The Role of Ontology in Systems of Knowledge Organizing of Digital Libraries

Atefeh Zarei, Firoozeh Dokhani
Faculty of Azad University of Hamadan
Tehran, Iran
Atefehzarei@gmail.com, fdokhani@yahoo.com

ABSTRACT: *In the world in which information are produced and are importance more than before and one of the bases of decision making _ depends on having information and being available. One of the human's procedures, to make information available is information organizing and knowledge management .So, since there have been libraries up to now that they are digital and under web, different procedures have been used to organize information. in accordance to the increasing growth of data and existing sources in web and web users need to have common perception of them, ontology has the main role in exchange of information and development of literal web to meaningful web .Also , ontology is a conceptional model which real existences models , relations between them in a special extent , explicitly. Therefore , by virtue of increasing growth of resources in digital frames and under web from one hand and increasing need of users to benefit from this resource on the other hand , the use of ontology that organize resources in these environments is needed. So, in this article, the researcher tries to inspect the procedures and different means of information organizing at digital and printed world under web by confirming the role of ontology in knowledge organizing at digital libraries.*

Keywords: Data organizing, Catalog, Lists, Thesaurus, Digital library, Ontology, FRBR, Semantic web

Received: 30 August 2011, Revised 29 October 2011, Accepted 3 November 2011

1. Introduction

Data has an "*entropic*" entity. Data couldn't organize itself, but has tendency to be disordered. When data don't be organized in a special manner, it will have tendency to "*choate*" explicitly. The main goal of libraries organizing is to provide the data needs of libraries users. (Rubin,2004) libraries have used different tools during history that are : lists of titles or subject of existing libraries or providing a list of books , such as IBnNadim's list / glossary in 3th century at Islamic countries which knowledge adjusting in ten techniques or essay shows a kind of subjective information organizing. But, we should mention that data organizing started from 17th century in the world formally by printed resource organizing and now, also we face a world of data at web so, organize information.

Information organizing is one of the main results in libraries and informing centers to open data. Organizing can be considered as a system of libraries information retrieval that organizes libraries in a manner by contain, in order to have a useful availability. Main purpose of knowledge organizing systems, is, facilitating and acceleration of availability, retrieval and management of collections. These systems can be as a bridge between users' information needs and information resource so, able the users to retrieval and recognize a containable object without previous information from its contain, whether through review (e.g. the subject of a website) or/and direct research (e.g. using a motor to research it in a website) of course, makes answering to the questions possible, about collection extent and ... for organizations. Today, besides using old organizing systems such as expressions (letters of expression), subject heading and … should use ontology, in meaningful web network and some of the same systems. Therefore, according to the development of digital resource and under web in the internet world what is more

important, is the use of ontology (because of data retrieval in digital resource and some how internet discusses) to organize the resources. Present article tries to inspect knowledge organizing systems at the libraries and the use of newer generation of these models in organizing of resources from digital kind and under web.

2. The Kinds of Organizing Information Systems in Libraries

Knowledge organizing is all kinds of description and contains organizing, specifications, and goals of documents in such a way that are available for those who search these documents or hidden messages at them. knowledge organizing is included of all kinds and procedures of indexing, abstract and abstract writing, classification, developments, cataloging , document's management, bibliography and making bases of textual data and bibliography for data retrieval.(Encyclopedia of Library and Information science,2005).

Hodge, in 2000, defines knowledge organizing systems in this way that "knowledge organizing systems term, includes all kinds of designs, documents, techniques and information organizing tools and development of knowledge management: include 1)classification schema that organize information resource in a general level (e.g. books at bookshelves), 2) subject heading which provide more availability, 3) Document archives which control different forms of key information (such as geographical names and real and legal persons) and 4) new plats of organizing included meaningful web network and ontology existence.

So, knowledge organizing focuses on tools which promote information refunding. In the following ,those knowledge organizing tools in the libraries that are provided by Hodge, 2000 and Hill… et al 2002 and Tylor 2003 , with expressions and totaling present.

2.1 Lists of Terms

- *Authority files*: It's a documentary list of names which are used for controlling different names (persons, organizations and…) one entrance or identifying specific area of names such as documentary list of names of well-known and compilers and organizations.

- *Glossaries*: They are a list of expressions which usually are associated with definition. These expressions may be are pecific in a subjective area or in a work area. The expressions become meaningful in a special sense and rarely have varietyof meanings. For example, EPA (Environmental Protection Agency). (Hodge, 2000, p5.).

- *Dictionaries*: Dictionaries in clued the words of one language or expressions of a subject, profession or occupation that provides the regular basis of alphabet and meanings, pronunciation, stress and usage of words/vocabularies.

- *Gazetteers*: Provides some definitions of geographical expressions, they, also help to recognize small and big cities and population's tables. (Kumar, 1933).

2.2 Classification and Categories Schemes

- *Subject Headings*: Is a list that is prepared on the base of a subject, whether alphabetical or/and classified in such a way that the reader can reach the main document by that subject. (Chapman, 1990).Subject headings include: LCSH (SubjectiveHeadings of Congress Library), MESH (Subject Headings of Medicine)

- *Classification schemes*: Hierarchy order and symbolic show with numbers or alphabet in order to show great subjects, are: DDC(Dewey Decimal Classification), LCC(Library of Congress Classification), NLM(National Library Medical), UDC(Universal Decimal Classification), CC(Colon Classification).

- *Categorization Schemes*:

- *Taxonomies*: Are used to show grouping of objects on the base of a specific characteristic. Taxonomies are used in making scheme of objectivism systems and knowledge management.

- *Folksonomy:* Vander said that Folksonomy is a bottom _ up social classification. (Vander Wal, 2004). In practical terms, a Folksonomy is the complete set of tags _one or two keywords _ that users of a shared content management system apply to individual pieces of content in order to group or classify those pieces for retrieval.(Sturtz,2004).

The first step is that, data be placed in web, in a form that can be understandable for machine or/and changeable to this form, this act causes semantic web. A web of data which can be processed directly or in directly by machine device. (Berners_ Lee, 2002)

- *Ontologies*: ontology is consist of meanings, relationship between them and their specification. In other word, ontology defines the relations between meanings at web documents and real world, so relevant documents become process able and understandable by machines, therefore, facilitates data association between agents. They can show complex relationship between objects and embrace lost rules and principles. Ontologies which describe knowledge in a specific area will join with research_ data systems and knowledge management.In the context of computer and information sciences, ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application. (Gruber, 2009).

2.3 Relationship lists

- *Thesauri*: is based on a structure which controls the complexity of terms and provides , meaningful relations which are specified by classification / ontology .Thesauri provides authorized descriptive for abstracting and indexing for example thesauri AGROVACRT , ERIC , AAT (thesauri of architecture and art). The relation between words in most of the word is as follow: 1) Equivalent relationship 2) Hierarchicalrelationship 3) Association relationship.

- *Semantic Networks:* These networks are a graphic process for showing the knowledge. This meaning of first, was determined by Quilian in 1968.knowledge can be showed in form of objects, meanings, results and specific relationship between them. This organizing system, organizes meanings and expressions in the network model. (Quilian, 1968) .Semantic web is not separate from web in now, but also is the future of present web which is changed.

The first step is that, data be placed in web, in a form that can be understandable for machine or/and changeable to this form, this act causes semantic web. A web of data which can be processed directly or in directly by machine device. (Berners_ Lee, 2002)

- *Ontologies:* ontology is consist of meanings, relationship between them and their specification. In other word, ontology defines the relations between meanings at web documents and real world, so relevant documents become process able and understandable by machines, therefore, facilitates data association between agents. They can show complex relationship between objects and embrace lost rules and principles. Ontologies which describe knowledge in a specific area will join with research_ data systems and knowledge management.In the context of computer and information sciences, ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application.(Gruber, 2009).

3. Metadata schema

Metadata is defining as a data for data or data about data. In another words metadata is a data which describes the content, form or specification of data or a data resource. Metadata can be used to describe structured resources or unstructured data such as textual documents. Also, in order to describe electronic resources, digital data (such as digital pictures) and printed documents such as book, magazines and records are used. It can be kept in a data resource like web or in a data base separately. (Haynes, 2004).

Metadata is one of the most primary meanings of librarianship's world, since, librarians make a list of resources and existing documents in their collection they made a metadata for their collection. Of course metadata is a subject that is placed in the center of digital libraries, because it's the key of resources' discovery and use of each document. Lack of ideal standards and common metadata that are defined for the use in a specific area yet, are considered as a barrier in front of availability to data and use of digital libraries, and / or in a coordinated scheme of digital libraries are considered as a problem. Same of metadata schema are as follow:

- AACR2(Anglo _ American Cataloging_ Rules)
- Dublin Core
- GILS (Government Information Description)

- EAD (Encode Archives Description)
- HTML (Hyper _ text Markup Language)
- SGML (Standard Generalized Markup Language)
- XML (Extensible Markup Language)
- RDF (Resource Description Framework)
- MARC (Machine Readable Cataloging)
- MODS (Metadata Object Description Schema)
- METS(Metadata Encoding & Transmission Standard)
- MADS (Metadata Authority Description Schema)
- MIX (Metadata for Image in XML Schema)
- Text MD (Technical Metadata for Text). (Tylor,2003)

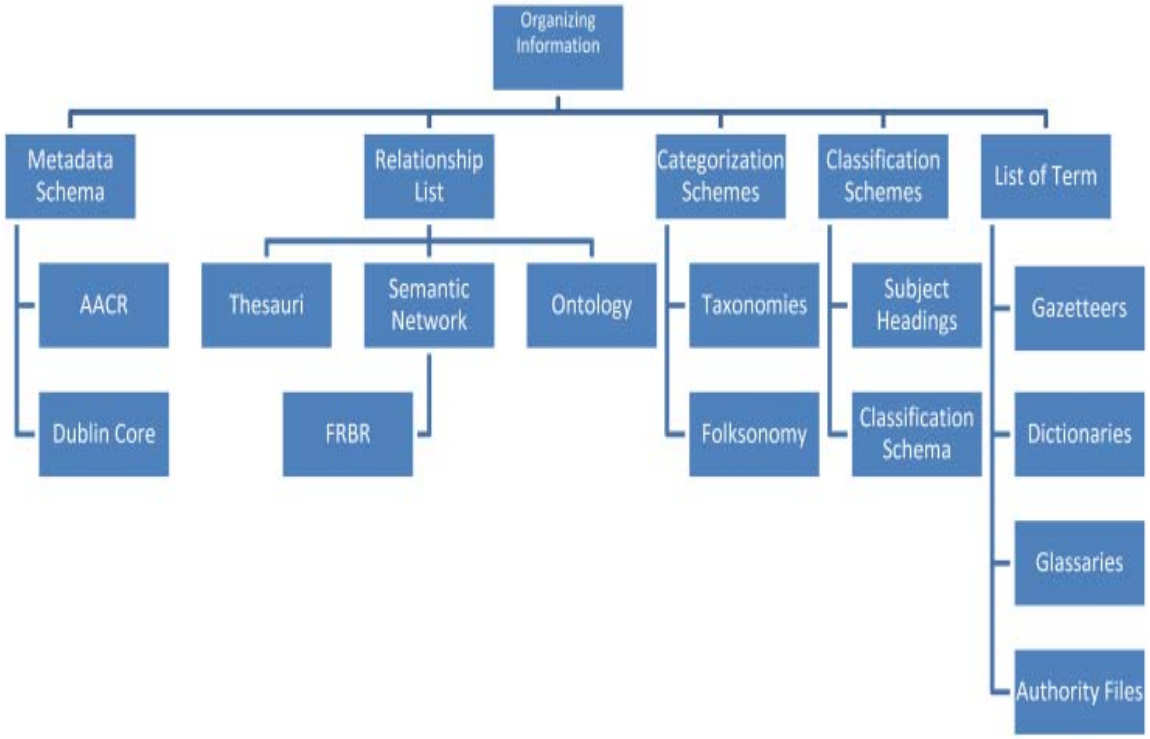


Figure 1. Knowledge organizing devices at libraries and digital and traditional information centers

What is discussed up to here about knowledge organizing at the libraries (traditional and digital), is shown in form of picture and in brief at figure 1. As, it is shown, the picture below shows knowledge organizing in digital resource, of course, with consideration to ontologies which will discuses more in following.

3.1 FRBR (Functional Requirements for Bibliographic Records)

As , shown in figure 1, easily can find the use and existence of semantic relationship between present contents. one of those models to show these relation, is FRBR. In fact said that one of the main elements in FRBR model definition, is the meaning of "*relation*". Relation is an important part of existential relationship model that these relations are in different forms such as: Hierarchical relationship, contented, general to part (association relationship) and detail to detail (serial relationship).

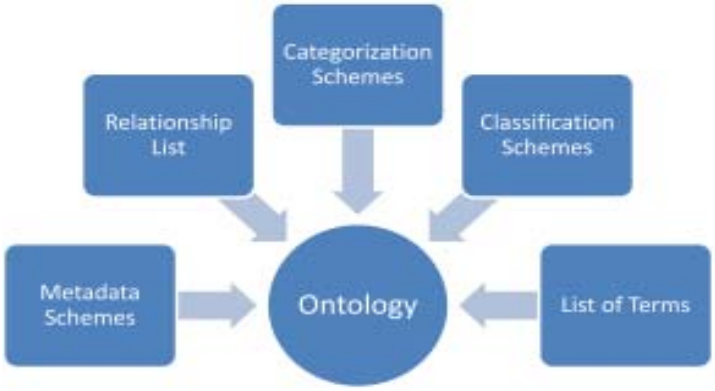
Yuan in 2010 divided all processes and knowledge organizing systems in two general groups: 1. Knowledge organizing presses are as: classification, indexing, subjective label and abstracting .2. Knowledge organizing systems are: classification systems, thesauri, ontologies and any meaningful form which is related to data refining. The second group mentions a kind of meaningful model and relations between them. (Yuan, 2010).

Considering to functional requirements for bibliographic records isn't only a production of hierarchical structure for all bibliographic existences, but also, it's explaining exact relations over these bibliographic existences which finally givesthis possibility to the list of computer, in order to change its level structure in to a multi _ layer structure and in fact has changed in to a kind of ontology so, gradually becomes closer to a web metadata structure. (Le Boeuf, 2005). With FRBR we look at entities, not at traditional catalog cards, bibliographic records, or authority records. (Brenndofer, 2011). At the end the changeof show procedure the result of research will be first step in this area.

3.2 Semantic web

Semantic web which is called expressive and expressionism imagined a global environment from machinery wise calculations nature in which all books, libraries, diplomas and universities (knowledge bases) in a form of expressive and with ability of meaningful understanding will be placed besides each other. Generally, semantic web is a network of data in global measure in such a way that their processing by machines will be possible easily. In other word, semantic web is consist of web intelligent data which can be processed by machine.

What is the relation between semantic web and libraries? Libraries have a unique situation, to prosper semantic web benefits; they have a great store of public information which has been coded on the libraries' lists semantically. Librarians' society, in



order to change in to a part of semantic web that is emerging, should act in two manners: first, bibliographic data of libraries should restore their structure from present reading machine structure in to a compatible from with semantic web. Secondly, this data shouldn't be enclosed in data _ base before, but should be placed in web in order to have compatibility with other web resource. This requires great changes in form and libraries' data storage. In a common scheme between co _ committee control on RDA and Dublin core metadata base, effort is done to produce the first semantic web comment about standard elements of libraries data.(Coyle, 2008).

Today, most of things that we do are under web such as: retrieval, comment, experience and analysis, records' search and which are related to the meaning of semantic network and development of semantic web network is related to ontologies, greatly. (Singh, 2004, 574). In order to justify this subject, should be said that, the goal of expressive web, is providing internet resource which is directly understandable and without any cause by machine, so, it can't be possible except by considering ontologies. Later, we will discuss about it.

3.3 Ontology

May be there are definition of ontology, but one of the most evident, was provided by Gruber: "*An ontology is a formal, explicit specification of a shared conceptualization*". (Gruber, 1993). Fensel has discussed, Gruber definition, in four main concepts.

1. An abstract model of a terminological phenomena with title of "*conceptualization*".
2. Mentioning to an exact mathematical description of word " *formal*"
3. Concepts' accuracy and relation between them, are explained Cleary by "*explicit* " expression.
4. And mentions the concord existence between user's ontology with expression "*shared* ". (Fensel, 2001).

Although, Gruber provided a general definition of ontologies, but Fencel(2000) knows ontology as a description of formal

important concepts, common and public in a specific area. Therefore, ontology can be defined in a specific concept for example: knowledge organizing in knowledge bases.

Ontology are used for a great area of research activities such as knowledge presentation, natural language processing, data retrieval, data bases, knowledge management , unifying the base of connected data, "*digital libraries*" geographical data system, multi _ agents system of connected data _refining.

(Nieto,2003). according to Starlab in 2003 viewpoint, in fact ontology is consist of expression's specifications which were used (terminology) and agreement for identifying the meaning of these expressions with their relations (Starlab,2003). Each ontology defines phrases and meanings which by them a part of knowledge can be described and to be shown. (Daconta, 2003). In 2004, Sigh said that ontologies are used in philosophy , knowledge management, library science, knowledge engineering and artificial intellect . Brank in 2005, knows ontology as a formal meaning which is complex from many interested areas that increasingly is used in different aspects such as knowledge management, data exploitation and semantic networks and generally, is used as a structure of data collection in a specific area by providing the relevant meanings and relations between them.(Brank,2005 and others). In web environment, ontology is a concise and regular device for defining web resource meanings (Jacob, 2003). Now according to a clear view of ontology which is provided we should say that, when the goal of present essay is inspecting and use of ontology at organizing digital libraries, so, later we will discuss about this part of ontology.

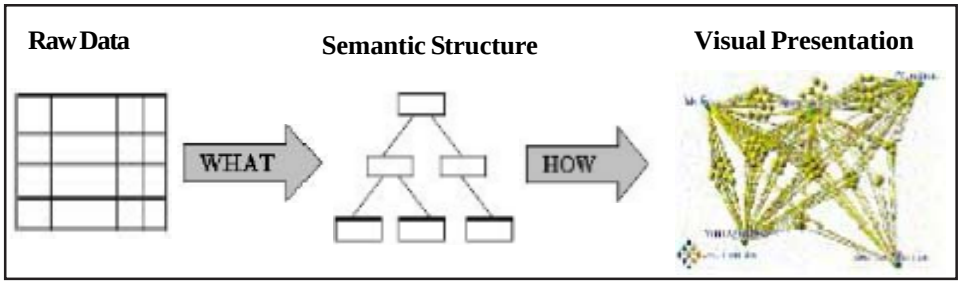


Figure 3. General Goal to visual show

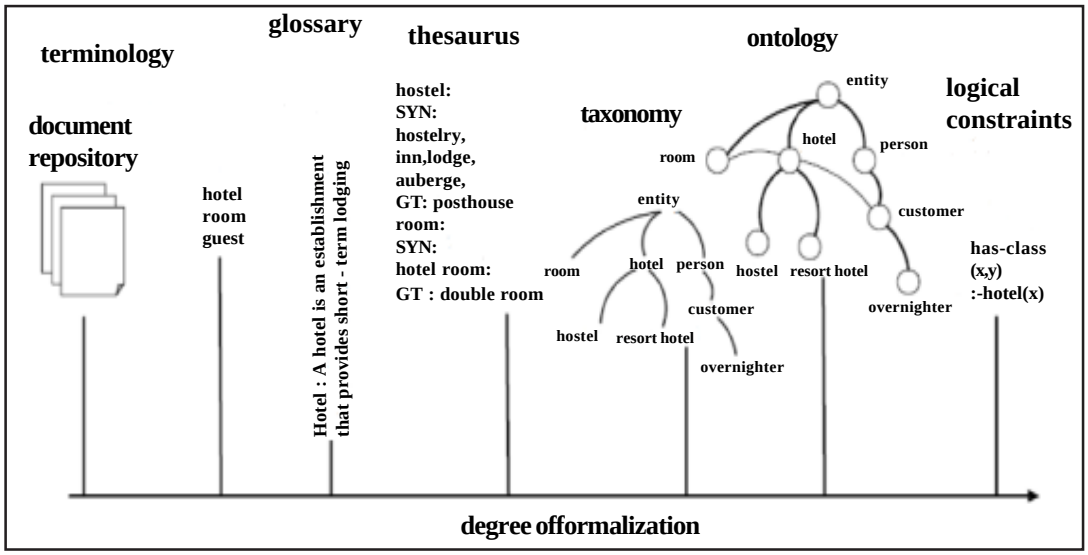


Figure 4. Different classifications of formalization

3.4 Ontologies and Knowledge Organizing

By virtue of extra growth of electronic resource, presentation of all kinds of metadata scheme, lack of professional subjective cataloger and a new goal to subjective availability to data in web and digital libraries and development of classification scheme of Taxonomies and personal subjective labels in availability to data (Folksonomy) and web network arising and establishment

of digital libraries, ontologies, can be considered as a intentional process. Ontology is a clear specification of meanings and their role at conversational area and providing structural and common viewpoints of data. (Singh, 2004).O ne of things which has been done in aspect of data organizing at bibliographic history in semantic network, is FRBR, conceptual model which makes access to the books' bibliographic data at RDF structure in semantic web network, possible ontologies presentation, in fact is the result of need to the semantic device which provides conceptual relations that are more clear and more defined than conceptual relation of Thesauri.

Rapid growth of data at web network has changed data organizing under web to a necessary case. During recent years, several efforts have been done for organizing internet resource such as, to profit from Dublin core metadata systems or using classification systems of libraries like Library Of Congress and Dewey and also goals based on users such as web 2 concept and producing social soft wares under web. Folksonomy (or public classification is one of organizing presses which during it different formats of data in web such as text, data , sound and picture in form of usual keywords which is called Folksonomy are labeling by users). These keywords, not only facilitate the process of web contain organizing, but also have provided the possibility of refining and finding knowledge resource, different data and web connections for all users.

In organizing area, ontology knowledge for different goals can be used including visual show in Figure 3 data changes in to a medium semantic structure, in this structure row data organizes in to a structure with meaning, in this step all operations are used for this structure that are: selection, transition, filtering, classification, combination and so on, in next step semantic structure is used as a bases for visual show. (VanHarmelen…etal.2000)

From unstructured text to logic roles and ontology

In Figure 4. Different levels of formal knowledge are shown which, is consist of a repository document to terminology, glossaries, thesaurus, taxonomy, ontologies and logical constraints. (Naigli and Velardi, 2008,P.72) .

In connection to subjective organizing, data processing are done very minutely and concisely on behalf of indexer, cataloger (in form of non _ automate) or/and by virtue of a concise and comprehensive processing by computer (automate) , in order to gain the selected labels with the most useful ingredient relevant to the document. This action at last makes the profound processing of data easier for audience, because as mentioned before, these labels include data and values which were experienced before to make decision _ making and outlook in audience simpler.

The usage of thesaurus / classification / knowledge base relevant to ontology at digital libraries are:

- Supporting of data composing and learning.
- Researchers and actives' cooperation for solving / making clear the problem.
- Supporting of data retrieval.

1. Providing some supports based on knowledge for final user in research.
2. Supporting of meaningful data show.
3. Providing a device for indexing.
4. Compound facilitating of several data base and massive access to several data base.
5. Supporting document processing after retrieval. (Soergel, 2002, P.25).

Organizing purposes and data classification, whether in traditional environments of libraries or in web, up to now this purpose was up to down, it means managers, indexers and experts of cataloging with present devices and also controlled results from present concepts in resource, in form of describers and terms organize data. In web environment, because of data resource variety and also subjective difference, not only compatibility of such result with potential needs of resource users isn't possible, but also it won't be in time and up to data.

Therefore, in recent years, goals of user _ base such as web 2 and producing social software under web have been more presented. Folksonomy is a process of down to up organizing in which users are labeling by goal of retrieving and their association , for regulating the resource of texts and electronic documents, pictures, video clips and sound files. In such

systems, users are encouraged to use knowledge organizing with its private process and compatible with needs and considered vocabularies. So, Folksonomy can be a potential device for data managing, especially in digital libraries.

4. Conclusion

Libraries and informing centers use some process such as descriptive and analytic cataloging, indexing and classification orderly for a long time, to organize knowledge. These traditional devices of knowledge organizing is a kind of ontology, because elements and concepts that describe, analytic and descriptive bibliography (descriptive cataloging rules, thesauri, classification and …) have made a network from relation between concepts for retrieving of data which is meaningful for librarians In data organizing and is a base for modern knowledge organizing languages. In modern knowledge organizing systems under web and digital libraries can mention this point that ontology has existed in library and information science area, but by considering that these relations in traditional process were done by human, but, by considering that resources become digital and organizing by machine should be done in a manner that has meaning for machine and establishes a network of relations, only it's possible by use of device of ontology. Of course, librarians in digital era have thought, elements and bibliography history in web environment are without output in aspect of relations between these elements. Therefore, they have provided concept of FRBR which is a kind of ontology, but with less area of output. Although knowledge organizing devices in semantic web environment, is ontology that provides samples of complete and perfect data with different subjects which can improve digital libraries organization so, makes availability to their contain facile and easy. In future, in semantic web networks which libraries are placed on web layers with inspiration of traditional devices and with use of ontology, will organize information.

References

[1] Berners _ Lee, T. (2002). Weaving the Web. San Francisco: Harper San Francisco.

[2] Brank. J .M. Grobelnic, Mladnic, D. (2005). A survey of ontology evaluation techniques in proceeding of the conference on data mining and data warehouses, SIKDD.

[3] Brenndofer, Thomas. (2011). Visualizing RDA (Guelph public library), CLA 2011. Halifax May 27.

[4] Chapman, Liz. (1990). How to catalog: a practical handbook using AACR2 2nd ed., New York: Ubs publishers.

[5] Coyle, Karen. (2008). Meaning, Technology and the semantic web, *The Journal of Academic Librarianship,* 34 (3) 263-264.

[6] Daconta, Michae, C., Kevin, T., Smith, Leo, J., Obrst. (2003). The semantic web: A guid to the future of XML, web services and knowledge management, John Wiley&sons, New York NY, USA.

[7] Elink, Jacob. (2003). Ontology and semantic web, *Bulletin of the American society for information science and technology.* 29 (4) (April/May 2003) 19_22.

[8] Encyclopedia of Library and Information Science, (2005).

[9] Functional Requirements for Bibliographic Records: final report, IFLA study group on the Functional Requirement for Bibliographic.Records.http://www.ifla.org/VII/s13/frbr.pdf

[10] Gruber, Thomas, R. (1993). A translation approach to portable ontology specification, Knowledge Acquisition, 5 (2) 199_220.

[11] Gruber, Thomas, R. (2009). Ontology, *In*: the Encyclopedia of Database Systems.Springer _ Verlay.

[12] Haynes, David. (2004). Metadata for information management and retrieval. London: Facet publishing.

[12] Hill, L. L., Buchel, O., Janee, G., Zeng, M. L. (2002). Integration of knowledge organization systems in to Digital Library Architectures: Position paper. *In:* J. E. Mai & C. Beghtol&J. Furner & B. KWASNIK (Eds), Advances in Classification Research. Proceeding of the 13th ASIST SIG/CR Workshop on " Reconceptualizing Classification Research. PP.62-68. Philadelphia, PA. HTTP://alexandria.sdc.ucsb.edu/1hill/paper_drafts/KOSpaper7_2final.doc

[14] Hodge, Gail. (2000). System of knowledge organization for digital libraries: Beyond traditional Authority Files, Massachusetts: The Digital Library Federation Council on Library and Information Resources.

[15] Kumar, Krishan. (1933). Reference services, Publisher house: Vikas publishing.

[16] Le Boeuf, Patrick. (2005). FRBR, Hype or cure all? Introduction . *Cataloging & Classification Quarterly*. 39 (3/4) 1-13.

[17] Navigli, Roberto., Velardi, Paola. (2008). Ontology Learning and Population: Bridging the Gap between Text and Knowledge: editors. Buitelaar, Cimiano, P, IOS press.

[18] Nieto, Mariaauxilio mediana. (2003). An over view of ontologies, Technical Reports, March, 1- 14.

[19] Quillian, M. R. (1968).Semantic memory, *In:* M.Misky (ed.) Semantic information Processing. Cambridge, Mass: MIT press.216-220.

[20] Rander Wal, T. (2004). You down with Folksonomy? Retrieved December 5,2004 from http://www.vanderwal.net/random/entrysel.php?blog_1529.

[21] Rubin, Richard. (2004). Foundation of library and information science.

[22] Singh, P. (2004).Web ontology to Facilitate Semantic Web2.2nd International Caliber, New Delhi, 11-13 February.574-585.

[23] Soergel, Dagobert. (2002) .Thesauri and ontology in digital libraries, *In*: Join conference on DIGITAL Libraries (JCDL, 2002), Portland, OR, USA, July14, 2002.

[24] Starlab. (2003). Systems Technology and Application Research Laboratory home page. Faculty of Sciences, Deparetment of Computer Science, VrijeUniversiteitBrussel. Available at http://www.starlab.vub.ac.be/default.htm.

[25] Sturtz, David, N. (2004). Communal Categorization: The Folksonomy.INFO 0622 content representation, December 16.

[26] Taylor, Chris. (2003). An international to metadata _ UQ library. Available at: http://www.library .uq.edu.au/iad/ctmeta4.html.

[27] Van Harmelen, F. (2001). Ontology_ based information visualization, London, UK, This paper appears in Information Visualisation, *In*:Proceedings. Fifth International Conference on : available at: http://www.ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber_942109.

[28] Ying Ding, Dieter Fensel, Stork, H. (2003), the semantic web: from concept to percept, Collected papers from the international conference of sustainable development of library, Beijing.

[29] Yuan, Ming_shan. (2010). The implementation of knowledge analysis and organization in the Professional Domain, *In*: International Symposium on Information organization Research,Taipei,Taiwan.