Semantic Web Query on e-Governance Data and Designing Ontology for Agriculture Domain

Swaran Lata[1], Bhaskar Sinha[2], Ela Kumar[2], Somnath Chandra[1], Raghu Arora[3]

Department of Electronics and Information Technology, New Delhi[1] , India
[1]{slata, schandra}@deity.gov.in
Gautam Buddha University ,Greater Noida, India
[2]bhaskar_sindel@hotmail.com, ela_kumar@gbu.ac.in
W3C India ,New Delhi, India
[3]ragsarora85@gmail.com

Abstract

*Indian agriculture has made rapid progress on the agricultural front during the past three decades and is in a queue of the major producer in the world. But still it has long way to go and meet challenges ahead such as communication, resources, and availability at right time at right place. The web has had an amazing existence and it has been the driving force for a cause to grow information across boundaries, enabling effective communication and 24x7 service availability all leading to a digital information based economy that we have today. Despite that, its direct influence has reached to a small percentage of human population. Since localization populated with India and the applications are translated and adapted for Indian users. With the possible localization of spread raw formatted Indian government data, at different locations are thought to have integrated with each other using the internet web technology as – **Semantic Web Network**.*

Keywords
Semantic web,   Resource Description Framework,   Ontology

1. Introduction

As we know that semantic web network technology has gained superiority over the previous web technology because of its added feature of intelligence in web. So, the ICT's role has advanced to get the benefit of its potential into respective domain where its use is inherent to facilitate service users/stakeholders. Agriculture is one of the important domain of any country on which people's food and livelihood depends. India is one of the country whose dependency on agriculture and its related activity is mostly sought. India as it contributes 16% GDP and provides employment to 52% of the Indian population. Timely access to information and service delivery is critical and of utmost important in this sector in view of time bound farm activities involved in all stages of the crop cycle [9]. Population and Agriculture is closely related according to FAO, UN and every country should make their strategy accordingly to reduce the cost of generation of census data and plan better. India is so vast and rich in varied agricultural product and its produces that distribution of products and service access remains untouched to some parts of the country. One major component of an progressive workplace is ICT accessibility - including websites and intranets. In this connection semantic web technology is well suited to facilitate the services and manage the resource distribution through proper management of knowledge base. Semantic web's support for linked data is one of the solution to integrate all data sets and information spread across the various URL sites, that promises the availability of information and data at any time any were in the web. Making a consistent data available in web is one of the challenge and difficult to maintain. Semantic web's RDF data support is the building blocks for developing ontology of any domain of interest. Our agriculture domain ontology is based on this RDF data and because of its inherent

powerful feature of machine readability, representation of information and dynamic linking with other information datasets is amazing over XML based datasets. We are using Protégé framework to design and develop the our agricultural model and generate RDF data set to further use in query building and searching techniques. Also Oracle database has been used to store RDF/OWL datasets to enhance the farmers database. We also used SPARQL end-point for querying the RDF datasets. Rest of the paper is organized as - In section 2, we first briefly describe in the Indian scenario and shortcomings in agricultural domain which eventually builds a motivation for building such integrated development environment. Section 3 describes methodology of design and development. Section 4 we have design a logical Agriculture Ontology model to generate RDF/OWL format. Section 5 defines the query processing over SPARQL end-point. Section 6 describes the future scope for enhancement and section 7 includes conclusion.

2. Indian Agriculture And Its Resource Distribution – A Challenge For Semantic Web Technology

Well this is fine to rate our agriculture in the world scenario, but still we have some of the unfinished agenda in land reform, quantity and quality of water available, technological fatigues, access to information and data, adequacy and timeliness of institutional credit, etc. Adverse meteorological factors add to these problems. The worst affected are small and marginal farmers, tenants and share croppers, landless agricultural labour and tribal farmers, since their coping capacity is very limited[14]. Women suffer more since they have little access to institutional credit or organized extension support. Under these initiatives, Ministry of Agriculture Govt. of India itself has number of websites and portals for different divisions, directorates and projects related to various departments is currently existing. However these websites do not share web services among them and hence contents are static, non-consistent, non-integrated. Many time farmers and other stakeholders in the agricultural sector have to visit multiple websites to trace the desired piece of information or to avail a single service. Because of the varied technology standards, look-and-feel and aesthetics involved in these sites resulting a lot of inconvenience to the user and requires a lot of learning on their part to access the information and services. This results in repeated efforts, obsolete content, multiple sources of information, mismatch of information and hence confusing the service consumers.

Table 1. Ref. *World Population* [15]	
China	19.4
India	17.5
U.S.A	7.26
Indonesia	15.05
Brazil	9.39
Pakistan	24.78
Bangladesh	16.76
Nigeria	26.84
Russian Fed.	-4.29
Japan	1.1
Other Countries	15.43
World	12.97

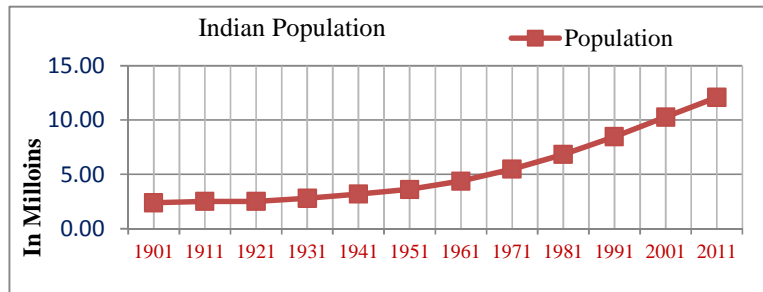**Figure.** 1. *World Population Graph*
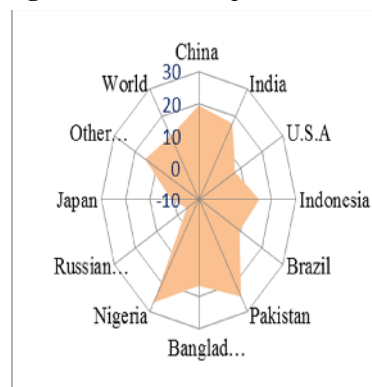
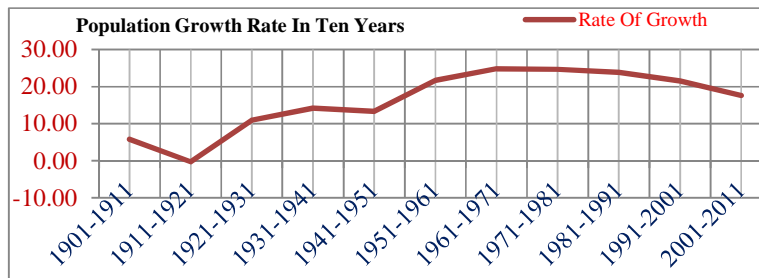**Figure. 2**. *Ref. Indian Population*[15]

Figure. 3. Ref. *Indian Population Growth Rate*[15]

3. Methodology

Semantic Data Model for Indian Agriculture Domain for e-governance is using Semantic Web Technology and is based on RDF/XML/OWL data format of raw unstructured government data spread across various sites, which is further analyzed by setting trend analysis and spline curve fitting method in Matlab for interpolation of data into matrix format to minimize the error. Protégé Framework is used to design and develop our logical model into RDF/XML/OWL format dataset, which is the immediate objective of our getting well-form validated datasets. Protégé helps in correlating datasets, defining and describing well among the piece of datasets and establishes relationship between Subject-Predicate-Object an inherent feature of RDF concept. It helps to reason as well as infers the class and subclass validity and their relationships.
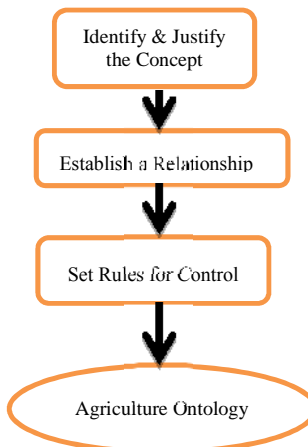

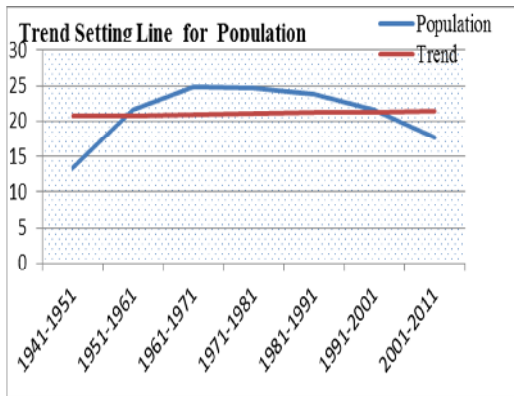**Figure.4** *Agriculture Ontology Design and Development Process Steps*

Figure.6 *Spline Curve fitting Analysis of Population Growth. X and Y-axis is divided by 10*

Figure .5. *Trend analysis Of Population Growth Rate* [15]

3.1   Identify & Justify the Concept

Logical model is identified on the basis of facts, information and data available or derived from facts. Present functioning system and its drawbacks, short comings help in analyzing and justifying better approach to a solution. Such as the current system is not able to fulfil the farmers /stakeholders query and their day-to-day problem which the semantic web technology promises to prove better solution in terms of cost/benefit and others with ease. Lot of distributed infrastructure support may resulting in cause of functional failure of the system. Optimized results through graph based search techniques justifies to adopt semantic web technology.

3,2   Establishing a Relationship

Semantic Web Technology's basic building block is finding relationship between subject-predicate-object, which forms a triplet and this triplet infers URI to the web sites so as to make relationship between other nodes of the graph network of interest. Query which is done through specific query handling process with valid criteria to get desired results.

3.3   Set a Rule for Control

Constraints are imposed to set SWRL axioms and laws to guide the search criteria or specific purpose and move to the right destination dynamically. We use OWL for describing rules to guide the search path, reasoning and inferring during clients request.

4. Agriculture Ontology Design - Logical Model

This model is an Agricultural Logical Model, there are various associated components attached with this overall model. We tried to generalized it by depicting its components. Other components are attached with its characteristic and functionality based on resources input /output and other concerning dependent bodies. Content management system supports handling all features and terms available including local datasets structured/unstructured as well as web related resources such as data available at URLs, which gets processed to get in RDF/OWL format and stored into database and accessed by various components to provide services to farmers/stakeholders. Ontology engg. Process helps to do proper analysis and reasoning /inferring of each component process with demand which will guide through OWL. Analyst /Expert can avail the resources and do

queries of interest. Also can get various generated data /information reports and other form of resources in return.
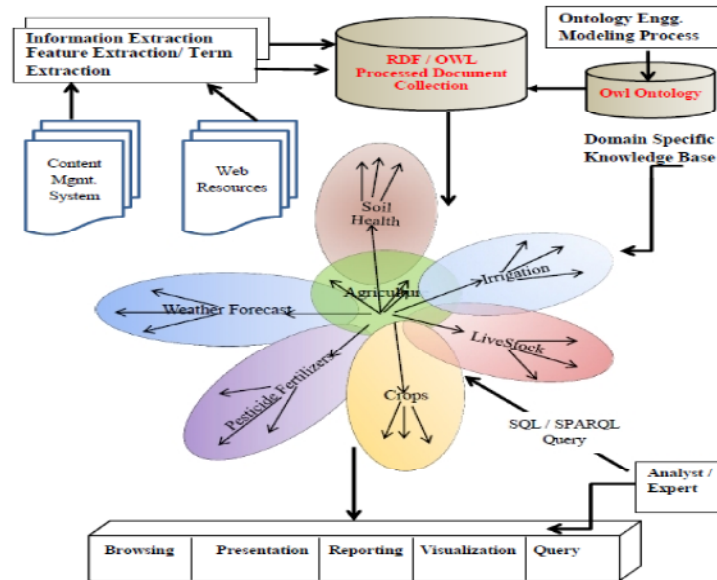

Figure. 7. *Agriculture Ontology Model and its various components.*

5. Query Processing

In this section we define the query processing done at SPARQL end-point which takes RDF format data and generating desired query result if any criteria set for. We have used open link's SPARQL end –point to generate query. Query in triplet is generated to see the desired result of specific interest.
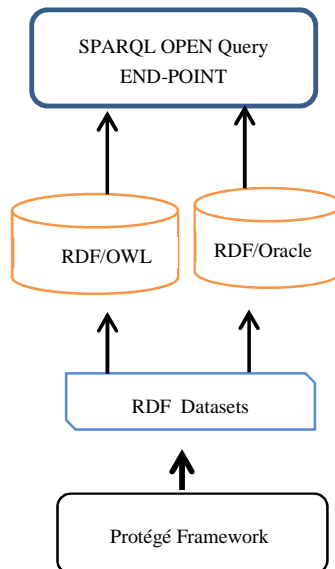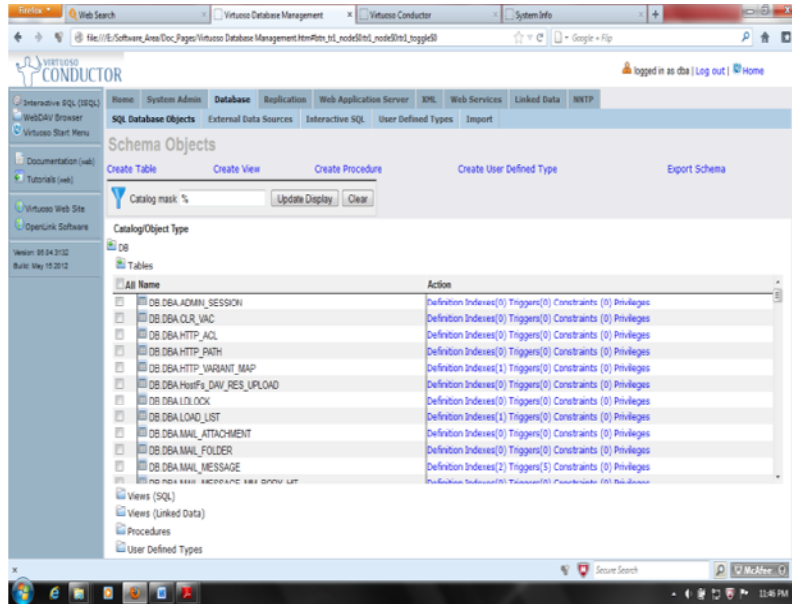

Figure.8 *Query Processing at SPARQL End-Point*

Here Ontology is reasoning and inferring for the concept and trust which has been set for verification and validation.
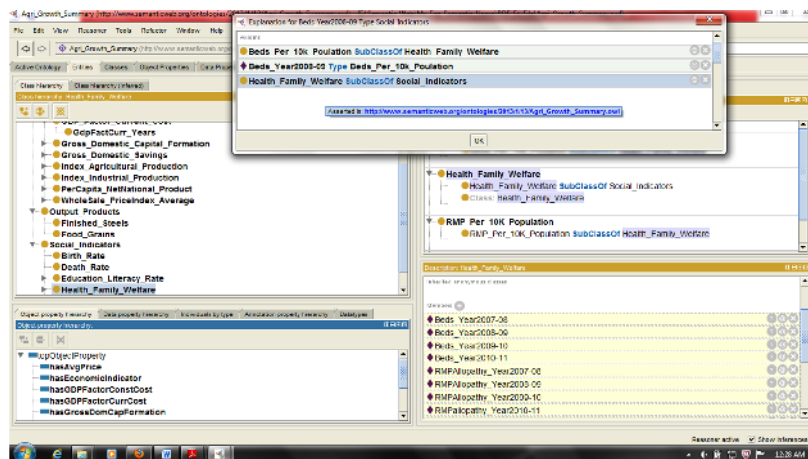

Figure. 10 *Protégé Framework used for Designing and Developing Ontology.*

6. Future Scope Of Work

Here in this paper we have drawn an overall idea of querying and searching RDF data by creating an ontology of agriculture domain through Protégé, there are various other way to develop and design ontologies using other frameworks. One of the way is to use of Jena/ Sesame that can be used to process RDF datasets for describing ontology better way. In our future work we are likely to include axioms and rule language of OWL to query in complex and specific query requirement to meet the desired result. Also, our effort will go to integrate other components of the agriculture domain and establish integrity and relations so that every component can share wherever it is needed.

7. Conclusions

RDF/OWL dataset has its inherent capacity to establish natural relationship between URLs. Although it is very unpredictable to draw idea about next URLs network with whom current URL is trying to make relationship. And there exist a dynamic complex model of class to handle this. We draw a conclusion by saying Semantic Web Technology has amazing and tremendous power of mapping complex logical model into graphical model which binds and describes to the machine to understand that reduces lot of intermediate intricacies to carry out web based linking URLs with ease. This technology opens a door for intelligence era in web technology.

8. References

[1]   Berners-Lee, T., Linked Data, July 2007
        http://www.w3.org/DesignIssues/LinkedData.html

[2]   Hendler, J. 2008. Web 3.0: Chicken Farms on the Semantic Web.
        Computer 41, 1(Jan. 2008), 106-108.

[3]   Bhall. G. S, Singh. Gurmail, Final Report on Planning Commission Project Growth of Indian Agriculture: A District Level Study, G.S .New Delhi (Mar 2010)

[4]   Accenture: eGovernment leadership: High performance, maximum value.
        http://www.accenture.com/NR/rdonlyres/D7206199-C3D4-4CB4-A7D8-846C94287890/0/gove_egov_value.pdf (2004).

[5]   Berners-Lee, T., Putting Government Data online, June 2009
        http://www.w3.org/DesignIssues/GovData.html

[6]   Li Ding, Dominic DiFranzo, Deborah L. McGuinness, Jim Hendler, Sarah Magidson, The Data-gov Wiki: A Semantic Web Portal for Linked Government Data, 2010.

[7]   United Nations Statistics Division. 2011. 2010 World population and housing census programme: census questionnaires. [Online] New York: UNS http://unstats.un.org/unsd/censuskb20/Knowledgebase.aspx.
.

[8]   Keita, N.; Srivastava, M.; Ouedraogo, E. & Kabore, M. 2010. Collecting agricultural data from population census: Overview of FAO recommendations and experiences of Burkina Faso and other countries. Fifth International Conference for Agricultural   Statistics, Kampala, Uganda, 13 - 15 October 2010)

[9]   Kumar. Ranjit, Ailing Agricultural Productivity in Economically Fragile Region of India: An Analysis of Synergy between Public Investment and Farmers' Capacity,(Jan. 2010).

[10]   http://www.semanticweb.com/

[11]   www.w3.org/standards/semanticweb/

[12]   www.w3cindia.in/

[13]   www.bbc.co.uk/

[14]   National commission on Farmers, 25th May2006, Serving Farmers and
        Saving Farming, Ministry of Agriculture, Govt. of India, New Delhi.
        http://agricoop.nic.in/Agristatistics.htm

[15]   http://www.censusindia.net/

[16]   http://www.semantic-web-book.org

[17]   http://www.agricoop.nic.in
[18]   W3C, SPARQL Query Language for RDF,
         http://www.w3.org/TR/2006/CR-rdf-sparql-query-20060406, 2004.
[19]   F. Manola and E. Miller (Ed.), "RDF Primer," W3C Recommendation, February 2004,
         http://www.w3.org/TR/rdf-primer/.
 [20]   M. Wylot, J. Pont, M. Wisniewski, and P. Cudr´e-Mauroux,"diplodocus[rdf] - short and long-tail
         Rdf analytics for massive webs of data," in International Semantic Web Conference
         (ISWC), 2011.
[21]   L. Zou, J. Mo, L. Chen, M. T. Oezsu, and D. Zhao, "gstore: Answering sparql queries via sub-
         graph matching," PVLDB, vol. 4, no. 8, 2011.
[22]   Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, "Foundations of Semantic Web Technolo-
         gies".Chapman & Hall/CRC, 2009.
         http://www.semantic-web book.org/wiki/FOST

Authors

Ms. Swaran Lata is presently heading Department of Information Technology's Na-
tional Strategic programme 'TDIL'. She is member of various international Standardi-
zation bodies ISO, UNICODE, W3C and ELRA. She is instrumental in getting all 22
officially recognized languages including Vedic Sanskrit incorporated in UNICODE.
She is also Country Manager, W3C India Office. She has a very vast and in depth ex-
perience towards proliferation of Indian Language solutions ,Key role in initiating mis-
sion mode consortium projects in the area of Development of English-Hindi Machine
Translation system and Indian language to Indian language Machine Translation System, Optical Character
Recognition System etc, Key role in establishing W3C India office and leading the activity of International-
ization requirements in W3C standards, Member of W3C Consortium's "Global Vision Task Force"

Dr. Somnath Chandra is currently working as Dy. Country Manger W3C India, C.G.O
complex, New Delhi, India. He has done his B.Tech, M.Tech. from IIT Kharagpur, and
Ph.D. from IIT Delhi. He has extensively worked on Optical Fiber and Web Technology.

Dr. Ela Kumar is working as Associate Professor, and Dean SOICT, Gautam Buddha
University, Greater Noida (U.P.). She has done her B.E and M.Tech from IIT Roorkee,
Ph. D. from Delhi University**.** She has worked lot in Artificial Intelligence, Soft Compu-
ting Technique and other Intelligence Systems with having number of research papers
and publications on her name.

Bhaskar Sinha has completed his M.tech. from GBU, Greater Noida, with Intelligent
System & Robotics and preparing for Ph.D. program. Did his M.Tech thesis in "Seman-
tic Web Technology" and willing to continue his Ph.D. in Semantic Web Technology.
Earlier he had worked in IT industry and have good experience in IT.

Raghu Arora is currently working as a Sr. Software Engineer. in W3C India, New Delhi
and has done his B.Tech from SKIT ,Jaipur. He is having Three years of Expeience in
Web Technology currently he is working on Semantic Web Technology.