Journal of Engineering and Applied Sciences 11 (7): 1568-1574, 2016
ISSN: 1816-949X

A Framework Exchanging Data from Database Resources to Ontology for Integration

Aftab Ahmed Abbasi
Faculty of Computer Science and Information Technology (FCSIT),
University Technology Malaysia, Sarawak, Malaysia

Abstract: Now a days, databases offer the best technique for storing, managing and retrieving data but traditional databases suffer from the absence of a semantic viewpoint in supporting the goal of databases integration from different databases resources into one unified form, in a common framework. For data integration Source models help in integrating distributed database sources. The autonomous and heterogeneous data resources required to represent in logical and conceptual models which provide common framework and reach global goal for databases integration. So, ontology is a best solution for databases integration. As ontology is a clear and detailed details as to how a thing is to be done of shared conceptualization can amount in better way for semantically united as complete thing of heterogeneous resources. Ontology constitutes a bridge between human and machine to machine understanding through formal and consensual terminologies. As a result of this key property of ontology, it can be shared and reused between computer to computer and human. In our proposed approach, the technique which is applied in this study is the rule based transformation on the relational databases to the OWL ontology. Generated ontology is completed with structurepoint of view from relational schema which is written in structure query language. This makes least inaccuracies with less work through manually building ontology. Generated ontology has been validated and enhanced with conceptual model extended entity relationship diagram. This study has reported about techniques applied with examples

Key words: Database integration, logical model, source model, conceptual modelling, ontology, OWL, semantic web

INTRODUCTION

At present time database integration is one of the research area that getting more and more attention from researchers. It has a goal to representing distributed database resources in a common framework where information can be shared from a distributed environment. Many approaches have been used to address database integration problem such as federated databases and warehouse. Problem in these approaches having lack of semantic in the area of data representation.

Semantic web technology can facilitate database integration problem because it has ability to solve complicated problem such as database integration. The ontology then at the core of the semantic web, supports the explicit specification of shared conceptualization (Gruber, 1993). The huge amount of datataking place on the web and kept in the databases can be considered as the backbone for the existing web; this large portion of data stored in databases is referred as deep web (Volz *et al*., 2004; An *et al*., 2004; Juric *et al*., 2008) and more than 73.3% data is stored in databases (Hu and Qu, 2007). Web search engines cannot access data which is stored in databases. To address these issues, a common framework is needed where information can be shared from different heterogeneous resources in aunified form.

The best solution for this is ontology that can provide shared setting where resources can either be combined or rationalized without changing entire structure. Two elements are obtainable including resource and global ontology where local ontology is missing. Many approaches can be used to create local ontology from databases but these approaches cannot differentiate among fragmented entities and entities hierarchy which is displayed as IS-A in relationship (Li *et al*., 2005; Benslimane *et al*., 2006) some of these approaches have not attended to it (Astrova *et al*., 2007a) in its construction. However, some of these are using simple examples as rules which can be misrepresentative relating to the outcomes. Some other approaches create unrealistic suppositions having no fragmentation, no N-ary relationship and no IS-A relationship. This study aims to cover various situations using proposed directions relating to the benefits of the relational and conceptual models and its analysis of approaches. The bases are integrated, effaced and rationalized with no transmuting the whole construction in Scalable database integration.

This needs a source model to provide amalgamated access to distributed data sources (Khan and Morvan, 2006). The source model can signify allsources equivalently in an ontological model. That will be in heterogeneous environment where information can be shared and reused by taking the advantages of ontology alignment technique. Data is stored independently from other resources which can create integration system very malleable and accessible. To deal with the heterogeneity nature of these sources, it is required that source model should be suggested in logical model and conceptual model (EER). Both EER and ontology models are conceptualized fill in as self-sufficient of the major syntactic representation confining thought based models of data sources. It handles semantic heterogeneity of the database sources which help association among them. It also allows them to increment shared information and reuse structure. The standard focus of this study is on physical (relational schema) data model which is comprehended and by and large recognized model to be used. The method used is made out of two stages for building source ontology.

First of all, local ontology by means of OWL was produced (Shen *et al*., 2006; Alalwan *et al*., 2009) from relational schema (in SQLDDL). Then, produced local ontology was validated and certified through conceptual model (EER) database. Moreover, the construction of ontology is a repetitious process in which anew discovered semantics is added to the originally construct ontology. The projected technique diminishes thelengthy task of constructing ontology by hand designed for its basis. This technique needs disseminated setting in which schematic relational (SQL-DDL) area professionals present to further research.

The technique presented in this study needs no prior awareness in terms of sources, principal relations and additional semantics implausible the existing structures (Buccella *et al*., 2004; Shen *et al*., 2006). However, extract ingmeta-data from relational tables also addressed in this study. Logical model of database is used to generate local ontology then validated and enhanced by using conceptual model (EER Model) torecognize the problematic area in order tore construct the system (Alhajj, 2003).

Literature revies: Astrova *et al*. (2007b) and others discharge the substances levels of leadership hierarchies or the break in their table mapping; ignore fragmentation. What's more,they center generally on mapping essentials. The past studies Li *et al*. (20085), Upadhyaya and Kumar (2005), Barrasa *et al*. (2003) and Xu *et al*. (2006) reported that the tenets for change to OWL are shown space subordinate. In the case, the techniques used as apart of a couple of systems are not worldwide and summed up and they neglect to address particularsorts of substance or relationship, for event, binary relationship (M:N) with additional behavior orunary or ternary and higher relationship affiliations. Stojanovic *et al*. (2002) used f-logic does not separatekey sort. It sees each one of them as objects. Attributes are their affiliations are not discernible. Distinctive strategies depend on upon a course of action of sources, for case, relational structure, part relationship model, extended entity relationship Model, SQL-DDL schema, tuples of database examination, customers' requesting examination of HTML pages, to make the standard for the change of a relational database into appeal. Most by a wideedge of these techniques use a mix of sources. The considered change was at first associated with that from R. schema Model to study of engineered model, then from a R. schema to (RDF) model which is an ontological model. In Bucella *et al*. (2004) present general global rules which can't be implemented. Arranged approaches (Li *et al*., 2005; Benslimane *et al*., 2006; Barrasa *et al*., 2003; Xu *et al*., 2006; Tirmizi *et al*., 2008) have misleading or wrong representation for fragmented entities and IS-A relationship between entities and subentities. There are many mistakes in their presented techniques-rules; they can't be associated with all cases which may be appeared indifferent databases. The techniques used in (Sonia and Khan, 2008; Upadhyaya and Kumar, 2005) light of tuples examination and concentrates on the related model-entity relationship diagram from the source and constructed transformation rules. The obstruction of such systems is that they construct the ontology without central metadatafrom relations.

Benslimane *et al*. (2006) studied the perspective of the examination of HTML pages and get R-schema layout, the frame work r elies on upon the probability that these mantics lose by restructuring the HTML pages structures and ontology can be breakdown, as HTML pages are regularly restructuring will beused to reproduce and drive the relational schema. The limitation of this research is that it combines agreat deal of human participation. What's more, ontologies can separate after any change to the structure of the HTML pages on which they are based. Our philosophy deals palatably with tangled issues, for instance, separate fragmentation and IS-An associations other than manage binary associations with additional characteristics, ternary associations, higher associations, in our method, however, most past structures have disregard to do in that farthest point and our framework successfully ace. This study has used three techniques as a touch of system the extended entity association outline, the R. schema example model and the examination of database tuples.

MATERIALS AND METHODS

Proposed system architecture: The proposed architecture is divided into 4 steps. The complete architecture is shown in Fig 1. In this architecture, step 1 stimulates the OWL ontology on relational schema in SQL. Step 2 validates andrefines the produced ontology from conceptual model (extended entity relationship diagram).

Step 3 uses alignment technique to exchanging information among different ontology in distributed environment. Step 4 maps the produced ontology or local ontology into comprehensive or world-wide ontology which provides common understanding of structure forsharing information from different ontology of different domains. Interestingly, this study bases on first two steps. This research has adapted database integration using ontology vision frame work from (Wache *et al*., 2001). The proposed approach in this study needs 2 main inputs. In the first step, input one is regardedas the SQL-data definition language which represents database structure and SQL-data manipulation language which represents database instances. In first step, generated ontology is completed with structure which included all the ontology definitions (classes, object properties, data type properties, instances). Similarly, the next input for this proposed approach is the conceptual model-extended entity relationship diagram of the database used for refining and validating the OWL ontology which is produced in the first step. The database holder can supply the complete system based on the EER Model at the second step. If not, the database administrator can analyse the system and can construct the ER Model. Otherwise, these database administrators can practice on the techniques available in (Sonia and Khan, 2008; Upadhyaya and Kumar, 2005) for its construction.

The technique in this study can be applied when the EER diagram is either supplied or produced. The produced ontology has validated and has enhanced at this step. The main objective of second step in the study is to achieve the ideal results. In the third step, this study takes the advantages of alignment technique for produced ontology to the local ontologies of different domains for sharing of information in distributed environment. In final step, ontology is mapped with global ontology which provides a common environment for information sharing from different domains. In this study, the first two steps are focused mainly while the step third and fourth stepare recommended for future work.

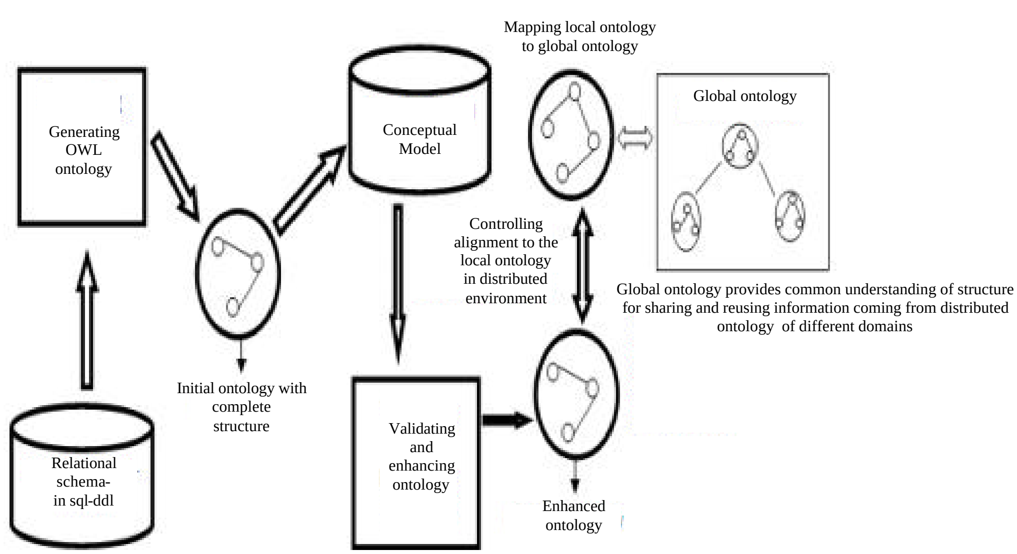

Fig. 1: Proposed framework for database integration

RESULTS AND DISCUSSION

Transformational technique: from relational schema to ontology: This approach has used the logical model (relational schema) for generating ontology which is written in SQL-data definition language showing database structure. Instead of the conceptual model, since the database conceptual model (ER model) is not implemented during database design process or the model had been lost. Another possible explanation may relate to undocumented changes altering the structure of the database. The logical data model isused to represent physical schema part of relational databases. As logical model is used for actual representation of updated data and able to get more semantics.

Identification of fragmentation in relation: When the data of one entity type is grounded on more than one relation can be best identified as disintegration with perpendicular splitting the tableby assimilating as one class. However, there are 3 ways to deal fragmentation1. Database specified correctly and completely then fragmentation could be differentiable from hierarchy (inheritance). If database is not completely specified but data is available then tuples in parent and child relations must be identical, considered as fragmentation. Considered fragmentation as inheritance if relations are in third normal forms and relationships type one to one. Fragmentation is used to improve the performance of the system, as database designers split one entity into more than one relation, important fields assigned to main table and other fields are assigned to other table. Similarly, another need for fragmentation can be felt when one entity has two parts of information, first part (confidential) fields need to be stored within high security and can only be accessed or manipulated by the administrator and the second part fields can be accessible by other users. The third need for the fragmentation technique is for the distributed databases. When part of one entity has been manipulated in the branches of the database, then the update could take place ata later stage in the main database. Fragmentation can be addressed when all tables participating infragmentations attain their primary keys as identical and the primary keys performing as foreign keys can refer to each other. The technique for addressing fragmentation is as:

{If Primary-Key (a,T1) and Primary-Key(b,T2) and Foreign-Key (b,T2,a,T1) and ForeignKey(a,T1,b,T2) then class (c) is created from the union of Table1 and 2}

where, T1 is table 1 and T2 is table 2. Primary keys of the Table (T1 and T2) act as foreign keys referring to each other. So, the above two tables (T1 and T2) are sharing same primary keys and also their primary keys are referring toeach other. So, here is a class-c through the integration of T1 and T2. Here, it can be indicated that the two primary keys are equal which means fully dependent and the condition does not acceptthe tables having partial primary keys. By applying this technique on university schema example, canget staff details and staff tables which satisfy condition of fragmentation. So by integrating these two tables, one staff class can be achieved. In fragmentation, tables are merged into one class. Data type properties are created for each attribute and repeating attributes appear once in staff class. So, this generate as:

Staff_class (staff_id, n_id, staff_ family_name,staff_first_name, staff_mid_name, DOB, address, email_addr, home_phone,university_extension_ phone)

Identification of entities hierarchy: It is one of the most important steps to construct hierarchy in ontology. The hierarchy means classand subclass relationship as in ISA relationship between entities in EER Model. In this approach our technique for hierarchy is used as:

{If Primary-Key(a,T1) and Primary-Key(b,T2)and Foreign-Key(b,T2,a,T1) then class(c1) fromT1 and class(c2) from Table T2 where class-c2 isthe subclass-c1}

Only one primary key of table can act as foreign key to other. Suppose, there are 3 tables including T1, T2 and T3 and each table have its own primary key. Here, T2 primary key referred to T1 and T3 primary key referred to T2. When this technique is applied between T2 and T3, then the result is class for T2 and subclass for T3. Similarly, if same technique is used for T1 and T2, then the result super class for T1 and subclass for T2. Third possibility regarding to relationship between T1 and T3 was not applicable. Some approaches Li *et al.* (2005) and Tirmizi *et al.* (2008) failed to compare the disintegration of one entity with many others instead of one relation. This failure came from incorrect specifications of the fragmentation case and therefore the fragmentation and hierarchy case have the same representation. However, the approach distinguished between the two cases by the foreign key restriction. Both fragmentation and hierarchy had the same first condition of tables sharing the same primary key. The fragmentation case required the primary keys to act as foreign keys referring to each other whereas the hierarchy case required only one of the primary keys to act asa foreign key to others. Even if there was a chain of class and subclass hierarchies for more than two levels; the condition remained true for inheritance, since the two tables do not refer to each other. There was clearly no restriction on the master relation primary key, whether it was a foreign key or not which allowed the master relation to be subclass from a super class. This meant the technique was applicable even if the master relation preceded its primary key as foreign key or not. These things provide power to our technique to build many levels of hierarchies (inheritances). Similarly another way through observation IS-A relationship represented as if foreign key disjoined from primary key. For example, we have two table person-table and student graduate table. In student graduate table foreign key has two restrictions NOTNULL and UNIQUE. Whenever these two characteristics combined in a foreign key it could form an IS-A relationship:

Person table (P_Id (Primary Key), DOB, Name,..)Student_graduate (St-Id (Primary Key), Research_topics, Id UNIQUE NOT NULL foreign key references Person-table (P_Id))

Similarly 2 classes in ontology are created as person table class and student graduate class which show hierarchy relationship. Further, Tirmizi *et al*. (2008) argued that if the IS-A relationship was modeled but the primary key of the super class did not become the primary key of the subclass, then, there could be no assumption of the existence of an IS-A relationship as discussed in the example of student and person tables in student graduate IS-A Person. It is because the prime key of the Person table (P_Id) did not become the principal key of the graduate student table; therefore, the existence of an IS-A relationship cannot be assumed. Conversely (P_Id) was placed in the Student_graduate table as a foreign key referencing person table at the same time declared to be the UNIQUE and NOT NULL in the table. This effectively declared that (P_Id) was an alternate key and this should declare an IS-A relationship. For multiple inheritances, this can be considered the only solution. Student table can be added in above example showing multiple inheritances as:

Student (St_Id (Primary Key), specialization, degree) Student_graduate (St_Id (Primary Key), research_topics, ID UNIQUE NOT NULL foreign key references Person-table (P_Id), St_Id foreign key references student (St_Id))

So, the class of student-graduate will inherit all properties of class person-table and class student. Ontology classes are created, for the attributes of the relations that are participating in hierarchy, data type properties are created.

Representing N-ary and Binary (M:N) relationships with attributes: When there is N-ary relationship where N>2, OWLdid not represent it. Some approaches Li *et al*. (2005) and (Benslimane *et al*., 2006) suggested decomposition of n-ary relationship or bond to binary bond or relationship. According to this approach when there is N-ary relationship orternary relationship, a class is created to deal such relationships with cardinality one (Noy and McGuiness, 2001). By this way such type of relationship exists as whole can been sured. Many approaches ignore binary (M:N) relationships along with supplementary characteristics. For example, there are two entities student and subject having many to many relationship and relationship having an attribute grade. A new class can be created including result class identifying such type of relationship by means of two pair of object properties. One pair of property with student class and result class where result is domain and range is student. Similarly, one object property is created for subject class and class result.Where class result is domain and subject class isrange. For attribute representation data typeproperty is created (Fig. 2).

Object properties and data type properties representation: OWL has 2 key categories of properties that include object and data type properties. The first issame in terms of relationships in entities which areused to connect an example of one class to the example of other class. Further, the properties data type can connect an instance to the value with XML data type schema. Each relationship is represented by pair of two inverse object

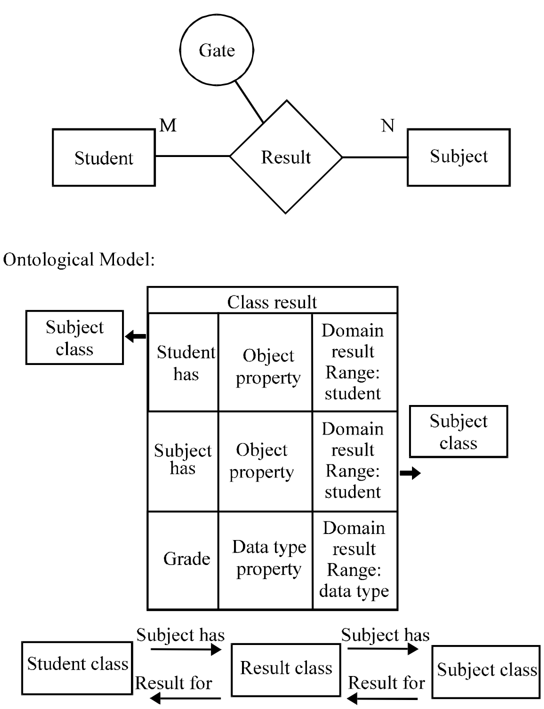

Fig. 2: Entity relationship model

properties which are represented asOP1 and OP2, where OP1 is reversed of OP2.If there are many classes, the object properties are calculated as n*(n-1). It is important to differentiate between degree of relationship and their cardinalities. Relationship between two entities is called binary relationship and has degree two. Similarly, relationship degree of one can exit in unary relationship. Relationships between three entities have the degree of relationship three. Cardinalities are type of relationship which can be represented as one to one (1:1), one for many (1:M) and many for many (m: n). Attributes of entitiescan be represented as data-type properties in OWL).

Validating and enhancing produced ontology: Extended entity relationship model and ontology have a comparative structure. Not with standing, beginning with ER outline is trouble some for some reasons, for instance, the ER graph is not accessible, might be lost, the qualities don't have spacen further more there is no examples (instances). Along these lines, OWL ontology can be produced from physical (relational schema) database constructionin SQL-DDL and made classes, object properties and information sorts or data type properties. As both model applied model and ontological model having comparative structure so the accepting procedure begin with these properties:

- Binary relationship (M:N) withsupplementary characteristics spoke to by aclass
- For N-ary relationship a class is made andafterward divided into binary relationship toguarantee that relationship exists as entirety
- In the instance of inheritance or is-arelationship spoke to by class and subclass
- For association such as many to manyconnections between two substancesentitie 2 reverse article properties exist

Also to validate and upgrading created ontology, a few ideas are not accessible in legitimate model of database, for example, composite attributes characteristics, multivalued properties, complex characteristics and multiple-inheritance we utilized extended entity relationship model for check of these concepts.

CONCLUSION

Various systems at present use assorted techniques to investigate the change of a relational model in to an ontological model. These studies has used relational outline or an ER-Model chart or expelling creation from HTML pages and a few procedures getting relational configuration from the examination of tuples. In this proposed approach, the upsides of the relationals chema model and the hypothetical model increased extended entity relationship diagram are successfully merged by applying rules based system to the SQL illuminations and the expelled of some meta data from the database tuples to made neighborhood reflective theory with complete method of ontologydefinition (Classes, object properties, data types properties and instances) to achieve databases integration. By then we have validated the produced ontology with EER model of the database. It also has secured circumstances, for occurrence, the relationship with itself (self-relation), N-ray relationship has fragmented entities which can bedisregarded in different other methodologies. Itlegitimizes saying that we have secured circumstances break, binary-relationship (N: M) with additional qualities (properties), fragment edentities and ISA relationship which is unattended by various philosophies. Dynamic parts of SQL, for event, triggers, refer ential action and assertions will be regulated in future research.

REFERENCES

Alalwan, N., H. Zedan and F. Siewe, 2009. Generating OWL ontology for database integration. Proceedings of the 3rd International Conference on Advances in Semantic Processing, October 11-16, 2009, Sliema, Malta, pp: 22-31.

Alhajj, R., 2003. Extracting the extended entity-relationship model from a legacy relational database. Inf. Syst., 28: 597-618.

An, Y., A. Borgida and J. Mylopoulos, 2004. Refining Semantic Mappings from Relational Tables to Ontologies. In: Semantic Web and Databases, Christoph, B., V. Tannen and I. Fundilaki (Eds.). Springer, Berlin, Germany, ISBN:978-3-540-24576-6, pp: 84-90.

Astrova, I., N. Korda and A. Kalja, 2007a. Rule-based transformation of SQL relational databases to OWL ontologies. Proceedings of the 2nd International Conference on Metadata & Semantics Research, October, 1-2, 2007, Corfu Island, Greece.

Astrova, I., N. Korda and A. Kalja, 2007b. Storing OWL ontologies in SQL relational databases. Int. J. Electr. Comput. Syst. Eng., 1: 242-247.

Barrasa, J., O. Corcho and A.G. Perez, 2003. Fund finder: A case study of database to ontology Mapping. Proceedings of the 3rd International Semantic Web Conference, Number 2870 in Lecture Notes in Computer Science, October 20-23, 2003, Springer, Sanibel Island, Florida, USA, pages: 17-22.

Benslimane, S.M., D. Benslimane, M. Malki, Y. Amghar and H. Saliah-Hassane, 2006. Acquiring OWL ontologies from data-intensive web sites. Proceedings of the 6th international conference on Web engineering, California, USA., July 11-14, 2006, pp: 361-368.

Buccella, A., R.P. Penabad, F.J. Rodriguez, A. Farina and A. Cechich, 2004. From relational databases to OWL ontologies. Proceedings of the 6th National Russian Research Conference, (NRRC'04), Puschchino, Russia

Gruber, T.R., 1993. A translation approach to portable ontology specifications. Knowledge Acquisit., 5: 199-220.

Hu, W. and Y. Qu, 2007. Discovering simple mappings between relational database schemas and ontologies. Proceedings of the 6th International The Semantic web and 2nd Asian Conference on Asian Semantic Web Conference, Volume 4825, November 11-15, 2007, Busan, South Korea, pp: 225-238.

Juric, D., M. Banek and Z. Skocir, 2008. Uncovering the Deep Web: Transferring Relational Database Content and Metadata to OWL Ontologies. In: Knowledge-Based and Intelligent Information and Engineering Systems, Iganac, L., R.J. Howlett and C.J. Lakhmi (Eds.). Springer, Berlin, Germany, ISBN:978-3-540-85562-0, pp: 456-463.

Khan, S. and F. Morvan, 2006. Data integration in distributed biomedical sources. Proceedings of ISCA 19th International Conference on Parallel and Distributed Computing System, September 20-11, 2006, ISCA, San Francisco, California, USA, ISBN:978-1-880843-60-4, pp: 1-1.

Li, M., X. Du and S. Wang, 2005. Learning ontology from relational database. Proceedings of the 4th International Conference on Machine Learning and Cybernetics, August 18-21, 2005, Guangzhou, China, pp: 3410-3415.

Noy, N.F. and L.D. McGuiness, 2001. Ontology development 101: A guide to creating your 1st ontology. Standard Knowledge Systems Laboratory Technical Report KSL-01-05 and standard Medical Informatics Technical Report SMI-2001-0880.

Shen, G., Z. Huang, X. Zhu and X. Zhao, 2006. Research on the rules of mapping from relational model to OWL. Proceedigs of the Conference and Workshop on OWL: Experiences and Directions, November 10-11, 2006, Athens, Georgia, USA, pp: 1-9.

Sonia, K. and S. Khan, 2008. R2O transformation system: Relation to ontology transformation for scalable data integration. Proceedings of the 12th international symposium on database engineering & applications, September 3-5, 2008, Munster, Germany, pp: 291-295.

Stojanovic, L., N. Stojanovic and R. Volz, 2002. Migrating data-intensive web sites into the semantic web. Proceedings of the ACM Symposium on Applied Computing, March 10-14, 2002, Madrid, Spin, pp: 1100-1107.

Tirmizi, S.H., J. Sequeda and D. Miranker, 2008. Translating SQL Applications to the Semantic Web. In: Database and Expert Systems Applications, Bhowmick, S.S., J. Kung and R. Wagner (Eds.). Springer-Verlag, Berlin, ISBN-13: 9783540856542, pp: 450-464.

Upadhyaya, S. and P. Kumar, 2005. ERONTO: A tool for extracting ontologies from extended E/R diagrams. Proceedings of the 2005 ACM symposium on Applied computing, March 13-17, 2005, Santa Fe, New Mexico, USA, pp: 666-670.

Volz, R., S. Handschuh, S. Staab, L. Stojanovic and N. Stojanovic, 2004. Unveiling the hidden bride: Deep annotation for mapping and migrating legacy data to the semantic web. Web Semant. Sci. Serv. Agents World Wide Web, 1: 87-206.

Wache, H., T. Voegele, U. Visser, H. Stuckenschmidt and G. Schuster *et al*., 2001. Ontology-based integration of information-a survey of existing approaches. Workshop Ontol. Inf. Sharing, 2001: 108-117.

Xu, Z., S. Zhang and Y. Dong, 2006. Mapping between relational database schema and OWL ontology for deep annotation. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, December 18-22, 2006, Hong Kong, China, pp: 548-552.