

IN SEARCH OF AN AXIOMATIC CONCEPT OF DOCUMENT

Bidyarthi Dutta

Librarian
St. Xavier's College
30, Park Street,
Kolkata 700 016
E-mail: bidyarthidutta@rediffmail.com

Anup Kumar Das

Faculty of Library and Information Science
School of Social Sciences
Indira Gandhi National Open University
Maidan Garhi, New Delhi 110 068
E-mail: anupdas2072@hotmail.com

Ordinarily, the word 'document' denotes a textual record that holds information. The term 'document' represents a discrete physical object in human perception. The role of document in the context of information retrieval has been discussed. The nature, form and scope of documents are also continuously changing. The feasibility and scope of considering natural objects as primary documents have been discussed. Various concepts of documents are presented from different viewpoints.

INTRODUCTION

The word "document" means '...a record, which contains information, originally an inscribed or written record but now considered to include any format in which information might be held (e.g. map, manuscript, tape, video, software)' [1].

Library and information centre may be viewed as a trinity or triangle of user (*information clientele*), document (*information*) and library staff (*information professional*). The macroscopic view of library and information science subtends the *library user-document-library staff* triangle, whereas the microscopic view of library and information science subtends the *information clientele-information professional* triangle. Information is the microscopic counterpart of document and vice versa. Macroscopically, a document is the fundamental unit or building block of a library and information is the fundamental unit of a document. Hence, a document may also be considered as an information centre in restricted sense. Document is a frame that holds up information. Information is everywhere in space and time and saying this is equivalent to saying that gravitation is everywhere. Information is the result of interaction between human perception, nature and the resident of human brain. Nature is the prime source of

information and most commonly and widely used information centre. The scope and feasibility of considering natural objects as most fundamental or truly primary documents have been discussed here. Organized information in some specified ways has been shaped into document or a group of documents or document clusters, which are secondary or tertiary documents. The document may exist in a number of physical forms in real world or in non-physical forms in abstract world, which is commonly known as Invisible Colleges. Paul Otlet [2] and Buckland [3] discussed nature and scope of documents and the physical forms of information, i.e., 'Information-as Thing'. These issues are vital because mechanical information systems can only operate on physical representations of information.

DOCUMENT AND DOCUMENTATION: SCOPE AND VARIOUS CONCEPTS

Two fundamental questions are very important: What is a document? And what could not be a document? Ordinarily information storage and retrieval systems have been concerned with text and text-like records. Text-like records generally describe some physical media of two dimensions (i.e. flat in Euclidean space) having inscriptions as carriers of information. Information is retrieved from that flat surface for processing. Textual records are linear in function, i.e. only a single mode of perception (reading, visualizing, hearing, calculating etc.) is possible at a time. But increasing interest in multimedia reminds us that not all information disseminating media are textual or text-like, i.e., linear. Multimedia presents various non-linear forms of consolidated information packages to observe events, processes, images as well as texts.

Documentation was a set of techniques developed to manage potentially significant documents, in practice, printed texts. The term 'Documentation' sketches the functional profile of the document. If document is a variable in information space then documentation is the function of that variable. Documentation tells the dos and don'ts with/by the document. But, there is no theoretical reason why 'Documentation' should be limited to texts or printed texts. There are many other kinds of objects other than printed texts. If scope of 'Documentation' includes printed or non-printed texts could it not also deal with non-textual documents? What is the extension of the twin concepts, 'Document' and 'Documentation'?

In the late 19th century, there was increasing effort on document organization and management with the rapid increase in the number of publications, especially of scientific and technical literature. Continued effectiveness in creation and dissemination of recorded knowledge needed new techniques to manage the growing literature. The traditional term for the technique of managing the literature was known as 'Bibliography'. Early in the 20th century, the word 'Documentation' was increasingly adopted in Europe instead of 'Bibliography' to denote the set of techniques needed to manage this explosive growth of documents.

Wollege provided a detailed account of the evolving usage of the term 'documentation' and its related words in English, French and German [4]. From about 1920, 'documentation' was increasingly accepted as a general term to encompass bibliography, scholarly information services, records management and archival work. There are numerous literatures on the definition, scope and nature of 'documentation'. Much of it is concerned with the relationship between documentation, bibliography and librarianship. Loosjes explained documentation in historical terms [5]. According to him, systematic access to written texts became more difficult after the invention of printing resulting in proliferation of texts. Scholars were increasingly obliged to delegate tasks to specialists. Assembling and maintaining collections was the field of librarianship and bibliography was concerned with

the description of the documents. The delegated task of creating access for scholars to the topical contents of documents, especially of parts within printed documents and without limitation to particular collections, was redefined as 'documentation'.

However, since beginning, the notion of document was extended beyond written texts. "Any expression of human thought" was a frequently used definition of 'document' among documentalists. In the U.S.A, the phrases "the graphic record" and "the generic book" were widely used. This was the convention for extending the scope of the field to include pictures and other graphic and audio-visual materials.

Paul Otlet pointed out that document could be three dimensional, which enabled the inclusion of sculpture. From 1928, museum objects were likely to be included by documentalists within definitions of document. Otlet extended the definition of document as graphic and written records are representations of ideas or of objects, and Otlet wrote: "...but the objects themselves can be regarded as 'Documents' if one is informed by observation of them" [6].

In 1935, Walter Schurmeyer wrote: "Nowadays one understands as a document any material basis for extending our knowledge which is available for study or comparison" [7]. Suzanne Briet pointed out that "A document is evidence in support of a fact". She elaborated a document as "Any physical or symbolic sign, preserved or recorded, intended to represent, to reconstruct, or to demonstrate a physical or conceptual phenomenon".

Briet enumerates six objects and asks if each is a document [8]:

Object	Document
Star in sky	No
Photo of star	Yes
Stone in river	No
Stone in museum	Yes
Animal in wild	No
Animal in zoo	Yes

On the basis of Briet's enumeration, Buckland drew the following inferences [9]

- There is materiality: physical objects and physical signs only.
- There is intentionality: it is intended that the object be treated as evidence.
- The objects have to be processed: they have to be made into documents.
- There is phenomenological position: the object is perceived to be a document.

S. R. Ranganathan viewed document as "*Embodied micro-thought on paper or other material, fit for physical handling, transport across space, and preservation through time*" [10].

DOCUMENT CLUSTER AND DOCUMENTATION

The term 'documentation' sketches the functional profile of the document. If document is a variable in information space then documentation is the function of that variable. Documentation tells the dos and don'ts with/by the document. The quark of the function 'documentation' is a bit of information; a set of information bits forms the document and a set of documents forms a 'document cluster'. Document clustering is thus a method of automatic generation of a set of documents around which all similar documents are available. A clustered file provides efficient file access by limiting the searches to those document clusters that appear to be similar to the corresponding queries. Conceptual and contextual nexus between documents, measured by document-document similarity coefficients, conveys information about the joint relevance of documents to queries in a collection. Two main strategies however can be adopted for generating hierarchical clustering [11]:

1. A complete list of all pairwise item similarities can be constructed, where it is necessary to employ a grouping mechanism capable of assembling into common clusters of items with sufficiently large pairwise clusters, and
2. Heuristic methods can be used that don't require pairwise item similarities to be computed.

The following are the basic steps involved in a hierarchical agglomerative clustering method [12]:

1. Compute pairwise document-document similarity
2. Place each of the documents into a cluster of its own.
3. Form a new cluster by combining the most similar pair of current clusters; update the similarity matrix; calculate the entries in the row corresponding to the new cluster.
4. Repeat step 3 if the number of clusters left is greater than one.

There are two strategies for searching clusters, the *top-down approach* and *bottom-up approach*. In the top-down approach, the query is first compared with the highest level centroids, and the search proceeds downwards in the cluster tree until eventually some individual items in the lowest level clusters are chosen for retrieval. In the bottom-up approach, only the lowest level centroids are stored, i.e., those containing only the specific items of the collection, and the higher-level cluster structure is disregarded. Starting with the low-level centroids, the best clusters are identified, and some of the documents located in these clusters are then retrieved.

These two functions, viz. 'documentation' and 'document cluster generation' are alike in nature. The function 'documentation' deals with organisation and consolidation of information, whereas, the function 'document cluster generation' deals with organisation and consolidation of individual documents. The term 'documentation' may be renamed as 'information cluster generation' also being conceptually in consonance with 'document cluster generation'. The former concept is operative at micro level while the later at macro level. 'Information cluster generation' may function over a wide range of information sectors, starting from traditional library aiming up to invisible colleges or virtual libraries. 'Document cluster generation', whereas is suitable only for traditional library systems.

PHYSICAL OBJECT AS DOCUMENT

If the nature is our prime source of knowledge then any natural object may be considered as a source

Table 1
Physical Object: STONE
Nature: Non-living
Residence: Inner Earth & Earth's Surface

Specific User Group (Information Clientele)	Concerned Specific Subject Area	Broad Area of Discipline	Concerned Features (Information in Context)
Archaeologist	History & Archaeology	History & Museology	Age & site of residence
Astrologer	Palmistry	Astrology	Compatibility in forecasting fate
Chemist	Chemistry	Chemical Sciences	Chemical composition
Civil Engineer	Structural Engineering	Engineering Sciences	Strength & concreteness
Craftsman	Arts & crafting	Fine arts	Cut & clarity
Geologist	Geology	Earth Sciences	Intrinsic nature
Geomorphologist	Geography	Earth Sciences	Shape & land of residence
Pharmacist	Pharmacology	Medical Sciences	Chemical composition for medical usage
Physicist	Physics	Physical Sciences	Physical properties

Table 2
Physical Object: STAR
Nature: Celestial Object
Residence: Space

Specific User Group (Information Clientele)	Concerned Specific Subject Area	Broad Area of Discipline	Concerned Features (Information in Context)
Cosmologist	Stellar Dynamics	Space Physics (Physical Sciences)	Stellar Origin, Big Bang Theory
Mathematician	Stellar Precession	Mathematical Space Physics (Mathematical Sciences)	Stellar Shift or Gravitational Red Shift
Musician	Music	Performing Art	Twinkling splendour in endless Black
Nuclear Scientist	Nuclear Astrophysics	Physical Sciences	Thermonuclear Reaction
Philosopher	Aesthetics	Philosophy	Symbolic isolation in fathomless eternity
Poet	Poems	Literature	Beauty in the black sky
Spectral Analyst	Stellar Spectroscopy	Physical Sciences	Stellar radiation & luminosity
Theoretical Astrophysicist	Black Holes	Astrophysics (Physical Sciences)	Stellar evolution

Table 3
Physical Object: ROSE
Nature: Living Being (Plant)
Residence: Earth's Surface

Specific User Group (Information Clientele)	Concerned Specific Subject Area	Broad Area of Discipline	Concerned Features (Information in Context)
Botanist	Botany	Biological Sciences	All botanical features of flower
Horticulturist	Floriculture	Agricultural Sciences	Breeding of rose plant
Painter	Drawing	Fine Arts	Colour & appearance of the petals
Photographer	Still Photography	Fine Arts	Photographic features
Plant Pathologist	Plant Pathology	Biological Sciences	Disease afflicting the plant
Psychologist	Emotional Bondage	Psychology	Credibility as a symbol of love

Table 4
Physical Object: TIGER
Nature: Living Being (Animal)
Residence: Earth's Surface

Specific User Group (Information Clientele)	Concerned Specific Subject Area	Broad Area of Discipline	Concerned Features (Information in Context)
Animal Physiologist	Animal Physiology	Biological Sciences	Physiological features
Anthropologist	Anthropology	Biological Sciences	Origin of species
Business Merchant	Export/Import	Business	Material values of skin & other limbs
Film-Maker	Media & Film Studies	Fine Arts	Different attitudes & actions
Photographer	Still Photography	Fine Arts	Valor in appearance
Veterinary Practitioner	Animal Health	Medical Sciences	Health condition
Zoologist	Zoology of Mammals	Biological Sciences	All zoological features

Table 5
Physical Object: BLOOD
Nature: Living Animal's Organism
Residence: Animal Body

Specific User Group (Information Clientele)	Concerned Specific Subject Area	Broad Area of Discipline	Concerned Features (Information in Context)
Biochemist	Biochemistry	Chemical Sciences	Biochemical properties
Biophysicist	Biophysics	Physical Sciences	Biophysical properties
Cardiologist	Cardiology	Medical Sciences	Circulation dynamics through the ventricles
Clergyman	The Crucifixion of Jesus Christ	Christianity/Religion	Christ's divine love & forgiveness
Clinical Practitioner	Clinical Sciences	Medical Sciences	Resident bio-ingredient's quantity
Forensic Expert	Criminology	Forensic Sciences	Clue to locate criminal
Haematologist	Haematology	Medical Sciences	Components like Haemoglobin, WBC etc.
Molecular Biologist	Molecular Cell Biology	Biological Sciences	Blood cell properties
Novelist	Fiction	Literature	Symbol of cruelty, crime etc.
Painter	Drawing	Fine Arts	Colour & appearance
Poet	Poem	Literature	Symbol of emotion, surrender etc.
Thespian Personnel	Theatre/Drama	Fine Arts	Mental shock & hearty emotion

of information or document. But a particular object appears differently to different users belonging to different areas of disciplines. Here, five such natural objects have been chosen (Stone, Star, Rose, Tiger and Blood) and their interpretations to different groups of information clientele are given in Table 1-5.

Thus a particular natural or physical object bears different interpretations in different contexts to different groups of information clientele. A natural or physical object may then be considered as a

multi-dimensional document. Human perceptions (seeing, visualizing, hearing, thinking, etc) are applied on these natural objects and the resulting outcome is first-hand nascent information or primary documents.

CONCLUSION

It may be concluded that any physical object should be considered as a generalised document. Generalised document is a document with multiple dimensions. Documents for different information

clientele and belonging to different information clusters are nothing but different specialized forms of generalised documents under certain conditions.

ACKNOWLEDGEMENTS

We are indebted to our mentors Prof. B.K. Sen, Information and Communication Society of India, New Delhi; Dr. Chaitali Dutta, DLIS, Jadavpur University, Kolkata; and Dr. Krishnapada Majumder, DLIS, Jadavpur University, Kolkata; for their valuable remarks and suggestions.

REFERENCES

1. Brown [D]. Dissemination of Information. In *Encyclopedia of Library and Information Science. Volume 68*. 2000. Marcel Dekker; New York; 25-45.
2. Otlet [P]. International Organisation and Dissemination of Knowledge: Selected Essays (FID 684). 1990. Elsevier; Amsterdam.
3. Buckland [M K]. What Is a "Document"? *Journal of American Society of Information Science and Technology*. 48, 9; 1997; 804-809.
4. Woledge [G]. 'Bibliography' and 'Documentation': Words and Ideas. *Journal of Documentation*. 39; 1983; 266-279.
5. Loosjes [T P]. Was Ist Dokumentation? In *Dokumentation Wissenschaftlicher Literature*. Pp. 1-8. (On *Documentation of Scientific Literature*. 2nd ed. 1973; London: Butterworths).
6. Otlet [P]. *Traite de Documentation*. Pp. 153. Amsterdam: Elsevier, 1990.
7. Schurmeyer [W]. Aufgaben und Methoden der Dokumentation. *Zentralblatt fur Bibliothekswesen*. 52: 533-543, 1935.
8. Briet [S]. *Qu'est-ce que la documentation*. Pp 7. Paris: EDIT; 1951.
9. Buckland [M K]. *Op. cit.*
10. Ranganathan [S R] Ed. *Documentation and its Facets*. 1963. Asia Publishing House; London.
11. Chowdhury [G G]. *Introduction to Information Retrieval*. 1999. Library Association Publishing; London.
12. Ibid