

# Semantic Data Mapping on E-Learning Usage Index Tool to Handle Heterogeneity of Data Representation

Arda Yunianta<sup>a,b\*</sup>, Norazah Yusof<sup>a</sup>, Mohd Shahizan Othman<sup>a</sup>, Abdul Aziz<sup>b</sup>, Nataniel Dengen<sup>b</sup>, Muhammad Ugiarto<sup>b</sup>, Haeruddin<sup>b</sup>, Joan Angelina<sup>b</sup>

<sup>a</sup>Faculty of Computing Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

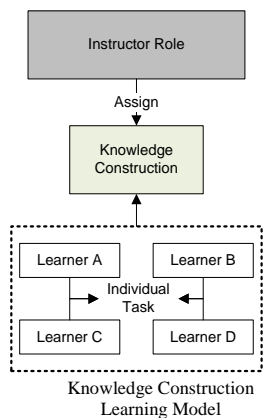
<sup>b</sup>Faculty of Information Technology and Communication, Mulawarman University, 75123 Samarinda, Indonesia

\*Corresponding author: yarda2@live.utm.my

## Article history

Received :5 March 2014  
Received in revised form :  
19 April 2014  
Accepted :3 May 2014

## Graphical abstract



## Abstract

Distribution and heterogeneity of data is the current issues in data level implementation. Different data representation between applications makes the integration problem increasingly complex. Stored data between applications sometimes have similar meaning, but because of the differences in data representation, the application cannot be integrated with the other applications. Many researchers found that the semantic technology is the best way to resolve the current data integration issues. Semantic technology can handle heterogeneity of data; data with different representations and sources. With semantic technology data mapping can also be done from different database and different data format that have the same meaning data. This paper focuses on the semantic data mapping using semantic ontology approach. In the first level of process, semantic data mapping engine will produce data mapping language with turtle (.ttl) file format that can be used for Local Java Application using Jena Library and Triple Store. In the second level process, D2R Server that can be access from outside environment is provided using HTTP Protocol to access using SPARQL Clients, Linked Data Clients (RDF Formats) and HTML Browser. Future work to will continue on this topic, focusing on E-Learning Usage Index Tool (IPEL) application that is able to integrate with others system applications like Moodle E-Learning Systems.

**Keywords:** Data mapping; D2RQ; learning environment; semantic ontology

## Abstrak

Proses pengagihan dan kepelbagaian data merupakan isu utama dalam peringkat implementasi data. Perbezaan dari segi perwakilan data antara aplikasi menjadikan masalah integrasi bertambah kompleks. Kadang-kala, proses penyimpanan data antara aplikasi berlaku berdasarkan persamaan maksud. Akan tetapi, faktor perbezaan dari segi perwakilan data menyebabkan aplikasi yang terlibat tidak dapat berhubung dengan aplikasi yang lain. Kebanyakan pengkaji menemukan faktor bahawa teknologi semantik merupakan penyelesaian terbaik untuk menangani isu terkini dalam integrasi data. Teknologi semantik mampu mengurus kepelbagaian data, perwakilan data yang berbeza dan juga data yang diperolehi dari sumber yang berbeza. Penggunaan teknologi semantik juga mampu untuk melakukan proses pemetaan data dari sistem pangkalan data yang berbeza, format data yang berbeza dengan maksud data yang sama. Kajian yang dilakukan ini ditumpukan kepada aspek pemetaan data semantik menggunakan pendekatan ontologi semantik. Dalam proses tahap pertama, enjin perhubungan data semantik akan menghasilkan pemetaan data dalam bentuk format turtle (.ttl) yang boleh digunakan dalam Local Java Application menggunakan Jena Library dan Triple Store. Dalam proses tahap kedua, matlamat utama proses ini adalah untuk mengeluarkan D2R Server yang boleh berhubung melalui persekitaran luar menggunakan HTTP Protocol untuk memperoleh akses kepada SPARQL Clients, Linked Data Clients (format RDF) dan HTML Browser. Titik tolak dari kajian terkini, kajian akan datang akan memfokuskan pada menghubungkan antara E-Learning Usage Index Tool (IPEL) kepada aplikasi sistem yang lain seperti sistem Moodle E-learning.

**Kata kunci:** Pemetaan data; D2RQ; lingkungan pembelajaran; semantik ontologi

© 2014 Penerbit UTM Press. All rights reserved.

## 1.0 INTRODUCTION

The heterogeneity of data is a common phenomenon in distributed information sources and ever growing with the development of

computer and information technologies that have created a huge amount of data and information [1, 2]. Heterogeneity of data, which is data with different representations and sources, is the other existing problem in current obsolescence management tools,

while data conflicts are also more common than data agreement [3, 4]. At the same time today's software systems develop are more distributed and more autonomous. Both of these trends are a natural reason for the intensive efforts in a domain of data integration.

Implementation of data integration still has a many problems to be solved. Exchanging and merging data from loosely coupled, heterogeneous data representation and mapping data on different data sources are the serious problem for data integration [5-11]. A lot of application integrations are implemented in the current days. Enterprise Application Integration (EAI) is the one of most famous integration application that is currently being implemented. EAI enables the enterprise to function more efficiently, provide better services for its customers and ensures faster realization of its business ideas. It also ensures quicker and more reliable communication of business information that supports the strategic and tactical business goals [12]. Enterprise Information Integration (EII) is the other integration application that has already been implemented in many organizations. EII is a service oriented architecture used to implement the integration process [13].

EAI and EII are developed under licensed software and still have a lot of weakness. Technical problems are the common issues encountered in the EAI and EII implementation [12, 13]. The first problem in the implementation of EAI and EII is closed platform that cannot be customized with specific institution. Second problem, EAI and EII are licensed software and very difficult to implement in education institutions. The most significant problem is that EAI and EII cannot do mapping on heterogeneous data in different data source that have a same meaning data/information.

This research utilizes E-Learning Usage Index Tool (IPEL) application. IPEL is a one of the application software on learning environment that is able to measure the e-learning usage for both the instructors and learners. In the learning environment, there are three parts; constructive alignment that have relationship with each other to support learning process, they are teaching and learning process, assessment task and learning outcomes [17]. To obtain learning knowledge from constructive alignment, different learning application system and different data learning source must be integrated. In this case, we face heterogeneity data representation in distributed data source. To solve this problem semantic data mapping is used to integrate numerous data sources. Semantic data mapping is able to handle communication and integration data/information with different data representation in different data source that have same meaning data/information [4].

Researchers use Semantic ontologies extensively in E-Learning Usage Index Tool (IPEL) application to annotate their data, to drive decision-support systems, to integrate data and to perform natural language processing and information extraction. Ontologies provide a means of formally specified complex descriptions and relationships about information in a way that is expressive yet amenable to automated processing and reasoning [14-16]. As such, they offer the promise of facilitated information sharing, data fusion and exchange among many distributed and possibly heterogeneous data sources [4].

However, the focus of this paper is to produce Semantic data mapping in E-Learning Usage Index Tool (IPEL) to be integrated with other applications that have the same data/information meaning. In the future, IPEL will be integrated with the other learning system to communicate and collaboration on specific data that have the same meaning to produce Decision Support System. We claim that the semantic data mapping can produce better data utilization. We also argue that semantic data mapping

can handle heterogeneous data with different representation that have the same meaning data/information.

In this paper, we produced semantic data mapping on E-Learning Usage Index Tool (IPEL) application with several parts of process. The first part is building semantic data mapping architecture and drawing the IPEL data source structure and relationship. The second part is creating data mapping language, creating D2R server to communicate and integrating with the other systems from outside environment and implementing D2RQ engine and Jena Library to communication with local application.

## 2.0 SEMANTIC DATA MAPPING METHODOLOGY

Generally, semantic data mapping is the relationship between four parts that are important to semantic data mapping and integration of data. The core part is the semantic data mapping that will handle communication and integration with the other three parts. The second part is IPEL data source that will be mapped in semantic data mapping. The third part is local application that uses semantic data mapping. And the fourth part is the other system that will communicate and integrated from outside environment using HTTP Protocol. Semantic data mapping architecture is illustrated in Figure 1.

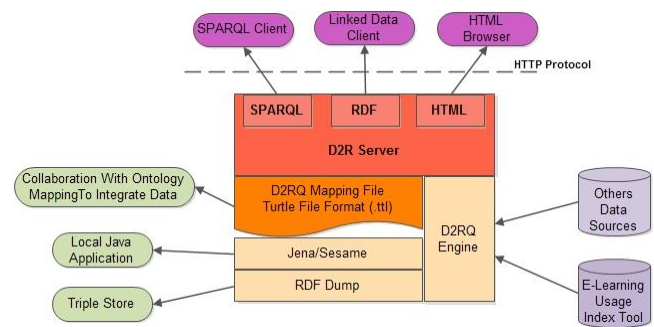


Figure 1 Semantic data mapping architecture [18]

The mapping defines a virtual RDF graph that contains information from the database. This is similar to the concept of views in SQL, except that the virtual data structure is an RDF graph instead of a virtual relational table. The virtual RDF graph can be accessed in various ways, depending on what is offered by the implementation. The D2RQ Platform provides SPARQL access, a Linked Data server, an RDF dump generator, a simple HTML interface, and Jena API access to D2RQ-mapped databases.

In the semantic data mapping, there are three important parts that we can see on Figure 1. The first part is D2RQ engine that is the core part in semantic data mapping process. D2RQ engine is responsible to communicate with local data source and produce D2RQ data mapping file that can be used to communicate with local application using jena library and RDF Dump. The second part is D2R server to communicate and integrate with the others system from outside environment using HTTP Protocol. This part will produce SPARQL that can be access from SPRQL Clients, RDF that can be access from linked data clients and HTML that can be access from HTML browser [18]. In the third part, which is D2RQ, the data mapping file is in text mode file with turtle file format (.ttl) that contains data mapping from local data source based on ontology based language. The D2RQ Mapping Language is a declarative language for describing the relation between a relational database scheme and RDFS vocabularies or

OWL ontologies. A D2RQ mapping is itself an RDF document written in Turtle syntax. The mapping is expressed using terms in the D2RQ namespace. Namespace is a domain that serves to guarantee the uniqueness of identifiers. It is written like uniform resource locator (URL). For example, <http://www.wiwiss.fu-berlin.de/suhl/bizer/D2RQ/0.1#>. The terms in this namespace are formally defined in the D2RQ RDF scheme (Turtle version, RDF/XML version).

## 2.1 Knowledge Construction Interaction

IPEL is a software tool that is able to measure the e-learning usage for both the instructors and learners. It is a tool for the administrator of the CTL (Center for Teaching and Learning) to filter the data log from Moodle e-learning system. IPEL system has seven main functions that produce decision support system

process. The first function is the overall result showing the overall hits owned by students and lecturers per course subject. The second is student function that shows the detail activities hits and actions hits per course subject owned by students. The third is the lecturer's function showing the detailed activities hits and actions hits per course subject owned by the lecturers.

Fourth is the activity score student function that shows the detail activities student hits after give a meaningful score weight. Fifth is activity score lecturer function to show detail activities lecturer hits after give a meaningful score weight. Sixth is Action score student that show detail activities and actions student hits after give a meaningful and active-passive score weight. Seventh is the action score lecturer that show detail activities and actions lecturer hits after give a meaningful and active-passive score weight. The data structure relationship can be seen in Figure 2.

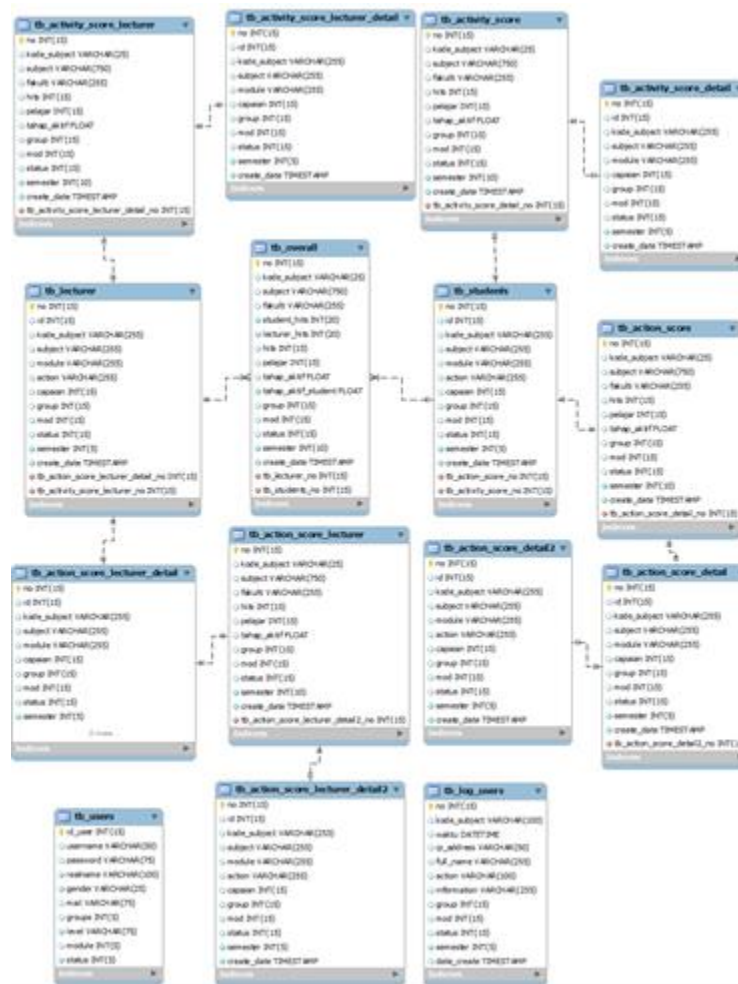


Figure 2 IPEL database structure and relationship [3]<sup>3</sup>

In the IPEL database development, there are thirteen main data tables that save personal data by using on the IPEL application. `Tb_overall` is the central table that is for the relationship with `tb_lecturer` and `tb_students`. `Tb_overall` contains general data about student and lecturer. In the `tb_overall`, we can see comparison data hits between students and lecturers. But if we want to see detailed activities and actions hits students and lecturer, we can go to `tb_lecturer` and `tb_students`. Each table, `tb_lecturer` and `tb_students`, has

relationship with two other tables. `Tb_lecturer` relationship with `tb_activity_score_lecturer` and `tb_action_score_lecturer_detail`. And `tb_students` relationship with `tb_activity_score` and `tb_action_score`.

Generally IPEL system contains four main data that give a significant result on that system. The first data is course subject data that contain subject code, name of subject and semester. The second is students' data that contains students' activities, students' actions and students' access hits. The third data is the

lecturer data that contains lecturers' activities, lecturers' actions and lecturers' access hits. That fourth data is the access hits owned by students and lecturers. These four main data are separated into thirteen tables to support a seven function on IPEL system.

### 3.0 SEMANTIC DATA MAPPING IMPLEMENTATION PROCESS

Implementation of semantic data mapping focuses on three parts. Data mapping language is the important part to create data mapping from data source into ontology language. D2R server is the other part to produce communication and integration with other system using HTTP protocol, whereas D2RQ engine is to communicate with local application using Jena API.

#### 3.1 Data Mapping Language

D2RQ data mapping file is text mode file with turtle file format (.ttl) that contain data mapping from local data source based on ontology based language. The D2RQ Mapping Language is a declarative language for describing the relation between a relational database scheme and RDFS vocabularies or OWL ontologies. A D2RQ mapping is itself an RDF document written in Turtle syntax. The mapping is expressed using terms in the D2RQ namespace. Namespace is a domain that serves to guarantee the uniqueness of identifiers, written like uniform resource locator (URL). For example <http://www.utm.my/exercise/ipel#>. The terms in this namespace are formally defined in the D2RQ RDF schema (Turtle version, RDF/XML version). This is the sample of data mapping file in turtle file format names Ipel.ttl

```
@prefix map: <#> .
@prefix db: <> .
@prefix vocab: <vocab/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix d2rq: <http://www.wiwiss.fu-berlin.de/suhl/bizer/D2RQ/0.1#> .
@prefix jdbc: <http://d2rq.org/terms/jdbc/> .
@prefix flight: <http://www.utm.my/exercise/ipel#> .
```

```
map:database a d2rq:Database;
  d2rq:jdbcDriver "com.mysql.jdbc.Driver";
  d2rq:jdbcDSN "jdbc:mysql://localhost/ipel";
  d2rq:username "root";
  d2rq:password "";
  jdbc:autoReconnect "true";
  jdbc:zeroDateTimeBehavior "convertToNull";
.

# Table tb_action_score
map:tb_action_score a d2rq:ClassMap;
  d2rq:dataStorage map:database;
  d2rq:uriPattern
"http://www.utm.my/exercise/ipel#Tb_action_score@@@tb_action_score.no@@@";
  d2rq:class vocab:tb_action_score;
  d2rq:classDefinitionLabel "tb_action_score";
.

map:tb_action_score_kode_subject a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:tb_action_score;
  d2rq:property vocab:tb_action_score_kode_subject;
  d2rq:propertyDefinitionLabel "tb_action_score_kode_subject";
  d2rq:column "tb_action_score.kode_subject";
.
```

In the first nine rows, ipel.ttl describes a prefix names to locate the address namespace of each part with different name prefix. The important part is prefix flight that describes the namespace of ontology language address. And after that, ipel.ttl describes the detail of database, including database driver, database connection, database username and database password. After finishing describing the database, the next rows describe the table mapping properties. On the table mapping, detail of field/column that owned of every table will be described.

#### 3.2 D2R Server

D2R server is used to communicate and integrate with the other systems from outside environment using HTTP Protocol. From this page, we can see direct table mapping by clicking the name of table that we want to see. From this page, it can also produce SPARQL that can be accessed from SPARQL Clients, RDF that can be accessed from linked data clients and HTML that can be accessed from HTML browser. Display of D2R server can be seen in Figure 3.



Figure 3 D2R server IPEL data mapping

### 4.0 RESULTS

#### 4.1 Visualization Data Mapping Result

Data mapping process produce two kinds of results, that are D2RQ mapping file with turtle file format and D2R server. D2RQ mapping file is used to implement local system and D2R server is used to implement external system via internet connection using HTTP protocol. D2RQ mapping file and D2R server is a representation of table scheme from IPEL database system to generate into standard formats that can be used by the other system to perform communication and integration with local database system. Figure 4 shows the visualization of data mapping process, for example one table on IPEL database system.

Figure 4 shows the data mapping result from table `tb_action_score` to represent D2RQ mapping file and D2R server. D2R server will communicate and integrate with the other systems through internet connection using HTTP protocol and D2RQ mapping file will implement using Jena API using Java language programming to collaborate with ontology knowledge to perform semantic data integration approach.

#### 4.2 Implement D2RQ Engine and Jena API

D2RQ engine is used to communicate with local application using Jena API. Jena API is the specific library for java programming language to develop application. D2RQ engine collaborates with Data mapping file (.ttl) and ontology (.owl) to



get better data utilization. Data mapping file facilitate local application to use the other data mapping from different data

sources. Implementation of D2RQ engine and Jena API can be seen in Figure 5.

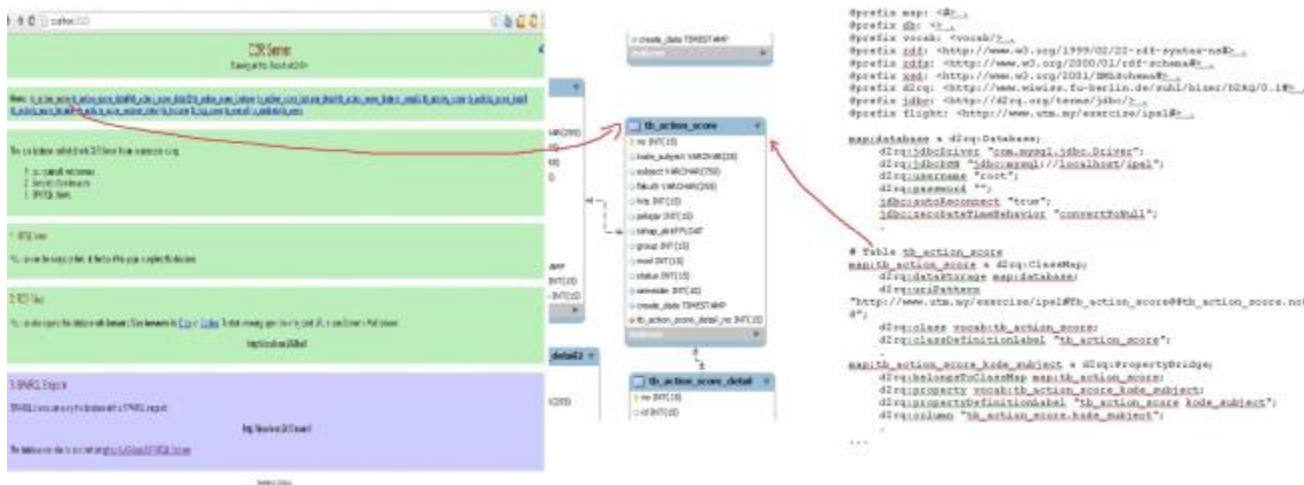


Figure 4 Table mapping result

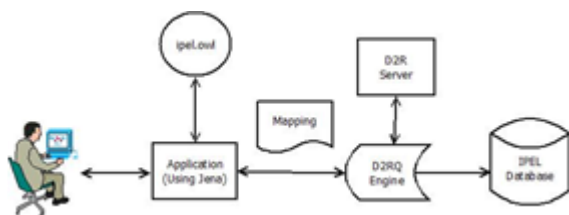


Figure 5 Using Semantic Mapping with Jena API

In this implementation, we can see the E-Learning Usage Index Tool (IPEL) application using PHP language programming to communicate and integrate with other applications using Java language programming. For future development, this application can be used to communicate and integrate with other data sources. This implementation can also solve a heterogeneous data representation that has the same data/information meaning. This is an example of java program that uses Jena API to implement D2RQ engine. We can see on this program, java using ipel.ttl and ipel.owl to using and manipulate data from IPEL data source.

```
package com.semantic.example.d2rq;
import com.hp.hpl.jena.ontology.Individual;
import com.hp.hpl.jena.ontology.OntClass;
import com.hp.hpl.jena.ontology.OntModel;
import com.hp.hpl.jena.ontology.OntModelSpec;
import com.hp.hpl.jena.ontology.OntProperty;
import com.hp.hpl.jena.ontology.OntResource;
import com.hp.hpl.jena.rdf.model.ModelFactory;
import com.hp.hpl.jena.util.FileManager;
import com.hp.hpl.jena.util.iterator.ExtendedIterator;

import de.fuberlin.wiwiss.d2rq.jena.ModelID2RQ;

public class D2RQModelExample {
    private static final String URI = "http://
www.utm.my/exercise/ipel#";

    public static void main(String[] args) {
```

```
ModelID2RQ dbmodel = new
ModelID2RQ("C:\\PersonalProject\\D2RQ\\d2rq-0.8.1\\ipel.ttl");
OntModel model =
ModelFactory.createOntologyModel(OntModelSpec.OWL_DL_ME
M_RULE_INF);
model.setNsPrefix("ipel", "http://
www.utm.my/exercise/ipel#");
model.add(dbmodel);

model.read(FileManager.get().open("data/ipel.owl"), null);
model.setDynamicImports(true);

model.write(System.out);
```

5.0 CONCLUSIONS AND FUTURE WORK

Application systems and growing data produced more complex heterogeneity of data. Optimization and efficiency on data utilization is now the current issue. Weaknesses that still exist on current data integration application makes semantic data integration more popular in these days. The one part in semantic data integration that has very important effect is semantic data mapping. Semantic data mapping provides the solution for heterogeneous data with different data representation in different data source that have the same meaning data/information. Semantic data mapping is one of the advantages of the semantic data integration technology. This paper produces semantic data mapping result that can be used to communicate and integrate internal and external applications to share, utilize and manipulate the data sources. Future work will continue on this research, focusing on the application of E-Learning Usage Index Tool (IPEL) that is able to integrate with other system applications on learning environment like Moodle E-Learning Systems.

References

[1] Kashyap, V., Sheth, A. 1997. Semantic Heterogeneity in Global Information Systems: The Role of Metadata, Context and Ontologies.

- In M.P. Papazoglou & G. Schlageter (Eds.). Cooperative Information Systems San Diego: Academic Press. 139–178.
- [2] Kim, W., Seo, J. 1991. Classifying Schematic and Data Heterogeneity in Multi Database Systems. *IEEE Computer*. 24(12): 12–18.
- [3] Sandborn, P., Terpenney, J., Rai, R., Nelson, R., Zheng, L., Schafer, C. 2011. Knowledge Representation and Design for Managing Product Obsolescence. In Proceedings of NSF Civil, Mechanical and Manufacturing Innovation Grantees Conference. Atlanta, Georgia.
- [4] LePendu, P., Dou, D. 2011. Using Ontology Databases for Scalable Query Answering, Inconsistency Detection, and Data Integration. *Springer Science Business Media*. 37: 217–244.
- [5] Arenas, M. and Libkin, L. 2005. XML Data Exchange: Consistency and Query Answering,” in Proc. of the 24th ACM SIGMOD Symposium on Principles of Database Systems, PODS 2005, ACM.
- [6] Bonifati, A., Chrysanthis, P., Ouksel, A. and Satter, K-U. 2008. Distributed Databases and Peer-to-Peer Databases: Past and Present. SIGMOD Record. 37: 1.
- [7] Bouquet, P., Serafini, L. and Zanobini, S. 2004. Peer-to-peer Semantic Coordination. *Journal of Web Semantics*. 2(1): 81–97.
- [8] Calvanese, D., Giacomo, G., Lenzerini, M. and Rosati, R. 2004. Logical Foundations of Peer-To-Peer Data Integration. In Proc. of the 23rd ACM SIGMOD Symposium on Principles of Database Systems, PODS 2004, ACM. 241–251.
- [9] Fagin, R., Kolaitis, P. and Popa, L. 2005. Data Exchange: Getting to the Core. *ACM Trans. Database Syst.* 30: 1
- [10] Pankowski, T. 2006. Management of Executable Schema Mappings for XML Data Exchange. In Database Technologies for Handling XML Information on the Web, EDBT 2006 Workshops, LNCS 4254, Springer. 264–277.
- [11] Pankowski, T. 2008. XML Data Integration in SixP2P-a Theoretical Framework. Data Management in P2P Systems. *ACM*. 11–18.
- [12] Ana, C., Kresimir, F. 2009. EAI Issues and Best Practices. Proceedings of the 9th WSEAS International Conference on Applied Computer Science. 135–139.
- [13] Kong, Z., Wang, D., Zhang, J. 2007. A Strategic Framework for Enterprise Information Integration of ERP and E-Commerce. International Federation for Information Processing. 254: 701–705.
- [14] Bellatreche, L., Dung, N. X., Pierra, G., Hondjack, D. 2006. Contribution of Ontology-based Data Modeling to Automatic Integration of Electronic Catalogues within Engineering Databases. *Computers in Industry*. 57.
- [15] Castano, S., Antonellis, V., Vimercati, S. D. C. 2001. Global Viewing of Heterogeneous Data Sources. *IEEE Transactions on Knowledge and Data Engineering*. 13(2): 277–297.
- [16] Chen, Y. 2010. Knowledge Integration and Sharing for Collaborative Molding Product Design and Process Development. *Computers in Industry*. 61: 659–675
- [17] Biggs, J. B. 1999. *What the Student Does: Teaching for Quality Learning at University*. Buckingham: Open University Press.
- [18] Cyganiak, R., Bizer, C., Garbers, J., Maresch, O., and Becker, C. 2012. The D2RQ Mapping Language. v0.8 – 2012-03-12. Retrieved 2, 2012.