Semantic Mapping Relational to Graph Model

[1,2]Dewi W. Wardani, [2]Josef Küng

[1]Informatics Department
Sebelas Maret University, Indonesia
[2]Institute for Application Oriented Knowledge Processing
Johannes Kepler University, Austria

***Abstract*— Making data to be more connected is one of the goals of Semantic Technology. Therefore, relational data model as one of important data resource type, is needed to be mapped and converted to graph model. In this paper we focus in mapping and converting without semantically loss, by considering semantic abstraction of the real world, which has been ignored in some previous researches. As a graph schema model, it can be implemented in graph database or linked data in RDF/OWL format. This approach studies that relationship should be paid more attention in mapping and converting because, often be found a gap semantic abstraction during those processes. In our small experiment shows that our idea can map and convert relational model to graph model without semantically loss.**

Keywords*— *mapping; converting; relational model; graph model; graph schema; semantic; big data.

I. Introduction

Semantic technology has been being discussed in many topics, in improving and supporting the current technologies since the latest decade. Some standards of this technology have been being implemented and progressed both in theory and practice. Since the researchers started the discussion, they have defined different point of view in this technology. In area of knowledge management research, recently this technology takes a big part not only practically in some domain [6-11], but also contributing in the core subject such in big data and new direction of data management [12-15].

In this paper we refer to the inventor of Web's idea, that one goal of Semantic technology is to create the web of data [16]. The data will be more equivalent with a thing, sometimes it's called the network of thing [17]. The idea is making data more connected, meaningful and understand able on the machine side. In both theory and practice, preparing data which ready to be used in semantic technology is the bottom block in this technology's stack. Basically, semantic technology uses graph graph-based model data [18-19] . The common and famous implementation is formed the data as linked data in RDF/XML [20-21]. Since 2003, W3C announced it as one of the standard of semantic technology.

In reality, a lot of data exists in many domains, in structured or unstructured type, but the model is not semantic technology friendly. Specially, the structured data in relational model schema, it has huge and broad-area user and has been used successfully for long time [1]. It usually contains high quality information, in term that contains main information in almost all application. This type of data can't be ignored to be part as data source of semantic technology. Concerning to its importance, it's needed to map and convert this relational model to graph model, which is semantic technology friendly. Therefore, in reaching the goal web of data which is more connected and meaningful, some previous works tried to map relational model directly to RDF linked data [2-5]. Even, W3C released the features of direct mapping relational model to RDF data model. It's really good way and framework, but in our opinion, one problem of direct mapping is the process often loss the semantic of the real world problem from relational model. Some other works, convert relational database model to graph database model. This approach usually uses naive converting process, which put all tuple in each relation as a node and the foreign key as an edge between two nodes, but not so many scientific papers are written for this approach. Almost all ended their work until relational database is converted in graph database and do not concern possibility to be extended as semantic technology's need. As far as our knowledge our work is the first which create property relationship in the result of the mapping and converting process and there were no previous works which consider relationship as an important in graph schema. The opposite, in our work we think relationship will be an important part in the network of data, and mining the relationship will be important in the near future.

Here, we propose mapping relational model to graph model which try to keep the semantic abstraction from the real world of relational model and also satisfied for the need of semantic technology. This approach can be said as the process to create semantic schema in graph model, which can be implemented both in graph database or linked data in RDF format. In this work we implemented it in graph database, but the schema will be easier to be mapped to linked data in RDF format because it uses property graph model. Once the data is

978-1-4799-4575-7/14/$31.00 ©2014 IEEE

connected, we don’t see data in relation format but in sub graph format. It means that we have not only network of data but also network of relationship. It also means, things are connected in semantic technology's point of view. Based on this step, with data which has semantically connected, we can build further works which explore not only the data but much more the relationship, such as Flexible Query Answering or General Recommendation.

This paper is organized as follows, after the introduction, in Section 2 presents a few closest related works then in Section 3 presents the proposed approach. In Section 4 explains the experiment and finally we summarize our work and proposes future work in Section 5.

II. Related Works

In this paper we focus in exploring schema of relational model and map it to graph schema without semantically loss and keep the possibility to be extended as linked data in RDF format. In [22], the author’s goal is improving the performance in querying the graph model data. The approach aggregates the populated data into one node as far as its possible in all possibility search query to cut the query traversal time. In our point of view, aggregating the data into one node might make the semantically loss. Node does not have specific semantic, node is only a bag for holding data. The other work [24], transform relational model to graph model based on dependencies between each attribute of the relation. In our point of view, the same with previous approach which the semantic of relation will be loss, and possibility of graph data redundancy. In general, we think that mapping process should not only focus in consuming the data but also can provide supporting for mining the relationship or other extended works that possibly only can work better by using the graph model data.

Some others works, convert relational model to RDF [25] [2], but almost all focus on converting to the populated data from relational database to RDF as linked data without concerning the global schema of relational database. This is what we mean that the converting process has potential of semantically loss. Moreover, RDF engines focus in managing individual triple, more than providing a graph model to represent the data. After some times and evaluation, W3C released the guide features of direct mapping in 2012, to cover some drawbacks, describes some potential conditions per relation or between relations. In our opinion, the rules still does not see a whole schema of relational model as internal schema. These works [3] [26], has implemented the recommendation, which also consider the foreign key. The other work [27], focus on maintaining the mapping evolution process, but the research's direction is not the concern of our work. In general, the previous works focus on mapping the data, attributes and relation but somehow, missed in considering the schema which is the representation of semantic abstraction of the real world.

III. The Approach

A. *Study of the Semantic of Relational Model*

It’s well known that the famous relational model is the result of mapping from Entity Relationship (ER) conceptual model or Extended Entity Relationship (EER) conceptual model. Basically, ER/EER is the representation of the real world problem. It has higher abstraction as the semantic of the real world. In the real world, entity is the representation of concept, thing or object and relationship is definition of connection among entities. In our opinion, it’s quite similar with graph concept, a node fits with entity, and an edge fits with the relationship. Hence, the concept of object-entity of ER/EER is much closer with the concept of thing in semantic technology terminology. In our opinion, mapping process from relational model to graph model also should consider ER/EER as higher abstraction conceptual design for making sure there is no semantically loss.

In database theory [23], the relation of relational model probably is mapped from entity side (strong entity and its attributes, multivalued-attributes, weak entity and its attributes, entity-subclass or entity-superclass), or from the relationship side (m:n-cardinality relationship, n-ary relationship). Only 1:1-cardinality and 1:n-cardinality relationships are mapped as the real relationship which is represented as Integrity Constraint (IC) in relational model. In mapping process from relational model to graph model, we think we also should consider this situation. In this beginning of our work, firstly we will focus studying the semantic abstraction from the relational side. As we mentioned before, there is no different semantically in 1:1-cardinality and 1:n-cardinality relationships, both are as relationship in ER/EER conceptual model and in relational schema model. We find semantically different between conceptual model and relational model, within m:n-cardinality and n-ary relationships. Both are relationship in ER/EER conceptual model, but as a relation in Relational Schema. Therefore, both should be mapped to graph schema differently. Figure 1 is the simple example of relational model database. We noticed the model is mapped from ER/EER conceptual model in Figure 2. The semantic abstraction in the real world, that CASTING is not a concept of thing, but the relationship between two things, ACTOR and MOVIE. The same situation will be found in n-ary relationship.

We also study from W3C recommendation's Direct Mapping and in our opinion some features of it are not

really possible happens in the real case, such as table which doesn't have primary key and reference table without primary key. Practically, if the relational database schema contains those features, it means that the relational database schema is not good enough, or not a valid schema and probably has some redundancies. Therefore, we put assumption that the relational database model is already valid and has a good quality schema.
We formulated what our approach in semantic of relational model within DEFINITION 1.

DEFINITION 1. Let ***S*** $(R_1, R_2 \ldots R_i)$ is a relational schema which consist a set of Relation, ***i*** is degree of relation, a set of Primary Key, ***PK*** $(PK_1, PK_2 \ldots PK_i)$ and a set of integrity constraints ***(IC)***. Each Relation consist of a set of attributes, ***R*** $(A_1, A_2 \ldots A_j)$, ***j*** is degree of attribute. ***r*** is relation instance of ***R***, is a set of tuple ***r*** $(r_1, r_2 \ldots r_i)$. Each ***n-tuple t*** is an ordered list of ***r*** values $t = \langle v_1, v_2 \ldots v_i \rangle$, each value v_n is an element of attribute A_j.

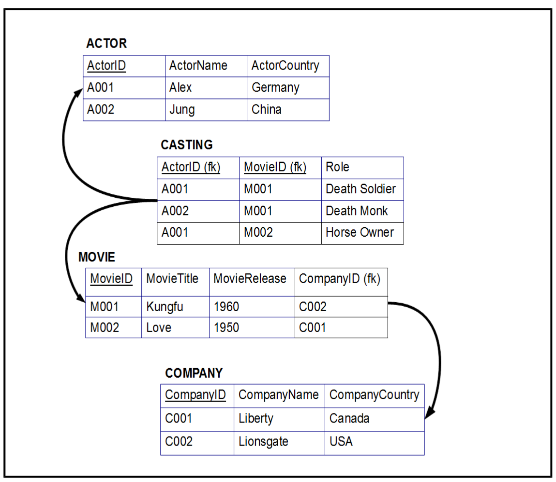

Fig 1. Relational data model schema.

B. Semantic Graph Model

From the example in Figure 1, the graph schema that commonly be used to map to graph schema is like in Figure 3, that a node is a relation and an edge is possible integrity constrains which connect between 2 nodes. In [22], might be different because it aggregates nodes and the schema much more describes all the possibilities in searching process. The goal of this work is improving the performance in consuming data. We think the approach breaks the semantic of the schema. It's hard to notice which type of semantic of the node belongs to. This will be another problem for further works which will not only consuming the data but also mining the data and relationship, or other further analysis which needs semantically consistent.

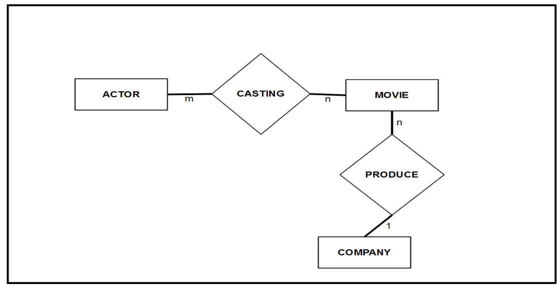

Fig 2. ER/EER conceptual schema of Figure 1.

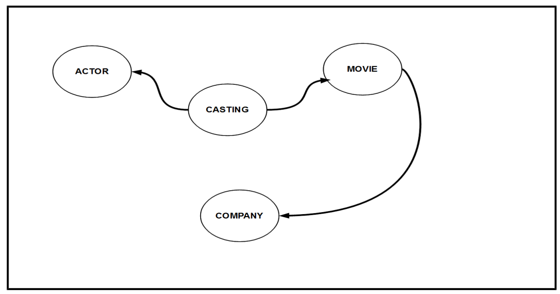

Fig 3. Naïve approach graph schema of Figure 1.

Considering ER/EER conceptual schema, we learned that m:n-cardinality is the relation which has more than one Foreign Key (FK), besides its own Primary Key (PK). Sometimes it has PK, sometimes doesn't have it. The same situation occurs in n-ary relationship. We formulated our approach in semantic graph model within DEFINITION 2 and DEFINITION 3.

DEFINITION 2. Let ***L*** is a set of link information of relation $L(L_1, L_2 \ldots L_i)$, each L_i is a list value L_i = **<*Inlink, Outlink*>**, and let ***SinkR, SourceR and HubR*** are a list of relation R_i which satisfy with ***L*** data. ***Inlink*** is the number of the relation which becomes referenced relation, and ***Outlink*** is the number of the relation refers to another relation. If ***Inlink*** = ***0***, the relation is Sink Relation (***SinkR***), if ***Outlink=0***, the relation is Source Relation (***SourceR***), otherwise the relation is Hub Relation (***HubR***). Each type of relation is a set of Relation R_i

Based on our DEFINITION 2, CASTING relation is a Sink Relation, ACTOR and COMPANY is a Source Relation and MOVIE is a Hub Relation

DEFINITION 3. Let ***G*** is semantic graph schema of relational schema ***S***, is a set of node ***N*** and edge ***E***, directed property graph G (N, E), where ***N*** is a set of node ***N*** $(N_1, N_2 \ldots N_i)$ and $N \in (SourceR \cup HubR)$. The edge, $(N_l, N_k) \in E$, if satisfied these situation: (i).there is FK between value of $A_j.N_l \xrightarrow{FK} A_j.N_k$, (ii). $(A_j.N_l \cap A_j.N_k) \in A_j.N_r$ and $N_r \in (SinkR)$

Our semantic graph for relational schema from Figure 1 is in the Figure 4. From DEFINITION 2, we can see that the graph is directed property graph. In practical, we can reduce some redundancy properties if it's needed. As shown in Figure 4, in edge **:casting** was not drawn as directed graph because it's bidirectional graph in the theory but in technical as directed graph. However, there is no conflict on it because, in practical even though directional graph but the process can work bidirectional and there is no reducing in performance. The PK or FK are used for creating connection between nodes, and normally, the real property information is existing in non-PK and non-FK. This is not a big problem, and only technical problem in implementation. The same situation but opposite, in edge **:movie_company**, basically doesn't have property, but we can add FK as a property if its needed depend on the needs. In Figure 5, is the example for n-ary relationship semantic graph schema.

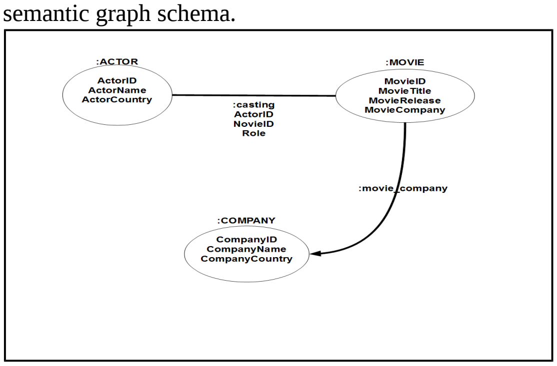

Fig 4. Semantic Graph Schema of Figure 1

C. Converting Relational Data Model to Graph Data Model

From the previous step, in Figure 4 we have obtained the semantic graph schema of relational model in Figure 1. Next step is converting the data of relational model to graph model. As far we already obtained the semantic graph schema, this step will not be complicated. We formulated DEFINITION 4 to explain the converting process of instance data from relational data to graph data.

DEFINITION 4. Let g is graph data, $\boldsymbol{g = (N,E)}$ which following G as a semantic graph schema. Each $\boldsymbol{N_r}$ is a set of name of key-value attributes of $\boldsymbol{R_i}$ and its attributes $\boldsymbol{N_r}$ ***(<A**$_j$ **>,<value>)***, and has a label ***:name of R**$_i$*. The edge, $\boldsymbol{(N_l,N_k) \in E}$, after satisfied the semantic graph schema then: (i). Edge is a key-value pair of from the attribute as a FK in $\boldsymbol{N_r}$, $\boldsymbol{E_m}$ ***(<A**$_j$**>,<value>)*** and edge has a label ***:name of N**$_l$**_name of N**$_k$*, (ii). Edge is a set of key-value pair of from $\boldsymbol{N_r}$, $\boldsymbol{E_m}$ ***(<A**$_j$**>,<value>)***, and has a label ***:name of R**$_i$*.

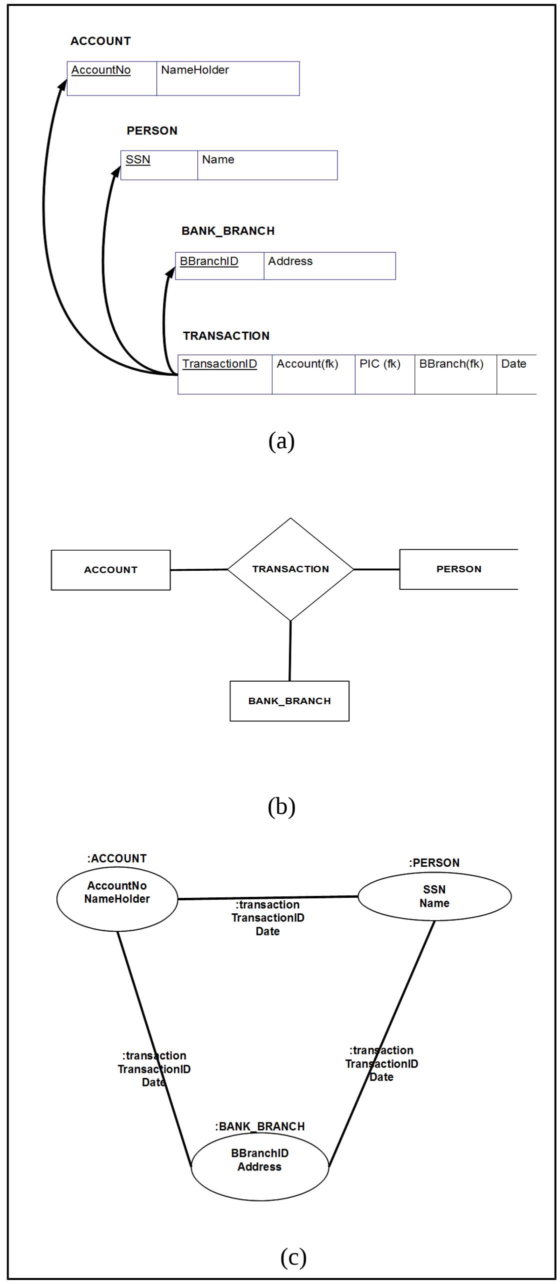

Fig 5. Example of for n-ary relationship (a) Relational Schema Model (b) ER/EER conceptual model and (c) Semantic Graph Schema

Figure 6 is our graph data, which the result of converting the relational data in Figure 1. We can see that this approach maintain the abstraction semantic of the real world and no semantically loss. There are 3 types of node: ACTOR, MOVIE and COMPANY, and in semantic abstraction those are the concept of thing as well. There are 2 types of edge: :casting and :movie_company and in the semantic abstraction both are relationship among things. We still see the act of redundancy properties edge :movie_company, but this redundancy isn't against the concept only causing the size is bigger. In practically, it can be eliminated easily, if its needed and necessary.

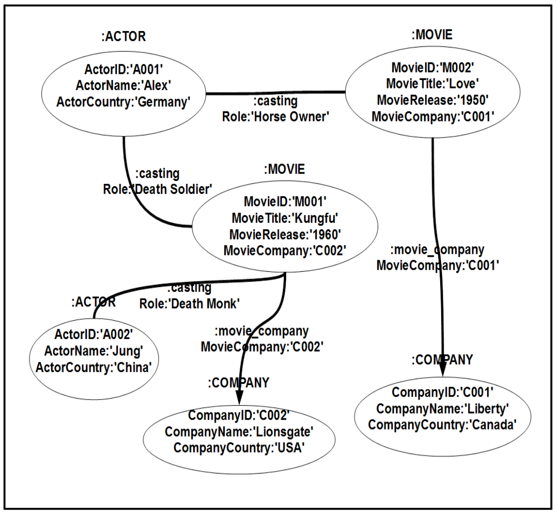

Fig 6. The Graph Data of Figure 1, using our proposed approach

As a property graph, both in node and edge, we can convert it to RDF linked data. Moreover we also consider the features of W3C recommendation. All we need is choosing the proper vocabulary or create our own vocabulary and change non-value string with URI. Some more works probably is in mapping relationships which has property.

IV. The Experiment

As our proposed idea is a new additional in term mapping relational to graph model, in this first step of our work we only want to show that our approach return the result of mapping without semantically loss. We compared with the result of example problem of this [22] work. It is shown in Figure 7, that relations USER, BLOG and COMMENT are nodes and FOLLOW and TAG are relationships. As we can see there is no semantically loss, no semantic gap with the real world as well. USER, BLOG and COMMENT are a thing and FOLLOW and TAG are much more an activity of thing, therefore it is much more as relationship. The result in [22] is the opposite, in our opinion the approach breaks the semantic of the schema.

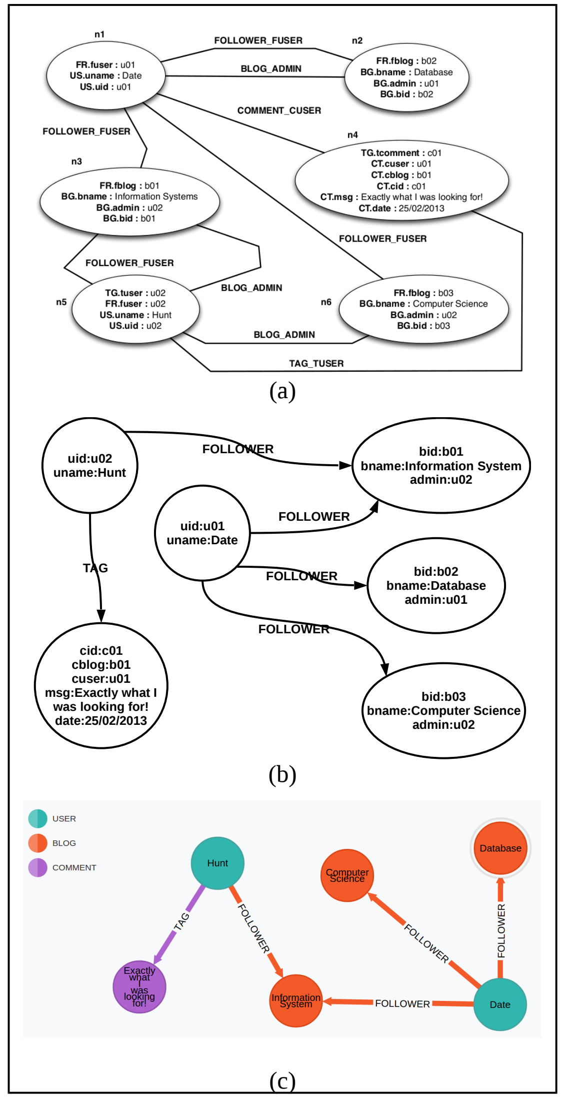

Fig 7. (a) an example from [22], and (b) is the result from our approach and (c) is the graph data model in the storage engine

We also tested IMDB dataset. It has 21 relations and the size is ~ 5 GB in plain text format. The result among those relations is 13 relations as node and 8 relations as edge. We did not use a whole instance data. In this first stage of work we implemented in localhost, and due the performance problem specially in relationship process we put the maximum number is only 1000 in each relationship. We used Neo4j engine as the graph storage. We tested 25 queries for making sure our approach is valid. Each query has various key searches, between 1 to 7. We used traditional matching query language. From 25 queries, returns 23 correct answers. 3 queries have no returns, because there is no data in our graph data which perfectly match with the queries.

V. Conclusions And Future Works

We have proposed semantic mapping, as a new simple approach in mapping and converting relational database schema to graph schema, which avoiding semantically loss, by considering semantic abstraction of real world. The other contribution is our approach will be also useful in making easier converting to linked data in RDF or OWL format. In this beginning work, the result of experiment using IMDB data shows that our approach can map and convert the relational data model to graph model without semantically loss.

Our near future works are divided into two directions. First, we are going to study more to improve this simple approach also cover situation from the entity side. We will also study to enrich meaning of the relationship by using inference approach included learning how it will support data stack of Semantic technology. Second, we are using this approach as fundamental of our further works which much more mining the relationship rather consuming the data. Our evaluation in testing queries, matching process is not the best way, along with complex queries and big data, the complexity process will too much costly. We will study how this approach will be useful in producing flexibility process in question answering. We believe, mining the relationship will give some insights in the mining process.

References

[1] Cafarella, M. J., Halevy, A. and Madhavan, J., "Structured data on the web", *Communications of the ACM* vol.54, no.2, pp. 72-79, 2011

[2] Zhou, S., "Exposing relational database as rdf", *Industrial and Information Systems (IIS) 2nd International Conference* IEEE, . Vol. 2, pp. 237-240. 2010.

[3] Sequeda, J. F., Arenas, M. and Miranker, D. P., "Directly Mapping Relational Databases to RDF and OWL", *In Proceedings of the 21st international conference on World Wide Web, ACM* , pp. 649-658, 2012.

[4] Bizer, C., "D2R MAP - A Database to RDF Mapping Language", 2003.

[5] Spanos, D.-E., Stavrou, P. and Mitrou, N., "Bringing relational databases into the semantic web: A survey", *Semantic Web* vol.3, pp.169-209, 2012.

[6] Antezana, E., Kuiper, M. and Mironov, V. . "Biological knowledge management: the emerging role of the Semantic Web technologies", *Briefings in bioinformatics* vol.10, no.4, pp.392-407, 2009

[7] Gábor, A. and Szabó, Z., "Semantic Technologies in Business Process Management", *Integration of Practice-Oriented Knowledge Technology: Trends and Prospectives*, Springer Berlin Heidelberg, pp. 17-28, 2013.

[8] Madkour, A., Aref, W. and Basalamah, S., "Knowledge cubes: A proposal for scalable and semantically-guided management of Big Data", *IEEE International Conference*, pp. 1-7, 2013.

[9] Samiotis, K. and Stojanovic, N., "Semantic-Based Knowledge Management in E-Government: Modeling Attention for Proactive Information Delivery"., *Semantic Technologies for E-Government*, Springer Berlin Heidelberg, pp. 237-260, 2010.

[10] Bösser, T., Casellas, N.; Melchior, E.-M., Thurlow, I. and Vallbé, J.-J., "User Quality and Business Benefit of Semantic Knowledge Management Applications", *Semantic Knowledge Management*, Springer Berlin Heidelberg, p.231-244, 2009.

[11] Xu, Y., "Knowledge Integration Based on Semantic Web for Drilling Risk Management"., *Information Computing and Applications*, Springer Berlin Heidelberg, vol.307, pp. 738-746, 2012

[12] Bäck, A. Vainikainen, S. Södergård, C. and Juhola, H.," Semantic Web Technologies in Knowledge Management.,"., ELPUB, 2003.

[13] Lim, L., Wang, H. and Wang, M., "Semantic queries in databases: problems and challenges", *Proceedings of the 18th ACM conference on Information and knowledge management*., pp. 1505-1508, 2009.

[14] Bouzid, S., Cauvet, C. and Pinaton. J., "A survey of semantic web standards to representing knowledge in problem solving situations", *Information Retrieval & Knowledge Management (CAMP), International Conference on IEEE*, pp. 121-125, 2012.

[15] Davies, J., Warren, P. and Sure, Y., "Semantic Technology and Knowledge Management: Foundations for the Web of Information and Services", Springer Berlin Heidelberg, 2011.

[16] Berners-Lee, T. and Fischetti, M., "Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor", Harper San Francisco, 1999.

[17] Berners-Lee, T., Hendler, J., Lassila, O. and others, "The semantic web", Scientific american, vol.284, no.5, pp. 28-37, 2001.

[18] Bizer, C.; Heath, T.; Ayers, D. and Raimond, Y., "Interlinking open data on the web", *4th European Semantic Web Conference, Innsbruck, Austria*, 2007.

[19] Bizer, C., Heath, T. and Berners-Lee, T., "Linked data-the story so far", *International journal on semantic web and information systems, vol.*5, pp. 1-22, 2009.

[20] Hayes, P. and McBride, B., "RDF Semantics. W3C Recommendation", *World Wide Web Consortium,* 10 February 2004.

[21] Manola, F., Miller, E. and McBride, B, "RDF primer, W3C recommendation", vol. 10, (2004)

[22] De Virgilio, R., Maccioni, A. and Torlone, R., "Converting Relational to Graph Databases", *First International Workshop on Graph Data Management Experiences and Systems*, ACM, pp. 1:1-1:6, 2013.

[23] Date, C. J., "An introduction to database systems. Addison-wesley Reading", MA, 1990.

[24] Bordoloi, S. and Kalita, B., "Designing Graph Database Models From Existing Relational Databases", *International Journal of Computer Applications, vol.*74, pp. 25-31, 2013.

[25] Rikacovs, S. and Barzdins, J., "Export of Relational Databases to RDF Databases: A Case Study", *Perspectives in Business Informatics Research*, Springer Berlin Heidelberg, pp. 203-211, 2010.

[26] Thuy, P. T. T., Thuan, N. D., Han, Y.; Park, K. and Lee, Y.-K., "RDB2RDF: Completed Transformation from Relational Database into RDF Ontology", *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication,* ACM , pp. 88:1-88:7, 2014.

[27] Konstantinou, N., Kouis, D. and Mitrou, N., "Incremental Export of Relational Database Contents into RDF Graphs", P*roceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14)*, ACM, pp. 33:1-33:8, 2014.