



# An Effective Knowledge Modelling of Distributed Databases using Ontology and Information Retrieval

Kranthi Kumar.R<sup>1</sup>,Dr. B. Padmaja Rani<sup>2</sup>,E.Hemalatha<sup>3</sup>,Dr.Harish Sharma<sup>4</sup>

<sup>1</sup>Research Scholar,JNTUH college of Engineering,Hyderabad,India,kranthikumar.rudrarapu@gmail.com

<sup>2</sup>JNTUH College of Engineering, Hyderabad, India,padmaja\_jntuh@jntuh.ac.in.

<sup>3</sup>JNTUHCollege of Engineering, Hyderabad, India,hemamorarjee@jntuh.ac.in.

<sup>4</sup>RajasthanTechnical University, Kota,India,hsharma@rtu.edu.ac.in.

## ABSTRACT

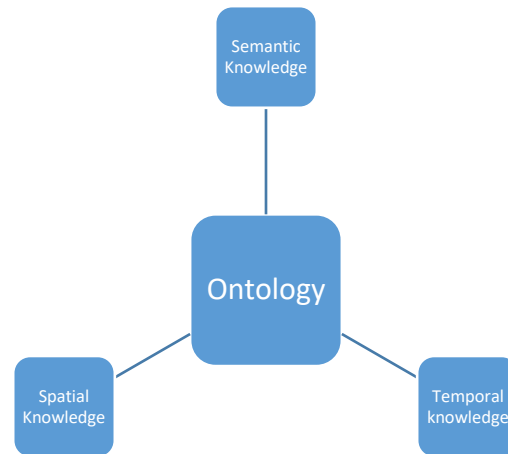
.There's massive amount of secret information in the internet. Numerous methods were developed to obtain the amount of quality information from the internet. This involves the execution of complex queries which could also yield insignificant results. To overcome these difficulties, this paper focusses on a method for obtaining information from meaningful data which is structured, for this researchers are focusing on the processing of queries based on ontology, where collected information is more relevant to the needs of users. Our proposed system gives more accurate results in retrieving the information required by the users using ontologies.

**Keywords:** Distributed databases, D2R,Knowledge representation, Ontology,SparQL.

## 1. INTRODUCTION

Structured application specification languages are one way of retrieving information in information management systems. Semantic-based methods using ontologies of the domain have been adapted for modeling data and retrieving information. Ontology [1]based information representation and reasoning strategies for the analysis of tasks or procedures provide advanced knowledge about the world. Ontology aims to reflect the information about the world, events and behavior that help make robots more autonomous while planning paths. In a nondeterministic world, information reasoning techniques can draw new conclusions and thus help to dynamically plan.

The below figure1 shows the ontology knowledge representation in terms of semantic knowledge, spatial knowledge and temporal knowledge.



**Figure 1: Ontology Knowledge representation**

Given that Distributed DataBases (DDBs) became a requirement for users and real time applications in the present world, still DDBs are not as perfect as centralized DBs, with reliability and efficiency in their search engines. In addition, due to the timing problem in the scenario of distributed environment and the quality of data, DDB could not benefit from text search engines.

In this paper, we propose a new way of dealing DDB by employing D2R mapping approach, as a contribution in this research area. We got good results when compared with traditional way of getting results.

Looking at the vast quantities of data [2] generated on a daily basis, there is a great need for advanced, strong and effective technology that can efficiently manage data storage, processing and analysis, thus saving hardware resources and time. Data delocalization[3] through multiple network-connected networks significantly reduces costs and ensures greater connectivity and efficiency.

A distributed database (DDB)[6] is a group of many logically interrelated databases that are distributed over a network of computers. A centralized framework for the management of databases. (DDBMS) is the program that manages the DDB and offers an access mechanism that makes it easy for people

## 2. LITERATURE SURVEY

Structured question in the information management systems gives us a response. Writing organized query is a way of accessing data [4], as it enables end users to create their own queries to extract information by getting awareness of different semantics of queries.

Nonetheless, the creation of queries by considering the visualization of results and optimization methods remains appreciably tough for different user rates of the system. Recovery of information in current years became very hard with growing usage of applications [6] for data mining, decision support, and business analytics.

Researcher's idea has been on methods that give more importance to visualization database techniques and bilateral forms of query in graphical mode, keeping eye on condition that reciprocal query results by considering natural language methodologies.

Lately, semantic-based perspective of using an area-wise ontology [5] utilized for data modeling and getting information. Ontology-based method of retrieving information is primarily aimed at enhancing the interaction in between requests and data retrieved to give the accurate results almost nearer to the needs of the users. An ontology typically implies a common, accepted, and comprehensive replica in a given area of problem.

The key benefit of utilizing a domain ontology, [2] describes a semantic data [7] copy combined with the awareness of a domain linked with that. So ontologies are utilized to formulate lean techniques for finding data.

The governance of big data [8] also involves various challenges including structuring, searching, analysis, visualization, etc. Because of the broad and complex data sets the retrieval of deposited information employing conventional data administration responsibilities is difficult. Currently a small amount of research is going to investigate accurate ways in which big data can be processed for obtaining outcomes as listed in terms of: search performance, getting linked data and enabling analysis.

Through this way, ontologies are worked out to find Big Data knowledge [9]. Considering these aspects, later efforts by researcher's efforts may aim at investigating the feasibility of using the structure of knowledge base of ontology to specify the metadata of big data as a basis for the efficient discovery of useful analytical information.

The advantages of RDF can be given as, RDF with a promising framework, encourages the giving of metadata of resources across Internet and also RDF contains a quality syntax to describe and query the data, any software that uses the metadata, the task becomes easy and fast to get results by exchanging information between different applications [10] which can also be used by sophisticated software agents.

The disadvantages of RDF can be given as, (OWL Ontology Web Language) features [11] cannot be queried using RDF. One cannot be able to query RDF document using XPath. Querying is complex in RDF since data in RDF is in graphical format (tree). RDF imposes constraints on design of XML, so designing own type of XML is very difficult. Without proper knowledge of namespaces, RDF can't be used because RDF utilizes namespaces to differentiate its objects and their properties along with resources. Finally knowledge of RDF vocabulary is required to use the RDF format.

The advantages of D2R can be given as, it facilitates RDF and HTML browsers to give best content to the users based on their input by employing SPARQL [12]. D2R stands in one among popular methods, which can publish the relational database into Linked data. Database information [13] browsing is done by employing a website which navigates according to the requirements given by the user and in getting the appropriate result set which can be understood by a human easily from RDF data. In simple words information about RDF [14] can be accessed through URI across the web.

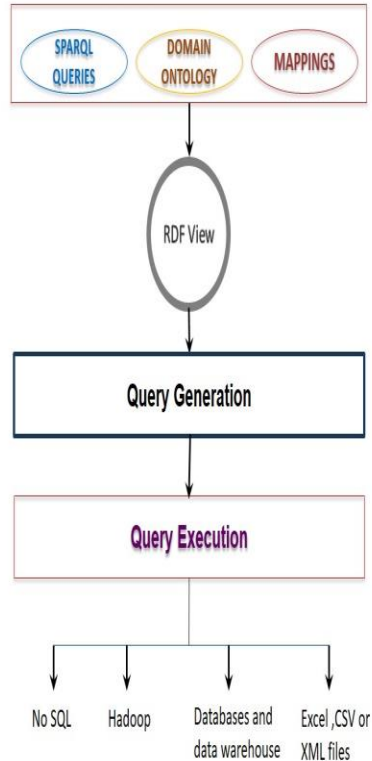
## 3. PROPOSED SYSTEM

The main objectives proposed in this paper are to define a semantic data model, specify knowledge of the domain, define relations between various forms of semantic information and build a unique ontological knowledge structure process for any dataset

The methodology proposed in this paper involves the following steps.

- Ontology generation by taking input from a database.
- Processing the acquired knowledge in the form of ontological knowledge.

- Finally getting relational database queries from ontological knowledge.



**Figure 2: Proposed Architecture**

The proposed architecture (figure2) retrieves the information from the databases using D2R mapping with the help of SPARQL query. In the process, we will use RDF Format in this paper even we can also use Virtual Graph mode. The Successful result, depends on the Query we have given and for this we can transform the query before execution so that our results are most appropriate.

#### 4. IMPLEMENTATION

The method of extracting information from ontology is carried out by following a few sub-processes keeping in view of distributed databases. The sub-processes include: Mapping databases to ontologies i.e. format: RDF. Retrieving Ontology details. A database mapping to ontology The data present in the form of tables i.e. rows and column format can be converted into RDF graphical format as well as TURTLE i.e. RDF textual syntax. "RDF is a system for resource classification" reproduces ontologies in graphic format. RDF follows tree-structured representation of the graphical results. TURTLE is "Terse Resource Description Framework "(Turtle) is textual representation "Resource Description Framework "(RDF) data model which is graphical. Data mapping is done by D2R mapping for which we should first install D2R

server in our network, and then Java MySQL connector.jar file as well. After performing the above operations, from the path where D2R server is mounted and its respective files are stored we need to go to command prompt. Its suggested to store downloaded D2R server folder in system drive D.i.e data drive and make sure Java MySQL connector jar file exists inside D2R server folder i.e" D:/d2rq0.8.1/mysql-connector-java-5.1.39".

Java has to be built in the system before any of this. Path configuration should be performed where D2R server directory is connected to windows route and checked as well. This is the preliminary setup needed to start the process itself. The next move is to map the database i.e. to construct an RDF mapping using D2R server for MySQL database.

With reference to we execute command "generate-mapping -o outputfilename.ttl-u youruserid -p yourpassword jdbc:mysql://localhost/yourdatabasename" The output of mapping is stored in .ttl format in "D:/d2rq0.8.1/outputfilename.ttl". After executing above steps now we have to start D2R server web interface by following below steps: Run command in command prompt "d2r-server outputfile- name.ttl". We can view D2R web interface at "http://localhost:2020".

Finally we dump databases by creating a RDF mapping of a MySQL database using D2R server. To perform this in Windows command prompt we run command "dumprdf -f RDF/XML -b http://localhost:2020/outputfilename.ttl & output- filename.rdf". After executing we can view our RDF dump file in "D:/d2rq- 0.8.1/outputfilename.rdf". B.

We execute command in SPARQL format to Retrieve Ontology information. SPARQL is a query language used to store and retrieve data in RDF format.

Finally, we link SPARQL endpoint to HTML web page where retrieving information based on the entered ID of each data becomes user friendly. HTML type connects the values entered to a Python script which plugs the values into SPARQL query and formats the result as HTML. The above steps include:

Firstly, write a SPARQL query requesting detailed details Next, create a site in HTML form Where the end-user is able to enter the question value.

Finally, add the SPARQL query to a popular Python gateway interface script which takes the passed values from the web form, connects them to the respective locations in the database, sends the database to endpoint, and then displays the result as HTML. The effects of our information extraction can be viewed in html web page with this.

The data is translated into RDF format and TURTLE format after data from relational database undergoes mapping into D2R. The data in the form of rows and columns is first transformed by D2R mapping to TURTLE(.ttl) format, where we create RDF mapping using D2R server for MySQL database. The welcome page of the D2R server is viewed as:

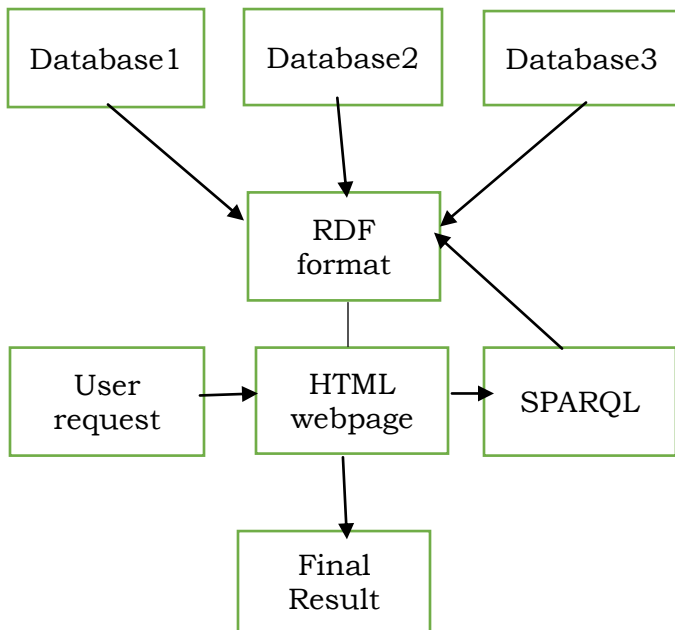
The TURTLE data structure can be interpreted as the following set of patient information that was previously of relational database structure.

For the above TURTLE format data in RDF graph is  
 To retrieve information from ontologies  
 We execute command " SELECT DISTINCT ?hasValue  
 ?isValueOf WHERE  
 http://localhost:2020/resource/patientdetails/23456  
 ?property?hasV alueUNION; http : //localhost : 2020  
 /resource/patientdetails/23456  
 ?property ?hasValue ORDER BY  
 (!BOUND(?hasValue)) ?property ?hasValue  
 ?isValueOf "

After performing this we connected this information retrieval process to HTML page and create a user friendly platform to retrieve information from ontologies.

**5.RESULTS AND DISCUSSION**

**5.1 Evaluation model**



**Figure 5.1.**Flowchart for information retrieval using D2R Mapping

The above figure explains the evaluation process of this model. Initially the data is collected from different databases located at various locations and later they are

converted into rdf formats for a better security and easy retrieval with the help of SparQL query. The user enters the required details with the help of a html (hypertext markup language) webpage. Once the user gives the inputthrough SparQL query and required information is retrieved from RDF Format with in very less time. The main advantage of this whole process is the system is more secured and reliable.

The sample rdf format can be represented as

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:vcard="http://www.w3.org/2001/vcard-rdf/3.0#">
<rdf:Descriptionrdf:about="http://127.0.0.1//621210">
  <vcard:LABEL>29.5</vcard:LABEL>
  <vcard:Other>No</vcard:Other>
  <vcard:Locality>hyderabad</vcard:Locality>
  <vcard:TEL>8975489962</vcard:TEL>
  <vcard:Region>TS</vcard:Region>
  <vcard:BDAY>23</vcard:BDAY>
  <vcard:UID>621210</vcard:UID>
  <vcard:TITLE>Female</vcard:TITLE>
  <vcard:EMAIL>abc@gmail.com</vcard:EMAIL>
  <vcard:FN>abc</vcard:FN>
  </rdf:Description>
</rdf:RDF>
    
```

**5.1.1. A retrieved RDF with patient details**

The above rdf format gives the patient information details including locality, mobile number, user-id, email etc.

**5.2 Comparison**

**Table 1:**Comparison of various approaches

S.No	Method Adapted	Performance	Conclusion
1	Traditional method	Takes more time	Not used frequently
2	SQL Query	Average	Used in simple databases
3	SPARQL Method using D2R Mapping Approach	Better	Preferred more especially when we have large databases

The Table1 shows the following information

- 1 The traditional method retrieves the values from the excel sheets and gives the results to the user
- 2 The SQL Query method takes the query as an input from the user and searches in the database for the appropriate results and gives it back to the user

3 The SPARQL method uses D2R mapping tool and searches in the RDF which preserves the data without any loss and gives the intended result to the user based on the input given by the user.

## 6. CONCLUSION

We identified a new framework for retrieving information where a front end web page takes input from the user and provides the necessary details from RDF formats in the back-end. The back-end process involves converting patient databases to ontologies using D2R mapping and querying these data using SPARQL query language after obtaining RDF data. Using python script, SPARQL endpoint is then finally linked to the HTML web tab. Each of this is provided on hospital patient servers. As technologies such as machine learning and deep learning are combined, it can be enhanced in future to distributed databases of any organizations.

## ACKNOWLEDGEMENT

This work is sponsored by a research project grant from TEQIP-III, JNTUH University, Hyderabad with title “Modelling Distributed Databases using Ontology”.

## REFERENCES

- [1]Escobar Esteban, Pilar and Roldan, Maria del Mar and Peral, Jes\_us and Candela, Gustavo and Garb a-Nieto, **An Ontology-Based Framework for Publishing and Exploiting Linked Open Data:A Use Case on Water Resources Management**, volume 10, Applied Sciences,2020.
- [2]Nai-wen Chi, Yu-Huei Jin, Shang-HsienHsieh, **Developing base domain ontology from a reference collection to aid information retrieval**,Automation in Construction, Volume 100, April 2019, Pages 180-189.
- [3]VanjaLukovi, SašaCukovi, DanijelaMiloševi, GoranDevedži, **An ontology-based module of the information system ScolioMedIS for 3D digital diagnosis of adolescent scoliosis**, Comput Methods Programs Biomed. 2019 Sep;178:247-263.
- [4]Kamran Munir, M. SherazAnjum, **The use of ontologies for effective knowledge modelling and information retrieval**, Applied Computing and Informatics Volume 14, Issue 2, July 2018, Pages 116-126
- [5] Chengbin Wang, Xiaogang Ma, Jianguo Chen, **Ontology driven data integration and visualization for exploring regional geologic time and paleontological information**, Computers & Geosciences, Volume 115, June 2018, Pages 12-19.
- [6] Gayathri R, V. Uma, **Ontology based knowledge representation technique, domain modeling languages and planners for robotic path planning: A survey**, ICT Express 4 (2018).
- [7]FadouaHassena ,GrissaTouziAmel, **An efficient synchronous indexing technique for full-text retrieval in distributed databases** International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France.
- [8]Naseebah H Alkahtania, Sarah Almohsen, Nasibah Mohammed Alkahtani, Ghadahabdullahalmalki, Safaa Samir Meshref, HebahKurdi, **A Semantic Multi-Agent system to Exchange Information between Hospitals**,The 8th International Conference on Ambient Systems, Networks and Technologies(ANT 2017).
- [9] Dimitrios A. Koutsomitropoulos, Aikaterini K. Kalou, **A standards-based ontology and support for Big Data Analytics in the insurance industry**, ICT Express Volume 3, Issue 2, June 2017, Pages 57-61.
- [10]Shengtao Sun, Weijing Song, Albert Y. Zomaya, Yang Xiang, Kim-Kwang Raymond Choo, Tejal Shah, Lizhe Wang, **Associative retrieval in spatial Big Data base spreading activation with semantic ontology**,Future Generation Computer, 2016.
- [11] GeorgiosSantipantakis, KonstantinosKotis, George A. Vouros, **Ontology-Based Distributed framework for Accessing Integration and Reasoning with data in disparate data sources**, Expert Systems with Applications, Volume 90, 30 December 2017, Pages 464-483.
- [12][https://www.cs.ox.ac.uk/files/5511/owled2013\\_wp3.pdf](https://www.cs.ox.ac.uk/files/5511/owled2013_wp3.pdf).
- [13]Prud'hommeax, eric and Seaborne, Andy **SparQL query language for RDF**,2010.
- [14]Yadav, Poonam and Singh RP, **Ontology-based intelligent information retrieval method for document retrieval**,International Journal of Engineering Science, volume 4, 2012.