

October 2011

Query Reformulation: Data Integration Approach to Multi Domain Query Answering System

Niharika Pujari

*Department of Information Technology Institute of Technical Education & Research Siksha O Anusandhan
University Bhubaneswar, Odisha, India, niharika.pujari06@gmail.com*

Debahuti Mishra

*Institute of Technical Education and Research, Siksha O Anusandhan University, Bhubaneswar, Odisha,
India, debahuti@iter.ac.in*

Kaberi Das

*Department of Computer Applications Institute of Technical Education & Research Siksha O Anusandhan
University Bhubaneswar, Odisha, India, kaberidas@rediffmail.com*

Follow this and additional works at: <https://www.interscience.in/ijcct>

Recommended Citation

Pujari, Niharika; Mishra, Debahuti; and Das, Kaberi (2011) "Query Reformulation: Data Integration Approach to Multi Domain Query Answering System," *International Journal of Computer and Communication Technology*. Vol. 2 : Iss. 4 , Article 1.

DOI: 10.47893/IJCCT.2011.1094

Available at: <https://www.interscience.in/ijcct/vol2/iss4/1>

This Article is brought to you for free and open access by the Interscience Journals at Interscience Research Network. It has been accepted for inclusion in International Journal of Computer and Communication Technology by an authorized editor of Interscience Research Network. For more information, please contact sritampatnaik@gmail.com.

Query Reformulation: Data Integration Approach to Multi Domain Query Answering System

Niharika Pujari

Department of Information Technology
Institute of Technical Education & Research
Siksha O Anusandhan University
Bhubaneswar, Odisha, India
Email: niharika.pujari06@gmail.com

Debahuti Mishra*

Department of Computer Sc. & Engineering
Institute of Technical Education & Research
Siksha O Anusandhan University
Bhubaneswar, Odisha, India
Email: debahuti@iter.ac.in
*Corresponding author

Kaberi Das

Department of Computer Applications
Institute of Technical Education & Research
Siksha O Anusandhan University
Bhubaneswar, Odisha, India
Email: kaberidas@rediffmail.com

Abstract: Data integration gives the user with a unified view of all heterogeneous data sources. The basic service provided by data integration is query processing. Whatever query posed to the system is being given to global schema which has to reformulate to sub queries that are to be posed to the local sources. Reformulation is being accomplished by mapping between global and local sources by Global-as-View (GAV), Local-as-view (LAV) and Global-local-as-view (GLAV) approach. When a query involves multiple domains, it is difficult to extract information in case of general service engines.

Keywords: Data Integration; GAV, LAV; GLAV

1. Introduction

Data plays a very crucial role in today's real world, but there exist an increasing number of data sources. Hence there is a need to integrate data and the system that handle those data. Data integration provides users with a collective view of the data by combining data from different data sources. In variety of situations the process of data integration has become noteworthy like in commercial sector when companies need to combine the data base and in scientific sector when there is a need to merge the results of different research of Bioinformatics [4]. With the increase volume of data and need to share data explodes, data integration has become appears to be more persistent. Data integration is frequently referred to as "Enterprise Information Integration" (EII) in management circles.

Issues with combining heterogeneous data sources under a single query interface have existed for some time. The increasing use of databases has led to the need to share or to integrate existing repositories. At several levels in the database architecture, merging of these repositories could take place. Data warehousing has been popular solution to above issue. The Data warehouse is refreshed using ETL (extract, transform, load) [1] from several sources and into representing in a single queryable schema. Since the data reside in a single repository at query-time, so it offers a tightly coupled approach. Tight coupling can arise a need for re-execution of ETL process since it arises a problem related to the "freshness" of data, that is when an updating is carried out in original source,

but the warehouse still contains the older data. Data warehousing is considered where Data warehousing is impractical, over expensive. Two main concepts related to architecture of data integration are mediators and wrappers. The two basic approaches of data integration are Global-as-view (GAV) and Local-as-view (LAV) [11]. There is a need to have relationship between global schema and local sources that is mapping between global and local schema. The mapping is made possible by GAV and LAV. GAV approach has better query processing capabilities than LAV but LAV has better extensibility than GAV. Ignoring the limitations GAV and LAV, considering merits of both GLAV approach best suitable for answering queries from multiple domains or cross domain queries. For example:

“Find all database conferences that are held within one year in locations whose seasonal average temperature is 25°C and for which a cheap travel solution exists”,

2. Preliminaries

2.1 Data Integration

Data integration can be expressed in terms of triple $\langle G, L, M \rangle$ where

- G is the global schema
- L is the local sources/source schema.
- M is the mapping between the global and local schema.

$$\begin{aligned} Q_L &\rightarrow Q_G, \\ Q_G &\rightarrow Q_L \end{aligned}$$

Where, Q_L and Q_G are the two queries that query over source schema and query over global schema G that has the same arity. Queries Q_L can be expressed in terms of a query language $L_{M,S}$ over the alphabet AS , and queries Q_G can be expressed in terms of a query language $L_{M,G}$ over the alphabet A_G . Intuitively, an assertion $Q_L \rightarrow Q_G$ denotes that the concept represented by the query Q_L over the sources corresponds to the concept in the global schema represented by the query Q_G (similarly for an assertion of type $Q_G \rightarrow Q_L$). Queries are posed in terms of global schema which is represented through mediator but extraction of data is actually done from the local sources where the real data exist. This process needs query reformulation of the user query to subdivide into parts in the form that could be distributed to the local sources. This reformulation and identifying data sources and distributing queries to their relevant data sources are carried out query planner, while this total process come into picture by execution engine.

2.2 Motivation for Data Integration

Computing technology has evolved as the basic operating prototype for data processing over past twenty years which itself has changed. There has been revolutionary changes from mainframe-based, centralized data management systems to networks of powerful PC clients, group servers, and the Internet, while currently res moves to more extreme ,peer-based model that provide data in a decentralized architecture. Motivation for these changes has come from the need of decentralize control and administration of data rather than from more advanced hardware and networking technologies [5]. In terms of scaling performance centralized control not only forms bottleneck but in terms of administration also centralized computing control becomes scalability bottleneck. It is very difficult to design centralized schema for the data from different heterogeneous data sources while at the same time decentralized collection of autonomous systems can be much dynamic since individual components can be designed and redesigned to make the users available with the required result [1].

Unfortunately, a decentralized control of heterogeneous data that a common data management model suffers from shortcomings that is there is no central place from which data could be at a single query that could be analyzed in a comprehensive form. Uniformity and global perspective is provided by the centralized computation model while flexibility is provided by decentralized computation model. Data warehousing and virtual data integration are two solutions that have been proposed to this

problem, both of which are end-points along a broad continuum of possible implementations: both approaches develop a single mediated schema for that domain by taking decentralized data sources. Then the relationship between each data source and the mediated schema can be described by specifying each data source and the mediated schema.

2.3 Query Processing for Data Integration

The researchers mainly emphasizes on developing models[2],[3],[4], mappings[5],[6],[7] and wrappers[8],[9] of data integration for data sources, along with the problem of reformulating queries over the mediated schema into queries over the local sources. There are existing algorithms and methods for data integration, and in order to make data integration a widespread technology, two challenges that are to be addressed are: correspondence between entities at different data source while other one is to develop techniques to enable the data integration to be most useful practice in today's real world. Our focus is on the second problem, building techniques for efficiently answering data integration queries that have been posed by the user. Traditional query processing is used to statistically optimize the query and also deals with an environment in which statistics are computed offline. This generally is effective in a standard database environment because the data and computing environment are under the strict control of the database, and they are thus fairly predictable and consistent. Yet even in this context, many simplifying assumptions must be made during optimization, and there are many situations in which the traditional model does poorly (e.g., in many circumstances, the optimizer accumulates substantial error in modeling complex queries)

The data integration system is harder than conventional database management system due to its advanced features. Data integration communicates with different heterogeneous data sources and external networks for extraction of data but evaluation of performance of the data is very difficult to model. Hence two norms of remote querying are unpredictability and inconsistencies. Traditional query processing optimize for complete results by batch oriented queries. Data integration applications suggest the need for a different metric in query optimization in order to make it more interactive and also suggest at least a different set of heuristics.

2.4 Data Integration Approaches

Data integration comprises of several different issues because of the complex task in its designing. One of the most important issues is building mapping between the global schema and local schema and answering queries that are posed in the global schema [8]. One of the basic services that the data integration provides is the query processing. In order to answer queries that are being posed in the global schema, redesigning of the query is needed in order to get sub queries in terms of local sources. Redesigning needs mapping between each element of global schema to the elements of local schema and mapping function could be accomplished by two basic approaches of data integration: Global-as-view (GAV) and Local-as-view (LAV) approach [11]. Global-as-view need that global schema are described as view over the local sources while Local-as-view need that local sources are expressed as view over global sources and global schema are placed independently of the local sources. The relationship between global and local sources is built by specifying information content of each source element as view over global sources. Global-as-view work through creation of mediator which acts as a interface between user and local sources. Global-as-view offers better query processing capability because of unique technique that is *Rule Unfolding* but very poor performance in scalability. Local-as-view provides easier extensibility since adding new data sources does not create problem but query processing is quite difficult in LAV.

A. Global-as-View (GAV) Approach

In GAV approach, the mapping between global schema and local schema can be established by mapping each element g of global schema as view over local sources. Mapping associates each element of global schema G as query Q_L over local sources L . Thus mapping is an assertion in the form of: $g \rightarrow Q_L$

In the modelling point of view, GAV is mapping each element g of global source G as a view over local sources. GAV approach tells explicitly how to retrieve data from local sources and have the advantage of query processing which is made possible through the rule unfolding but adding new sources is difficult to achieve since it need to redefine the views of global schema.

Example of GAV: Different local sources are integrated in mediator and we have to decide what data interface or schema needs to offer to outside world.

Local sources: $S1(Cid, name, Branch)$, $S2(Cid, name, Branch)$ and $S3(Name, Review)$

Sources: $S1(Cid, name, Branch)$, $S2(Cid, name, Branch)$ and $S3(Name, Review)$

Global view

$S1$ union ($S2$ join $S3$)

SELECT * FROM $S1$

UNION

SELECT $S2.Cid, S2.name, S2.branch, S3.review$

- A global relation “containing” Customer name and their Branches:

- $CustBr(Name, Branch) \leftarrow S1(Cid, name, branch)$

- $CustBr(Name, Branch) \leftarrow S2(Cid, name, branch)$

- A Disjunctive Query (defined as disjunction of conjunctive query).

Defined by two Datalog rules

$CustBr$ defined as the union of two projections, of $S1$ and $S2$ on attributes $Name, Branch$, i.e.

- $CustBr := \Pi Name, Branch(S1) \cup \Pi Name, Branch(S2)$

- Another global relation “containing” Name, Branch and rev

- $CustRev(Cid, Name, Review) \leftarrow S1(Cid, Name, Branch), S3(name, Review)$

SELECT $S1.Cid, s1.name, s3.Review$

FROM $S1, S3$ WHERE $S1.name = S3.name$;

- A Conjunctive Query (defined in terms of conjunction)

- In relational algebra:

$CustRev(Cid, name, Review) := \Pi Cid, name, Review(S1_name JOIN S2)$

$CustRev(Cid, Name, Review) \leftarrow S1(Cid, Name, Branch), S3(Name, Review)$

defines a conjunctive query, i.e. it is expressed in terms of conjunctions (or joins) projections.

- With the basic concept of Global-as-view, it can be interpreted to find the solutions to a query posed by the user to sub queries in terms of local sources

- Query: Customer of Branch ‘Perryridge’ with their reviews

$Result(Name, Review) \leftarrow CustBr(Name, 'perryridge'), CustRev(Cid, Name, Review)$

(Query is expressed in terms of global schema)

- Query can be rewritten in terms of local schema which is simple under GAV: rule unfolding “Unfold” each global relation into its definition in terms of the local relations

$Ans(Name, Review) \leftarrow S1(Cid, name, 'perryridge'), S2(Cid, name, 'perryridge'), S3(Name, Review)$

These new queries do get answers directly from the sources

- Final answer is the union of two answer sets, one for each rule by rule unfolding

$Result(Title, Review) \leftarrow S1(Cid, name, 'perryridge'), S2(Cid, name, 'perryridge'), S3(Name, Review)$

B. Local-as-View (LAV) Approach

In LAV approach [11], the mapping between global schema and local schema can be established by mapping each element l of global schema as view over local sources. Mapping associates each element of local schema L as query Q_G over global sources G . Thus, mapping is an assertion in the form of: $l \rightarrow Q_G$

In the modelling point of view, GAV is mapping each element l of local source L as a view over global sources. LAV approach is somewhat impractical since it is the situation where view contains the real data rather than global virtual and is poor in query processing since it does not the

simple rule unfolding as GAV. Extensibility is best suitable in LAV since adding new sources will not affect the views of local sources.

Example: Global schema offered by mediated system G:

$Cust(Cid, Name, Branch, Gender), Loan(Cid), Review(Name, Review)$

Sources $S1, S2$ are defined as views by means of conjunctive queries with built-ins:

$S1 : V1(Cid, Name, Branch) \leftarrow Cust(Cid, Name, Branch, Gender), Loan(Cid) \geq 1020.$

$S1$: Has a relation $V1$ containing customers with customer >1020 that has taken loan.

$S2 : V2(Name, Review) \leftarrow Cust(Cid, Name, Branch, Gender), Review(Name, Review), Branch = 'perryridge'.$

$S2$: Has a relation $V2$ containing Customers of branch 'Perryridge' with their reviews,

- Sources does not depend on other sources but in the example $S2$ there may be other sources that will be containing information that has been later added and the $S2$ does not contain. Hence $S2$ may contain incomplete information: sources are incomplete.

Query to G: *Customer of female gender of branch 'Perryridge' with their Reviews?*

$Ans(Name, Review) \leftarrow Cust(Cid, Name, 'Perryridge', 'F'), Review(Name, Review)$

Query is expressed over global (mediated) schema but data really exist in sources that are defined as views, so it is difficult to extract data directly as can be done by GAV.

Query is rewritten in terms of the views; and can be computed:

- Extract values for $Name$ from $V1$
- Extract the tuples from $V2$
- At the mediator level, compute the join via $Name$

C. Global and Local-as-View (GLAV)

Limitations of GAV and LAV [11] [8] are overcome by Global and Local-as-view (GLAV) that is both views over global and local schema are established. Powerful features of GAV and LAV i.e scalability and extensibility of LAV and query processing of GAV are integrated in GLAV.

- Gives more expressive power,
- Natural source descriptions are offered.
- Mediated schemas found are more accurate.

Example of GLAV: Source Relations: $Branch(Name, Branch), Cust(Cid, Name)$
Global Relations: $CustRev(Cid, Name, Review), CustBr(Name, Branch)$

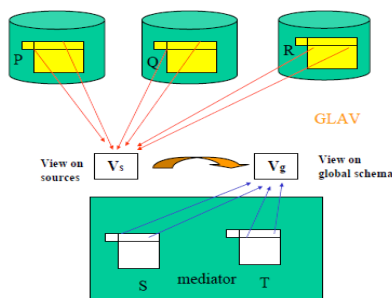


Figure 1: Global and Local-as-view

View over the local schema: $Cust(Name) \leftarrow Branch(Name,Branch),Cust(Cid,Name)$
 View over global Schema: $Cust(Name) \leftarrow Cid,Review \quad CustRev(Cid,Name,Review)$
 Equivalently: $Cid,Branch(Branch(Name,Branch),Cust(Cid,Name)) \leftrightarrow Cid,Review, CustRev(Cid,Name,Review)$

3. Related Work

Table I: Related Chronological Survey on Data Integration

| AUTHOR | TITLE | YEAR | DEVELOPMENT | ADVANTAGES |
|----------------------------|---|------|--|--|
| Sanjay Madria et al.[21] | An XML Schema integration and query mechanism system | 2007 | Integrates different XML schemas with different ontology by using a model called XSDM and graphical XML schema | XML integration process can integrate multiple DTD because of the expressive and capability of DTD structure |
| Pakornpong Pothipruk [22] | Query Answering for Multiple Complex Resources: Description Logic in the Semantic Web Context | 2007 | It proposes space reduction method resulting in combination two logics called DDL | A non-monotonic rule system has the ability to handle information that are incomplete in more easier way than description logic system |
| Abir Qasem et al. [23] | An Efficient and Complete Distributed Query Answering System for Semantic Web Data | 2007 | Answering Query using distributed Semantic web data sources by using OWLII(subset of OWL) which is compatible to GAV ,LAV and more expressive than DHL | It is very efficient since system identifies minimum set of relevant data sources and time to load sources is main factor of performance |
| Aditya Telang et al.[24] | Querying for Information Integration: How to go from an Imprecise Intent to a Precise Query? | 2008 | Introduces a framework InfoMosaic for query re formulation in the context of multi domain integration | Query re formulation makes use of minimal input and minimum interactions by using refine-as-you-input and refine-after input |
| Zijing Tan [25] | Answering XPath Queries in Virtual XML Data Integration System | 2008 | Define views supporting edge path mapping, data value bindings. Gives an algorithm to answer XPATH query | Better understandable since global instances are not materialized. Extensibility is easier to achieve |
| D. Braga et al. [26] | NGS: a Framework for Multi-Domain | 2008 | Proposed a framework NGS for answering query | Query optimization stage of NGS helps in |

| | | | | |
|--|--|------|---|--|
| | Query Answering | | in multiple domains and use xml related technologies | effective cost metrics and capability of joining ranked result list |
| Jia-Lang Seng,I.L. Kong[27] | A schema and ontology-aided intelligent information integration | 2009 | Provides generic construct transformation and achieving query reformulation by GAV approach. Handles structural heterogeneity. | Use of intelligence of XML and ontology. Enhancing the interoperability between different sources by XQuery. |
| Gösta Grahne ,Alex Thomo[28] | Bounded regular path queries in view-based data integration | 2009 | Recursive regular paths queries over views in data integration are decided by query equivalence and boundedness and also proves k-boundedness is PSPACE hard. | It expresses view based query rewriting without recursion and query optimization with recursion |
| David Kensch et al. [29] | Generic schema mappings for composition and query answering | 2009 | Uses mapping composition for rewriting in source-to-target mapping. Uses GeReMo and OWL ontology for schema mapping between models | It introduces an algorithm that translates source side in generic mapping to a query in XQuery. |
| Xiong Fengguang, Han Xie, Kuang Liqun[30] | Research and Implementation of Heterogeneous Data Integration Based on XML | 2009 | Based on XML schema, implementation through java API, Netbeans as development environment. | FLWOR feature of XQuery helps in binding and integrating. Use of java as it cross platform independent language. |

The chronological survey of work related to Data Integration is given in Table I. After the study of above papers of data integration, we have chosen to work on query processing, that is one of the important aspects of data integration. Query processing could be done in single domain or multi domain approach. Hence processing query from multiple domains can be accomplished by different data integration approaches GAV, LAV. But both have some limitations which could be overcome by another approach.

4. Conclusion And Future Work

In this paper we conclude by the fact that data integration has been best possible one to integrate data from different heterogeneous data sources along with web as data source. Query processing has been one of the basic service of data integration, so processing query in single domain is possible through different techniques and software like DBMS software etc and how query reformulation is to be done in order to sub-part the query posed to global schema to sub queries in terms of local schema. But query processing in Multi Domain could be achieved in different ways. The way we have chosen for Multi-Domain or Cross-Domain query processing for our future work is to use GLAV approach of data integration and using XML standard as exchanging and representing data from web.

References

- [1] The Meta Group. www.metagroup.com, 2001.
- [2] Laura Bright, Louiqa Raschid, Vladimir Zadorozhny and Tao Zhan,: Learning response times for web sources: A comparison of a web prediction tool (WebPT) and a Neural Network. In *CoopIS 1999*, 1999..
- [3] Vladimir Zadorozhny, Louiqa Raschid, Tao Zhan, and Laura Bright.: Validating access cost model for wide area applications. In *CoopIS 2001*.
- [4] Mary Tork Roth, Fatma Ozcan, and Laura M. Haas. Cost models do matter: Providing cost information for diverse data sources in a federated system. In *VLDB'99, Proceedings of 25th International Conference on Very Large Data Bases*, Edinburgh, Scotland, pages 599-610, 1999.
- [5] AnHai Doan, Pedro Domingos, and Alon Y. Halevy. Reconciling schemas of disparate data sources: A Machine-learning approach. In *SIGMOD 2001, Proceedings ACM SIGMOD International Conference on Management of Data*, May 21-24, 2001, Santa Barbara, California, USA, 2001.
- [6] Jayant Madhavan, Philip A. Bernstein, and Erhard Rahm. Generic schema matching with Cupid. In *VLDB 2001, Proceedings of 27th International Conference on Very Large Data Bases*, September 11-14, 2001, Roma, Italy, 2001.
- [7] Mauricio A. Hernandez, Renee J. Miller, and Laura M. Haas. Clio: A semi-automatic tool for schema mapping. In *SIGMOD 2001, Proceedings ACM SIGMOD International Conference on Management of Data*, May 21-24, 2001, Santa Barbara, California, USA, 2001
- [8] N. Kushmerick, R. Doorenbos, and D. Weld: Wrapper induction for information extraction. In *ICJCAI '97*, 1997.
- [9] Robert Baumgartner, Sergio Flesca, and Georg Gottlob. Visual web information extraction with (lixto). In *VLDB 2001, Proceedings of 27th International Conference on Very Large Data Bases*, September 11-14, 2001, Rome, Italy, 2001.
- [10] D.Braga, D. Calvanese, A.Campl, S.Ceri, F.Daniel, D. Marchtinenghi, P.Merialda, R. Torlone, NGS a Framework for Mutli-Domain Query Answering, (2008), in the *proceeding of International Council of open and distance education (ICDE) Workshop 2008*, 254-261.
- [11] Andrea CaliReasoning in Data Integration Systems: Why LAV AND GAV are siblings, *Foundation of Intelligent Systems*,2003, Vol 2871/2003,562-371.
- [12] M. Friedman, A. Levy, T. Millstein , Navigational Plans for Data Integration. *Proc. AAAI'99*.
- [13] Leopoldo Bertossi, Virtual Data Integration, Carleton University, *School of computer Science*, Ottawa, Canada.
- [14] Serge Abiteboul, Richard Hull, and Victor Vianu: *Foundations of Databases* , Addison Wesley, Publ. Co., Reading, Massachusetts, 1995.
- [15] Bertossi, L. Consistent Query Answering in Databases : *ACM Sigmoid Record*, June 2006, 35(2):68-76.
- [16] Bertossi, L. and Bravo, L. Consistent Query Answers in Virtual Data Integration Systems. In *'Inconsistency Tolerance'*, Springer LNCS 3300, 2004, pp. 42-83.
- [17] M. Gruninger and J. Lee. Ontology applications and design. *Communications of the ACM*, 45(2):39-41, 2002.
- [18] S. Abiteboul and O. Duschka. Complexity of answering queries using materialized views. In *Proc. of the 17th ACM SIGACT SIGMOD SIGART Symp. on Principles of Database Systems (PODS'98)*, pages 254-265, 1998
- [19] G. Grahne and A. O. Mendelzon. Tableau Techniques for Querying Information Sources through Global Schemas. In *Proc. of the 7th Int. Conf. on Database Theory (ICDT'99)*, Volume 1540 of Lecture Notes in Computer Science, pages 332-347. Springer,1999.
- [20] A. Y. Levy: Obtaining Complete Answers from Incomplete Databases. In *Proc. of the 22nd Int. Conf. on Very Large Data Bases (VLDB'96)*, pages 402-412, 1996.
- [21] Sanjay Madria, Kalpdrum Passi, Sourav Bhowmick : An XML Schema integration and query mechanism system, *Data & Knowledge Engineering* 65 (2008) 266-303,2007
- [22] Pakornpong Pothipruk: Query Answering for Multiple: Complex Resources: Description Logic in the Semantic Web Context, A thesis submitted to the University of Queensland for

- the degree of Doctor of Philosophy Department of School of Information Technology and Electronic Engineering January 2007.
- [23] Abir Qasem, Dimitre A. Dimitrov, and Jeff Heflin: An Efficient and Complete Distributed Query Answering System for Semantic Web Data Lehigh University Technical Report LU-CSE-07-007, Lehigh University, 19 Memorial Drive West, Bethlehem, 2007
 - [24] Aditya Telang, Sharma Chakravarthy, Chengkai Li: Querying for Information Integration: How to go from an Imprecise Intent to a Precise Query?, Department of Computer Science and Engineering University of Texas at Arlington Arlington,
 - [25] Zijng Tan: Answering XPath Queries in Virtual XML Data Integration System, in the proceeding of 2008 International Conference on Computer Science and Software Engineering ,IEEE computer society, 978-0-7695-3336-0/08 \$25.00 © 2008 IEEE..
 - [26] D. Braga, D. Calvanese, A. Campi, S. Ceri, F. Daniel, D. Martinenghi, P. Merialdo, R. Torlone:NGS: a Framework for Multi-Domain Query Answering, ICDE Workshop 2008,
 - [27] Jia-Lang Seng, I.L. Kong: A schema and ontology-aided intelligent information integration, Expert Systems with Applications 36 (2009) 10538–10550, 0957-4174/\$ 2009 Elsevier.
 - [28] Gösta Grahne , Alex Thomo: Bounded regular path queries in view-based data integration, Information Processing Letters 109 (2009), Elsevier.
 - [29] David Kensché, Christoph Quix , Xiang Li , Yong Li, Matthias Jarke: Generic schema mappings for composition and query answering, Data & Knowledge Engineering 68 (2009), Elsevier.
 - [30] Xiong Fengguang, Han Xie, Kuang Liqun: Research and Implementation of Heterogeneous Data Integration Based on XML, in the proceeding of The Ninth International Conference on Electronic Measurement & Instruments, ICEMI'2009.