

Implementing an Improved Mediator Wrapper Paradigm for Heterogeneous Database Integration

Adeyelu, A.A., and Anyebe, E.O.

Department of Maths/Computer Science,
Benue State University, Makurdi, Nigeria

Abstract

This study developed and implemented an improved mediator wrapper approach to addressing the challenges of integration of semantic heterogeneous databases. It employed Local as View (LaV) paradigm of database integration so as to reduce its cost as well as offer the local sources a degree of independence. The developed model was implemented as web service using JEE and some software development tools with a number of heterogeneous databases as a case study.

Key words: mediator wrapper, semantic, Local as view, heterogeneous



Introduction

*The rapid growth of database technologies and networking have had a major impact on information processing requirement and methods in organizations. Information systems for large organizations today are mostly implemented on a distributed architecture, using a number of different computers interconnected through the Internet. Different data are usually stored in various databases based on different departments in the organization. For these reasons, organizations may have incompatible databases and their associated database management systems. A good example is the group of national databases in Nigeria. Each of these databases have unique functions and are usually created and managed internally by one organization established by the Federal Government. For example, the national identity card database which is used to give each registered citizen a national identity number is managed by the National Identity Management Commission (NIMC), SIM registration databases is managed by Nigerian Communication Commission (NCC), and BVN (Biometric Verification Number) database managed by Central Bank of Nigeria (CBN), just to mention a few. Integrating these databases will offer many benefits to the nation like access to accurate and complete information about citizens, aid in criminal investigation as well as aid planning for national growth.

Database integration is the process of taking databases (data sources) and making the data in them work together as if they were a single database. The aim of this is to achieve easy access to information from different databases so as to avoid duplication of data acquisition and processing efforts, and enable efficient use of resources. Integration models for heterogeneous databases are challenged with how autonomous the source databases are, resolving data semantics from various databases and the cost of implementing and managing an integrated system (Shi, 2002); hence this study. The study sought to tackle these problems by using the mediator-wrapper approach to harness data from different sources and platforms for the purpose of management.

Review of Related Literature

Most database systems form foundations were designed without integration in mind. Whenever there is desire for integration, the sources and their data (that do not fit together) need to be combined using additional adaptation and reconciliation functionality; hence the development of research in database integration.

According to Almanhy and Salim (2013) database integration is a process involving multiple steps in finding similar objects in two or more databases to create a non-redundant, unified view of the databases. The databases are often unrelated and this make the task of integration difficult. Dissimilar sources means that the databases may differ in some respect such as the Database Management application each database is using (Ali, 2009). Heterogeneities between data sources (databases) are broadly classified into syntactic, platform, schematic and semantic heterogeneities (Xiao and Cruz, 2004). Syntactic heterogeneity is caused by the use of different models or languages. In today's computing there are large amounts of unstructured data in form of XM and pdf in different formats, thus different type of data models are used to store these data. Some may be document oriented, graph based arts, relational data models, or object oriented, every data model having its own query syntax, query language and no knowledge about other databases (Garg and Kaur, 2015). Therefore harnessing of data in this is a big problem. Schematic heterogeneity results from the different data organizations or conceptual design. This implies different data values can be used for the same information and kept at multiple sites but of different values. Sometimes data is repeated on multiple sites but due to synchronization problems same copy has different values. Semantic heterogeneity is caused by different meanings or interpretations of data, for example in representing gender as male or female, one database design may use m or f and another may use 0 or 1. Sometimes a database design can use Naira as currency while the other use Dollar, Euro and so on. Furthermore Ceruti, Kemal and Thuraisingham (1986) pointed out that platform heterogeneity, constraint heterogeneity and transaction processing heterogeneity were issues that have to be resolved to achieve the goal of integration.

Database integration can be accomplished either by physical or virtual approach. In the physical approach, copies of the original heterogeneous databases are combined to form a single database called the integrated database or data warehouse, while the virtual approach deploys a multi-database or mediation system to support queries on an integrated view constructed upon the local (original) databases (Mohd, Yazid and Wan, 2013). This approach solely depends on local databases thus providing a closely coupled paradigm and consequently making the approach very slow. Though on this warehouse, data can be extracted, transferred and loaded as in data mining, it has a big setback in that the data will not

always be up to date. The other approaches are Federated and mediator-wrapper system which are virtual forms of integration (Ullman, Molina and Widom, 2002, Almahy and Salim, 2013).

A federated database achieve uniform access solution by logically integrating data from underlying local DBMS. The component databases are integrated by each having its own database management system (DBMS) and a management system that controls the coordination of the databases, called a federated database management system. The databases themselves and the associated DBMS's play a pivotal role in deciding the overall structure of the federated system, since these can vary greatly. Federated database systems are fully-fledged DBMS in that they implement their own data model, support global queries, global transactions, and global access control. However queries may be costly and more difficult to optimize (Ullman, Molina and Widom, 2002).

In the mediator wrapper approach, queries can be performed directly on original databases. Here a query is posted to global schema called a mediator which reformulates it according to formats of local systems with the assistance of wrappers and then carry out a direct query on databases. This is a loosely coupled architecture and faster approach which overcome the up to date data challenge of data warehousing. The mediator exploit encoded knowledge about certain set of data to create information for higher layer of applications (Herschel, 2014) and processes queries, merge results and keeps track of available data sources. It consists of a query processor for constructing query trees and merging data. It is used to determine relationships between different objects which may be located in different data sources. Here each object must also contain information about the data source it belongs to. A wrapper is a software component that establish communication and the data flow between data sources and mediators. It has a homogeneous interface for working with heterogeneous data sources and provides frameworks to extract the schema from a data source in a format that is usable by the repository (Garg and Kaur, 2015). It also uses a parser to execute queries in its data source, using one wrapper for each data source and integrating its schema with the repository. A mediator may either be designed using object oriented language or XML schema and XQuery (Ulman, Molina and Widom, 2002). Either way, a database integration system *I* based on mediator-wrapper approach can be defined as a tripple (G,S,M) (Lerzerini, 2002), where:

i. G is the global schema

ii. S is the source schema(s)

iii. M is the mapping between G and S.

The global schema is usually developed using a process called schema matching and the process of mapping is referred to as schema mapping. Schema matching is the process of generating correspondence between elements of two schema while schema mapping determines how data from the local or different sources can be mapped to the global schema while keeping data consistency (Almahy and Salim, 2013).

An integration system based on mediation technique is also defined as a system characterized by an architecture based on a global schema that provide a reconciled and integrated view of the underlying sources Xiao and Cruz (2004). This means the integrated system is concerned with two key issues: system modeling and query processing.

System modeling deals with modeling the relation between the sources and the global schema. Two basic approaches have been proposed for system modeling as listed below:

i. Global as View (GAV): This approach expresses the global schema in terms of the local sources.

ii. Local as View (LAV): This approach, requires the global schema to be specified independently from the sources, and the relationships between the global schema and the sources are established by defining every source as a view over the global schema

Query processing require a query reformulation step: query over the global schema has to be reformulated in terms of a set of queries over the sources. In the GaV approach, every entity in the global schema is associated with a view over the source local schema, therefore query processing in this case uses a simple “unfolding” strategy. In contrast, query processing in LaV can be complex, since the local sources may contain incomplete information. Query processing in LaV is called view-based query processing. Common mediator wrapper systems include TSIMMIS (The Stanford-IBM Manager of Multiple Information Sources) and GARLIC System. (Shi, 2002).

Introduction to National Database Integration Project

In Nigeria, there are a number of databases in use. These include: National Identity database that is managed by the National Identity Management Commission (NIMC), SIM

registration database an initiative of Nigerian Communication Commission where data is collected by various telecommunication companies (MTN, Airtel, GlobaCom, 9mobile) in the country for submission to the SIM registration database, BVN (Bank Verification Number) database meant to unify bank accounts belonging to an individual, Tax identification database managed by Federal Inland Revenue service just to mention a few. These databases resulted from polices by the government to gather information about citizens so as to enable smooth functioning of the activities of these agencies and commission. Since these databases are managed by different organizations, there is need for them to be autonomous so that the various applications that use them can still function properly. All the information gathered about citizens are from strategic areas and as such one outcome is glaring: the need for access in order to have detailed complete and accurate data about a person in such a way that information gathering will not be disrupted. A commission, the National Identity Management Commission (NIMC) was given the mandate to carry out integration of national databases by the government. An attempt is made in this study to adapt the mediator wrapper model of database integration to solve this problem. The results will be relevant to developers of integrated database systems.

Methodology

This section discussed the methodology adopted in undertaking this study. It presented detailed analysis of the proposed systems and the model which was later implemented using Java Programming language and some database managers.

Analysis of the integration system

An integration system should be able to combine data from different sources to form a single record such that it appears to the user as if the data is from a single source. In the case of this system, the sources were databases. The integration system behaved like a database management system in the sense that it receives query from applications requesting for data. This implies that the integration system will be a server program. The system was designed to enable the sources to be autonomous and resolve discrepancies in data about individual record from the various sources. To be able to perform this

function, the system was developed with the following:

- I. a virtual database also called global schema that can be queried,
- ii. a language for encoding ontologies,
- iii. a protocol for query and a means of processing queries,
- iv. a way to acquire data from different data sources,
- v. a technique of harnessing data from each database, and
- vi. a means of sending back report in a comprehensible form to the requesting application.

Description of the Proposed System

This study adapted the Mediator Wrapper paradigm (Seksun, Djamaal, Champin and Thairin, 2002) of data integration and uses Local as View approach of schema mapping. The mediator provides a virtual database in the form of classes to applications and the wrappers. The applications send query asking for objects of a class at a particular time. The mediator does not need to understand the underlying structure of each database. All it does is to know which local table(s) or entity from a database form its class. This information is encoded in a global ontology. The global ontology then enable the mediator to know which wrapper to call in the event of a query. When a mediator gets a query, it simply decodes it to form a query message and send it to the wrappers. The wrappers interpret the mediator schema and query message from the mediator via the use of a local ontology. When a wrapper gets a query, it converts it into a local query that its source database can understand. Afterwards, it connects to the database, execute the query and get the result. The result it receives is converted to a format the mediator can understand. To do this, the wrapper uses the local ontology to convert the data into objects of the mediator class. As such the result from the wrapper is an array of object of a particular mediator class. The mediator receives the various results from the wrappers, use the integration system to combine them and send the information to the application. To be able to communicate effectively as proposed by the model, the integration system will be a web service. The proposed integration model is illustrated figure 1.

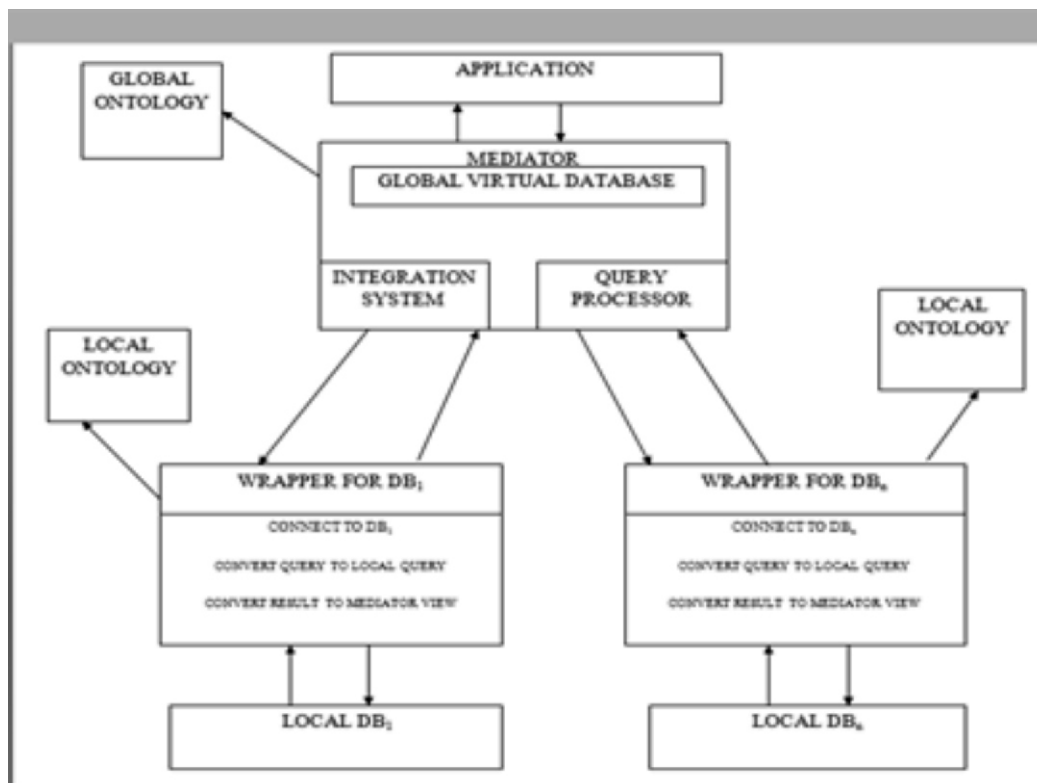


Figure 1: The Proposed Architecture, Adapted from (Seksun, Djamal, Champin and Thairin, 2002).

The architecture was implemented using JEE platform for the reasons that:

- i. It is a single standard that sits on top of a wide range of existing enterprise systems, and database management systems. In addition, JEE specification ensures Web services interoperability,
- ii. it can be deployed on most web application servers, many operating systems and can also be used with many Relational Database Management System (DBMS), and
- iii. its implementations on multiple hardware and software platforms have been successful in high performance and throughput system space.

The integration system was developed as a web service; a collection of open protocols and standards used for exchanging data between applications or systems which will be available over the Internet or private networks using standardized XML messaging for communication. Web services allow various applications to talk to each other and share data and services among themselves, share documents such as XML making it most suitable choice for implementing the proposed model. It also use standardized protocols like SOAP over HTTP protocol. This is low cost because existing low-cost internet services is used for implementing web services.

Data Design

The Database Management Systems

(DBMS) used were MYSQL, ORACLE and POSTGREL. These database management systems are robust and reliable with widespread usage. Although all of them make use of Structured Query Language (SQL), there were some variations in the syntax for writing SQL in each of the DBMS. Also the underlying data structure as well as data types used in each of them are different. Hence making the system heterogeneous. The data in the five databases were simulated in this project. To do this, data collection forms used by the various agencies were gathered and studied. The tool used for the designs is MYSQL workbench as such the data types are of MYSQL. Each database was implemented randomly in one of the three database managers mentioned earlier, for example National ID management database implemented in Oracle, BVN database implemented in MYQSL, SIM registration database implemented in POSTGREL, National Health Insurance database (NHIS) implemented in MYSQL and TAX database implemented in POSTGREL.

Global Virtual Database

A structure-based schema matching was used to design the virtual database. The process involved determining the relationship between entities within a database or between databases: equivalence, synonymy, containment, homonymy (Ali, 2009) and antonym. Customer and Customer from BVN and SIM registration databases respectively were equivalent. Person, Taxpayer,

Personal Data from NIN, Tax and NHIS databases respectively were synonymous. Taxpayer contains workdetails and Taxdue contained tax type from TAX database. Secondly a schema matching was performed using operators. The operators applied here were union, generalization, specialization and import (Ali, 2009). The rules used were that:

i. All equivalence and synonymy can use

union.

- ii. Generalization can be used for equivalence, synonymy or containment where possible.
- iii. Specialization can be used for equivalence, synonymy or containment where possible.
- iv. Use import for homonymy and antonymy (Ali, 2009).

The algorithm for the global database is as

- i. *Let n=number of databases*
- ii. *let i=1*
- iii. *Repeat (iv) until i> n*
- iv. *Identify the relationship between the entities in each database. The relationships could be binary or n-ary in degree.*
- v. *let i=1*
- vi. *Repeat (vii) until i> n*
- vii. *Compare the entities from each database and identify the relationship between them. The relationships could be binary or n-ary in degree.*
- viii. *Let r=number of relationships identified.*
- ix. *For each r, perform (x to xii).*
- x. *Get pair of entities in the relationship.*
- xi. *Decide on the operator(s) to use in integrating the entities.*
- xii. *Compare the attributes of the entities and apply operator to form new entities and attributes. /*New name or adapted names may be used to represent new entities and their attributes*/*
- xiii. *Let NewEntities=total number of new entities formed.*
- xiv. *let i=1;*
- xv. *repeat (xvi to xvii) until i>NewEntities*
- xvi. *Compare and identify the relationship between the new entities created.*
- xvii. *perform (x to xii)*
- xviii. *Repeat (xiii-xviii) until the integrated schema cannot be merged any further*

Figure 2: Algorithm for the Global Database

The global ontology was specified using xml document. It was used by the mediator to know which wrapper a query message is to be sent. The global class name is written in the tag <classname> </classname>. The <database> </database> tag contains data required by the mediator to be able to connect to a database wrapper. Each wrapper has its own local ontology.

This is a well formed XML document and is meant to enable the wrapper convert a query message from the mediator to local query. The local ontology is also used by its wrapper to convert the result from its database to the required mediator class. The local ontology help solve semantic problems. Figure 2 shows the data flow diagram of the model.



Figure 3: Data flow diagram of the Model

The query received from the application is first decoded by the mediator to get the name of class requested. The selector is a simple “where” clause

that specifies only one field value as search criteria. The syntax of the query from an application is shown in figure 4:

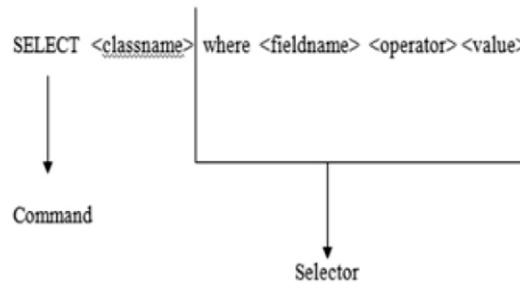


Figure 4: The syntax of query message sent by an application to the Mediator

The class diagram also has a class called query message which defines how query to be sent to the Wrappers will be. An ID was given to the query message. Parameters such as command, classname, fieldname, operator and value were the arguments list respectively. This enabled the wrapper to easily parse and convert a message from the mediator. The query processor validates a query and provides the list of wrappers to which message will be sent. The message is sent to the wrappers. On receipt of the result, the mediator perform integration using methods from the integration system. Upon receipt of records from each wrapper, the mediator calls the third component which is the integration system. This component is responsible for combining the

records from each wrapper to form a single result set. The wrapper receives query messages from the mediator and executes them. It does so by converting the query message to its local query with the help of the local ontology, connect to its local database and execute the local query, get the result and convert it to the mediator class specified in the query message by using the local ontology then send records to mediator.

Results and Discussion

The integrated system used the concept of distributed processing to perform its task. In this regard, it is assumed that the source databases are always running on servers at different locations and internet connection is available at the

integration system site at all times. The design was fully implemented as a web service in JAVA and tested using the proposed client application and test samples from different databases on different platforms. The client application is assumed to be a search application. At the client application home page, an authorized personnel was presented with a search box requesting for the name of the entity to search for. The user inputs the required data and on click the search button, the client application connects to the integration system

requesting for information about the specified individual using the proposed name.

Records in the Test Table of the Individual Databases

The records in each table from each of database are shown in Figure 5. This class relates to Person, Customer, Customer, Taxpayer and Personal Data tables of NIN, BVN, SIM, Tax and NHIS databases respectively.

NIN	TITLE	LASTNAME	FIRSTNAME	MIDDLENAME	OTHERNAME	MAIDENNAME	PSURNAME	PFIRSTNAME	PMNAME	DOB	COUNTRYOFBIRTH	STATEOFBIRTH
431057TeTa	Mrs	Tevase	Tabitha	Member	-	Tedoo	Sughter	-	-	15/5/1989	Nigeria	Benue
746909AKTE	Mr	AKIGHIR	TERSOO	TERSOO	-	Tedoo	Sughter	-	-	19/7/1993	Nigeria	Benue
244234IKJO	Mr	IKANYI	JOHN	JAMES	-	Adanu	-	-	-	25/2/1993	Nigeria	Lagos
5018AdBI	Miss	Adoga	Blessing	Inori	-	Inori	-	-	-	31/1/1995	Nigeria	Benue
640789AnEv	Miss	Anyebe	Evelyn	Onyi	-	Idoko	-	-	-	19/12/2003	Nigeria	benue
714ANMA	Miss	Anyebe	MARY	Emani	-	Idoko	-	-	-	15/5/1996	Nigeria	Benue
558453IdAd	Mr	Idah	Adah	Isiah	-	Onyanobi	-	-	-	31/12/1986	Nigeria	Benue
626839AFAN	Mrs	APEH	ANN	IKOCHE	-	Ada	Agada	-	-	23/11/1982	Nigeria	Benue
812813AnPe	Mr	Ankali	Peter	Peter	-	Oko	-	-	-	2/3/1992	Nigeria	Benue
850023TSSA	Mr	TSAVUJA	SAMUEL	AONDOFA	-	odidoo	-	-	-	16/10/1990	Nigeria	Benue
623366ORTO	Mr	ORIOYE	TOSIN	MARK	-	Oluwa	-	-	-	16/11/1990	Nigeria	Lagos
447851AGVI	Mr	AGADA	VICTOR	VICTOR	-	Ada	-	-	-	16/11/1980	Nigeria	Benue
241927JUFI	MR	JUNIOR	FIDELIS	ASAMOAH	GANYO	nana	asamoah	fidelis	junior	21/1/1994	Nigeria	Benue
898924AGZA	Mr	AGADA	ZAINAB	OBUNU	-	Ata	-	-	-	31/11/1988	Nigeria	Benue

Figure 5: Screenshot of data contained in the table person from NIN database.

NAME	BVN	NIN	MARITAL STATUS	GENDER	STATE OF ORIGIN	RESIDENTIAL ADDRESS	STATE OF RESIDENCE	LGA OF RESIDENCE	LANDMARK	EMAIL	Home	Customer	Bank	Help
MISS BLESSING INORI ADOGA	Ad98250BI	5018AdBI	SINGLE	FEMALE	Benue	Behind modern market	Benue	Makurdi	Modern market	adogaBlessing@gmail.com				
MRS ZAINAB OBUNU AGADA	AG77868ZA	898924AGZA	MARRIED	FEMALE	Benue	Behind former Eunny land Bakery	Benue	Makurdi	Former Eunny land bakery					
MR VICTOR VICTOR AGADA	AG89612VI	447651AGVI	MARRIED	Male	Benue	behind Eunny Land bakery	Benue	Makurdi	Former Eunny land bakery					
MR TERSOO TERSOO AKIGHIR	AK27218TE	746909AKTE	SINGLE	Male	Benue	Ahmedu Bello University zaria	Kaduna	Zaria	Ahmedu Bello University Zaria	Tersoo@gmail.com				
MR PETER PETER ANKALI	AN16629PE	812813AnPe	SINGLE	Male	Benue	Railway Quaters	Benue	Makurdi	Railway market					
MISS ONYI EVELYN ANYEBE	An62003On	640789AnEv	SINGLE	FEMALE	Benue	No 12 Bem Hotel road	Benue	Makurdi	Bem Hotel	anyebeevelyn@gmail.com				
MISS MARY EMANI ANYEBE	An94165MA	714anma	SINGLE	FEMALE	Benue	ANKpa queters	Benue	Makurdi	Bem Hotel	anyebeemani@gmail.com				
MRS ANN EKOCHE APEH	AP49340AN	626839APAN	MARRIED	FEMALE	Benue	No 12 Bem Hotel road	Benue	Makurdi	Bem Hotel					
MR FIDELIS JUNIOR ASAMOAH	As58024Fi	241927JUFI	SINGLE	Male	Ghana	NO 8 UKE WENDE STREET HIGH LEVEL MAKURDI	Benue	Makurdi	High level round about.	asamoahfidlisjr@gmail.com				
MR ADAH ISALAH IDAH	ID69307AD	558453IdAd	SINGLE	Male	Benue	Railway Quaters	Benue	Makurdi	Railway market	adah@gmil.com				
MR JAMES JOHN IKANYI	IK89589JA	244234IKJO	SINGLE	Male	Benue	Wurukun Makurdi	Benue	Makurdi	Benue State Library Makurdi	jsquare@gmail.com				

Figure 6: Screenshot of data contained in the table customer from BVN database.

SIM REGISTRATION											
Home New Regs											
NEW CUSTOMER FILL PHONE VIEW CUSTOMER											
S/NO	NAME	MOTHER'S MAIDEN NAME	DATE OF BIRTH	STATE OF ORIGIN	LGA OF ORIGIN	NATIONAL ID	DATE REGISTERED	ADDRESS	STATE OF RESIDENCE	LGA OF RESIDENCE	EMAIL ADDRESS
1	MARY ANYEBE EMANI	IDOKO	1/1/1996	Benue	Ogbadigbo	714ANMA	18/1/2016	ipole igwu	Benue	Ogbadigbo	anyeevelyn@gmail.com
2	FIDELIS ASAHMOAH JUNIOR	Onyanobi	21/1/1994	Benue	MAKURDI	241927JUF1	24/2/2017	NO 8 UKE WENDE STREET high level makurdi	Benue	MAKURDI	asamoahfidisjr@gmail.com
3	BLESSING ADOGA INORI	Inori	31/1/1995	Benue	MAKURDI	5018AdB1	23/2/2017	Behind Modern market	Benue	MAKURDI	adogaBlessing@gmail.com
4	ISALAH IDAH ADAH	Onyanobi	31/1/1995	Benue	MAKURDI	558453IdAd	24/2/2017	Railway Quaters	Benue	MAKURDI	adogaBlessing@gmail.com
5	JOHN IKANYI JAMES	APEH	1/1/1992	Benue	Ogbadigbo	244234IKJO	17/5/2007	LAGOS STREET	LAGOS	Ogbadigbo	jsquare@gmail.com
6	EVELYN ANYEBE ONYI	IDOKO	19/4/1994	Benue	Ogbadigbo	640789AnEv	27/2/2017	No 12, Bem Hotel Road	BENUE	Ogbadigbo	anyeevelyn@gmail.com
7	TABITHA TERWASE MEMBER	Terdoo	15/5/1989	Benue	Konshisha	431067TeTa	27/2/2017	Opposite BSU Fast gate	Benue	Konshisha	TerwaseTabitha@gmail.com
8	TERSOO AKIGHIR TERSOO	Terdoo	19/7/1993	Benue	Konshisha	746909AKTE	27/2/2017	Askpa ward wadata	Benue	Konshisha	Tersoo@gmail.com
9	ANN APEH EKOCHÉ	Ada	23/11/1982	Benue	Ogbadigbo	626839APAN	27/2/2017	no 12 Bem Hotel Road	Benue	Ogbadigbo	

Figure 7: Screenshot of data contained in the table customer from SIM database.

TAX PAYERS									
S/NO	TIN	NATIONALID	NAME	NATIONALITY	CITY	GENDER	DATE OF BIRTH	MARITAL STATUS	MOTHERS NAME
1	JA964		Agada Jane	Nigerian	Makurdi	Female	18/11/1986	Ihotu	Married
2	TT400	431067TeTa	TERWASE TABITHA	Nigerian	Makurdi	Female	15/5/1989	Sughter	Married
3	TA37	746909AKTE	AKIGHIR TERSOO	Nigerian	Makurdi	Male	19/7/1993	Terdoo	Single
4	EA765	640789AnEv	ANYEBE EVELYN	Nigerian	Makurdi	Female	17/04/1994	Idoko	Single

Figure 8: Screenshot of data contained in the table taxpayer from Tax database.

national health insurance scheme												
Home About Registration Setup												
REGISTERED PERSONS												
S/NO	PERSON ID	NAME	MARITAL STATUS	DATE OF BIRTH	GENDER	NATIONAL ID	PHONE NUMBER	HEALTH CARE PROVIDER	ALTERNATE HEALTH CARE PROVIDER	EMAIL	ORGANISATION	REG DATE
1	A1081982-11-23	APEH ANN EKOCHÉ	married	1982-11-23	female	626839APAN	08078563423	Federal Medical Center	Federal Medical Center	ann@gmail.com	toriloye	2017-02-28
2	A3401990-10-16	TSAVWUA AONDOFA SAMUEL	single	1990-10-16	male	850023TSSA	08078563423	Federal Medical Center	Federal Medical Center	aondofa@gmail.com	toriloye	2017-02-28
3	A5851986-12-31	IDAH ADAH ISALAH	single	1986-12-31	male	558453IdAd	08161500750	Federal Medical Center	Federal Medical Center	adah@gmil.com	toriloye	2017-02-28
4	B3831995-1-31	ADOGA BLESSING INORI	single	1995-01-31	female	5018AdB1	08078563423	Federal Medical Center	Federal Medical Center	adogaBlessing@gmail.com	BSU	2017-02-28
5	E1931994-4-17	ANYEBE EVELYN ONYI	single	1994-04-17	female	640789AnEv	08135596516	Federal Medical Center	Federal Medical Center	anyeevelyn@gmail.com	toriloye	2017-02-28
6	F519--	ASAMOAH FIDELIS JUNIOR	married	0000-00-00	male	241927JUF1	08078563423	Federal Medical Center	Federal Medical Center	asamoah@gmail.com	BSU	2017-02-28
7	J2371993-2-25	IKANYI JOHN JAMES	single	1993-02-25	male	244234IKJO	08165076450	Federal Medical Center	Federal Medical Center	Jsquare@gmail.com	BSU	2017-02-28

Figure 9: Screenshot of data contained in the table personal data from NHIS database.

Looking at the records and the fields of the table from each database, they exhibit different levels of heterogeneity. The field names are different and the number of fields in the table of each database differ depending on the kind of information required about a person by the

designers of the database. For example: sname, surname, lname, lastname refer to a person's surname. Also gender, sex and dob, dateofbirth. Figure 10 showed the result of searching for a person with surname "Anyebe" initiated by the client application.

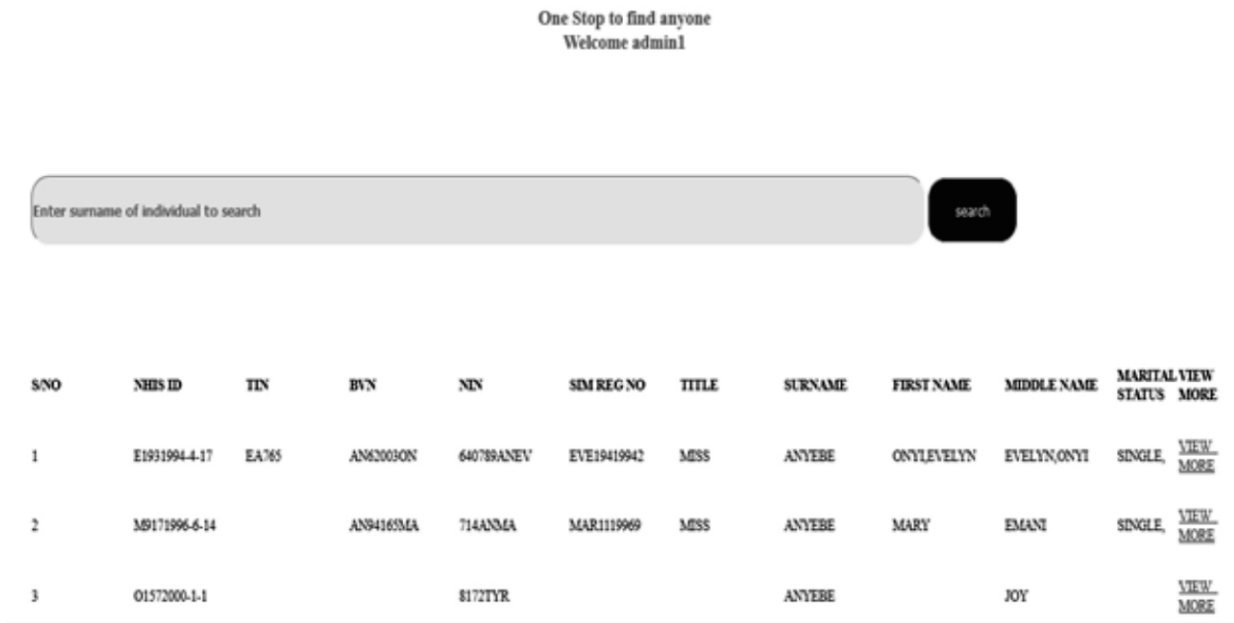


Figure 10: Client application page showing list of records displayed after search.

After comparing with the records in the various databases, only the records of individuals with surname "Anyebe" were displayed. Also, the system was able to merge the record of "Anyebe Evelyn Onyi" found in all the databases and "Anyebe Emani Mary" in all the databases except

in tax database and as such there was no TIN number. Also "Anyebe joy's" record which was found in only NHIS databases was displayed. The next search is done using national identity number as shown in figure 11.



Figure 11: Client application page showing record after search.

Only one user exist in all the data sources with national id " 746909AKTE ". As such the application correctly retrieved its data. The national id field has the property of uniqueness thereby making search more efficient.

The application correctly assessed the record for individual "AGADA ZAINAB" (figure

12) from the databases using her national id and was able to highlight a disparity in her title. With the use of the national id, unique records were retrieved and combined efficiently because the attribute has the property of uniqueness.

Search Wizard

One Stop to find anyone
Welcome admin!

S/NO	NIBS ID	TIN	BVN	NIN	SIM REG NO	TITLE	SURNAME	FIRST NAME	MIDDLE NAME	MARITAL STATUS	VIEW MORE
1	Z2761988-11-31		A077868ZA	898924AGZA	ZAD11119889	MR	AGADA	ZAINAB	OBUNU	MARRIED	VIEW MORE

Figure 12: Client application page showing record after search for user with national id "898924AGZA"

The integration system was able to combine data from different sources to form a single record such that it appeared to the user as if the data was from a single source. The system facilitated easy access to information from different databases of diverse platforms.

Conclusion and Recommendation

The study attempted to address the challenges of autonomy in database integration using the different databases as case studies. A modified mediator wrapper using Local as View method of integration was adopted. The paradigm when implemented was able to combine data from different databases, offering the local sources a degree of independence and some level of cost effectiveness for the reason that that it requires little technical knowledge to be maintained and due to the absence of a central database, there is reduced cost of data center setup and management. The system can be easily developed by programmers and deployed for usage. It can be made efficient by enforcing a unique attribute in all the local sources to be integrated. It also required very little storage capacity due to the absence of a central database. During the implementation of this work, it was observed that achieving an efficient integration system require the use of a field with the property of uniqueness for search. Therefore the use of unique attributes about an individual to search is recommended. Also network instability might make the developed system have too many down time. To minimize this, the use of a central database to carry out periodical update via the integration system is recommended.

Adding the use of a central database and agents to the integration system, using unique attributes like biometric information and images of people for search parameters and using pattern

matching and natural languages for resolving discrepancies in the records acquired during merging are suggestions for future studies in this work.

References

- Ali M.G. (2009). Object-Oriented Approach for Integration of Heterogeneous Databases in a Multidatabase System and Local Schemas Modifications Propagation. In: *International Journal of Computer Science and Information Security* Vol. 6, No. 2. India.
- Almahy I. & Salim N. (2013). Database Integration: Importance and Approaches. In: *Journal of Theoretical and Applied Information Technology* Vol. 54 No.1.
- Ceruti M. G., Kemal M. N. & Thuraisingham B. M. (1986). Object-oriented Technology for Integrating Distributed Database Systems. In: *Proceedings of the Twelfth Annual Department of Defense Database Colloquium "Emerging Technology for Database Interoperability and Data Administration"*, San Diego, CA, August 28-30 1995. pp 79-97.
- Garg B. & Kaur K. (2015). Integration of Heterogeneous Databases in: *Proceedings of International Conference on Advances in Computer Engineering and Applications (I C A C E A) I M S* Engineering College, Ghaziabad organized by IEEE. pp. 1033-1038
- Herchel M. (2014). Information Integration Architectures. (online). Retrieved from: melanie.herschel@ipvs.uni-stuttgart.de. [Accessed: 28, June 2016].
- Lerzerini M. (2002). Data integration: A theoretical perspective. In: *Tutorial at PODS 2002*. Wisconsin: USA.

- Mohd K. Y., Yazid M. S. & Wan N. S. W. N. (2013). Comparative of Mediator Approach for Database Integration. In: *Journal of Computer Science 2015*, 11 (1): 204-217. Malaysia.
- Seksun S., Djamel B., Champin P. & Thairin P. (2002). Wrapping and Integrating Heterogeneous Databases with Owl.
- Shelake, V.M. & Bhojane, V.S. (2013). A Novel Approach for Multi-source Heterogeneous Database Integration, Dec. 184-190, 10.1109/ICMIRA.2013.42
- Shi G. (2002). Data Integration using Agent based Mediator-Wrapper Architecture. Tutorial report: for agent based software engineering. [online]. [Accessed: 9th Sept. 2016].
- Ullman J. D., Molina H. & Widom J. (2002). *Database Systems. (2nd.Ed)*. pp 1037-1091. Pearson Education Inc. New Jersey: USA.
- Xiao H. & Cruz I. F. (2004). Integrating and Exchanging XML Data using Ontologies. [online]. Retrieved from: <http://wordnet.princeton.edu>. [Accessed 11, August, 2016].