic algorithm:** Both the primary key and Foreign key represent the standard way for relationship building by following relational data model. Before applying any of attributes characteristic for foreign keys needs to exclude all the foreign keys belonging to the binary relation of (Many-To-many) cardinality and a self-relation. The approach would then create two inverse object properties (OP1 and 2) representing each Foreign key:

- If the attribute \( x \) is a FOREIGN KEY and NULL and not UNIQUE then make the OP1 and OP2 correspond to the attribute \( x \) with functional (OP1) and minimum cardinality OP2 = 0
- If the attribute \( x \) is a FOREIGN KEY and NOT NULL but not UNIQUE then make it’s cardinality (OP1) = 1 and minimum cardinality OP2 = 0.
- If the attribute \( x \) is a FOREIGN KEY and NULL and UNIQUE then make it functional (OP1) and functional (OP2)
- If the attribute \( x \) is a FOREIGN KEY and NOT NULL and UNIQUE but not a primary key then apply hierarchy technique (class-subclass relationship)
- If the attribute \( x \) is a FOREIGN KEY and part of the PRIMARY KEY for weak entity table or N-ray relations or (Many-To-Many) relationship relation with additional attribute, then assign, cardinality (OP1) = 1 and cardinality (OP2) = 1

**General data type axiom algorithm:** All attributes below, firstly satisfy the two conditions:

- Not a Foreign key
- Not a multivalued attribute

The SQL aspects could be applied as outlined:

- If the attribute \( x \) is NOT NULL then gloss the attribute \( x \) with \( \text{Min C} = 1 \)
- If the attribute \( x \) is UNIQUE then make it \( \text{Max C} = 1 \)
- If the attribute \( x \) is a primary key or (NOT NULL and UNIQUE) then make it \( \text{Crd} = 1 \)
- If the attribute \( x \) is DEFAULT then make it Functional and has Value
- If the attribute \( x \) is CHECK IN then make it one of
- Otherwise any attribute annotate with functional

**Algorithm for translating an EER to OWL ontology:**

The algorithm of creating OWL ontology from an Extended Entity Relationship model contained two steps. Firstly, to deal with the entities and attributes (Fig. 1) and secondly, to handle the relationships. Thus, EER Model can be obtained from logical model of database by applying reverse engineering. Before carrying out these steps, there was one requirement in order to ensure successful implementation of fixing the model representation; since EER was a graphical representation and it was difficult to parse unless we reformatted it into a formal definition. Therefore we chose to present the EER diagram in terms of facts (schema way) which were easy to parse (Xu et al., 2004).

![Fig. 1: Translating entity, attributes and relationship of the EER model to OWL ontology](image-url)

**Employing different SQL features to data type attribute practice:** Specifying the SQL constraints was possible (like DEFAULT, NOT NULL, UNIQUE, etc.) to OWL language. In fact, our approach attempted to map database constraints into OWL axioms while preserving the semantics of the original database schema.
RESULTS AND DISCUSSION

it has been observed that both the entity relationship diagram and ontology seem similar in structure but it is very difficult to start with ER diagram due to different reasons that include firstly, the ER diagram is unavailable and the attributes do not access easily to the domain without instances. Thus, OWL ontology can be generated from relational database schema which is written in SQL-Data definition language. The OWL ontology based on classes, object properties and data-types properties were generated using protege editor for this approach.

The ontology model can be validated in order to ensure the correctness of the transformation method. This means all the conceptual parts of the database should correctly correspond in the ontology.

The validation contained two stages. The first stage was positioned at the concerns leading to incorrect ontology representations that include multivalued attributes, fragmented entities, class hierarchy and relationships of (1-1).

These concerns can easily be obtained from an EER model. All entities in the EER model were now represented by tables in the RM model with the exception of the vertical partitioning for tables. To verify the correctness of mapping in the model EER compared with that of the model relational, the number of all entities might be accumulated in a higher degree relationships; (Many-To-Many) relationships and multivalued attributes and this number should have matched the number of tables. Otherwise, there were fragmentation tables for some entities. The database designer had to specify the superfluous tables and their cross-holdings at this stage.

For validation process, it can be proposed to involve the domain expert for validation of produced ontology. That process can bring a comparability of EER model of databases with that of the OWL ontology in terms of these characteristics:

- Clear hierarchical tree can disjoint entities and the shared sub-class of multiple inheritances
- Easy to identify the kind of attributes including composite or multivalued
- Easy to generalize from the group of entity sets to one super entity
- Naturally translating EER models to ontology can preserve more information (Stojanovic et al., 2002) through comparison between entities and relationships
- Explicit participation and cardinality of relationships

Due to these reasons translating from the EER model to ontological model would preserve more semantics:

- Additional attributes along with binary relationship can be considered as an entity
- Different entities can relate with the n-ray type whereas (n>2) in the ER diagram can be considered as equivalent to class numbers in the ontology model
- The term IS-A can develops relationship in two entities which can be used to compare the classes in these entities through sub-class of ontology
- Two reciprocal object properties are existed for each one-to-many relation between two entities
- In every, from many to many relations, two radically different objective properties are existed having no classes
- If there is any n-ray relation is existed; where n>2, it can be two inverted objective properties for each and every foreign key. The same can be applied to any binary relation with an additive attribute

Table 2 below gives an example of counting the entities in the EER model and in the ontological model as equivalent. It is because the both models are drawn from the conceptual model which is shown in Fig. 2 indicating no. of entities in EER model which are equivalent to the no. of classes in ontology model. Following are given some features of semantic which can manually be shifted from ER diagram and can be placed to the ontology. If one entity is based on composite attributes (or complex attributes) then the user should construct a class and should add an object property for connecting the sub-attribute with that of the main class (Upadhyaya and Kumar, 2005).

<table>
<thead>
<tr>
<th>Table 2: The EER and ontology elements</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EER model</strong></td>
<td><strong>Ontology model</strong></td>
</tr>
<tr>
<td>Entity</td>
<td>Class</td>
</tr>
<tr>
<td>6+1</td>
<td>7</td>
</tr>
<tr>
<td>Ternary relationship (Foreign Keys)</td>
<td>2</td>
</tr>
<tr>
<td>Ternary relationship (Foreign Keys)</td>
<td>3</td>
</tr>
<tr>
<td>(1-M) Strong and weak entity</td>
<td>1</td>
</tr>
<tr>
<td>(1-M) Foreign key</td>
<td>4</td>
</tr>
<tr>
<td>IS-A relationship</td>
<td>2</td>
</tr>
<tr>
<td>Classes</td>
<td>7</td>
</tr>
<tr>
<td>Two inverse object properties</td>
<td>2</td>
</tr>
<tr>
<td>Two inverse object properties for each Foreign key in the n-ary relationship</td>
<td>3</td>
</tr>
<tr>
<td>Two inverse object properties</td>
<td>1</td>
</tr>
<tr>
<td>Two inverse object properties</td>
<td>4</td>
</tr>
<tr>
<td>Sub-class</td>
<td>2</td>
</tr>
</tbody>
</table>
Fig. 2: The ER diagram for university database

The cardinality participation can be added for each object property as the cardinality constraints cannot be found from SQL (Buccella et al., 2004).

The related work for this approach is obtained easily through semantic web community and can be identified as a semantic annotation. Conversely, the community of database is reflected as the inverse engineering for such database (Stojanovic et al., 2002).

CONCLUSION

Different methodologies as of now are utilized distinctive procedures to inspect the change in relational model within ontological model and they utilize either a schematic relation or an ER diagram. In this proposed approach, the advantages of the relational model are interconnected with that of conceptual model (EER diagram) by applying some certain rules to the SQL statements by extracting a number of metadata commencing from the database tuples in this study. In short, this study has engendered algorithms for object properties, data type properties and algorithm for making a comprehension of EER parts to ontology classes. In addition, produced subsequent ontology is validated in line of conceptual ER model of the database in this study. It also has secured circumstances, for occurrence, the relationship with itself (self-relation), N-ray relationship has disjointed entities and extra attributes of binary or parallel relationships which can be disregarded in different other methodologies. As far as SQL and its different aspects including prompts, approvals and model of referential activities are concerned are recommended for future research work.

REFERENCES


