International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 4 (2015) pp. 11555-11563

http://www.ripublication.com

Interlinked Online Knowledge Search Data Hub By Leveraging Semantic Data Integration

N.Gomathi Kalyani
U.G Student, Department of Computer Science
Sathyabama University, Chennai-600119, India
Kalyanirvs2000@gmail.com
A. Dhivya Priyaa
U.G Student, Department of Computer Science
Sathyabama University, Chennai-600119, India
aashikabama@gmail.com
B.Ankayarkanni.M.E.,(Ph.D).
Assistant Professor, Department of Computer Science
Sathyabama University, Chennai-600119, India
ankayarkanni@yahoo.com

Abstract

Data Integration is the method of providing a unified read of the info unfold across totally different sources. In Data warehousing and Data mining the Data Integration plays an important role. Our paper deals with, "LINKED DATA INTEGRATION FRAMEWORK" of Semantic Data Integration. This Linked Data has a wide range of different domains like Media, Music, Library, Life Science data, Government Data, Geographical Data. Here the Music Data has been integrated by Linked Data integration of Semantic Data integration.

Keywords: Semantic data integration, Linked Data Integration Framework (LDIF), Resource Description Framework (RDF).SPARQL.

Introduction

Fundamentally, Data Integration involves combining the info from totally different sources with the similar read of information. The info integration method that has both the COMMERCIAL (Merging a database for a similar of two companies) and SCIENTIFIC (Combining results from different bio-informatics domain).Data integration evolves as a frequency increases in volume. The data integration mainly focused on that theoretical work and diverse drawback remains unresolved. In management space, individuals often consult with the info integration as associate Enterprise info Integration (EII).The joined information Integration Framework

(LDIF) incorporates an information supply use a large vary of various sources of the RDF tools.

Data integration

Data integration is employed to fetch the knowledge simply From an outsized set of sources. Victimization the information integration approach the sources are often viewed in a much materialized manner. The large amount of data in an online networks tempts the organization to perform data mining for commercial purpose to increase their revenue. For this purpose the data integration plays a larger role in online networks.

Semantic Data Integration

The semantic data is a software engineering model [2] based on the stored system and real world. The smantic data organized without human intervention it can be interpreted meaningfully. Back to the 1970's the semantic data has a history, the data management and application system are used widely.

History of Semantic Data Integration

In 1970's the US Air Force computer aided manufacturing program integrated for applying technology to increase manufacturing productivity.

Goals of Semantic Data Integration

Define and explore modern data integration methods for the data, which is organized linearly and hierarchically. Increase the efficiency of data integration for distributed systems. Describe the new data integration methodology (Semantic Data Integration) to virtually integrated data sources.

Linked Data Integration Framework (LDIF)

The linked data has a rapid growth and contains a data of different domains like media, music, library, life science data, geographical data and government datasets as Dbpedia. As of September 2011, this knowledge house is calculable to contain thirty one billion RDF triples and around 504 million RDF links between knowledge sources [6]. For instance, the $64000 world entity of country or place is known with totally different URI's of various knowledge sources. A similar URI's might contain conflicted values the usage of different URI's makes the application developer to difficult write SPARQL queries. Till now there has not been any integrated tools to develop the tasks. For this tasks the LINKED DATA INTEGRATION FRAMEWORK has helps the application developer to develop.

System Architecture of LDIF

This architecture implement the information reposting pattern. This diagram highlights the steps of knowledge integration method victimization the connected knowledge INTEGRATION FRAMEWORK (LDIF). They has four layers:

a) Application layer
b) Data access, integration and storage layer
c) Web of knowledge
d) Publication layer.

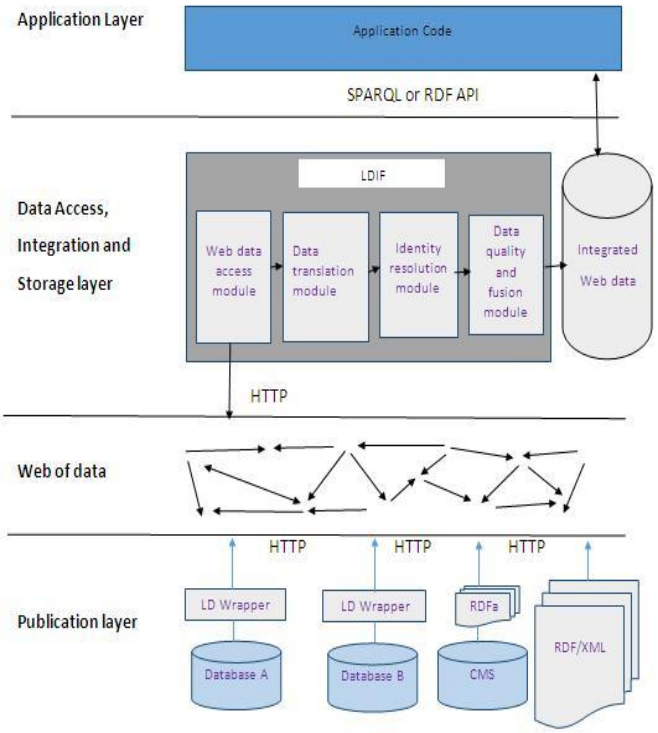

Figure 1: System architecture of LDIF

Fig 1. Shows the architecture diagram for the coupled information Integration Framework that implements the info storage pattern and indicate the steps involved in the pattern [1, 3].

A. Application layer

The application has an application code. The code has a components of the STRUCTURED QUERY LANGUAGE (SPARQL) and RESOURCE DESCRIPTION FRAMEWORK (RDF).

A. Data access, integration and storage layer

This layer has a LDIF of separate process. The application code must cross each process. The processes are,

- Web knowledge access module
- Data translation module
- Identity resolution module
- Data quality module
- Fusion module

B. ***Web of knowledge***

This layer has integrated code and the HYPERTEXT TRANSFER PROTOCOL (HTTP).

C. ***Publication layer***

This layer has a data base storage and Content Management System (CMS), LD Wrapper (Long Dwell Wrapper) and RDF.

Components oF LDIF

The Linked Data Integration Framework has a components. There are,

- Scheduler
- Data import
- Integration Run time Environment

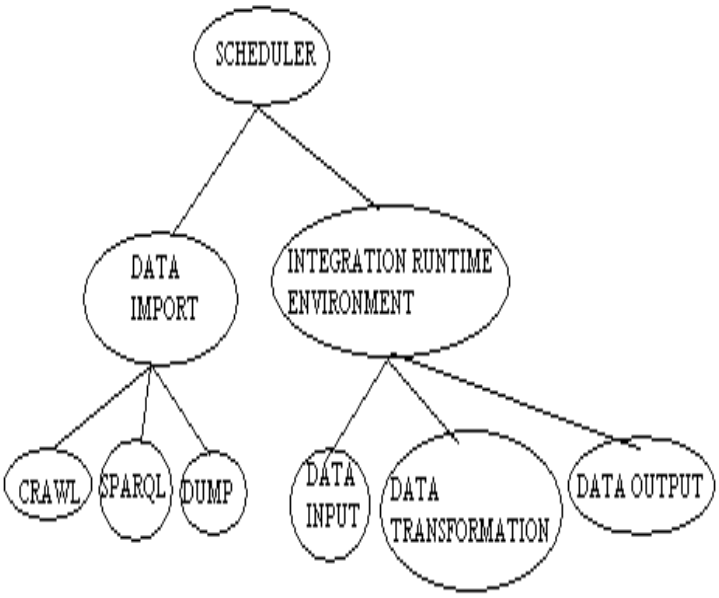

Figure 2: Components of LDIF

Fig 2. Shows the components of Linked Data Integration Framework (LDIF) consists of a Scheduler. The scheduler has modules like Data import and an Integration components [1].

A. Scheduler

The computer hardware is employed to unfinished knowledge import (or) the combination jobs. It specific once and the way a precise job be dead, configured with an XML document. To run the integration periodically, the scheduler component is useful. Otherwise it can just run the component.

B. Data import

The Linked Data Integration Framework can access the modules for the data sets replicated locally by the SPARQL (or) the file download. These different types of the imported jobs can generate a Meta data, throughout the process. These imported jobs are managed by the scheduler to refresh (Hourly based mostly or daily based) the native cache for every supply. The information import has 3 method. They are,

1. Triple / Quad dump import
2. Crawler import
3. SPARQL import

1. Triple / Quad dump import

Downloading a file containing a knowledge set is that the easiest way to urge a knowledge set from internet. This Triple / Quad dump import can do that method properly. The joined knowledge Integration Framework generates a graph for a triple dump import however this import can take the given graph from the quad dump import as origin graphs. This import can support RDF / XML, N-Triples, N-Quads and Turtles.

2. Crawler import

Data sets that area unit accessed solely by the differentiable URI's area unit sensible approach for a crawler. The Spider area unit used as associate degree import for crawler import jobs. Every and each crawled URI's contains a separate named graphs for the pursuit.

3. *SPARQL import*

The data sources that area unit accessed by the SPARQL area unit replicated by the coupled knowledge Integration Frameworks for a SPARQL modules. The SPARQL gets tracked by its own name graph.

C. Integration Run time Environment

The integration runtime has various modules for caching of the results and the execution of different modules of each stages. The stages are,

1. Data input
2. Data transformation
3. Data output

1. Data input

The integration parts expects the input file diagrammatic because the Named Graphs and keep in N-Quads format.

2. Data transformation

Linked Data Integration Framework has many transformation modules:

a. Data translation
b. Identity resolution
c. Data quality and fusion.

a. Data translation

The R2R framework [7] encompasses a net knowledge to represent the various vocabulary into one target vocabulary. These vocabulary languages square measure shown victimization the R2R Mapping Language [1].This language encompasses an easy transformation, advanced transformation and worth transformation for normalizing totally different units of measurements. The syntax for the R2R Mapping is analogous to a SPARQL Language.

b. Identity resolution

The joined knowledge Integration Framework has the Silk Link Discovery Framework [8] to search out the URI's area unit used inside the various supply to spot by an equivalent globe entity. The every set of duplicates area unit known by the Silk, Linked Data Integration Framework replaces the URI's with this single target URI's within the data output. It includes the OWL of the original URI's which makes the application to refer the data source on web.

c. Data quality and fusion

The Linked Data Integration Framework Sieve [9] has to provide the data quality evaluation and cleaning. This process has two steps:

- The knowledge quality module has assigned every names graph at intervals the processed knowledge supported the user quality policies. This policy combined a top quality connected Meta knowledge info employed in the method.
- The knowledge fusion has input as quality scores and reanalyze the information supported scores. Sieve provides the set of quality perform and fusion perform for the implementation of the value-added domain specific functions.

3. Data output

The connected knowledge Integration Framework output results is hold on (or) written to file (or) to the Quad store.

1. File output
2. Quad store output

1. File output

The output files supported by Linked Data Integration Framework are of two types:

a. N - Quads
b. N – Triples

a. N – Quads

It dumps the information into one file N- Quads. This file features a version that area unit translated of the graphs from the input graphs of the contents of rootage graphs.

b. N – Triples

It dumps knowledge the info the information into one file N – Triples however there exit no association to the information any longer and when it offers the output as N – Triples the birthplace data is rejected rather than giving output.

2. Quad store output

The data is written during a Quad store as an inventory of the SPARQL streams. The LDIF provides the implementation of the integrated runtime setting by 3 ways:

- In-memory version
- RDF hold on version
- Hadoop version

Proposed Work

Our paper deals with the method of LDIF and with its examples to integrate the different data sets. The example data sets are:

1. Music
2. Life science
3. Library

In this paper we have a tendency to area unit progressing to use the music information sets. Its shows however information sources area unit accessed for the various music connected information mistreatment the LDIF internet information parts and conjointly the opposite steps like information translation and identity resolution.

Data integration of music by LDIF

The LDIF integrate the info is applied to the music albums, artists, labels and genres mistreatment a number of the remote sources. They are:

- Dbpedia [1]
- BBC music [3]
- Music brainz [4]
- Free base [5]

Each remote sources can be accessed by the module. The Dbpedia can download the dataset [1]. The BBC and the music brainz are accessed by the SPARQL because of the unavailable data set. The free base are crawled because of the lack of possibilities.

Imports

The imports of the music data sets of LDIF are:

- Dbpedia – properties dump , types

- BBC music - SPARQL – artist,birth,death,record
- Music Brainz – SPARQL – artist,label,record
- Freebase – Crawl

Mapping
The mapping file provides for the interpretation supply information into the target information R2R Mapping file.

Silk Identity resolution
This resolution square measure won't to realize the music artists, label and album.

- Silk link specification : musicLinkSpec.xml
- The music brainz has not support the information of genre.

Exception method
To run the instance, transfer the LDIF and run the commands:

1. Change the LDIF root directory
2. Bin/ldif-integrate.bat examples/music/integration-config.xml [1]
3. Bin/ldif-integrate examples/music/integration-config.xml [1]

Related Work

Vigneshwari et al. [10] has created a work based on the Web Ontology Language(OWL) which is based on the particular user. The semantic similarities are also proposed by the the system model.

Mary et al. [11] has a proposed concept of clustering with XML documents.By these all the documents are clusted in a group repeatedly.

Gomathy et al.[12] has a work on agile integration of different business silos using the "Service Oriented Architecture" to enable the business enterprise systems flexible,loosely coupled and improve agility.

References

[1]. http://ldif.wbsg.de/#example1
[2]. Christian Becker :How to integrate Linked Data into your application.Semantic Technology & business conference,San Francisco, June 2012.
[3]. Andreas Schultz, Andrea Matteini,Robert Isele,Pablo Mendes,Christian Bizer,Christian Becker:LDIF – A Framework for Large scale Linked Data Integration.21st International World Wide Web Conference (WWW2012), DevelopersTrack.Lyon,France.April 2012.
[4]. Christian Becker, Andrea Matteini, "LDIF – Linked Data Integration Framework". SemTechBiz 2012,Berlin,February 2012.

[5]. Andreas Schultz,Andrea Matteini,Robert Isele,Christian Bizer,Christian Becker(October 2011),"LDIF- Linked Data Integration Framework"2nd Internation workshop on Consuming Linked Data,Bonn,Germany.

[6]. Heath,T.,Bizer c:Linked Data:Evolving the web into global data space.Morgan& claypool Publishers,ISBN 978160845431,2011.

[7]. Bizer,C.,Schultz,A.The R2R Framework:Publishing and discovering mapping on the web.1st International Workshop on Consuming Linked Data(COLD 2010),Shanghai,November 2010.

[8]. Isele,R.,Jentzsch,A.,Bizer,B.:Silk Server-Adding missing Links while consuming Linked Data.1st International Workshop on Consuming Linked Data(COLD 2010),Shanghai,November 2010.

[9]. Mendes,P.,Muhleisen,H.,Bizer,C.:Sieve-Linked Data Quality Assessement and Fusion.2nd International Workshop on Linked Web Data Management(LWDM 2012),Berlin,March2012.

[10]. Vigneshwari,s.,Aramudhan,M.,"ATechnique to user profiling ontology mining and relationship ranking."Journal of Theoretical and applied Information Technology,Vol.58.No.3,Dec.2013.

[11]. Mary posonia,A.,Jyothi,V.L."Context-based classification of XMLDocuments in Feature Clustering."Indian Journal of Science and Technology,Vol 7(9),1159-1162,Sep.2014.

[12]. Gomathy,C.K.,Rajalakshmi,S.,"A Efficient Business Process Integration and quality service for Service-Oriented Architecture."International Journal on Information Science and Computing,Vol.6.No.2.July 2012.