International Journal of Humanities and Social Science Invention
ISSN (Online): 2319 – 7722, ISSN (Print): 2319 – 7714
www.ijhssi.org ||Volume 6 Issue 5||May. 2017 || PP.16-22

Information Organisation for the Future Web: with Emphasis to Local CIRs

Radharani Rit[1], Juran Krishna Sarkhel[2]
[1]*(Library, Bijoy Krishna Girls' College, Howrah, India)*
[2]*(Department of Library and Information Science, University of Kalyani, Kalyani, India, Retired Professor)*

***Abstract*:** *Semantic Web is evolving as meaningful extension of present web using ontology. Ontology can play an important role in structuring the content in the current web to lead this as new generation web. Domain information can be organized using ontology to help machine to interact with the data for the retrieval of exact information quickly. Present paper tries to organize community information resources covering the area of local information need and evaluate the system using SPARQL from the developed ontology.*

***Keywords*:** *Community Information Services, Community Information Resources, Semantic Web, Ontology, RDF, SPARQL, Apache Jena Fuseki.*

I. INTRODUCTION

Community Information is considered as problem-solving bustle which helps the people to take right decision at right time [1]. In the late 1980s Community Information Services (CISs) were a well established feature in Britain, the United States, Canada, Australia and New Zealand [2]. Though the concept of CIS bloomed in the west but dissemination of community information can be inspected in the ancient period of India through the inception of dandora, clay tablets, palm leafs and edicts of Ashoka. With the advent of time local formal and informal system of Panchayat also played a vital role to provide community information. Mela and Parvana, the age old tradition in India may be claimed as CIS though it was not same in concept with the modern western concept of CIS [3]. From gram sabha to CABX , from community information centre to web enabled Community Information Services, from public library based service to community welfare schemes, the need of dissemination of Community Information has been felt worldwide.

Though in present day people can seek CI independently by searching the websites of community organisations, but due to lack of enough knowledge about effective "Google" search, they lost in the huge web of information. Even though when one can perform particular web search, search engines will not always rank the most relevant websites highly enough due to the growth of information on the web. A few years ago such source of information on community interest was less, so managing that information was easier. Now with huge availability of community information resources there is an urge to find a technique to community information resources for better dissemination. As Davenport and Prusak's maxime explain, "people can't share knowledge if they don't speak a common language", but with the evolution of technology we need to think that knowledge is not for sharing within people or a particular community but between people and machines as well. Here comes the point to discuss how semantic web and its related technologies which can be used to organise community information resources which will allow knowledge sharing at a conceptually concise and elaborate level. It might be possible, ultimately, to provide data about community services that can be linked in a systematic way to other data as part of the Semantic Web [4]. In the future, this may greatly enhance the accessibility of CI.

It is known fact to all that users are completely dependent on Google, Bing, Yahoo! Search and many other search engines to locate resources available on the web. Though raditional search tools uses keywords based search technique rather searching web documents based on their content. As a result it retrieves information from the Surface Web rather digging in the Deep Web. "Deep Web pages don't exist until they are generated dynamically in response to a direct request. As a consequence, traditional search engines cannot retrieve their content and the only manageable way of adding semantics to them is attacking directly its source: the database."[5]. Henceforth with the keyword based searches these above mentioned search engines retrieve results from unauthenticated source, on which most of the time user cannot rely. These limitations may be overcome by new intelligent web architecture called Semantic Web. As Semantic Web allows domain specific organisation of information and their retrieval, present study tried to manifest it on the domain of organisation of community information resources.

II. SEMANTIC WEB

Undoubtedly semantic web is getting recognition in every aspect. It is the extension of present web where information is organized in well defined manner to help computer to understand the meaning of it and retrieve the required answer. To extract meaningful information from millions of web pages is really difficult, and to overcome this problem semantic web technologies are playing an important role. The semantic web technology in semantic web will organize the content in a structured manner, enable software agent to carry the search for the users query. In a simple term it can be said that semantic web technology will help to make web content understandable to computers allowing software agents to get the desired or meaningful information.

The W3C defines the main goals of SW as, "The Semantic Web is a Web of Data. The vision of the Semantic Web is to extend the principles of the Web from documents to data. Data should be accessed using the general Web architecture e.g. Uniform Resource Identifier (URI), data should be related to one another just as document (or portions of documents) are already. This also means creations of common framework that allows data to be shared and reused across application, enterprise, and community boundaries, to be processed automatically by tools as well as manually, including revealing possible new relationships among pieces of data. In order to achieve these goals, it is necessary to define and describe the relations among data on the Web. This is not unlike the usage of hyperlinks that connect the current page with another page: hyperlinks define a relationship between the current page and the target. On the Semantic Web, such relationships can be established between any two named resources or values and the relationship itself (i.e. the link) is also named.. By contrast, a link on the traditional Web is not named, which means that the significance or meaning of that link needs to be deduced by the human reader. The naming and defining those relations explicitly enable better and automatic interchange of data. Resource Description Framework (RDF), which is one of the fundamental building blocks of the Semantic Web, gives a formal definition for that interchange". Some of the advantages of achieving this goal include [6]

- The ability to locate information based on its meaning, e.g. knowing when two statements are equivalent, or knowing that a reference to a person in different web pages are referring to the same individual.
- Integrating information across different sources-by creating mappings across applications and terminological boundaries we can identify identical or related concepts.
- Improving the way in which information is presented to a user, e.g. aggregating information from different sources, removing duplicates, and summarizing the data.

Though Semantic Web is relatively new concept but it is very dynamic in nature and present day it has created its own importance. Over 15 years since the inception of the concept Semantic Web potentially being developed as the next generation of the World Wide Web – or, sometimes. Web 3.0. Figure 3 tries to depict the evolution web through the time period

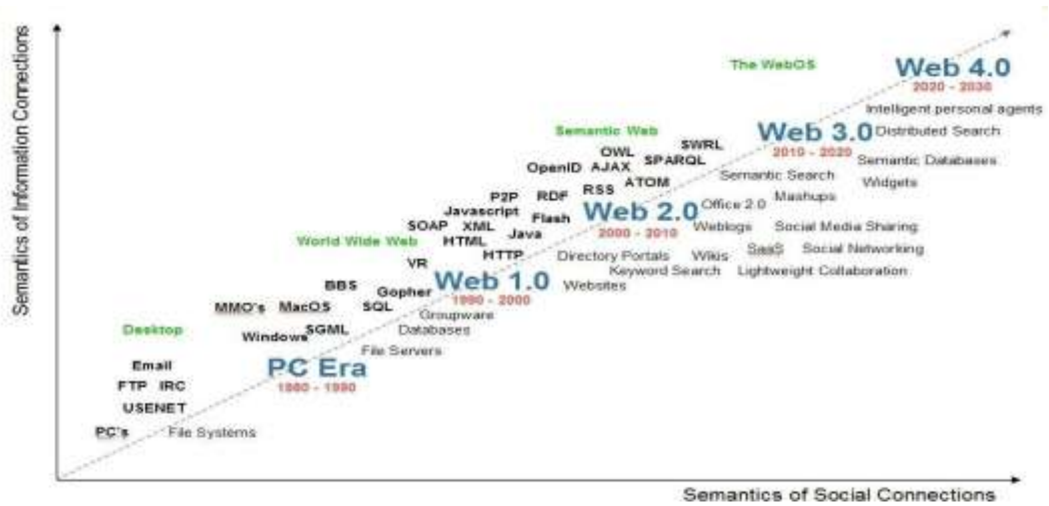

Figure 1 Evolution of Web

Semantic Web can tentatively be described as the Web of Openness. This refers to a Web that has the ability to break down old silos, link everyone everywhere together and make the entire Internet potentially smarter [7]. It is expected that the third generation web will aggregate and structure data in such a manner that an internet search will retrieve the specific information search for rather delivering hundreds or thousands or millions of responses.

III. WHY WE SHOULD USE SEMANTIC WEB

We use Web as a global database first of all for search. Keyword based search engine like "Google" is the main tool to retrieve data with issues associated with them like, high recall/low precision, low or no recall. To overcome these issues it is expected that semantic web will be available with following use:

- More specific information will be available.
- Search results will be more relevant.
- More personalized internet.
- It will act as automated tool.
- It will enhance web services.
- Quality issues.
- Trust issues.

Increased rate of data generation, access, flow, integration, and comprehension are the fundamental advantage of SW. Basically, this is the only way to achieve useful AI, based existing Web infrastructure.

IV. SEMANTIC WEB TECHNOLOGIES

To convert today's web into semantic web technological progress is required in the areas of integration, standardization, development of tools, and adoption by users. Few technologies which may help to achieve Semantic Web are: explicit metadata, ontologies, logic, agents etc.

V. ONTOLOGIES

In general, ontology is the study or concern about what kinds of things exist-what entities there are in the universe. The concept of ontology is having roots in philosophy where it signified the nature and the organisation of reality, ie. concerning the kinds of things that exist, and how to describe them. Then the concept of ontology has been adopted mainly in computer science as AI. Tom Gruber in explain ontology as "an explicit and formal specification of a conceptualization". Borst and Top have elaborate Gruber's definition as "Ontologies are defined as formal specification of a shared conceptualization." Studer, Benjamins and Fensel (1998) [8] have combined both Gruber and Borst's definition as "Ontologies are explicit formal specification of a shared conceptualization." Where RDF and RDFS helps to describe aspect of a domain in restrictive manner, ontology allows to describe the taxonomic structure of the domain, to model constraints or restrictions of the domain, and to state and reason over a set of inference rules associated with the domain. Ontology which is the heart of Semantic Web typically describes the hierarchy of concepts in a particular domain.

In a nutshell ontology can be expressed as

- Ontology is a term in philosophy and its meaning is ``theory of existence".
- Ontology is an explicit specification of conceptualization.
- Ontology is a body of common sense knowledge domain.

VI. COMMUNITY INFORMATION SERVICE (CIS)

Community Information Services in short CIS is a combination of Community, Community Information, and Community Information Resources. The main ingredient of CIS is Community Information Resources or component or categories of Community Information. Present study will try to explore various facets of CIS.

Community Information Service is a wrapped up term of 'Community', Information' and 'Information Service'. 'Community Information' has become one of the essential elements of the community development.

Library Association, UK, 1980 "Community Information Services can, therefore, be defined as services which assist individuals and groups with daily problem-solving and with participation in the democratic process. The service concentrates on the needs of those who do not have ready access to other sources of assistance and on the most important problems that people have to face, problems to do with their homes, their jobs and their rights ... this would mean a positive decision to concentrate on enabling people, particularly those in lower socioeconomic groups, to act either individually or collectively on their problems in the fields of housing, employment, family and personal matters, consumer affairs, household finance, education, welfare rights and civil rights".

According to Durrance [9] most acceptable definition of CI was given by Donohue, who identified two types of information to be provided by such a service "survival information such as that related to health, housing, income, legal protection, economic opportunity, political rights, etc., and citizen action information, needed for effective participation as individual or as member of a group in the social, political, legal, economic process." Durrance calls the former 'information and referral' and the latter 'public policy information'. He adds to these items a third component called 'local information' which according to him is information appropriate and useful to the community, including a calendar of local events, courses and other educational opportunities, and basic information concerning government agencies, local organizations, fraternal groups, and clubs [10].

VII. COMMUNITY INFORMATION RESOURCES (CIRS)

From the earlier section one can draw some idea about the periphery of CIS. The more structured CIS the more enriched society. Now we need to have little idea on about Community Information Resources. CIR can be defined as categories and subcategories of a particular CIS system of a particular community based on their information needs. After studying various papers and projects conducted to categorize information needs we found few basic thirst area such as Agriculture, Health, Education, Employment, Legal Information, Government related information etc. Present study tried to emphasize on very basic local information need i.e. educational information or precisely college admission information organization in ontological level. As we have already discussed that present web does not allow to retrieve structured information in a concise manner. So here we tried to organize a small area of local information needs of CIS to portray the power of Semantic Web in a complex manner.

VIII. STEPS FOLLOWED TO STRUCTURE THE INFORMATION AND ITS EVALUATION

Phase I

Step 1**:** Selection of domain i.e. CIS and its small part college admission information.
Step 2: Collection of Domain Information about 24 colleges in Hooghly District.
Step 3: Identification of various classess related to college admission information and categorized it. Here we organized the information in spreadsheet as flat data.
Step 4: Uploaded the spreadsheet into Open Refine.
Step 5: Mapped each head with standard like vcard, vivo, foaf, airs etc. Some cases created a new property which will be defined within the default namespace specified by the base URI.
Step 6: Export it in RDF format

Phase II

Step 7: Validated the RDF file using RDF validator.

Phase III

Step 8: Uploaded the RDF file into Apache Jena Fuseki Server 2.5.0 version.
Step 9: Executed SPARQL query out of it.

8.1. Phase I: Execution of RDF file

As discussed above for this phase we have selected a domain i.e. subcategories of Community Information Service (CIS) such as college admission information which lies under the category of Education. For this purpose we have selected twenty four colleges of district Hooghly affiliated to CU and BU. We need to study various sources to gather information which can be answered by the present model. To make the dataset and QA system more complex than present web we added few field in the dataset which is important at the same time. After collecting all the data and posting into spreadsheet we needed to convert it into RDF format. There are many tools to convert flat data into RDF data format, in present study we used OpenRefine (formerly known as GoogleRefine). It is very powerful tool to work with messy data. In the present project we have used openrefine-2.6-beta.1 and uploaded excel file containing college information. After uploading the file we mapped each column with rdf namespace prefixes like foaf, vivo, vcard, airs, dc, aiiso etc. Some cases created a new property which will be defined within the default namespace specified by the base URI.

Figure 2 Mapping of rdf namespace

8.2. Phase II Validate the RDF

After exporting the RDF file validation of saved rdf file was done using RDF validator from https://www.w3.org/RDF/Validator/. RDF validation is required for Developers who can understand the contents they are going to produce; they can ensure the correctness of expected structure, advertise the structure and generate interfaces. For RDF consumers it is required to understand the contents, verify the structure before processing it and query generation & optimization.

Figure 3 Data input window for RDF validation

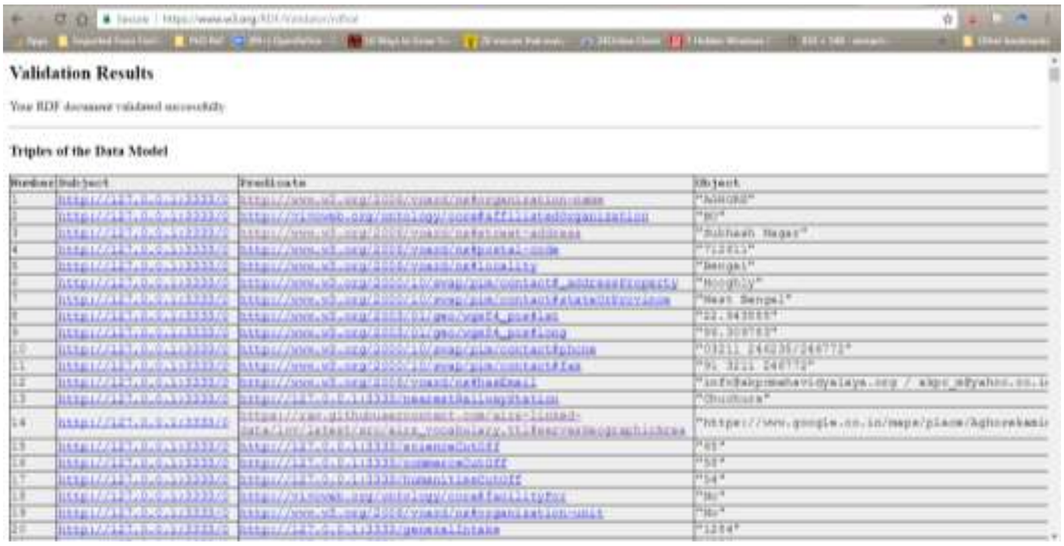

Figure 4 Validation report

8.3. Phase III SPARQL over dataset

In this phase we have executed SPARQL queries through Apache's Jena Fuseki Server to fetch the desired results from the CIR ontology. Jena is an open source Semantic Web Framework for Java. API (Application Program Interface) of Fuseki facilitates the writing and extraction of data to/from RDF graphs. Fuseki is an open source http based interface for RDF data which supports SPARQL for querying RDF. In this present study we used Apache Jena Fuseki-2.5.0.

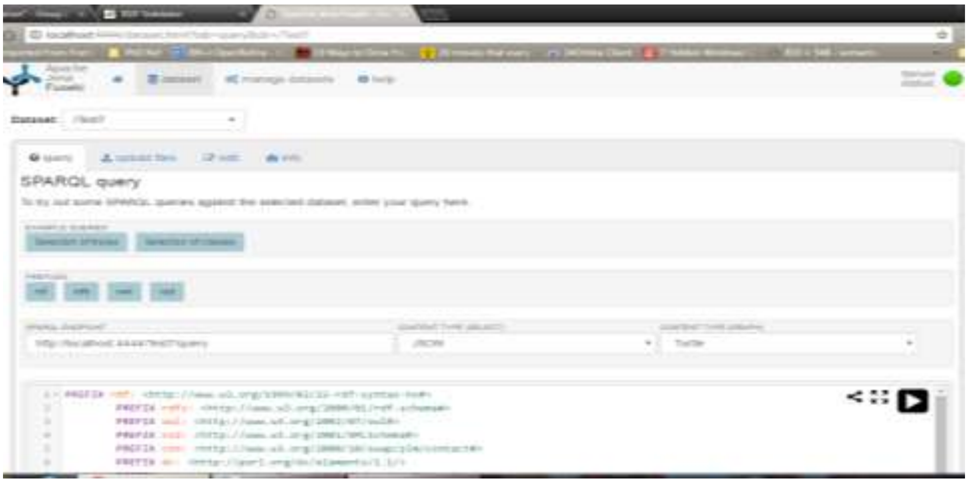

Figure 5 Interface of Apache Jena Fuseki Server

Query 1: In which college History (Hons) with Education (Pass) is offered?
SPARQL for Query 1:

```
SELECT ?OrgName
        WHERE {
          ?subject foaf:topic_interest ?a .
          ?subject vcard:organization-name ?OrgName .
          #FILTER (?a="History, Education") .
         FILTER regex(?a, "History, Education")
        }
        GROUP BY ?OrgName
        LIMIT 1000
```

Output of this query is shown in figure 5. The query extracts the college name where the combination of “History, Education” taught.

Figure 5 SPARQL query and result through Apache Jena Fuseki Server

Query 2: Which colleges are affiliated to “University of Calcutta (CU)”?
SPARQL for Query 2:

```
SELECT ?name ?affiliation
        WHERE {
          ?subject vivo:affiliatedOrganization ?affiliation .
 ?subject vcard:organization-name ?name .
          FILTER regex(?affiliation, "CU")
                               }
```

Output of the above query shown in below figure 6.

Figure 6 Output of how many colleges are affiliated to CU

IX. CONCLUSION

Present study tries to demonstrate the use of ontological concept in structuring domain knowledge to overcome keyword based search system. The proposed system tries to retrieve relevant information besides fetching the whole documents containing the particular keyword. As we know ontology is the backbone of Semantic Web and can be employed to organize any domain knowledge which machine can interpret. That way present prototype tries to convey that ontology can be used in large scale building of domain ontology in very complex manner to implement Semantic Web in place of present web.

REFERENCES

[1] J. J. Joselin, and P. Panneerselvam, Public Libraries as Community Information Centres: A futuristic approach. International Journal of Library and Information Studies. 62(2), 2016, 58-63.

[2] A. Bunch, Community information. Mousaion Series 3, 5(1), 1987, 2-15.

[3] B. Chakrabarti, Concepts of community information service in the West and India, in S. K. Sen (Ed.), Politics culture and society, (Kolkata: New Age, 2005) 279-291.

[4] B. Yoose and J. Perkins, The linked open data landscape in libraries and beyond, Journal of Library Metadata, 13(2-3), 2013, 197-211.

[5] N. Konstantinou, D.E. Spanos, M. Chalas, E. Solidakis, and N. Mitrou, VisAVis: An Approach to an Intermediate Layer between Ontologies and Relational Database Contents, Proc. International Workshop on Web Information Systems Modeling, Luxembourg, Europe, 2006, 1050-1061.

[6] J. Davies, D. Fensel, and F.Van Harmelen (eds), Towards the semantic web: ontology-driven knowledge management (England: John Wiley, 2003)

[7] K. Kelly, We are the Web, Wired magazine, 13(8), 2005, 113-23.

[8] R. Studer, V. R. Benjamins, and D. Fensel, Knowledge engineering: principles and methods, Data & knowledge engineering, 25(1-2),1998, 161-197.

[9] J.C. Durrance, Community information services—an innovation at the beginning of its second decade, Advances in Librarianship , 13, 1984, 99–128.

[10] V. V. Mohan, and A. A. N. Raju, Community information concepts, DESIDOC Journal of Library & Information Technology, 19(1), 1999, 5-14.