Annals of Library and Information Studies
Vol. 61, December 2014, pp. 286-293

Semantic solutions for the digital libraries based on semantic web technologies

Shriram Pandey[a] and K. C. Panda[b]

[a]Asst. Professor, Dept. of Library and Information Science, Mohanlal Sukhadia University, Udaipur, Rajasthan, India, Email:ram.mlib@gmail.com

[b]Prof and Head (Retd.), PG Dept. of Library and Information Science, Sambalpur University, Sambalpur, Odissa, India, Email: krushna52@yahoo.co.in

Received: 2 May 2014; revised: 24 November 2014

Using semantic web technologies in digital library offers a new level of flexibility, interoperability and way to enhance peer communication and sharing knowledge. It expands the usefulness of the digital libraries that will contain majority of data in future. Paper outlines the emerging concepts of social semantics digital library and explores the potential of using semantic web technologies for the digital libraries. Paper also provides an overview of semantic tools and services available and categorization of different types of semantic tools and services. A checklist for further evaluation of these semantic tools, services and projects is proposed.

Keywords: Semantic Web; Semantic Digital Library; Digital Library; Social Semantic Digital Library

Introduction

The semantic web adds meaning to information so that humans and computers can work together better. Berners-Lee et al.[1] describe the semantic web as: An extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

Currently information representation on the web has been designed for human understanding but not by machines, which leads to increased recall and lack of precision in results. Due to the lack of machine-readable structure and knowledge representation in web documents, programs are unable to comprehend web page contents precisely and hence semantic information from web documents cannot be extracted. However, there are significant numbers of semantic web applications either in terms of tools/portal/application being developed and used in diverse areas, including electronic commerce, digital libraries and e-learning, among others. All of these endeavors share the common belief that semantic web technology would significantly impact the use of the web, essentially in terms of increased task delegation to intelligent software agents[2].

The transition phase from web to semantic web may be seen as passing through different landscapes which can be divided under PC Era (1980-90), web 1.0 era (1990-2000), web 2.0 era (2000-2010), web 3.0 era (2010-2020) and web 4.0 era (2020-2030) respectively. The main difference between Web 2.0 and web 1.0 is that web 2.0 allows flexible web design, creative reuse, updates, collaborative content creation support collaboration and to help gather collective intelligence while main difference between web 2.0 and web 3.0 is that web 2.0 targets on content creativity of users and producers while web 3.0 targets on linked data sets[3].

The major difference underlying among these transitions phase is technological development and changing nature of web from 'Read only' web to 'Read-Write-Share-Execute' web busted by social media which has become a buzzword in the last couple of years. While analyzing this transition phase from web to semantic web and beyond i.e. Web Operating System (WebOS) one thing is clear that there are two major manifestations. First, communication media development i.e. from email to intelligent personal agent and second one is storage media i.e. from file system to semantic database. These developments trend towards semantic based web OS where users are facilitated by not only semantic based applications but a WebOS where everything will be in virtual world from applications

to operating systems. The implementation has already begun with introduction to cloud computing. The future of semantic web applications will be incorporated and supported with intelligent personal agents that will support powerful semantic based search and retrieval.

The web is turned into a dominant source of information where users search information by employing the available search engine, or by going directly to a site they tend to rely on. At conceptual level, there is significant difference between web and semantic web. The semantic web is an information space in which the information is expressed in a special machine-targeted language, whereas the web is an information space that contains information targeted at human consumption expressed in a wide range of natural languages. The semantic web is a web of formally and semantically interlinked data, whereas the web is a set of informally interlinked information. In nutshell, web was designed for much more than accessing the documents and semantic web has significantly developed and emerges as a platform where one can read, write, collaborate, and share. The vision of semantic web underlines in this belief and presented a framework where objects can be presented with well defined attributes with a network of relationship among named objects by adding semantics[4]. Semantic web is able to describe things (people, organization, documents etc.) in a way that computers can understand.

Semantic web has today become a reality. There are many applications developed in the semantic search domains which offer very innovative ways to search the web using semantic search engines. A semantic search engine ensures more closely suggested relevant results based on the ability to understand the definition and user-specific meaning of the word or term that is being searched for. Semantic search engines are able to better understand the context in which the words are being used, resulting in smart, relevant results with more user satisfaction. Apart from the semantic search engines there are varieties of enabling technologies developed that serve different purposes like for RDF storage : Sesame, Jena, YARS; for Reasoners: KAON, Racer ; Editors: Protege, SWOOP, MarcOnt Portal; Editors: Protege, SWOOP, MarcOnt Portal; Semantic wikis: Makna, SemperWiki etc. These solutions are deployed in practice to build digital libraries that provides variety of semantic digital library services powered by semantics.

Challenges in semantic web

There are many challenges associated with semantic web some of them are discussed here.

Focused search engines

Evolution in the domain of search engines that have adopted facet based search process significantly changed the approach of information seekers to access their resources; however they often lack domain specific search and the result is general rather than specific. Some of the search engine follows knowledge processing mechanism eg. Wolfram|Alpha and proven their capability to meaningful and highly relevant results, but such search tools are very restricted to their contents. Not many search engines are available in public domain which follows semantic based approach of searching and browsing data.

Organization of the web

The information representation on the web has been made much simpler by HTML. However, HTML cannot preserve the context of the term which plays major role at the time of retrieval on internet. This is the reason computer cannot understand the meaning associated with the concept. Web is not currently organized contextually; there is still a component mission i.e. ‘semantic’. Thus, it is a big challenge to have standard semantics available for general usage which can be applied directly to define ‘things’.

Lack of global standards and proven frameworks

There are only a few international standards/Schemas available for domain mapping and knowledge representation in the web. RDF (Resource Description Framework) provides standard platform for knowledge representation of various resources using ontology languages. Therefore, library specific data are ready to convert into RDF and enable library resources to get access and retrieval using semantic enabled technologies or tools because library resources have enriched metadata. As web resources lack metadata, it is difficult to convert them into RDF; however society-scale applications solve this purpose to some extent by applying semantic web agents

and/or services and by using consumers and processors for semantic data, and more advanced collaborative applications.

Lack of availability of formal domain specific ontology

Domain ontology (or domain-specific ontology) models a specific domain, which represents part of the world and developing the ontology for specific domain is a challenge by itself. There are several tools already developed which are available for use to developed domain specific ontologies including Dublin Core, SUMO, DOLCE and WordNet but formal domain specific ontologies are very few in numbers and only few applications have been developed using ontologies engine[5] e.g. SAPPHIRE (Situational Awareness and Preparedness for Public Health Incidences and Reasoning Engines) is a semantics-based health information system capable of tracking and evaluating situations and occurrences that may affect public health.

Undisclosed page rank algorithms

Another major problem associated with existing web is page ranking. Ranking algorithms are usually not publicized, giving rise to many speculations. Like, if an item A is listed before an item B, is it really because it is the better hit, or is it that whoever is behind A has a better relationship with the search engine, may be even to the extent of paying for preferred treatment? It the sometimes spreads rumors that sites using Google Analytics are on purpose better indexed by Google as bonus for using another Google product true or not?[6]

From libraries to social semantic digital libraries

Contemporary digital library is not merely considered as a digitized collection with information management tools rather the digital library creates an environment to bring together collections, services, and people in support of the full life cycle of creation, dissemination, use, and preservation of data, information, and knowledge. This notion opens a door to a new kind of digital libraries popularly known as 'Semantic Digital Library (SDL)' which integrate information based on different metadata, e.g.: resources, user profiles, bookmarks, taxonomies, provides interoperability as well as delivering more robust, user friendly and adaptable search and browsing interfaces empowered by semantics. By using social media tools and applications to the digital libraries, the ever changing relationship between a library and its users who are highly social media literate can be facilitated. This may be seen as real transition of library to the new social media era where users have options to create, annotate, share and collaborate. Table 1 presents the transition phase from libraries to social semantic digital libraries.

Social Semantic Digital Library (SSDL) is basically an outcome of the synergy between digital libraries, the semantic web, and social networking with aim to improve, among other things, usability of information discovery. Social networks are explicit representations of the relationships between individuals and groups in a community and provide the backbone for SSDL. Several of these social network based virtual communities have begun to publish members' public profile information, including social links, using the semantic web language resource description framework (RDF)[7]. Such RDFs can be reused and deployed to SSDL for better visualization of friends and community profiles as well as sharing and creation of knowledge within user communities.

SSDL brings out several key features to the end users/readers that are not available to the traditional

Table 1—Libraries to social semantic digital libraries

	Storage	Metadata	Interface
Library	Archive based on physical location	Bibliographic cards	Librarian
Digital Library	Database and archive	Digital bibliographic descriptions	Full-text search
Semantic Digital Library	Database and archive	Semantic bibliographic description	Search and browsing on ontologies
Social Semantic Digital Libraries	Database and archive	Bibliographic descriptions with annotations provided by users	Collaborative search and browsing

digital library where focus is on delivering content/information and not on knowledge sharing within a community of users. SSDL make users/readers involved in the content annotation process and allow users/readers to share their knowledge within a community as well as provide better communication between users in and across communities. Social semantic digital libraries follows the ideas of semantic web and extend the digital libraries by describing and exposing its resources in a machine 'understandable' way and enforce the transition from a static information to a dynamic (collaborative) knowledge space.

The idea of semantic web technologies for the digital libraries is easy to implement rather than its application to the web because web does not have metadata but libraries have metadata to catalogue its resources. We simply must make them available in a machine understandable format. Semantic web provides the format RDF and it is easy to convert library specific metadata in RDF by using several semantic web tools such as DMoz, WordNet , MarcOnt , SKOS, OWL ,SPARQL etc.

Semantic tools for the digital libraries

Semantic tools are designed to support semantic functionalities which can deploy or integrate with digital library software in order to support semantic features. It expands the usefulness of the digital libraries that will contain majority of data in future e.g. machine navigation resources etc. Semantic digital libraries cannot stand without technologies and combine many semantic tools solutions, such as semantic integration of information based on different metadata, interoperability with other system, user friendly and adaptable search and browsing interfaces empowered by semantics. These solutions are deployed in practice to build digital libaries that provides a variety of semantic digital libraryr services. These semantic solutions can be categorized based on their purpose and function it supports as under:

i. RDF conversion/Visualization tools: Used for documenting and explaining RDF mapping. It is also used for the GUI, if we can make some useful displays.

ii. Treasures/KOS tools: Used for representing and sharing knowledge organization systems over the web.

iii. Metadata schema/Standards: Helps in transforming a flat metadata schema to a semantic web ontology

iv. Supportive tools -Plug-in: Helps to add schema which adds a specific feature to an existing application.

v. Fully flashed project/Separate portal: These projects/applications incorporate features of semantic search and browse as well as follows semantic architectural models for storage of information.

vi. Open source-Java software/Program/ Searching tool: These tools serve different purposes of semantic applications from providing standard terminology to searching.

vii. Interoperability/Harvesting tool: Interoperability tool enables application to communicate with other applications at cross-linguistic interoperability and metadata interoperability levels. While metadata harvesting tools help to gather metadata from individual repositories.

viii. Knowledge extraction tools: Helps in creation of knowledge from structured (relational databases, XML) and unstructured (text, documents, images) sources.

ix. Ontology engineering tool: Any tool used for creating ontologies or similar semantic documents.

x. Semantic measures tools: Used for computation and analysis of semantic measures, e.g. semantic similarity, semantic relatedness, semantic distance, etc.

There are many semantic applications developed in recent past which can be classified under different semantic tools listed above. These tools are summarized in Table 2.

Evaluation criteria of semantic tools

Fuhr et al.[6] presented a comprehensive study of evaluation formwork for the digital libraries. They determined the attractiveness of the collections and ease to use of the technologies as the key factor in assessing the quality of digital libraries. The above study suggested that those current used relevance matrixes are not enough, since they do not take into account user satisfaction, the quality of information

Table 2—Categorization of semantic tools

Sl. no.	Categories	Tools
1	RDF Conversion/Visualization tools	a) FOAFcalm(Friend of a Friend) b) MARCont c) Drupal –Site vocabulary, CCK, LODR.info, Are RDF Store, SPARQL d) Knoodl-Mulgara(RDF Store) e) SIMILE-RDfizer, Gadget, Welkin, Longwell, Piggy Bank, Semantic Bank, Timeline, Timeplot f) RDF-DC Interoperability-RDF Schema g) RDF Topicmaps- Ontopia, webpage harvester, Topic map generator h) MARC Edit, MARCXML Converter i) 4store j) IsaViz k) Sesame l) Semantic JavaAnzo m) PoolParty Extractor n) VirtuosoRdfViews o) Krextor
2	Treasures/KOS tools	a) NCI Thesaurus b) SWED-Semantic Web Environmental Directory c) HIVE-Metadata Generator d) HIVE-SKOS e) eURl, wget by LOC f) PREMIS data dictionary for Preservation Metadata
3	Metadata Schema/Standards	a) Content model for RDA b) FRBR c) TextMD-JHOVE ASCII and UTF-8 Modules d) METS-Archivist Toolkit e) EAD- Archivist Toolkit f) MADS g) MODS h) MARCXML i) DOAP
4	Supportive Tools –Plug-in	a) Greenstone Plug-in for PDF, Post Script, Mp3 and Audio b) Greenstone Plug-in for XML, MARC, METS, SRW,OAI c) Zotero-Firefox Pluin
5	Fully Flashed Project/Separate portal	a) Semantic Medline b) PubMed c) JeromeDL d) LIBRIS e) BRICKS f) Europeana g) NSDL 2.0 h) Finction Finder-OCLC Wordcat-FRBR Based i) SPAR Project of National Library of France j) Talia by European Commission k) SKUA Project BY Joint Information System Committee l) SWAD-Project m) E-DVRA Project n) Living Knowledge o) FEDORA p) Google Books q) CACAO Project r) Semantic GrowBag s) STITCH@CATCH t) Whi (Semantic Web for History)

Contd—

Table 2—Categorization of semantic tools

—Contd

Sl. no.	Categories	Tools
6	Open Source-Java Software/Program/ Searching tool	a) Semantic Personnel Digital library b) Terminology services Project by OCLC c) Wordcat Identity-SRU searches, Name finder, OpenURl d) OCLC Connexion e) CoinS Generator f) CiteSeerx g) Amazon A9 h) Lucene i) Swoogle j) SDLIP-Core k) FUSION Semantic Registry l) GFacet m) GoNTogle n) JOWL o) Linked Media Framework p) OpenLink AJAX Toolkit q) OpenLink Data Explorer r) PoolParty Semantic Search s) RelFinder t) SPARQL-RW u) Twinkql
7	Interoperability/Harvesting	a) ECHODEP Hub and Spoke Interoperability b) JHOVE by JSTORE/Harvard object validation Environment
8	Knowledge extraction tools	a) PoolParty Extractor b) Triplify.
9	Ontology engineering tool	a) RDF2Go b) Fluent Editor c) Protégé d) EulerGUI e) Semantic MediaWiki
10	Semantic measures tools	a) Semantic Measures Library

and the relation between the documents. The study further suggested that evaluation schemata must be designed especially for the digital library field. Dion Hoe-Lian Goh et al[7]. in their study also presented a comprehensive checklist for DL evaluation and uses this checklist on four DL software packages namely, Greenstone, CDSware, Fedora and EPrints.The above study developed a checklist consisting of 12 categories of items. Bishop et al[8]. in their study also stressed the importance of social aspect in the design and evaluation of digital libraries. The authors argue that the usability of a digital library depends heavily on a user's ability to map their goal and system capabilities.

Although the scope of these studies is limited to evaluation of digital library software, it can be extended to evaluate semantic tools. While going through the literature, it was observed that much effort has been put on implementations of semantic tools into the practice but not so much on evaluation. This gap led to the present work. Most significantly, we have gone through large number of semantic tools available against a set of predetermined criteria that were deemed to be fit under some parameters listed in Table3. Therefore, our objectives were to map semantic tools which serve different purposes and develop a checklist for their evaluations. Here, we identified essential categories of features of semantic tools should possess. The evaluation criteria presented in Table 3 also followed guidelines presented by Fuhr et al[6], Dion Hoe-Lian Goh et al[7] and Bishop et al[8]. The evaluation criteria proposed to evaluate semantic tools are given in Table 3.

The evaluation criteria presented in Table 3 to evaluate semantic tools can be useful to answer the three key questions to evaluate a system i.e. why

Table 3—Criteria of evaluation –Semantic tools

Sl. no.	Parameter(s)	Criteria of Evaluation
1.	Community	1.1 Motivation of developers
		1.2 Roles of the member
2.	Release Activity	2.1 Testability
		2.2 Modifiability
		2.3 Portability
3.	Longevity	3.1 Longevity(*length of service*)
4.	License	4.1 GPL(*General Public License*)
		4.2 LGPL(*Lesser General Public License*)
		4.3 BSD(*Berkeley Software Distribution*)
		4.4 NPL(*Netscape Public License*)
		4.5 MPL(*Mozilla Public License*)
		4.6 Public Domain
		4.7 Price
5.	Training and Support	5.1 Documentation/manual
		5.2 Mailing List /discussion forum
		5.3 Offline/Automatic Bug Reporting
		5.4 Commercial Support
6.	Security	6.1 Encrypted Security Method
		6.2 Password based access
7.	Functionality	7.1 Functionality
8.	Integration	8.1 Integration with OSS software
		8.2 Integration with Commercial Software
9.	User Opinion solicitation	9.1 User Satisfaction
		9.2 User feedback
		9.3 Contact Information

evaluate, what to evaluate, and how to evaluate. The evaluation criteria proposed is very initial only and can be suitable for evaluation of some tools which work as independent software applications such as Open Source-Java Software/Program/Searching tool, Ontology engineering tool etc. discussed in Table 2. In order to evaluate the full fleshed projects, Metadata Schema/Standards etc, a different set of parameters can be developed, keeping the view of their special attentions to give semantic services. These proposed criteria set forth these limitations discussed above and presented a background to future direction of research to develop a checklist to evaluate for various other semantic tools categorized in Table 2.

Conclusion

The emergence of Social Semantic Digital Libraries (SSDL) offers a new level of flexibility, interoperability and way to enhance peer communication and sharing knowledge. It expands the usefulness of the digital libraries that will contain majority of data in future e.g. machine navigation resources. One thing is clear that Social semantic digital libraries cannot stand without technologies and combine many technologies solutions, such as semantic integration of information based of different metadata, interoperability with other system, user friendly and adaptable search and browsing interfaces empowered by semantics. These solutions are deployed in practice to build digital libraries that provides variety of semantic digital library services powered by semantics.

Plethoras of semantic solutions for the digital libraries are available on the web in the form of separate portal or software solutions or tools that are geographically distributed. Most of these have some kind of collections that user can browse and search. There are wide range of software solutions available

supporting the creation and maintenance of digital libraries and often comes with powerful features supporting semantic search, such as linguistic processing, entity and relationship extraction, and automatic classification. Apart from the software solutions, some semantic and social tools are available that can integrate with existing software solutions or semantic enterprise portal to enable powerful information discovery and social communications (by user profiling, semantic tagging, rating, comments). Some tools which are used to manage bibliographic description also plays vital role, such as MARC 21 , MARC-XML, BibTeX, FRBR and RDA has potential as a semantic bibliographic description in wide variety of digital library applications. Integration and employment of these technological semantic and social solutions in digital libraries can change the way of the resource discovery, improved precision and enable user to find information more easily.

References

1. Berners-Lee, T Hendler J, and Lassila O, The Semantic Web, *Scientific American*, 284(5) (2001) 34-43.
2. Sycara K, Giampapa, J A Langley, B K and Paolucci M , *The RETSINA MAS: a case study*, In Garcia, A., Lucena, C., Zambonelli, F., Omici, A. and Castro, J. (Eds), Software Engineering for Large-Scale Multi-Agent Systems: Research Issues and Practical Applications, LNCS 2603, July, 2005,Springer-Verlag, Berlin/Heidelberg, p. 232-250.
3. Aghaei S, Nematbakhsh M A, & Farsan H K, *Evolution of the World Wide Web: from Web 1.0 To Web 4.0.*International Journal of Web & Semantic Technology, 1-10. Available at http://airccse.org/journal/ijwest/papers/3112ijwest01.pdf (Accessed on 25 August, 2014).
4. Berners-Lee T, *Semantic Web Road map*, Available at http://www.w3.org/DesignIssues/Semantic.html (Accessed on 25 August, 2014).
5. Manuja Manoj and Garg Deepak. Semantic Web Mining of Un-structured Data: Challenges and Opportunities. *International Journal of Engineering*, 5(3) (2011) 268-276.
6. Fuhr N, Tsakonas G, Aalberg, T Agosti, M Hansen, P Kapidakis, S, et al., Evaluation of digital libraries, *International Journal on Digital Libraries*, 8(1) (2007) 21–38.
7. Dion Hoe-Lian Goh, Alton Chua, Davina Anqi Khoo, Emily Boon-Hui Khoo, Eric Bok-Tong Mak, Maple Wen-Min Ng, A checklist for evaluating open source digital library software, *Online Information Review*, 30 (4) (2006) 360 – 379.
8. Bishop, A P, House N A and Van Buttenfield B (Eds), *Digital Library Use: Social Practice in Design and Evaluation*, (MIT Press; Cambridge, MA.) 2003.