Nature of Complexities in a Document: Content Digitization Aspects with Special Reference to Indian Heritage Knowledge Domain

Ramdasi Nagnath Ramchandra[1] and Patil Suresh K.[2]

[1]Knowledge Management Cell,
Centre for Development of Advanced Computing, Pune India
ramdasi@cdac.india.com
[2]Librarian & Head of Library Information Science,
Jaykar Library, Pune University, Pune
patilsk@lib.unipune.ernet.in

Abstract. Since ancient days Knowledge Management has remained a topic of interest to the researchers, scientists, scholars and knowledge organizers. Especially, librarians have definite role to play when it is concerned with huge content digitization under the process of digitization. However, when certain documents are considered for digitization, it is necessary to study the complexities involved in them. Such complexities are inherent part of original work of the author or in relation with the digitization process of the content in that document or work. Under both aspects, logical complexities and physical complexities play important role. Especially, it can be observed that ancient documents under Indian Heritage have a typical pattern for such complexities. This may be true for general documents also to certain extent. Present article focuses on the nature of availability of such complexities and their impact on content management under digitization process. Study of such complexities in the content has definite impact over the knowledge representation in various digital forms.

1 Introduction

Knowledge Management is not a new concept in India. Oral disciplic tradition of knowledge in *Gurukul* System in ancient Indian history is explained in many authentic documents. Since ancient time, various branches of knowledge were persuaded and preserved and taught through these universities in the form of *Gurukul* run by many *rishies* and *acharyas*. This large oral tradition has done an important job of preserving knowledge for future generations in spite of the non-availability of the writing resources and material. Some portion of knowledge was made available in manuscript forms. About such tremendous efforts, it can be imagined from the only available rare ancient manuscripts, today. This observation does not limit to these sources only, but many other areas under Indian Heritage also have their roots in original authentic works in ancient India. Dance, music, drama, architecture, ayurveda, mathematics, astronomy and astrophysics, Astrology, Sports, Defense study, and many other knowledge branches they are referred under classified 64 puran vidyas.

Maharshi Vyasa for Veda, Upaishada and Puran-vidhya; Aryabhatta, Bhaskaracharya, Lilavati for mathematics; Bharatmuni for drama and music; Aswinikumaras, Susrut, charaka for Medicinal practice; Panini for Sanskrit Grammar; and so many….there will be a big list of such stalwart and scholars, rishis and sages, who have remained milestones in the odyssey of Indian Heritage knowledge pursuit. Ashram shiksha tradition in ancient India is one of the finest examples of preserving and practicing knowledge. The pursuit of knowledge, systematic oral preservation, and teaching methodologies would be a separate research topic. Progress in ancient Indian Universities and Societies, as well as technological developments are well written and documented in various works. Not only Hindu culture, but also Jainism, Budhism etc. cults have, their own knowledge traditions. However most of the works have disappeared in the course of timeline because of natural calamities, destruction, invasions, and theft. Whatever exists can be preserved and made available to the humanity with the help of digital technology.

2 Knowledge Management Today

Knowledge creation is often a result of processes like; brainstorming in day-to-day practice, concept mapping, creativity, mind mapping problem-solving etc. Knowledge involves the transformation of information into capabilities for effective action. This can only be done through human understanding. The human mind can create unique connections between seemingly unrelated information to come up with breakthrough ideas, but to do this; the mind must be trained to see beyond the obvious.

With the advances in Information Technology, digital world evolved as a parallel world with existing world in which we are living. There are continuous efforts allover the world to make the use of digital technology for wide-ranging applications. In the area of knowledge management, capturing, organizing and dissemination of knowledge, has got priority. Many traditional libraries and knowledge managers are engaged in putting the content in digital form. From the observation it seems that there are five prominent reasons for digitization:

1.Putting knowledge into practice 2.Professional/persnal Communication or Services

3.Preservation of knowledge and cultural heritage, 4.Educational applications

5.Research Community Publications

Digital Content and Complex Aspects

Digital world is brainchild of the human being. There are continuous efforts all over the world to add beauty and systematization of the information in it. The term 'digitization' has got special importance in this world. Process of digitization of documents is not exception to this. There are some simple documents can be digitized easily. For example; scanning a picture, typing a text on PC, sampling or capturing audio/video, etc. There are various forms to present and publish information in digital form. Information can be presented on web-site, or compiled in CD-ROM, stored in any magnetic storage media for presentation or towards the archival or for the sake of preservation and so on. Some of the intentions above do not require proper analysis or homework before the digitization process as the further intense be well met with this primitive digitization process. When it comes to the digitization of larger or compound documents mere digitization is not sufficient. Mere digitization becomes just a first step. However, such first step is also important but just not sufficient. Compound document (**R1**) is document having different type of content in it. If it has to be presented properly in digital form there are certain complex aspects, which need to be considered before the process of digitization. In the present article effort is to:

1. understand the origins and nature of such complexity in document and document under consideration for digitization;
2. study the complex nature of documents, possibly to find the conceptual co-relationships between (among) two (more) elements within and outside the documents in the same group under one subject area-Indian heritage
3. visualize the impact of these complexities on the process of digitization of the document belonging complexities

To find such relationship in a selected source-document for digitization is itself (a sort of analysis) a complex work.

It is necessary to explain more in the specified direction by answering some questions relating to the existence and nature of complexity in Indian heritage documents, if it does exists then the elements behind such complexity and so on. Necessary to consider such complexity before the process of digitization and all such issues are obvious. For the sake of convenience let us hereafter refer such documents as 'Complex Documents'. As the term is not in use in regular practice, it is obvious that, the information on is hardly available.

3.1 Meaning of the word 'Complex'

Let us understand the word 'complex' before discussing on complex documents. While using the word 'complex' in this article, in relation with complexity, to know about its origin will be worth before further analysis. From the original Latin word complexus, which signifies "entwined", "twisted together". This may be interpreted in the following ways:
In order to have a complex you need; (R2)

1. Two or more different parts or elements;
2. These parts must in some way be connected or knotted together, so that it is difficult to separate them.

Here we find the basic duality between parts, which are at the same time distinct and connected. Clearly a complex cannot be analyzed or separated into a set of independent elements without destroying it. Hence we must conclude that the reductionistic method cannot be used for understanding complex entities. This accounts for the connotation of difficult, which the word complex has received in later periods. Indeed the general rational method of tackling problems consists in analyzing the problem domain. If this domain is complex, then by definition it will resist analysis.

In the Concise Oxford Dictionary (9th edition.) term 'complex' has different meanings. Among all the meanings following two meanings are relevant and can be discussed:

i. A building, a series of rooms, a network, etc. made up of related parts i.e. arts complex.
ii. Psychologically related group of grammatical information i.e. repressed feelings or thoughts, which cause abnormal behavior or mental states.

Metaphorically, above two meanings can be correlated or applied to some of the documents to understand the scope and content. There is similarity to (i) above, nature of the complex document is with different parts, interrelated with each other but a single homogeneous document with a specific theme. As per (ii) above, documents also consist of feelings or thoughts, which cause abnormal behavior for creation or mental states in the artificial world, created by concerned author/s. If we see the meanings; a former deal not only with physical forms of the object, but it also deals with logical behaviors of the document and later deals with logical or meta-form of the objects.

3.1 Existence of Complexities

To elaborate more on the nature of complexities in any complex document; there are two aspects: i. Complexities embodied in the document itself , and ii. Complexities in relation with digitization

Two Types of Complexities - Further, from the above dictionary meanings, in relation with the above aspects (i.) and (ii.) of complexities, two subtypes of complexities can be thought under each aspect:

a. Logical Complexities; and b. Physical complexities

The document having physical and logical complexities in it is referred here as 'Complex Document'. For example: Document like 18 puranas under Indian Heritage. Complex of puranas is divided into 18.They are classified by sage Shri Vyasa. There are several stories and sub-stories. These stories/substories are interrelated, repeated in the whole edifice. The logical discourse and it's flow is narrated with the help of conversations. Eighteen puranas are full of different characteristics and wonderful/mystical environments, still having certain common features.

i. Complexities embodied in the document itself

Even without thinking that of digitization of document one can observe numerous types of complexities in a particular document (less or more way). It also depends upon the readers' interests, approaches and specified scope existing in his/her mind. This is a relative process and may vary person to person. However, in general sense, with some common observations (inferences) one reaches to some conclusion about a particular document. For example; sentences like 'that subject is not easy' or 'that book is so difficult to understand' or 'that part is very complex to grasp because it has several aspects' etc. are sort of broad measures to rate the complexity in a particular document/subject. Let us explain it with another example; average students in a school find mathematics or foreign language as a difficult subject. Though the grasping is relatively different and hence, relative feeling about the subject also varies, however, there is common consideration of complex nature for certain subject. Further, impact of such complexities in a subject can vary with the way that such document is interpreted by the author. The description about the document mentioned so far is about the complex nature of that particular document, that is relative from the both angles; readers and author.

Some of the popular terms like; 'Reference Document' or 'Meta-Document' are described because of definite qualities that are different than other document. To explain it better way, reference document is not for A-Z reading. Reference document provides information in a specified subject domain authentically. Hence for the sake of information finder, document is arranged in a specific order – Alphabetically, Subjectwise, Classified way etc. with a provision of subject-headings, term index. On the contrary a simple book that is not made for reference purpose may also have a reference value. Even though we differentiate in these two types of books.

To find a word meaning occasionally in a dictionary, consideration of search sequence arrangement in such document is very essential. It is an expert work and lexicographers decide on such arrangement. The structure of such reference documents is based on the necessity and as per the routine practice or recommended standards for doing that work. Such reference documents are also complex in nature. Thus, complexity is more comprehensive term than a reference document. To conclude; complex document can be a reference document and vice a versa. Although; 'complexity' is far broader term than the term 'reference'.

i.a Logical Complexities

These complexities are because of logical (development) nature of the source and logical forms taken into account while writing the document by an author. Sometimes they are easy to understand and many times apparently cannot make out. For example; type of language or use of difficult language, use of grammar, use of some syntax, presentation flow of the document, metaphorical explanations, meta-document nature, source document has strength or volume that can be interpreted in various ways.

Approach I -Concept of Anubandha Chatustaya

Since ancient time many authentic works are considered on the criteria of 'anubandha chatustaya'. Anubanndha chatustaya is defined in Sanskrit as – *"Purusham anubadhnaati-swadnyanen prerayati itee anubandhah"*

Meaning of the above definition is: Four components in the composition of written work normally used by an author to narrate the theme of the source, that also enable readers to learn and grasp the knowledge/message conveyed in the work. Four components together i.e. *chatustays* are:

1. authority of the author and capacity of the reader (*Adhikari*)

2. subject of the title (*Visaya*)
3. purpose of the source (*Prayojana*)
4. relationship with each other and bonding of all three above

Reader undertakes the study of such work is expected to be well verse with the logical structure of the title in the form of *anubandha chatustaya* mentioned as above. This may be noted that almost authentic works / document sources in the area of Indian Heritage are peered in the thread of *anubandha chatustaya.* No work could be considered a perfect work without this consideration. At the same time reader without the consideration of *anubandha chatustaya* will not understand and grasp the source subject properly, systematically and thoroughly.

Perhaps, in the light of these four qualities, possible complexities can be visualized within the source document and before taking the source for digitization. There is possibility of exploring these complexities not only from source document angle but also from potential readers' point of view. Such observation can make it possible to study the complex nature of the document. Thus, any title can reach out to its readers after the process of digitization, with a pre-study based on this approach of *anubandha chatustaya.*

Approach II – Six Tests as per Ancient Hindu Theory

The matter is not quite so simple when we take up the question of logical complexity in a document. It is difficult to make out the intentions of the author. If we leave aside commentaries, there are six well-known tests according to the Hindu theory of work. They (**R3**…) are:

Upakramo upasamharo abhyasau apoorvata phalam,
Arthavadau upapattich lingam taatparya nirnaye……..(Upanishada reference)

1. ***Upakrama*** (Commencement) & ***Upasamhara*** (Conclusion); 2. ***Abhyasa*** (reiteration); 3. ***Apurvata*** (novelty); 4. ***Phala*** (fruit); 5.***Arthvada*** (explanatory statements); 6. ***Upapatti*** (illustration)

In view of these tests, Samkara believes that Badarayan (Badarayan has commented on Sakara's literary work) had inview *Advaitism* of the type advocated by himself. This is in accord with the accepted position that the Vedanta Sutra sums up the teaching of Upanishadas…."

If we think on the same authentic pattern of Sankara, these six tests can be applied to find out the exact message of the author of the work. These six parameters can be helpful to map the process in author's mind. That will enable to tap the possible knowledge flow in his creation. Meta-theory understanding will help to select right metaphor for expression or for visual interpretations. Knowledge representation can be peered into a ontological meta-data and classification from the user end. Inherently a proper structure, navigation and functionality can be achieved in designing the information architecture. Even if such huge work is further taken for the process of digitization, it might give certain inputs towards the meta-structure of the source document. Such inputs can useful in the proposed core information architecture.

Approach III-Knowledge Classification

For knowledge retrieval, following factors are very important: 1.classification of knowledge in various logical and physical forms, 2. Knowledge exists in various languages, and 3. Document management

Conventional systems stresses upon the user to select query terms that match the system's index terms. But one cannot always assume that the user knows precisely what he/she needs to investigate, or that the user is able to express need in the system

language. With the availability of computers the concept was changed. To search linear way either remained one approach or was thought as absolute one. However, provision of non-linear search was properly taken care by computerized searches. As long as, knowledge domain was limited it worked well, however, today this has proven just not sufficient. Searches on web-site are very good example to experience this. It is not surprising; therefore, that consideration is being given to the use of classification for online subject access. In a ocean of full-text information that is automatically indexed and keyword searchable, classification is one of the useful logical solutions for subjects.

Classification as a guiding lamp helps human to identify focused experience and helps them associates thoughts and things. If we observe many search engines like yahoo, classification works as a gifted tool to full-text natural language terms. Ongoing meta-data projects like 'Dublin core project' classification is considered as systematic onlines of subjects from classification schedules can assist a user in retrieving information which he/she would unlikely to retrieve by using conventional subject heading or keyword search methods. The integrated subject tool online can accommodate individual learning styles of users by combining classification with authority controlled subject headings as an index to the classification. Human brains learn and remember information by categorization and by association, not by alphabetic means. Users look for unknown subjects by associations and by grouping like things together. Classification gives link and clues to unknown subjects. The use of a classification scheme helps in solving problem of users search in natural language and possible ambiguity in its meaning. Liu and Svenonius observe that: "Classification, however, go beyond thesauri by semantically structuring not only the vocabulary associated with concepts but also the concepts themselves. Classifications have sometimes been likened to semantic nets, in which concepts are linked by meaning relationships. They have also been likened to knowledge trees, in which each concept is comparable to every other concept, in the sense that its position in the scheme is defined with respect to every other position. While thesauri consist of….term clusters, classifications attempt to integrate these clusters into meaningful monolithic wholes….". Classification schemes not only assist user in locating his topic of interest within a hierarchy or cluster, but also help that user identify histopic of interest from different perspectives.

Colon Classification

Above-mentioned approach is soul for most of the existing classification systems. However, Dr.S.R. Raganathan's Colon Classification is something special. In justifying this statement it would be interesting to study 'Canons of Classification' described by Ranganathan. Pnemonicity and brevity achieved there in considering the chronological evolution and categorization in five fundamental facet shows close relationship with ontology of knowledge. We need not to go in details of that but, complex nature of the document is also an origin for the well defined five facets of Dr.S.R.Ranganathan in his 'analytico-synthetic colon classification scheme'- P Personality, M Matter, E Energy, S Space, and T Time. While classifying any document as per this scheme all these facets are taken into account. Further, cross-language existence of knowledge and document management is possible with the scheme systematically. This classification has got special place in Dublin Core Project of meta-data for subject headings description.

Technical processes in traditional library practices like; classification and cataloguing helps in preparation of meta-data in a standard fashion, so that such meta-data becomes a tool for users of that library for searching or locating a particular document. If we study and observe properly these processes, such meta-data is extracted from the basic qualities of the book/document (refer topic – How to read a book technically?). In a regular practice, many documents can be classified easily and they are simple to understand (this may not be true always..). Some rare documents

expect more attention before classifying them. They are complex in predicting right subject for a classifier because of their interdisciplinary or abstract nature. Such documents are complex documents. Such factors are sorted out while classifying the document ontologically, i.e. in natural order. Ultimate purpose is to make the document in the collection of library locatable and accessible. If we compare various schemes of classification Dr. S. R. Ranganathan's colon classification seems more natural in the light of 'cannons of classification'.

Ontology

Storing and recalling explicit information has come a long way from the handwritten indexes, card catalogs and other manual devices of a less information-dependent age. In knowledge management terms, classification involves the logical arranging of information for purposes of finding it quickly when it is needed. This is a great deal more difficult than it sounds, as language has many ways of expressing the same or similar ideas, and each individual mentally organizes his or her thoughts in ways unique to their understanding and vocabulary. The arrangement structure, itself, is usually referred to as taxonomy or ontology.

In the area of Indian Heritage a categorical schemes exhibits a hierarchical structure. Hence, there certain meta-data inputs are possible from the ontological structure of it. Knowledge under Indian Heritage is prominently placed under three main aspects: *Adhibhoutik*, *Adhidaivik* and *Adhyatmik*. This can be well observed in an authentic and prominent, single-handed work "Bharatiya Sanskriti-Kosh" by Pt. Mahadeoshastri Joshi.

The definition and concept of 'heritage' is fundamentally different in the Indian context, from what is understood from the western point of view. In the Indian context '*sanskriti*' is the true expression of the English word 'heritage'. '*Sanskriti*' means the actions and interactions, which are continuously done by humans, not only to prevent distortion but also to preserve and enhance it as the time progress. Man lives through interaction **(R7)** with nature and its constituents and transforming the environment in a manner beneficial to him. Man can succeed only if he can control the surrounding environment and also his own body, mind and intellect. In this process he changes the environment and also changes himself in a visible as well as invisible way. This process of continuous change through time results in '*sanskriti*'.

Thus, '*Sanskriti'* is a result of philosophy and practices of humans through successive generations. '*Sanskriti'* does not belong to only one individual, but to society as a whole. Individual lifecycle is short but the lifecycles of societies or civilizations are long. '*Sanskriti'* is embedded by one generation into the next generation and lives on. '*Sanskriti'*, thus decides the direction of the society or the civilization to which it belongs.

As referred in above paragraph, from Indian philosophical point of view Bharatiya *Sanskriti's* three dimensions, namely ***adhibhautic, adhidaivic*** and ***adhyatmic*** are narrated below***:***

*adhibhauti**c*** - Understanding and transforming the outer environment for man's progress and development is the *adhibhautic* dimension of *sanskriti*. In this manner he creates agriculture, animal husbandry, architecture, cities, water and wastewater management systems, hospitals and healthcare systems, and science, engineering and technology, in general. All these belong to the *adhibhautic* aspect of *sanskriti*. *Karma* (Action) is the inherent important facet under this dimension of *sanskriti*.

adhidaivic - Man often finds that he does not have complete control on either his destiny or fate or the surrounding environment, He thus perceives the existence of invisible divine (or supportive) as well as evil (or obstructing) forces. He therefore

performs worships, prayers, rituals, *mantras* and *tantras*, etc. to succeed in his endeavours or to remove obstacles in his paths. These invocations and practices belong to the *adhidaivic* dimension of *sanskriti* , where, *bhakti* (devotion) has prime importance.

adhyatmic - Finally, he aspires to discover the nature of his own self (*jiva*), the cosmos (*jagat*) into which he exists, as well as the nature of the supreme creator (*isvara*), who is the primal cause of the creation. This process of understanding the nature of *jiva*, *jagat*, and *isvara* is the *adhyatmic* dimension of the *sanskriti*. This search leads to pursuit of philosophy, religion, art, culture, and knowledge, all of which are the *adhyatmic* aspects of the *sanskriti*. Thus, *Dnyan* (Knowledge) is the inherent part of this dimension.

The foundation of Indian heritage is the ancient treatise of knowledge what is called as *sastra*-s. The entire edifice of Indian heritage is built on this foundation of knowledge. According to Indian tradition, knowledge pursuits are classified into three levels of learning, namely, *upasana,* meaning experience (consciousness), *jnana* meaning knowledge (pure science) and *kausala* meaning skill (applied science). Accordingly, *sastra*-s dealing with them are called *Para vidya*, *Apara vidya* and *Kala*. From Indian philosophical point of view *sanskriti* has *Para vidya* or *Brahma vidya*, numbering 32 and *nyasa vidya*, deals with meditation and selfrealisation on the basis of *Upanisadic* portion of *Vedas*. These are elaborated in *Brahma-Sutras* of sage *Vyas*. The *Apara vidya*-s (*sastra*-s) are classified into 14 subjects of study as *Vidya-stanas* (or *Dharma sastra*-s) which include the 4 *Veda*-s (scriptures), 6 *Vedangas*-s (*Vedic* and auxiliary sciences), 4 *Upanga*-s (supplementary subjects), and the nature of reality, itself has been apprehended in 6 *Darsana sastra*-s. *Natya sastra* of *Bharat* and *Alankara sutra*-s of *Vamana* deal with arts and there are 64 *Kala*-s (applied sciences) covering a wide range of subjects. These 4 *Veda*-s, 6 *Vedanga*-s, and 4 *Upanga*-s, 6 *Darsana*-s, and 64 *Kala*-s from the *saastric* treatises of Indian heritage. There are several major interpretations, commentaries, critiques, and translations of these primary texts which builds the edifice of Indian heritage.

i. b Physical Complexities

These complexities are related with the physical form of the document. There are various types of documents classified on the basis of physical form:

1. Printed Material
2. Handwritten. Manuscripts are handwritten not only on paper but also on cloth, parchment papers, papyrus, leaves of tree and many such natural processed or unprocessed resources.
3. Text or messages carved on stones,
4. Paintings, pictures, on walls of old monuments
5. Embroideries or artworks on the cloths
6. Electronic (i.e. Analog documents Microfilms/fiches, audio/video cassettes etc.)

All above types may or may not have sub categories in the forms of documents; like; Maps, Pictures, Tables etc. are some of the physical forms of the documents. There are some more features of the documents like; value and rare nature of the documents, clarity of the document message, authenticity, historical value and details, ownership, etc. are some of the attributes, which add complexity.

As per the time facet in Colon Scheme, certain document in 11th century may have different script, font, language, are also some of the physical aspects. Thus, not only time facet but personality facet also adds some such aspects. Matter and space can be thought on the same pattern. Thus, Colon classification makes us to think on logical ground, and throws focus on physical attributes of complexities. Thus, culture, country, civilizations, group of people etc. are some more factors, which have impact on physical forms.

3.11 Examples of some Complex Documents

There are several such documents in various areas. From the area of Indian Heritage itself or ancient Indian literature, there are several such ancient titles. Some such examples are:1.Four Vedas 2. Ten Upanishdas 3. Eighteen Puranas 4. Mahabharata 5. Ramayana 6. Shrimadbhagvadgita 7.Shri Dnyaneshwari, Shri Dasbodha etc. Saint literature.

Above referred all documents have a specific structure under the canons of '*anubandha chatustaya'*. Many scholars have authentic commentary on this and we need not to go into details. From such observations it is well-known that Upanishadas are further interpretations of four Vedas. *Purana-Vidhya* is also apparently interesting and entertaining but basic principles of Vedas and *Upanishadas* are systematically peered in its illustrations and narration of eighteen *Puranas*. This deductive logical flow is further reflected in *Shrimadbhagvad Gita.* Further the same flow is percolated in Saint literature. These original canonical principles are basic soul of the Hindu literature under Indian Heritage.

ii Complexities in relation with digitization

Process of digitization should be thought for to acquire maximum benefits out of the technology. To interpret or at least to present the message of original document effectively. Following questions need to be answered out of the digitization process: Whether the message of the document is disseminated properly? Whether the document is digitized with specified technical standards? Does the document have scope to present some elements in multimedia form? In case of large document, whether linear/nonlinear access structure has taken into consideration before presenting the document? Is it easy for navigation? Does it meet the necessary functioning for a reader? …. and so on….

In real sense to make the use of digital technology strength, all above questions will possible give trans-document experience to its reader after true sense of digitization.

ii. a Logical or Notional Complexities

In digitization process above referred only approaches are not just sufficient. Basic facets referred PMEST invite more complex challenges. For example; in case of time facet; document work in 12^{th} century e.g. Mahanubhao Cult's literature on Chakradhar Swamy - language / work in 12^{th} century requires to see a secondary document source. Such secondary source can be a guide, a translation or commentary on the original one. To maintain authenticity in text, addressing/reciting practices in the form of poetic meters like; Fonts with Ascents, *chhandas, vritta, sloka, ovi* etc. There are many more. Now question arises in our mind that; just scan this document in the form of an image or type the script/text to make it available in digital fashion. Document is digitized. This is just a primary process of digitization. i.e. capturing data. The document can be organized to make it available on the network. There are some more advantages except this one. It is very crude to say that the message of the original document is conveyed to target audience in this way. A book in print form will also serve this purpose to maximum extent.

Software Tool Development

Knowledge discovery is made through books, magazines and e-zines, organizations, software development and text or data mining and information visualization in day-to-day practice. As the volume of information surges, it becomes increasingly difficult to rescue interesting and thought-provoking concepts from among the undifferentiated flood of words and images. The use of automated methods and tools to encourage and/or confirm new ideas by analyzing bulk information can help individuals and

organizations to uncover heretofore-unsuspected patterns and connections. Software development is a solution to the process. Knowledge management has a wide variety of aspects, most of which are implemented via some sort of technology. Numerous companies have set up suites of interacting tools to cover a wide range of KM functions. This category was originally devised to hold web-sites that deal with these interacting, mostly large suites of tools. Cross-references have since been set up to link software & tools that relate to single or limited aspects of KM which are listed under that purpose.

Proper design and structure of the document for presenting it in digital form can be unveiled after this analysis. Proper meta-data cannot be visualized without this. This is not an easy task. After pre-digitization precaution only actual selection of software tool, and platform, storage of libraries for multimedia text, parallel efforts without redundancies, standardization and uniform integration of all such elements into a cohesive digital form can be achieved. Use of proper standard and copyright issues and earlier issues also have profound impact on post-digitization. Addition of new features in next upgrade or new version, migration of the platform, regular maintenance of digitized document in possible digital forms like; website, CD-ROM, Digital Library or archival centres, preservation digital content are some of the post-digitization issues.

Information Architecture

Knowledge Management **(R4)** is the science of storing, indexing, and retrieving information to add value to a culture, civilization or heritage. It focuses on creating a "repository" of collective knowledge, intellectual capital, skills, and experience. In the process of data to knowledge value of information plays important role. In this context Information architecture is a key issue under knowledge management. Information architecture is the process of defining, organizing, and developing navigation systems that helps people manage information more successfully. Information assets created in numerous forms in practice like; professional practices, books, business intelligence, human skills, intellectual property, learning organizations, etc. These are just some of the origins for information asset. As the world enters the ‘Knowledge Age’, information on virtually every aspect of our lives and businesses is becoming available at an increasing rate. Content search on the web has become a jargon task, because of getting number of hits. These hits are many time not relevant and time consuming to find the specific information out of the massive information. When the availability of information outstrips the time and energy of those who could potentially use it, frustration can result. This condition is often referred to as information overload. This can be resolved with proper knowledge retrieval with proper representation.

Knowledge Representation

Knowledge representation covers techniques for representing and storing knowledge in a way that is usable by a machine, for instance, programming languages specifically formulated for working with knowledge (knowledge representation languages), and databases for storing facts about the world.

There are various ways in representing knowledge like; Semantic Web, The Resource Description Framework (RDF), Topic Maps etc.

ii.b Physical Complexities

Physical complexities in relation with digitization are outcome in the process of human-computer interactions, at interface level. Physical experience with computerized information. Though, there is logic behind the process, existence of experience at physical level can be separated. Digital multimedia technology has made tremendous impact over the organization and presentation of information. At

interface level role of human sensory organs cannot be ignored. Human experience is based upon five basic sensory organs:

1.Word 2. Touch 3. Face or personality 4. Taste 5. Smell

Human-computer interactions are trying towards achieving real-resembling experience out of the tools developed on computers. Human can interact as much as natural way with computer is the ultimate goal. Taste and smell experiences are at virtual level and there is research going on around the world. However, first three organs are discussed below:

1. **Word-** Aristotle says, "there can not be a word without an image in mind". Words are in the form of text, audio, image and video (animated images). Under textual expression of words; Fonts, size, colors, layout, and character animation are important attributes. Fonts and colors interpret certain meanings. Color is an integral part of most visual representations. Color is an attribute that is not strictly necessary in order to distinguish shapes or to perceive the real world in a sufficiently operative way, as anybody that has seen black and white TV or photography can state. Labeling, measurement, representation, decoration, emotions and moods visual attention etc. are things related with use of colors. Though they are differently related as per the culture, some common meanings are common allover the world. At physical interface, way of presenting these has got special importance. Hence, selection of tools, formats, memory size, storage media, requirement of apparatus and many other factors are based upon this physical consideration. Multimedia has helped reverse way to saying of Aristotle. Words again can be interpreted pictorially.
2. **Touch**- visual metaphor plays an important role to acquire a sense of touch. Image, Audio, animation, are core elements for these visual metaphors. Various buttons, icons, functions, instructions, help in spreading sense of touch for users. Use of touch screens, mouse, keyboards, sticks and other controls are also important role players in concerned with sense of touch. This also rises further into complexities like; size, medium, of the information.
3. **Face or Personality**-Overall look of presented information on desktop is personality. It matters lot for further interactions with PC. Eye catchy or fascinating and pleasant personality of information presented meet with the soul purpose behind presenting certain information. All elements text, image, audio, animation play important role in personality. Resolutions, Color combinations, appropriate visual metaphors, natural flows of information most appealing to the users are important elements of personality. This physical dimension has profound impact on logical complexities like; Information architecture and knowledge representation.

Effectiveness of any communication depends upon the combination of (two or more) of sensory organs. The way of presentation is obviously important. If such combinations are used effectively, it results into a fascinating and eye-catchy as well as interactive presentation. Taste of presentation is further dependent sensory organs on these factors. Virtual creation of taste out of virtual word, touch, face is interesting to think, but it is digital world reality. Thus, multimedia plays very important role to reach out to the target audience in communication of information.

Multimedia basic elements have their own natures, which also add complexities. Language, large text size, quality and size of images, audio, video, compression techniques, concerned issues like; storage, organization of data, format, retrieval or dissemination of the information are some of the examples of these complexities.

3.2 Complexity in Knowledge Flow

Following figure indicates the flow of knowledge source for the process of digitization. Pre-digitization to Post-digitization complexities- from top to down direction can be observed. Similarly, Logical to physical complexities from top to

down and left to right direction can be observed. However, looking at the digitization process box, to demarcate distinct boundaries between the two is difficult. Also, depends upon the form of the content. Earlier infinite domain of knowledge is limited to a source document. The source document form also gets transformed into digital form later. It will be interesting to study the pattern of changes in the complexities related to both forms. That is beyond the scope of the present article. Continued human-computer interaction further provides the improvement feedback that enables to improve the whole process and related tools for digitization. Possible spin-offs and other factors mentioned in the figure makes tremendous impact on the post digitization complexities. **See the figure at the end**

3.3 Advantages of studying complexities

Thus, study of such complexities is very important before going for a knowledge management in a digital form. Above both types of complexities are the root cause for technical solutions and further developments. Analytical study of the document helps further to visualize a viable picture after digitization. Following are some of the areas, where such analysis will have advantages like;

- Estimation of expenditure
- Volume of the work and possible time schedule with logical stages
- Requirement of manpower
- Meta-data study and search algorithm
- Organization of the document with proper structure, navigation and functioning that meets users needs and expectations
- Special font development or selection of available fonts
- Possible spin-offs or by-products and intellectual property issues
- Size of the multimedia content and storage requirement
- To achieve standardization and uniformity in overall efforts
- Human-computer interactions
- Future compatibility and portability of the content

3.4 Conclusion

In a nutshell, one needs to accept the complex nature of the documents under Indian Heritage, in terms of original discourse of the documents and in relationship with digitization aspects. Understanding the logical and physical types of complexities will help in giving proper direction and shape to the digitization process of such complex documents. Whole knowledge reflected in the form of document is a metaphor of actual world with hazardous complexities. Putting it in the form of digits –0s & 1s, is just a simulated effort of human beings.

R1. D'Alleyrannd and Mark, R.D. 'Components of compound document' Image Storage and Retrieval System, 1989, pp.58.

R2. Complexity article…see Francis Heylighen 'Building a Science of Complexity' http://pespmcl.vub.ac.be/papers/Building_Complexity-fn.html\1"fn0" : Invited paper, to publish in: Proceedings of the 1988 Annual Conference of the Cybernetics Society (London)

R3. S. Radhakrishnan, Indian Philosophy Vol.II, Blackie & Sons Pub. Pvt. Ltd., Mumbai, pp.469-70. (Upkramo….)

R4 Open Directory Project Site: "http://dmoz.org/Reference/Knowledge_Management/ ")

R5.Sonqiao Liu and Elaine Svenoniums, "DORS:DDC Online Retrieval System." Library Resource and Technical Services 35, no 4 (October 1991): 360p.

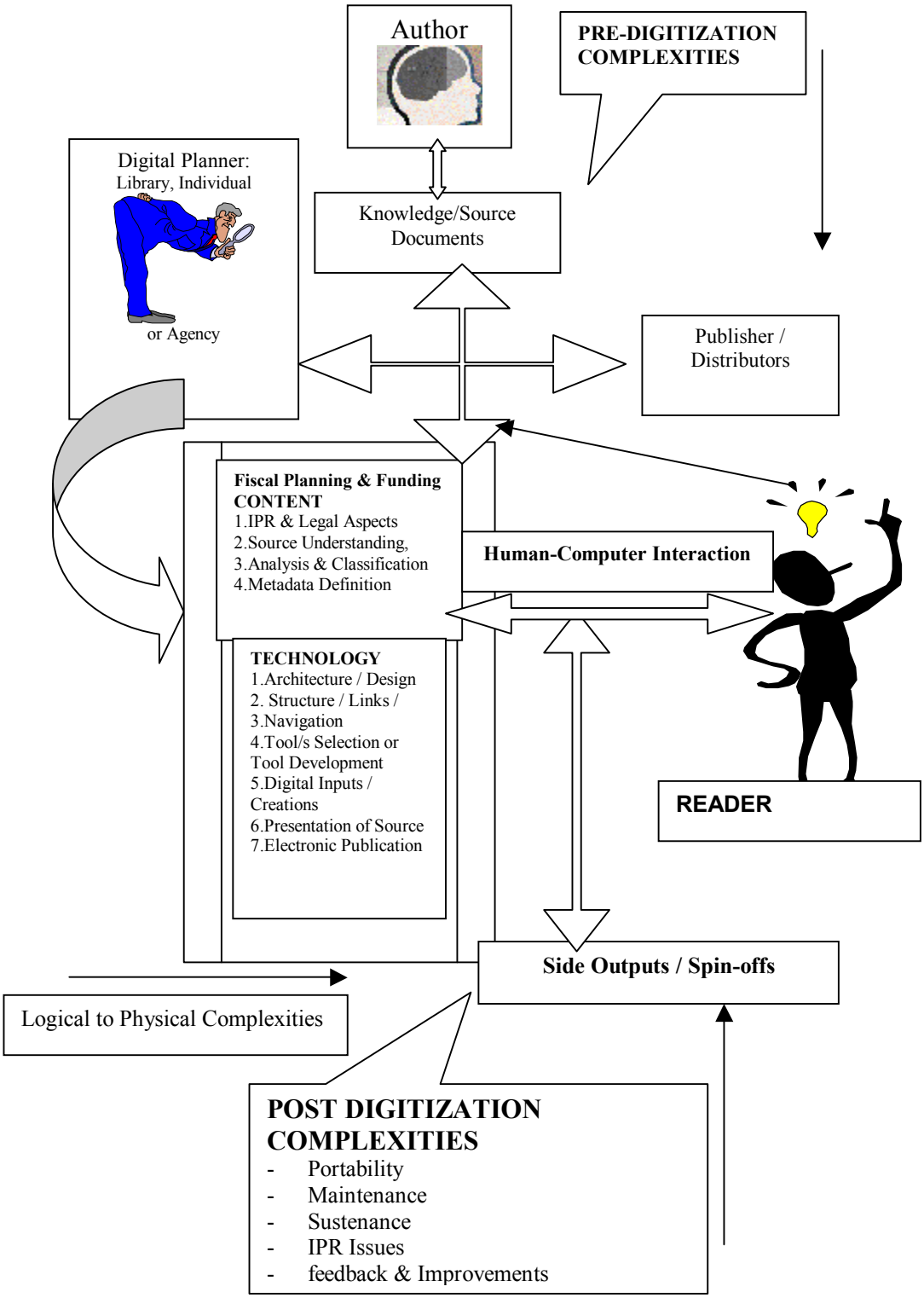
R6.Koh, Gertrude S. "Options in Classification Available Through Modern Technology" Cataloguing and Classification Quarterly 19(3/4) 1995, The Haworth Press, New York, 195-211p.

R7.Joshi, Mahadeo, Bharatiya Sanskriti Kosh

R8.Joshi, Vinayak Narayan, ' Vrutti Prabhakar Rahatsya' Vol 1, Sakhare Prakashan, Alandi (Pune) pp.141-142.

R9.Ranganathan, S.R., Colon Classification

R10.Profium, Michael Auffret, "Content Management makes Sense- Part I: Delivering Increased Business Value through Semantic Content Management and Part II: Technical Impact of Semantics on Content Management, Journal of Knowledge Management Practice, December 2001. (http://www.tlaic.com/article28.htm and article 29.htm


Author
PRE-DIGITIZATION COMPLEXITIES
Digital Planner:
Library, Individual
or Agency
Knowledge/Source Documents
Publisher / Distributors
Fiscal Planning & Funding
CONTENT
1.IPR & Legal Aspects
2.Source Understanding,
3.Analysis & Classification
4.Metadata Definition
Human-Computer Interaction
TECHNOLOGY
1.Architecture / Design
2. Structure / Links /
3.Navigation
4.Tool/s Selection or Tool Development
5.Digital Inputs / Creations
6.Presentation of Source
7.Electronic Publication
READER
Side Outputs / Spin-offs
Logical to Physical Complexities
POST DIGITIZATION COMPLEXITIES
- Portability
- Maintenance
- Sustenance
- IPR Issues
- feedback & Improvements