DESIDOC Journal of Library & Information Technology, Vol. 31, No. 2, March 2011, pp. 90-102

A Passage to Ontology Tool for Information Organisation in the Digital Age

Vivek Patkar

D-1, Anita C.H.S., Plot 612, Sector 6, Charkop, Kandivali (West), Mumbai-400 067
E-mail: vnpatkar2004@yahoo.co.in

ABSTRACT

To facilitate access to relevant documents and information has been the core of the library and information science (LIS) profession. In this regard tools like classification, cataloguing, and indexing formed the basis of library practice for a long time. These served particularly well for the material that was predominantly in the print form and required physical location for storage. New information sources, however, in contrast are increasingly in the electronic or digital form and stored on medium like computer hard disks requiring completely different strategy for access and management. Extension of the traditional bibliographic control tools as well as construction of new tools has therefore become pertinent. Ontology is one of the latest tools in this context. The paper discusses progress of information organising tools culminating in ontology, highlights the commonality of the concept of ontology and its applications among the fields of philosophy, computer science and LIS. It also discusses the select features of ontology development in practice and directions for features of ontology development in practice and directions for further work.

Keywords: Artificial intelligence, bibliographic control, information architecture, knowledge representation, ontology evaluation, semantic Web, social tagging

1. INTRODUCTION

The information world is now expanded and enriched by two producers, namely, traditional scholarly and professional publishers employing both print and electronic media, and any lay person who can access the computer and Internet. The latter is now assuming a significant portion thanks to the relentless advances in the information and communication technology (ICT). It is no wonder that similar variation is being witnessed in the information organisation practice.

Conventional tools of library science like cataloguing and classification and their newer versions are considered too cumbersome by the non-professionals in general. Instead tools like folksonomy, social tagging and wiki, which could be termed as structure-free and non-standard in their operation, are gaining acceptance particularly among non-specialist producers and users of digital information. This social classification approach is implemented through a collaborative process among amateurs rather than capturing relationships between entities through objectively developed classification structures by LIS professionals.

Basic premise of the new thinking is that knowledge filtering and organisation through the individual choices would be more useful, and a well-developed folksonomy could act as a shared vocabulary across the users. Some widely cited examples using folksonomic tagging services websites are: furl (www.furl.net), flickr (www.flickr.com) and del.icio.us (http://del.icio.us). It is to be noted that folksonomy creation and searching tools are not part of the underlying World Wide Web (WWW) protocols. The special provisions at the website level for creating and using tags generate folksonomies. The liberation thus offered from somewhat rigid classification structures and also from the portal engines is making it popular[1]. To some extent this reflects the self-service ethos of the new society, which is desirous of reducing its dependence on the intermediaries by exploiting say the ICT. Rapid growth in utilisation of e-banking, e-learning and online bills paying facilities lends support to this trend[2]. In a way, a challenge to the philosophy of LIS is posed by the ICT and Web driven digital information world. It is however apprehended that for serious applications and in-depth coverage of information and knowledge this new philosophy and practice would be of limited help. The non-standard approach for information organisation would be useful perhaps for narrow purpose such as sharing personal photographs and travel experiences. It is to the credit of LIS professionals that they have not given in panic. Instead, they have engaged themselves in devising novel schemes and tools keeping in view the changing information scenario and user needs with time[3]. Since there are some fundamental differences between the print and electronic media in information creation, storage, distribution and preservation, the information organisation methods will have to address the new requirements either by modifying the existing tools extensively or developing fresh ones.

To reflect the efforts made in this direction, a brief review of developments in the field of information organisation for the print and digital material is given in the following two sections of the paper, respectively. Ontology, one of the latest tools in this context is next presented with its conceptualisation and applications in the fields of philosophy, computer science, and LIS to highlight its commonality. It is followed by discussion on ontology development process in the information environment. Directions for further work are given to round off the paper.

2. TRADITIONAL INFORMATION ORGANISATION TOOLS

As the counting and communication (both oral and written) became the driving forces for individual and societal progress, the tools of mathematics and information organisation, in one form or the other, got firmly established. Both became sophisticated over the aeons and have no doubt paid rich dividends. With the advent of Gutenburg's printing press the proliferation of information and knowledge in print form like book, journal, magazine, and newspaper gained a new momentum. By the 18th century, it was well recognised that systematic methods for information organisation are necessary to make the best use of growing information sources. Depending on the human memory for comprehensive retrieval of information and documents alone will no longer work became clear. That set in motion construction of different information organisation tools in the library science. A brief review of the major tools thus developed for bibliographic control is presented below. These largely cover managing the print information sources with a few extensions for handling digital material.

2.1 Cataloguing

Though some form of cataloguing of records is found existing in the old library of Alexandria in ancient Egypt, the first national level cataloguing code was the French Code of 1791[4]. Credit, however, goes to C.A. Cutter who first formulated the objectives of a library catalogue in 1904 and laid the foundation of modern bibliographic system[5]. Library catalogue is basically a tool to describe a document and assist the user to know about the existence or otherwise of a document matching his selection choice like author, title or subject.

To create a catalogue record for every document in the library therefore becomes essential. Of course, there could be more than one record for a document depending on its nature like multiple authors. Ensuring as many searchable fields as possible in such record is helpful as that would provide more access points for searching. To maintain consistency of the records several catalogue codes have been developed. Most popular among these is Anglo-American Cataloguing Rules (AACR) with its evolved versions. For a book, for example it covers accession number, title, author, edition, publication details, physical description, series, notes, identification number (classification and ISBN), and terms of availability.

Traditionally hand-written or printed catalogue cards were prepared and arranged alphabetically in the array of card-holding cabinets. However, with the advent of computer technology in mid 1970s the digital version of the catalogue records, which could be accessed from any computer terminal within the library or from the website of the library from anywhere, gained wider acceptance. This system termed as Open Access Public Catalogue (OPAC) has also undergone several improvements over the last few years. Current OPAC provides search facilities akin to that of the WWW.

Cataloguing of the Internet resources is a huge challenge. This is mainly on account of variety in the content format like text, picture, audio and video of the digital material. Similarly the authorship or ownership of the digital document is not always clear. Special bibliographic format like Machine-Readable Cataloguing (MARC) was devised by the Library of Congress, USA in 1965-66 to supplement the AACR for cataloguing the digital objects[6]. Its latest version called MARC 21 and its allied versions (e.g. UNIMARC) specify record structure, data holding format, and standards for exchanging the data. An alternative Common Communication Format (CCF) was developed by UNESCO in 1984 and revised in 1988 and 1992, but its use has declined.

2.2 Classification

Classification of a document based on its subject content has been the core of the library profession. A class number is assigned for this purpose reflecting the subject matter. It helps in physical placing the documents of similar content together and different ones apart. M. Dewey developed, a scheme known as DDC Scheme, for library classification in 1876 using decimal number notation. It has naturally expanded as the knowledge domain over the years and run into more than 20 revisions and is still used widely. There are several other library classification schemes as described below.

A classification scheme has generally three components, namely, main classes and sub-classes, notations and indexes. The subjects are divided in main and subsidiary classes and suitable notations are assigned to distinguish these. To facilitate access to a specific topic or class number an index is provided in a classification scheme. Openness of the structure to cover new subjects, easy to use and remember notation, and ensuring uniqueness to the class number are some of the basic requirements of a library classification.

A classification scheme can be enumerative where every subject and class is listed with predefined notation system like in DDC and Library of Congress (LC) scheme. On the other hand, a classification scheme can be faceted where a set of rules is to be employed to construct a class number. Colon classification (CC), developed originally by S.R. Ranganathan in 1930s, and its subsequent revisions are examples of this approach[7].

To fit any facet or component of a subject in one of the fundamental categories like personality, matter, energy, space, and time (PMEST) construed the basis of this classification scheme. In between these two is analytico-synthetic classification scheme. Here the subject of a document is divided in constituent elements and classification scheme is used for assigning notation for each element to be joined according to the prescribed rules to prepare a class number for whole document. Universal Decimal Classification (UDC) scheme is one example of this approach. For classification of electronic resources several schemes are in operation.

WebDewey, which is an extension of the latest DDC version, incorporates additional access points and also provides a way to move to related records through hyperlinks (http://connexion.oclc.org).

Others are the BUBL (www.bubl.ac.uk) and the ACM classification scheme (www.acm.org/class/1998/ccs98.html). Construction of Web directories is one tool for organising Web resources. A faceted classification scheme like CC is considered quite useful for this purpose because it facilitates browsing, navigating and retrieving Web-based information and has been adopted by a few search engines[8].

2.3 Indexing

Since classification scheme employs symbols and notations, which at times look artificial, it was considered necessary to supplement it with natural language tools like subject heading lists and thesauri. Such indexing approach with vocabulary control helped in linking the information items according to their meaning and avoiding synonyms.

Library of Congress Subject Headings (LCSH) list is very extensive and widely used list for representing the subject of the document[9]. The most specific term representing a subject is used as its heading. Along with that 'broader term (BT)', 'narrower term (NT)', 'related term (RT)', 'see also (SA)' and 'used for (UF)' to denote non-preferred headings, are also provided. Unique headings, consistency and stability are the other hallmarks of the LCSH list. Due to such comprehensive coverage it is being used for organising both print and digital information to a large extent. For small to medium-sized libraries however the Sears List of Subject Headings is found useful. Attempts were made to build indexing-based automated information retrieval systems which try to derive meaning of the text from the observed syntax employing vector space model and latent semantic analysis[10, 11].

A thesaurus contains a controlled set of terms related to a given subject and linked by hierarchical or associative relations. It helps in building standardised vocabulary for information storage and retrieval systems. Thesauri are used widely for indexing purpose ever since first international standard for building thesaurus was established in 1974. Application of subject headings and thesauri has been made for organising Web resources too. For example, INFOMINE is one tool employing LCSH list for indexing databases, electronic journals and other online material. Intute: Health & Life Sciences (www.intute.ac.uk/healthandlifesciences/) is another tool giving over 30,000 resource descriptions in these two areas. Similar service is available for social sciences through Intute: Social Sciences (www.intute.ac.uk/socialsciences/). Numerous subject gateways and digital libraries are employing these tools for indexing and retrieval purposes.

In sum, the bibliographic control for the printed material was fairly achieved by the above tools which are viewed as 'class marking and document shelving' systems. Overall working of these tools for printed document say a book is shown in Fig. 1. It is clear that by examining various components of a document, necessary bibliographic information is derived that would facilitate its appropriate physical placement and retrieval with minimum efforts and expertise. For example in case of a book, the front and back covers, inner title page, imprint details, table of contents, preface, introduction, author and subject index (if available) are generally found adequate to build the bibliographical database for either manually or computer search.

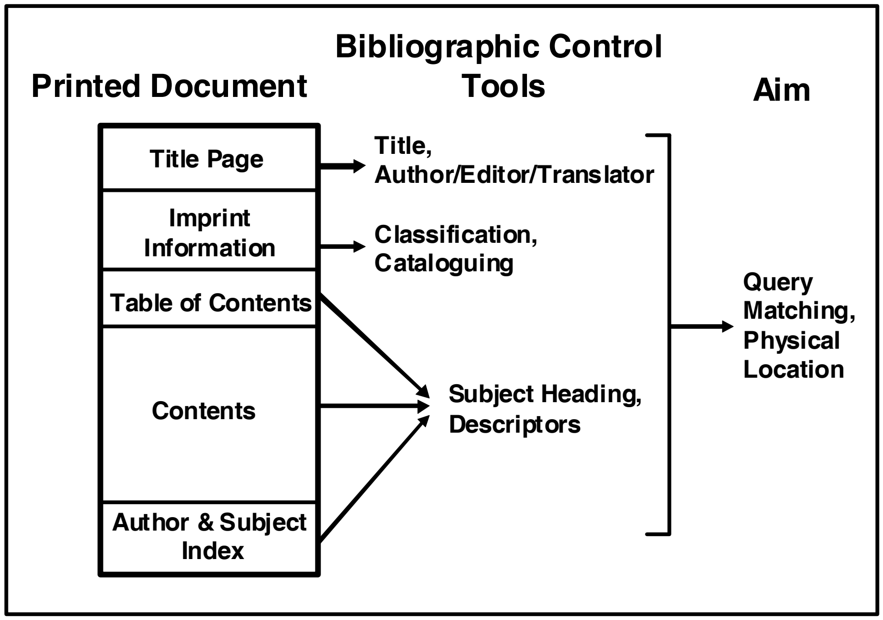

Figure 1. Traditional approach for information generation.

3. INFORMATION ORGANISATION TOOLS FOR DIGITAL MATERIAL

Information and Communication Technology tools like Internet and World Wide Web (WWW) have changed the information scenario to such an extent that even some new terminology such as e-journal, e-book and digital object has come into existence to identify the information material which could range from a single webpage to a pixel-based photograph to a digital piece of music. It would be worthwhile to compare the features of the print and digital material as shown in Table 1 to understand the need for differing organising tools.

The characteristics presented in Table 1 depict that the tools for bibliographic control of the print era would not be adequate to handle the digital material. For instance, though a bibliographic format like MARK 21 prescribes a number of fields and sub-fields for each record apart from consuming a large amount of time and resource, it cannot handle many issues typical of the Web and digital resources like the multiple dates of creation and revision, credit assignment and use rights management.

New tools have, therefore, been developed for this purpose. Let us take a brief review of these.

Table 1. Major differences between the print and digital material

Feature	Print Material	Digital Material/Object
Type	Paper based and only varying in size	Form varies from a simple text file to web page to satellite imagery to digital audio-video film with different formats
Size	Document like book or newspaper is restricted in number of pages	Web pages can increase limitlessly and are embedded with variety of formats and their continuity cannot be assured
Change frequency	Contents in a print document and its physical form do not change till the next edition	Web pages are volatile both in their content and location; difficult to track them for updating bibliographic details
User type	Users of the document can be profiled to guide the depth of classification process	Web pages in the public domain can be accessed universally making it difficult to decide the level of bibliographic control
Information resource control	Since the document is in fixed location like library the bibliographic processing can be done centrally	As anyone can publish any piece of material on the web, its organisation by one agency becomes impossible and thus making standardisation difficult
Resource required	Human-dependent library system could by and large handle the bibliographic processing	No single agency is there to handle the web resources and recourse to use of some ICT tool is necessary for organising such voluminous distributed information

3.1 Metadata

Although the term metadata is generic in nature covering description of any data, it is now used mainly to describe structured data about electronic resources or e-documents like digital text, image and audio-visual film. It may be noted that print material is not removed from its purview. Its definition can vary according to one's objective and viewpoint. Metadata essentially help in describing the resource, discovery, rights management, and long-term preservation[12].

Computer programs make use of metadata to facilitate location, use, sharing and reuse of a resource. Metadata are classified on the basis of their following principal uses for information resource handling: (i) administration, (ii) description,(iii) preservation, (iv) technical, and (v) usage[13]. Several standards or formats for metadata have been developed suiting to domain needs such as e-GMS (e-Government Metadata Standard), EAD (Encoded Archival Description), and TEI (Text Encoding Initiative). However, to serve all types of users a minimal set of data elements have been agreed upon by the international LIS professionals at the meeting first held in Dublin, Ohio and is known as Dublin Core.

This standard consists of 15 elements divided among three groups: (i) content consisting of title, subject, description, type, source, relation, and coverage, (ii) intellectual property consisting of creator, publisher, contributor and rights, and (iii) instantiation consisting of data, format, identifier and language[14].

Metadata can either be put along with the information resource itself or can be organised separately in a database. Normally user is not concerned with the metadata, but professionals and computer programers use the metadata extensively for resource identification and exchange. Since lot of flexibility is exercised in selecting metadata elements it is found necessary to standardise them to facilitate the interchange. The METS (Metadata Encoding and Transmission Standard) is one such standard designed to encode metadata for electronic sources[15].

3.2 Taxonomy

Taxonomic classification systems were classically developed for structuring a body of knowledge. Origin of taxonomy can be traced from the domain of life sciences to classify plants and animals in a form of a hierarchy based on their assumed natural relationships. Thus taxonomy, which classifies according to properties internal to data, differs from classification, which depends more on the external and sometimes arbitrary grounds[16]. Main objectives of a taxonomic system are: (i) to lend a structure to a mass of facts, and (ii) to build a unified and consistent view of the domain of interest. A Web directory is one example of modern taxonomy. Human genome project is another notable example of taxonomy application. Taxonomy applications are found across subjects ranging from botany to zoology to manufacturing to inventory control to supply chain system[17].

Elaborate methods drawing upon biological classification, systems theory, and object-oriented modelling have been developed to design taxonomy for specific purpose. Taxonomies have been used for information modelling, decision making, and performance measuring in different fields. Visualisation of system taxonomy by normally available software products provides immense advantage. For instance, in manufacturing organisations standardisation of terms and vocabulary and information unification have led to unambiguous communication among numerous production units (internal and external suppliers), streamlining transactions and efficient sharing of resources[18].

3.3 Ontology

As the Web technology is entering in its next stage called Semantic Web, automatic processing of information to represent semantic relationship between entities or objects in the given context becomes critical. To meet this emerging challenge a comprehensive information processing tool called ontology was developed which allows computer to process information resources based on their contents. It incorporates most of the above tools.

It is clear that there is some overlap between the operations of thesaurus, taxonomy and ontology. A thesaurus is essential for indexing and searching of information. Taxonomy in general, provides a classification of topics in the framework of laws and principle of a given domain and employs controlled vocabulary to construct hierarchical structure. Ontology is seen as describing subject knowledge matter using concepts, relations, functions and assumptions in addition to a taxonomy. A typical ontology for the Web will have a taxonomy and set of inference rules. No wonder taxonomy is treated as 'simple ontology' and subject heading lists and thesauri as 'lightweight ontologies' in some quarters. Classification provides say container to put the information items, i.e., where the data are, but ontology provides further information about their relationships, i.e., what the data are[19]. Explicitly defining the concepts used and constraints on their use and application of formal logic are the distinguishing features of ontology.

For the bibliographic control of digital material the above tools namely, metadata, taxonomy, and ontology have been developed specifically. Broad scheme of these tools for a digital object like electronic database and webpage is shown in Fig. 2.

To a large extent this process can be automated by extracting designated control records from the various sections of such an object. Aim of this process is to locate the relevant object matching the user query and describe its creation date, ownership, and user rights. For example, in case of a webpage the universal resource locator (URL), universal resource identifier (URI), universal resource name (URN), date of creation or update, owner, and free or paid access status are displayed.

It may be noted that with a passage from the manual to digital tools for information organisation the scope, speed and level of standardisation have increased. Figure 3 for example, shows the expanse in information parameters, complexity reflected by time and energy for development of bibliographical control tools symbolically (not to scale). Ontology is the most resource intensive due to its comprehensiveness. Dotted lines in Fig. 3 reflect the overlap between classification, thesaurus, taxonomy, and ontology.

4. DIVERSITY AND UNITY OF ONTOLOGY CONCEPT

Term ontology has a long history with its origin in philosophy. Of late the concept of ontology was adopted by the workers in the area of AI and knowledge engineering in computer science with specific applications in view. In the field of LIS it has been implemented to extend the bibliographic control of information and knowledge employing ICT tools.

A common thread, however, is found running in the conceptualisation and application of ontology in all these three streams. In one sense, development of the concept of ontology has some parallels with that of the concept of entropy, which basically implies increase in disorder in any closed system with operations over time. For example, the concept of entropy originated in the field of thermodynamics (study of heat phenomenon in physics) in the 19th century and has been used across a number of disciplines like metaphysics, economics, and information theory providing new insights[20].

Let us review briefly the concept of ontology as perceived and developed in philosophy, computer science, and LIS to understand its broader perspectives and concordant points.

4.1 Philosophy Standpoint

The question: When does entity or object exist or cease to exist? is one of the fundamental questions of philosophy. In other words, “Is existence defined through its relationship with some other entity is sufficient?” In essence seeking systematic explanation of this philosophical issue regarding ‘nature of being’ constitutes the topic of ontology. It has received attention ever since the time of Aristotle who attempted to classify the things in the world. Existential ontology is thus the long studied subject and is treated as a sub-discipline of philosophy and metaphysics concerned with the nature and relations of being. Ontology, as traditionally

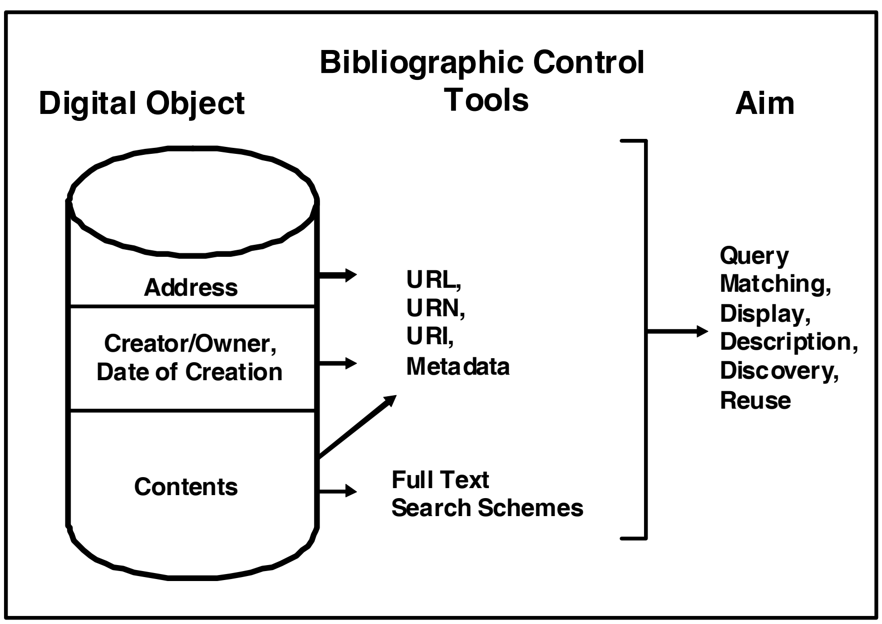

Figure 2. Approach for digial information organisation.

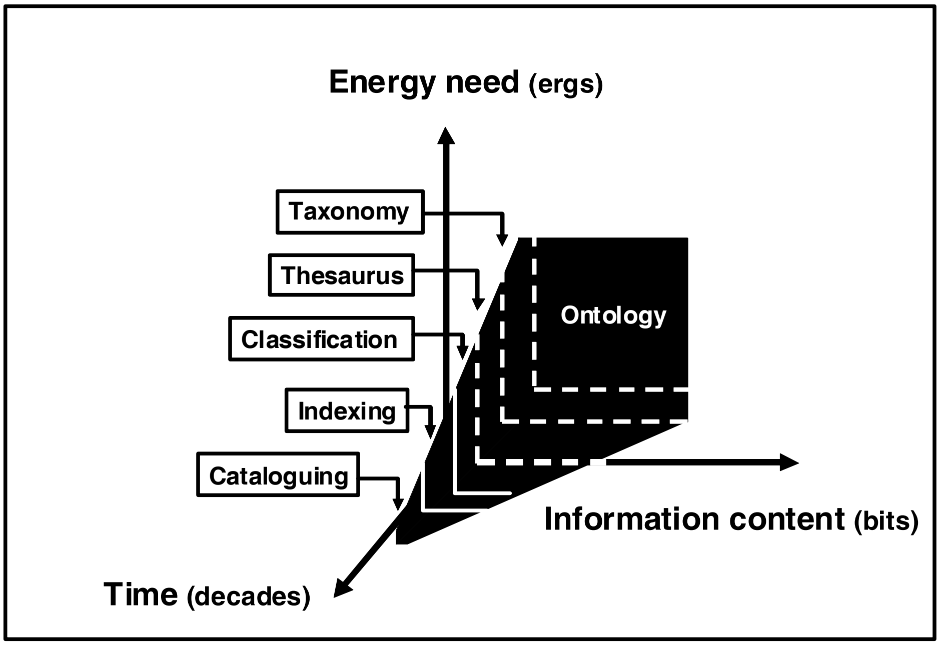

Figure 3. Development of information organisation tool.

conceived, is not a description of how we conceptualise but rather a description of the world itself.

Does the world make sense or do we make sense of the world? This is the related philosophical question. There is a shift here in semantics to the users from the system. So from identification of entity by a binary choice between saying two tags are the same or different we move to the multiple options of 'kind of is/somewhat is/ sort of is/overlaps to certain degree'. That provides a lot of freedom and leads ontology in practice to be treated as an approximate linguistic representation of agreed conceptualisation about a subject matter. It seeks description of a reality in terms of exhaustive classification of entities to help understanding the things happening in the universe and even making prediction about the future. Power of ontology to generate new categories through equivalence relationships is the key to extend it in new domain and its descriptiveness helps further in binding them.

In the late 19th and early 20th century the concept of ontology was extended notably by the philosophy of mathematics in the form of 'formal' ontology. It is used there to construct artificial language whose syntax reflects in some systematic way the structure of universe. Different syntactic types corresponding to different formal-ontological categories and relations between the symbols of each type corresponding to relations between entities of the corresponding categories were thus established. As a result development of quite a few ontological languages to formalise this process and derive further representations became possible leading to many applications[21, 22].

'Material' ontology is another development which is concerned with necessary and sufficient conditions for something to be a particular kind of entity within a given domain. In other words, what does actually exist becomes intelligible. For example, a description of necessary and sufficient criteria for something to be law or to be a legal object constitutes the ontology of law. This provides one approach to construct domain specific ontology in practice.

4.2 Computer Science Perspective

Use of the term ontology in computer science was first introduced in the context of data processing in 1967 where the representation of data entity or of 'what exists' was discussed[23]. As the database management technology advanced it was realised that different databases may be using different data labels to denote the same data entities or may tag identical labels with different meanings. Ontology or construction of reference taxonomy to include definitions along with supporting axiomatic framework to facilitate the management of database systems was therefore advocated[24]. Paying attention to the content along with the process is the key feature of ontology adoption in computer science.

Ontology naturally received more attention in other branch of computer science namely, 'artificial intelligence'. It may be recalled that with the development of first fully working modern computer in 1940s the dream of building a system as smart as human being was revived. Efforts to that effect were initiated under the AI and several specialised systems for game playing, theorem proving, vision or pattern recognition, robot building for a set of predefined tasks, and expert systems construction to store and process the information with rule-based algorithms for a particular subject to produce knowledge and address queries have been developed with indifferent success ever since. In course of theses developments it became imperative for a computer system to identify the presence or existence of an object. So the concept of ontology referring to the subject of existence needed practical treatment. That was done by modifying ontology as an explicit specification of a conceptualisation. It means for AI-based system what *exists* is that which *can be represented.* Considerable research is being carried for fine tuning this ontological aspect in order to enhance the capability of a robot, particularly for dealing with non-programmed situations. In real-life application, a robot is expected to identify the 'existence or presence' of new object in its environment and establish its association with other objects to draw suitable inferences about it such as useful, harmful and neutral and initiate action to deal with it accordingly.

Knowledge engineering is one more area of computer science where ontology is making significant contribution. Use of ontology to facilitate knowledge representation employing computer system is becoming common[25, 26]. Here ontology is seen as abstract model underlying a knowledgebase to aid information and knowledge processing and knowledge sharing. Ontology-based information extraction, ontology-based decision support provision, ontology-based human-computer interaction and natural language processing are some examples of its applications. Development of a special ontology framework called DOGMA (Developing Ontology-Guided Mediation for Agents) for this purpose represents one such endeavour[27].

The systemisation and elaboration of representations of entities and their associated reasoning techniques constitutes the 'formal' ontology in the field of computer science. It has been applied widely in diverse domains like bioinformatics, computational linguistics, GIS and information retrieval to state a few. Realising the importance of studying ontology as a general theory of the types of entities and their relationships to provide solid basis for the work in the respective domain, the international conference on Formal Ontology in Information Systems (FOIS) is held regularly to share the experience across various disciplines. The latest FOIS Conference, sixth in the series, was held at Toronto, Canada during 11-14 May 2010 to discuss the recent advances in the field.

4.3 Library and Information Science Adoption

Classical First Law of Cybernetics or Ashby's Law of Requisite Variety states: "Variety is required to address the variety"[28]. This practically means as the information sources have become complex and multi-dimensional the tools to organise these must be equally versatile. Development of ontology in the LIS field is precisely one such tool to manage the emerging information and knowledge in newer domains and that too in numerous forms and formats.

B.C. Vickery first drew attention to the concept of ontology for organising knowledge in the wake of its increasing complexity[29]. Ontology in the field of information management basically defines a common vocabulary for users who need to share information in a domain. The distinguishing feature is that it includes machine-interpretable definitions of basic concepts and relations among them. Ontology is thus taken as a formalised representation of the knowledge in a domain taken from a particular perspective or conceptualisation. Use of ontology has become common on the WWW due to sheer volume and variety of information items. The ontologies on the Web currently range from large taxonomies categorising websites to grouping of products according to their nature to facilitate search. Support of ontologies for both browsing and searching the digital information is of tremendous assistance. Several domain specific as well as general purpose ontologies have been developed to help the subject experts and others to share and annotate information in the corresponding field. Aim is to share and communicate knowledge, both between people and software agents[30].

The ontological scheme of organisation of information in different categories and enumeration of links between them is expected to play a crucial part in the next phase of web development namely Semantic Web[31]. Role of ontology there will be in building technologies, standards and tools to create information resources on the Web in such a way that computer software can read and process information from those documents easily for search and retrieval on a global scale. The new Web technology would analyse the user query not only for its syntax as done presently, but would further interpret its meaning (semantics) and that too in the given context. The retrieval of information from the Web by the software will thus be more precise and relevant for the user. Semantic web services will thus involve application of ontologies and knowledge markup (using say eXtensible Markup Language (XML) or Resource Description Framework (RDF)) and intelligent man-machine interfacing.

Here again the need of developing formal language tools for ontology construction has emerged. To that end the Semantic Web Advance Development (SWAD), the DARPA Agent Metadata Language (DAML) and Ontology Interchange Language (OIL) are some popular tools now available. The WWW consortium (W3C) semantic web activity has taken it further and with the base of DAML + OIL it has developed the Web Ontology Language (OWL).

Considering its wider use the OWL may emerge as the *de facto* language of the semantic web[32].

As the knowledge, its cross-fertilisation and applications are expanding in almost all disciplines, several types of ontologies have appeared on the scene. For instance, terminological ontology, informational ontology and knowledge modelling to deal with specific objective are becoming common[33]. On the level of generality these can be classified under domain ontologies, generic ontologies and representational ontologies, while for the problem solving, particularly in knowledge engineering, they are put under task ontologies and method ontologies.

4.4 Common Features

It emerges from the above discussion that despite wide differences among the subject matters of philosophy, computer science, and LIS ontology has commonly received attention. The meeting point is the similarity of the perspectives taken for the study, namely, cognition, usability, and economy of efforts.

On that basis the following three common features in respect of the concept and application of ontology can be elucidated:

(i) Existence of entity, whether physical, linguistic or abstract is studied and established with the help of its relations with other entities in a given domain.

(ii) Numerous branches have espoused the concept of ontology according to the respective operational objective.

(iii) Formal linguistic tools have been developed to implement the concept of ontology in practice.

It is envisaged that advances in development of ontologies in any of these three fields can help the others gainfully.

5. SELECT FEATURES OF ONTOLOGY DEVELOPMENT

Use of ontology in the LIS field is multipurpose as can be seen from its handling of the range of activities listed below.

- Representing and storing data
- Sharing common understanding of the structure of information
- Analysing domain knowledge
- Making explicit the domain assumptions
- Separating the domain knowledge from the operational knowledge
- Enabling reuse of domain knowledge
- Facilitating search and retrieval
- Enforcing uniform policies for classification to a large extent

Developing any ontology is similar to defining a set of data and their structure for other computer programs to emulate in practice.

Some of the currently used popular ontology-editing software tools are: Protégé-2000 (http://www.protege.stanford.edu), Ontolingua (http://www-ksl-svc.stanford.edu:5915/doc/frame-editor/index.html), and Chimaera (httpwww.ksl.stanford.edu/software/chimaera). Since formal ontologies are a form of software, methodologies for software engineering can be tailored for developing ontology. Basic steps for developing ontology are[34]:

- Perform requirement analysis involving the stakeholders to decide the scope
- define classes in the ontology
- Arrange the classes in a taxonomic (subclass or super-class) hierarchy
- Define slots and describe allowed values for these slots
- Fill-in the values for slots for instances
- Explore various ontological designs by re-factoring the classes
- Check the consistency and validate requirements

There is no one correct method for ontology modelling because it depends on the application in view and extensions expected in future. However, an iterative incremental change process is recommended for development of ontology, i.e. start with a rough classification scheme and apply it. Refine it with the consultation with the subject experts and by considering various implications of inclusion or non-inclusion of concepts in various slots and repeat the process till fairly stable and acceptable ontology is formed. Obviously this would be a continuous process as new information and knowledge in the subject concerned will be constantly streaming in. A typical ontological lifecycle involves the following phases: creation, substantiation, validation, application, maintenance, and revision. In real life, ontology need not be perfect and complete to commence its use. External domain experts should be consulted in its development without fail. Communication among the developers also plays a major role in its success.

A bottom-up approach for ontology medelling for web-driven service applications is gaining ground. According to this strategy ontologies close to information sources form the first and bottom layer. Next higher layer abstracts the information into more wide ranging concepts. The third layer provides high level semantics and allows defining main characteristics of domain and thereby represents some meaning element of the same. This process is facilitated by the application of OWL[35].

There are libraries of reusable ontologies on the Web and in the related literature which can be utilised for ontology development. For example, the Ontolingua ontology library (http://www.ksl.stanford.edu/software/ontolingua/), DAML ontology library (http://www.daml.org/ontologies/), (http://www.schemaweb.info) and a number of publicly available commercial ontologies such as UNSPSC (http://www.unspsc.org), RosettaNet (http://www.rosettanet.org), and DMOZ (http://www.dmoz.org) can be used profitably.

Apart from its development, analysis of ontology for its consistency has become important too. For instance, Chimaera provides diagnostic tools for analysing ontologies. It includes both a check for logical correctness of an ontology and diagnostics of common ontology-design errors[36].

6. DIRECTIONS FOR FURTHER DEVELOPMENT

Digital age would be driven by more advanced ICT in decades to come. One immediate prospect is that semantic aspect of the information would receive specific attention as against the syntax or string matching strategy for information organisation and retrieval employed currently by the search engines. Ontology will assist this process of multidimensional searching amicably due to its formal representation of information and knowledge by employing controlled terminology thereby facilitating machine-driven evaluation of documents for weighing the relevance[37]. Designing suitable procedures for managing ontologies, especially, for interdisciplinary and special subjects would need attention. Developing multilingual ontologies to discover knowledge is one more challenge ahead.

Ontologies are foreseen as instrumental in emergence of ‘semantic digital library’ in future. Different constituents of ontologies like bibliographic ontologies will facilitate managing different metadata standards employed by various information sources, ontologies for content structure will help efficient retrieval of digital content and community-aware ontologies would help tracking the information usage and behaviour of users[38]. Considerable work is needed to sharpen this process.

It is clear that user-driven social construction or classification of Web resources would continue despite its limitations. Regular empirical studies surveying views of the users about the usefulness, relevance and completeness of the information retrieved through the social tagging process is suggested in this regard[39]. That could lead in designing a new structural form for information organisation to supplement the ontologies.

Integration of ontologies in a distributed environment and multimedia driven databases will pose a big challenge. Development of bridging ontology could be one approach if completely meshed ontology could not be produced. Apart from the field of medical informatics international legal field is expected to benefit by development of such ontologies. This is essential because with the increasing pace of globalisation drafting of patents and contract agreements compatible with transnational legal systems will become vital. Suitable ontology can help building necessary structure for such documents and also searching.

Organisation of mammoth government information for efficient navigation and retrieval by developing appropriate ontologies is another major task. That would help government departments and agencies to locate and supply the time bound information effectively as is required to reply Parliamentary and Legislative Assembly or Council questions, submit documents in court cases and provide information to address the queries made under the “Right to Information” Act in our country. Applications in fields such as crime detection, judicial performance appraisal, medical diagnostics and financial market operations analysis will help refining the ontology building methodology besides delivering domain specific benefits. This is well illustrated by an ontology development in the context of e-learning employing the Bayesian Networks that led to effective tutoring and better adoption of the learning process suiting to demands of students[40]. Such ontology generation process and its wider use will also help developing more tools for analysing ontologies.

Due to proliferation of ontologies across Web servers and developed in different languages with diverse objectives it becomes a Herculean task to evaluate and select the appropriate ontology for a new project. Three major aspects of such evaluation are: (i) structured properties determined by concept structure graphs, (ii) usability covering metadata and annotation, and (iii) functionality measuring the purpose serving potential of ontology. A number of strategies and methods for this selection task are being evolved. For example, one methodology incorporating experience and expectation of the selector together with technical factors called ONTOMETRIC has been built up[41]. Further work to develop automated ontology selection will prove beneficial.

Institutions in general are either facing or will face in future the problem of organising huge amount of digital information created by the use of their Intranet and Internet for efficient and relevant retrieval. Systematic efforts for this purpose are being developed under a generic term of "Information Architecture" (IA)[42]. To guide the IA development process, the Information Architecture Institute has been set up. According to its definition IA covers labelling websites, Intranets, online communities, and software to support usability and discovery (http://www.iainstitute.org/). Extensive contribution by ontologies in strengthening IA of an organisation is anticipated because providing users a standard vocabulary for search actions is a key factor.

Training the practicing LIS professionals to undertake appropriate ontology construction for a given subject or field is urgently needed. Incorporation of necessary advanced LIS tools and computer software skills should form one part of the training programme. Practical exercises in ontology development in collaboration with subject specialist should form its other part. Ontology development either as a practical or project should become an integral part of the LIS curricula at the Master's level at the earliest.

Emergence of a knowledge society and that transforming further into a wisdom society is envisaged as a logical progression of the present information driven society. Information and knowledge needs of these new forms of the society would certainly be of very high order and cover many dimensions including ethical aspects[43]. Organising information and making it available in concise and sharable form through automated systems for wider use in that context would entail development of complex ontologies on regular basis.

7. CONCLUSIONS

Central philosophy of LIS is to organise and structure the bibliographic universe and track domain knowledge as well as knowledge in general. To retrieve information about relevant sources in minimum time and efforts matching the given query from the vast amount of print and digital information sources is its implication. To that end of managing information overload, development and use of a comprehensive information organisation tool like ontology becomes central.

It is surmised that the success of emerging semantic Web would largely depend on the advances in the ontology development to help automated systems in retrieving precise chunks of information from various sources and proactively integrating them. Novel strategies for this purpose have to be designed and for that matter accommodating the social tagging practice in the ontology construction should also receive attention.

Developing subject specific ontology is however necessary because the variety and terminology vary a great deal among the subjects now. How the ontology construction process could be simplified and both the subject specialist and LIS professional can make maximum use of the end product needs attention. Integration of a number of ontologies will become necessary, particularly in the interdisciplinary research and applications. Designing suitable methodologies to address this need emerges as one important area of research in the LIS field.

Like digital divide, ontology would not lead to 'fluency' divide among LIS professionals is to be ensured. Suitable training and availability of requisite computer resources would be important inputs in this regard.

Constructing or revising ontology would be an ongoing process because new concepts, methods, and knowledge are making constant appearance. It means that the LIS professionals should work continuously in collaboration with subject experts to revise, disseminate and test application of ontology to keep it relevant. Tremendous scope is therefore seen for research in this area. New ontological tools so developed to process complex information and guide the automated systems will certainly enhance the prestige of the LIS profession in the digital era.

REFERENCES

1. Patkar, V. Transforming the library to cope with the new web technology and social challenges. *In* Knowledge, library and information networking, edited by H.K. Kaul & S. Kaul. DELNET, New Delhi, 2007. pp. 23-35.

2. Patkar, V. E-Learning: Liberation of education and training with evolving library and technology support. *DESIDOC J. Lib. Inf. Technol.*, 2009, **29**(1), 14-22.

3. Chowdhury, G.G. & Chowdhury, S. Organising information from the shelf to the web. Facet Publishing, London, 2007.

4. Hunter, E.J. & Bakewell, K.G.B. Cataloguing, Ed. 3. Library Association Publishing, London, 1991.

5. Cutter, C.A. Rules for a dictionary catalog, Ed. 4. Government Printing Press Office (reprint Library Association), London, 1962 (original 1904).

6. MARC 21. MARC 21 specifications for record structure, character set, and exchange media. Library of Congress, Network Development and MARC Standard Office, Washington, 2004. http://www.loc.gov/marc/specifications/spechome.html

7. Ranganathan, S.R. & Gopinath, M.A. Colon classification. Ed. 7. ESS ESS Publications, New Delhi, 1987.

8. Broughton, V. The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings: New Information Perspectives*, 2006, **58**(1), 49-72.

9. Library of Congress. Library of Congress subject headings, Ed. 29. Washington, 2006 (www.loc.gov/cds/lcsh.htm).

10. Salton, G. Automatic text processing. Addison Wesley, Wokingham, 1989.

11. Deerwester, S.; Dumais, S.T., Furnas, G.W., Landauer, T.K. & Harshman, R. Indexing by latent semantic analysis. *J. Amer. Soc. Inf. Sci.* 1990, **41**(6), 391-407.

12. UKLON Metadata. http://www.ukoln.ac.uk/metadata/

13. Gilliand-Swedand, A. Defining metadata. *In* Metadata applications and management: International yearbook of library and information management 2003-2004, edited by G.E. Gorman & D.G. Dorner. Facet Publishing, London, 2004.

14. Dublin Core Metadata Initiative. Dublin core metadata element set, version 1.1. 2006 http://dublincore.org/documents/dses/.

15. METS. Metadata encoding and transmission standard. 2006 http://www.loc.gov/standards/mets/.

16. Rees, R.V. Clarity in the usage of the terms ontology, taxonomy and classification. *In* Construction informatics digital library, 2003. http://itc.scix.net/paper w78-2003-432.content.

17. Chandra, C. & Tumanyan, A. Supply chain system taxonomy: A framework and methodology. *Human Sys. Manage.*, 2005, **24**(4), 245-58.

18. McCarthy, I. & Ridgway, K. Cladistics: A taxonomy for manufacturing organisations. *Integrated Manufact. Syst.*, 2000, **11**(1), 16-29.

19. Gilchrist, A. Thesauri, taxonomies and ontologies: An etymological note. *Journal of Documentation*, 2003, **59**(1), 7-18.

20. Rifkin, J. Entropy: A new world. Bantam Books, Toronto, 1981.

21. Ingarden, R. The literary work of art: An investigation on the borderlines of ontology, logic and theory of literature. Northwestern University Press, Evanston, IL, 1973.

22. Smith, B. An essay in formal ontology. Grazer *Philosophische Studien*, 1978, **6**(1), 39-62.

23. Mealy, G.H. Another look at data. *AFIPS Conf. Proc.,* 1967, **31**, 525-34.

24. Guarino, N. (ed.). Formal ontology in information systems. Amsterdam, IOS Press, 1998.

25. Chandrasekaran, B. & Josephson, J.R. What are ontologies and why do we need them? *IEEE Intell. Sys.*, 1999, **14**(1), 20-26.

26. Gruber, T.R. Towards principles for the design of ontologies used for knowledge sharing. *Inter. J. Human-Computer Stu*, 1995, **43**(5/6), 907-28.

27. Zaho, G., Zheng, J. & Meersman, R. An architecture framework for ontology development. *In* Proceedings of IADIS International Conference of e-Society 2003, 745-49.

28. Ashby, W.R. Requisite variety and its implications for the control of complex systems. *Cybernetica*, 1958, **1**(2), 1-17.

29. Vickery, B.C. Ontologies. *J. Inf. Sci.,* 1993, **23**(4), 276-86.

30. Gruber, T.R. A translation approach to portable ontology specification. *Knowledge Acquisition*, 1993, **5**(2), 199-220.

31. Berners-Lee, T.; Hendler, J. & Lassila, O. The semantic web. *Scientific American*, 2001, **284**(5), 34-43.

32. W3C semantic activity. Semantic web activity statement., 2002. http://www.w3c.org/sw.

33. Gokhale, P.A. Ontology: A tool for organisation of knowledge. *Information Studies*, 2009, **15**(4), 233-41.

34. Noy, N.F. & McGuinness, D.L. Ontology development 101: A guide to creating your first ontology. Knowledge Systems Laboratory and Stanford Medical Informatics, Stanford, 2001. Technical Report, KSL-01-05 and SMI-2001-0800, .

35. Georgiev, I. Ontology modelling for semantic web-driven application. *In* Proceedings of the International

Conference on Computer Systems and Technologies (CompSysTech'2005). 2005. pp. II-8.1-II-8.6.

36. McGuinness, D.L., Fikes, R., Rice, J. & Wilder, S. An environment for merging and testing large ontologies. *In* Principles of knowledge representation and reasoning. Proceedings of the seventh International Conference (KR-2000) edited by A. G. Cohn, F. Giunchiglia & B. Selman. Morgan Kaufmann Publishers, San Francisco, CA, 2000.

37. Lausen, H.; Ding, Y., Stollberg, M.; Fensel, D., Hernandez, R.L. & Han, S.K. Semantic web portals: State-of-the-art survey. *J. Know. Manage.*, 2005, **9**(5), 40-49.

38. Kruk, R.S.; Haslhofer, B.; Piotrowski, Westerski, A. & Woroniecki, T. The role of ontologies in semantic digital libraries. Networked Knowledge Organisation Systems and Services. *In* 5th European Networked Knowledge Organization Systems (NKOS) Workshop, 21 September 2006, Alicante, Spain.

39. Gilchrist, A. Structure and function in retrieval. *J. Docum.*, 2006, **62**(1), 21-29.

40. Colace, F., Santo, M.D. & Gaeta, M. Ontology for e-learning: A case study. *Interactive Technol. Smart Edu.*, 2009, **6**(1), 6-22.

41. Lozano-Tello, A. & Gomez-Perez, A. Ontometric: A method to choose the appropriate ontology. *J. Database Manage.*, 2004, **15**(2), 1-18.

42. Gilchrist, A. & Mahon, B. (Eds.). Information architecture: Designing information environments for purpose. Facet Publishing, London, 2004.

43. Patkar, V. Perspectives on role of the library in the emerging wisdom society, *In* Knowledge, library and information networking edited by H.K. Kaul & M. Ishwara Bhat. DELNET, New Delhi, 2010. pp. 113-34.

About the Author

Dr Vivek Patkar, an independent researcher, worked as an Operations Research Specialist in Mumbai Metropolitan Region Development Authority (MMRDA) for 25 years and later as a Faculty at the ICFAI Business School, Mumbai. To his credit are seven books and over 250 research papers and articles published in various journals and magazines. He is a member of the Editorial Board of the International journal, *Human Systems Management*, IOS Press, the Netherlands and *Journal of Geomatics* of the Indian Society of Geomatics.