

# Chapter One

*If a discipline is defined by the nature of its problems, then library science must be the discipline to end all disciplines. We have more problems per square head than almost any other field. (Phyllis Allen Richmond, 1977)*

## INTRODUCTION

### Background

We are living in an age of change unparalleled in history. Technological, cultural and social upheavals have impacted us with regularity and increasing ferocity, radically changing the way we live, work and learn. In particular, these changes have been all pervasive in fields like education, research and industry. The traditional frameworks of learning, research and development are being challenged by technological innovation. The academic system is continuously changing. The libraries are also undergoing through so many evolutionary changes like digital library and virtual library etc from the traditional close-walled book-store concept. The traditional library system today can hardly fulfill the needs of the students, faculty members, research scholars and scientists. Both the information need and information seeking behaviour are also considerably changed. Information professionals need to know more than just their subject to serve information clientele. They need to know the ways to be properly informed, the ways not to be misinformed; they need to know how individuals experience their information need. In traditional learning system several students gather together into a lecture room to sit for some hours to listen to lectures and after that they visit library to collect some notes from some recommended text books. Some time later they are required to restate the information they have copied onto paper in a manner that is pleasing to the examiners. If they keep to the rules they earn a pass mark. This is the prevalent model of higher education and traditional library system, which are evidently inadequate to meet many of the needs of the contemporary information society. In spite of so many initiatives,

antiquated practices are still very much in evidence, underlining the fact that generally the traditional learning system is not moving with the times. The traditional library cannot survive in its present format, because global trends are ensuring that alternative methods are gaining ground. The traditional library and information system mainly deal with macroscopic documents, which include both book and non-book material in print format. Today libraries are developing collections in electronic media in various formats like CD-ROM, online database, online e-journal consortia etc. The transition from print to electronic media caused a drastic change to the present-day library scenario. But, still the macroscopic documents are the main concern of the electronic media. Information professionals and clientele always look for book, journal-article, conference-proceedings etc. through different search parameters like author, title, subject, publisher, call no. etc. which are quite inadequate for information seekers in the context of their changed information need. For instance, a research scholar, whose area of research is, say, Fermi liquid (a specific subject under condensed matter physics); hardly asks about all papers published on that subject in a particular year. But he'll make query about the thrust areas of research of the subject for a stipulated time-period, or he may ask about extinct areas in a particular subject, or he may ask about stable areas of research over a long time, or he may ask about current trend of research, or latest state-of-the-art etc. These types of queries could be answered only if the information service provider considers not only the macroscopic documents, but also collect the keywords from the published literature to examine it critically, so that the information at microscopic level could be retrieved. The thorough look at the microscopic level only contemplates the current modes of changing information need.

However, it is not only technology that is driving the changes in library and information systems and services. The other reason is changing trends in our educational systems. There is growing opinion that the current educational system must change, or society must face the consequences. Some have

argued that the education system must change purely because it is a system originally set up to meet the needs of the industrial revolution, and is now therefore outmoded. The industrial revolution demanded the synchronisation of human behaviour with machinery, and mass education was instrumental in preparing generations of regimented workers. Information and Communication Technologies (ICT) have already made a significant impact on higher education to change forever the face of traditional education. The contemporary information society doesn't demand a flock of closely controlled workers with mere mechanized monotonous efficiency, but groups of fresh well-trained professionals with acute skill of finding out the right information with right bit at right instant in right way for the right clientele from the ocean of information. The demand for more knowledge in the information society has resulted in an increase of subjects offered in higher education. A glance through a decade-old prospectus and a comparison of the same with this year's from the same university will undoubtedly show that there has been a big increase in the number of courses offered, and the diversity in range of subject matter. This is a consequence of universities diversifying in an attempt to capture new student populations. Furthermore, the subjects are becoming so specialised, that the academic viability of traditional subjects are continuously squeezing. The traditional subjects are proliferating continuously in several specialised components due to information overload in a particular discipline. The universities must keep pace with this changing environment by delivering courses that are both economically viable and high in quality, or they must diversify using a strategy based on distributed learning. The library and information resource centres should also amend strategies of services to the information seekers. In this way they can maintain the quality using technology whilst optimizing the time and resources of well-qualified information professional to support services.

## Universe of Knowledge

From the primitive days of human civilization to the present day knowledge has always been a component of growth and improvement in life. Knowledge is an organized set of statements of facts or ideas, presenting a reasoned judgment or an experimental result. The entirety of knowledge forms the universe of knowledge that is highly relevant in library and information perspectives. All subjects originate in the universe of knowledge and broaden the confine of knowledge. The process is continuous and dynamic. The continuous addition to the knowledge confine results in consecutive birth of new subjects through various mechanisms of subject formation, viz. fission, fusion, loose assemblage, clustering etc. along with proliferation of subjects and mushrooming of inter- and multi- disciplinarity of subjects. Library classification theory includes the concept of universe of knowledge and library classificationists have expounded on the universe of knowledge extensively. Sayers remarked in the 1920s, that no matter how the universe of knowledge is structured, that structure is ultimately strictly logical in character (Miksa, F 1998). Later, Bliss expanded Richardson's idea of the relationship of classification categories to the "things" of actual existence into a grand theory of Nature (Miksa, F). As the universe of knowledge has become an important object of investigation, classification theorists have discovered that it is a much more complex phenomenon than was previously supposed. In fact, Richardson, Bliss, Sayers, Ranganathan, and those who have investigated them to get at the phenomenon, the more complex it appeared and the more difficult it was to describe. Much of this new sense of complexity was due to the new approach to subjects that arose from the documentation movement and elsewhere, where subject complexity is a direct reflection of the way in which scientists, engineers, and specialists of all kinds search for information.

Recently, Sen stated some characteristics of Universe of Knowledge presented below (Sen, B.K. 2007):

- 1) The Universe of Knowledge (UoK) is ever-expanding. As such, the component of time automatically creeps in there. If the time component

becomes zero, expansion automatically stops. Similarly, if the space harbouring the UoK becomes finite, expansion stops. Hence, time and space both will have to be infinite for the eternal expansion of knowledge. Expansion without energy is incomprehensible. It is obvious that energy is but an integral component of the universe of knowledge.

- 2) According to Hindu philosophers the universe is composed of five primary elements, i.e. *kshiti* (earth), *apa* (water), *teja* (energy), *marut* (air), and *byom* (space). In the UoK we find all the elements. Some segments of the universe are like *kshiti* having solid state, identifiable shape and measurable size. The segments of second category are like *apa* still not having any solid foundation and well-defined shape but likely to have a definite shape and size in future. The rest of the segments are nebulous like *marut* not having any fixed shape or size. They may or may not attain the *apa* and subsequently *kshiti* stage. Some may vanish forever from the *marut* stage itself.
- 3) In UoK, time is but an indispensable component. The concept of time is missing in the five primary elements described above. This we can find in the Hindu philosophy of *srishti* (creation), *sthiti* (stability) and *laya* (destruction). If the time component is zero, then there won't be any *sthiti*. Hence, it can be easily assumed that the time component is embedded in *sthiti*.

According to Sen, of the eight components of Hindu philosophy described above, Ranganathan in his PMEST has taken five. The presence of E(nergy), S(pace), and T(ime) is quite clear. Possibly, P(ersonality) and M(atter) take care of the primary elements *kshiti*, *apa*, and *marut*.

The type of information need and, consequently the information seeking behaviour of various information clienteles has radically changed today. Two derivative results subsequently arose from this effort to characterize the universe of knowledge. One result has been an ongoing attempt to characterize the universe of knowledge in all of its complexity by means of some suitable analogy.

Various analogies have been propounded in this respect, some of which have been fairly traditional. For example, Richardson, Bliss, and Ranganathan at one time or another resorted to the analogy of a growing tree as a way to speak of the universe of knowledge, the tree's various branches representing its divisions and the tree's system of branches representing its complexity. Since their writings, other analogies have been put forward, one of the most striking being that of E. J. Coates in his *Subject Catalogues* (1960, p. 32) where he pictures the universe of knowledge in terms of a geographical metaphor in order to show the growing complexity of the universe of knowledge since the nineteenth century (Miksa, F).

Ranganathan fundamentally changed the way the universe of knowledge (i.e., Ranganathan's universe of subjects) is viewed in library classification theory and technique by shattering the previous view of it as a singular, one-dimensional hierarchical structure and replacing that view with one that is structured in a complex, modular, faceted way (Ranganathan, S.R. 1967). Since Ranganathan's day library classificationists have viewed the universe of knowledge in this new arrangement of faceted structures. At the same time, it should also be noted that hardly any library classificationists have adopted Ranganathan's picture of the universe of knowledge in terms of the mathematical analogy behind it. Library classificationists appear to have adopted Ranganathan's view chiefly because it offered a more satisfactory way to view and deal with subject complexity than was offered by any previous approach. The universe of knowledge is an assembly of the classified entities with a somewhat metaphysical sound. The classes in classification systems are mainly based on the literature that they classify.

## Information Explosion

It is a cliché today that we are all drowning in too much information. Information explosion is an obvious phenomenon today and, also it is the direct consequence of the incessant expansion of the Universe of Knowledge, that results proliferation of subjects with non-stop up rise of multi-disciplinarity and inter-disciplinarity of subjects. Many people have tried to quantify the size of the problem and statistics around, often with no firm basis but, nevertheless, scary. Let us read the following statements, taken from various websites and articles:

*“As much new information will be available in the next decade as has been discovered in the whole of human history”*

*“It is estimated that it would take ... seven centuries to read a year’s chemical literature”*

*“There are 550 billion web-connected documents”*

*“Fifty thousand new book titles appear annually, with 1.5 million books now in print from 20,000 different publishers”*

*“People could read 24 hours per day, 365 days per year, and never catch up with what is written”*

The web site [www.sims.berkeley.edu/research/projects/how-much-info-2003/](http://www.sims.berkeley.edu/research/projects/how-much-info-2003/) contains an article titled “How Much Information?” from which comes the following:

“ . . .almost 800 MB of [new] recorded information is produced per person each year. It would take about 30 feet of books to store the equivalent of 800 MB of information on paper”. An interesting chart stating quantitative measurement of information is also given there, which is presented below:

### **Title: How Big is an Exabyte?**

#### **Kilobyte (KB)**

*1,000 bytes OR  $10^3$  bytes*

2 Kilobytes: A Typewritten page.

100 Kilobytes: A low-resolution photograph.

### **Megabyte (MB)**

*1,000,000 bytes OR  $10^6$  bytes*

1 Megabyte: A small novel OR a 3.5 inch floppy disk.

2 Megabytes: A high-resolution photograph.

5 Megabytes: The complete works of Shakespeare.

10 Megabytes: A minute of high-fidelity sound.

100 Megabytes: 1 meter of shelved books.

500 Megabytes: A CD-ROM.

### **Gigabyte (GB)**

*1,000,000,000 bytes OR  $10^9$  bytes*

1 Gigabyte: a pickup truck filled with books.

20 Gigabytes: A good collection of the works of Beethoven.

100 Gigabytes: A library floor of academic journals.

### **Terabyte (TB)**

*1,000,000,000,000 bytes OR  $10^{12}$  bytes*

1 Terabyte: 50000 trees made into paper and printed.

2 Terabytes: An academic research library.

10 Terabytes: The print collections of the U.S. Library of Congress.

400 Terabytes: National Climatic Data Center database.

### **Petabyte (PB)**

*1,000,000,000,000,000 bytes OR  $10^{15}$  bytes*

1 Petabyte: 3 years of EOS data (2001).

2 Petabytes: All U.S. academic research libraries.

20 Petabytes: Production of hard-disk drives in 1995.

200 Petabytes: All printed material.

### **Exabyte (EB)**

*1,000,000,000,000,000,000 bytes OR  $10^{18}$  bytes*

2 Exabytes: Total volume of information generated in 1999.

5 Exabytes: All words ever spoken by human beings.

---

Source: [www.sims.berkeley.edu/research/projects/how-much-info-2003/](http://www.sims.berkeley.edu/research/projects/how-much-info-2003/)

The phenomenon, "Information explosion" is a consequence of industrial revolution, and the advent of "Industrial revolution" was signaled with the invention of printing. The first milestone of information explosion was probably laid down when the broad discipline of natural science split up in physics, chemistry and biology in nineteenth century. Derek J. de Solla Price measured the exponential growth, doubling every 15 years, of science and technology by counting the annual production of science papers. He noted that aim of the primary journal, which was developed in 1665 and abstract journals that were introduced around 1830, was to enable researchers to keep up with the findings of other research workers.

Conyers Herring remarked in 1968 that the information explosion sparks a need for creative synthesis of facts and ideas (Herring, C). For efficient access to good science literature new schemes for compression of subject headings should be devised. The researchers are increasingly dependent on review materials to evaluate prior research and thereby plan their own work. Much research work was compromised by an overload of information that is more widely scattered than ever. Half the articles he sampled on solid-state physics reported research that he found to be "trivial," "outdated," duplicative, or "wrong" five years after publication. As time passes nearly all regress in value. The papers of negative value become less harmful as positive papers become outdated (Herring, C).

The essential dismay emerging out of this phenomenon of information explosion includes the shielding of necessary information from the information seekers. Information shielding tends to information loss and information pollution at the same time. Information explosion causes growth of irrelevant information which in turn creates loss of relevant information due to information overload, which again creates information fatigueness over the brain and mind. Too much information brings mental exhaustion, and affects imagination to reduce creativity. David Lewis of the International Stress Management Association coined the phrase "Information fatigue syndrome" (<http://www.gdrc.org/icts/i-overload/infoload.html>). The symptoms include paralysis of analytical capacity,

increased anxiety, greater self-doubt, and a tendency to blame others. When people are faced with more information than they can process, they become unable to make decisions or take action. There are several aspects of this problem.

- One is information overload due to overwhelming complexity of the situation. Typical examples include network troubleshooting, complex programming and/or networking problem when complexity of the stuff is above human capacity to comprehend.
- The second situation is "junk" information overload when civilization produces more information than necessary for normal functioning, with most information of low quality. This is kind of new type of pollution, *information smog*.
- What started out as a liberating stream during the Renaissance has turned into a deluge of chaos. In the USA, for example, there are ten thousand of newspapers and magazines. There are also more than 50,000 new book titles published every year and just for the record, over 60 billion pieces of advertising junk mail come into our mail boxes every year. Everything from telegraphy and photography in the 19th century to the silicon chip in the twentieth has amplified the stream of information, until matters have reached such proportions today that for the average person, information no longer has any relation to the solution of problems (Source: <http://www.gdrc.org/icts/i-overload/infoload.html>).

The barrage of data, which we are constantly exposed to, carries a cost, both physically and mentally. David Shenk cites psychological studies spanning thirty years and lists several of the symptoms which accompany information overload (Source: <http://www.gdrc.org/icts/i-overload/infoload.html>):

- Increased cardiovascular stress, due to a rise in blood pressure,
- Weakened vision, citing a Japanese study which predicts a nearly universal near-sightedness in the close future,

- Confusion and frustration,
- Impaired judgement based upon overconfidence,

It is not enough to flee from the problem. Researchers at Israel's Tel Aviv University studied 76 electronics industry clerks and found that the sense of relaxation and happiness derived from a vacation began to fade only three days after returning to work. Pre-vacation levels of stress and burn-out returned within a mere three weeks (*Journal of Applied Psychology*, August 1997). For those who remain "in touch" via pagers, cell phones, laptops, radio, and TV, the results will be even more dismal. On a society-wide level, the dangers of information overload are enormous. The engendered feelings of helplessness, confusion, and anger will erode work efficiency, family functioning, and most likely increase crime rates. We will lack the information-processing skills needed to elect responsible leaders or counter the myriad waves of propaganda pushing our money this way and that. The effective ways of information processing and organisation only can make a control over such chaotic situation. Proper information control should be the prime function in such situation. The possible steps towards an effective information control may include following steps (Source: *Fighting data asphyxiation is difficult but possible ...* by William Van Winkle; Collected from a website: <http://www.gdrc.org/icts/i-overload/infoload.html>):

#### 1. *At the office*

- Be careful with your phone time. Leave a short, efficient message which indicates precisely what action you want to be taken.
- Reduce paper. Either use and file it or toss it in the recycle bin. There is no paper involved, and the "delete" key really can be your best friend.
- Get organized. Corel CENTRAL or Microsoft's Outlook are good examples of utilities which will structure your time.
- Keep meetings short, sweet, and focused.

#### 2. *At home*

- Reduce television time. Television programmes are mostly fluffing designed more to sell commercials than to educate the public. Commercials now account for half of each hour's broadcast.
- Keep your phone number unlisted to reduce solicitation calls.
- Prioritize your phone time. It's taken years, but friends and family have learned to call me with planned discussion items and then not take it personally when I shove the call to a conclusion.
- Develop a hobby. Many of us feel that we don't have the time or talent for a hobby, or may be that was something our parents did -- and God knows we don't want to be like them. A hobby, however, besides having its own inherent rewards (not to mention a second possible source of income) will take time from otherwise wasteful brain drains like TV.

### 3. *On the Net*

- E-mail can be a virus in its own right. Only drop your address when essential, because software robots will see it and automatically add you to marketing lists.
- Newsgroups can consume your life. Be aware of them.
- Remember the library when doing research; you may save innumerable hours forsaking the Web altogether and logging into your local library's server.
- Use your printer. When you find information that you need, print it. This saves both on reading time and the need to find the page again later. (Source: <http://www.gdrc.org/icts/i-overload/infoload.html>)

Besides, another major cause of information overload is the ignorance about information need. The majority of information seekers don't know what they actually want, or what their proper need is. On the other hand, a large number of information seekers can't express their need through a suitable keyword, or keyphrase. The choice of most suitable keyword is the major criteria to increase

precession of information retrieval results and to reduce junk information minimizing information overload simultaneously. This work suggests a model for selection of most appropriate keyword from the large ocean of information of a subject.

## Subject: Fundamental Concept

A subject can be represented by one or a set of keywords. The keyword comprises the fundamental building block of a subject. A set of keywords describes the content of a subject. The keyword provides the content-information of the subject. A keyword may thus be looked upon as the fundamental quanta of information of a subject. The selection of appropriate keywords from a subject is a logical function. The first and fundamental assumption of this study is to describe the subject as a set of keywords. As the keywords are representatives of contents, the basic assumption is compatible with the classical picture of a subject, i.e. a subject is a set of well-defined inter-related contents. Two contents 'A' and 'B' are said to be inter-related, if there exists a subject-domain 'D' such that 'D' is classified both under 'A' (or one of its descendants) and under 'B' (or one of its descendants). This is second fundamental assumption of the study.

The word "subject" has several meanings. The synonyms or near synonyms for the word "Subject" used in the literature are, among many others, "aboutness", "content", "theme" and "topic". It is in itself controversial which words are synonym or near synonyms. The subject of, for example, a document is often evident in a way that even makes it difficult to rise as a problem or to define. Ranganathan (1963, p. 27) for example, declared "subject" an "assumed term" implying that its definition or analysis is not possible or necessary. However the meaning of "subject" is important in indexing and classification of documents and in information retrieval. If subject descriptions of the same document differ or if different kinds of retrieval systems selects different sets of

documents, what exactly then is it that the subject analysis of documents is to identify? What is the subject of a document?

From an epistemological point of view the important question is: Is the subject of a document something subjective or objective? Is it something inherent in the documents or something that the indexer produces from an interpretation of the document? What is the ontological nature of subjects? In both cases we need an understanding of how subjects should be determined in order to produce fruitful subject analysis and document representations. A deeper understanding of this issue is extremely important for all theories and practices of knowledge organization as well as for information retrieval.

Hjørland (1992) demonstrated that implications of different kinds of indexing and classification systems (manual as well as automatic) are based on quite different understanding of the ontological nature of subjects. Systems such as facet analysis, bibliometric coupling, vector space models, user based indexing and so on are based on different implicit notions of “the subject of a document”. Such systems can only be compared if the concept of subject has been properly defined. Contributors to the theory of subject analysis include Cutter (1904), Drake (1960), Wilson (1968), Hutchins (1975, 1977, and 1978), Maron (1977), Miksa (1983), Soergel (1985), Hjørland (1992, 1997, 2001 and 2002) and Molina (1994). The view proposed by Hjørland emphasizes that subject analysis is always done from a given perspective and purpose. The goal of subject analysis is to support some activities of users, which are defined by the (explicit or implicit) purpose of the information service which undertake the subject analysis. Thus two different services, say a physical science database and a public library need different kinds of documents and different kinds of descriptions and subject analysis. When doing subject analysis information professionals rely on conceptions both in the document (derived indexing) and from elsewhere (assigned indexing). Those conceptions may be more or less in harmony or in conflict with the purpose of the indexing service. A document may be wrong

about its own subject, for example by claiming a broader coverage than actually provided. Any document has given potentials in relation to user's task and the purpose of the information service. Whether or not users may recognize those potentials it is the job of the subject analyzing process to describe those potentials as realistic as possible. This may be done, for example, by including methodological and epistemological issues in the analysis. Users may just ask for an effective cure against cancer. Many different documents claim that they describe such a cure (or at least that they apply the proper methods for finding one). The indexing must make discriminations between different bases for claims. Subject analysis is thus not user-oriented but rather "task-oriented". In the ISO-standard for topic maps the concept of subject is defined in this way: "Subject is anything whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever." This definition may work well with the closed system of concepts provided by the topic maps standard. In broader contexts, however, it is not fruitful because it does not contain any specification of what to identify in a document or in a discourse when ascribing subject identification terms or symbols to it. If different methods of subject analysis imply different results, which of these results can then be said to reflect the (true) subject? Different persons may have different opinions about what the subject of a specific document is. How can a theoretical understanding of the term "subject" be helpful deciding principles of subject analysis?

Metcalfe (1973) provides an overview of the history of the concept in libraries for almost hundred years. Metcalfe concluded, the subject of a document often seems so obvious, that it is hard to imagine alternatives or to understand that deep theoretical problems should be or could be involved. However, the most important thing to realize is probably that different persons may have good reasons to ascribe different subjects to the same document that it is illusory to speak of the one true subject of a document disregarding the situation and the purpose of the describing activity. It is thus better to say anything whatsoever

may be ascribed a subject by somebody for some purpose (concerning non-documents). If considered this way then the subject is something that is ascribed to documents or to other objects, but not something with an independent existence beyond this ascribing activity. But then what is it that is being ascribed? And that obvious question still remains, "What is a subject?"

Miksa (1983) discussed Charles A. Cutter's concept of subject (Frohmann, 1994, p. 112-113). According to Cutter the stability of subjects depend on a social process in which their meaning is stabilized in a name or a designation. A subject "referred . . . to those intellections . . . that had received a name that itself represented a distinct consensus in usage" (Miksa, 1983, p. 60) and: the "systematic structure of established subjects" is "resident in the public realm" (Miksa, 1983, p. 69); "subjects are by their very nature locations in a classificatory structure of publicly accumulated knowledge (Miksa, 1983, p.61). Bernd Frohmann adds:

"The stability of the public realm in turn relies upon natural and objective mental structures which, with proper education, govern a natural progression from particular to general concepts. Since for Cutter, mind, society, and SKO [Systems of Knowledge Organization] stand one behind the other, each supporting each, all manifesting the same structure, his discursive construction of subjects invites connections with discourses of mind, education, and society. The DDC [\*Dewey Decimal Classification], by contrast, severs those connections. \*Dewey emphasized more than once that his system maps no structure beyond its own; there is neither a "transcendental deduction" of its categories nor any reference to Cutter's objective structure of social consensus. It is content-free: Dewey disdained any philosophical excogitation of the meaning of his class symbols, leaving the job of finding verbal equivalents to others. His innovation and the essence of the system lay in the notation. The DDC is a poorly semiotic system of expanding nests of ten digits, lacking any \*referent beyond itself. In it, a subject is wholly constituted in terms of its position in the system

[Emphasis added, BH]. The essential characteristic of a subject is a class symbol which refers only to other symbols. Its verbal equivalent is accidental, a merely pragmatic characteristic...

....

The conflict of interpretations over "subjects" became explicit in the battles between "bibliography" (an approach to subjects having much in common with Cutter's) and Dewey's "close classification". William Fletcher spoke for the scholarly bibliographer.... Fletcher's "subjects", like Cutter's, referred to the categories of a fantasized, stable social order, whereas Dewey's subjects were elements of a semiological system of standardized, techno-bureaucratic administrative software for the library in its corporate, rather than high culture, incarnation". (Frohmann, 1994, 112-113)

Cutter's early view on what a subject is, is probably wiser than most understandings that dominated the 20th century, and also the understanding reflected in the ISO-standard quoted above. The early statements quoted by Frohmann indicate that subjects are somehow shaped in social processes. When that is said, it should be added that they are not particularly detailed or clear. We only get a vague idea of the social nature of subjects.

A system, which has an explicit theoretical foundation, is Ranganathan's Colon Classification. As far as known Ranganathan is the only researcher who have earlier given an explicit definition of the concept of "subject":

"Subject - an organized body of ideas, whose extension and intension are likely to fall coherently within the field of interests and comfortably within the intellectual competence and the field of inevitable specialization of a normal person". (Ranganathan, 1967, p. 82).

Another definition is given by Gopinath:

"A subject is an organized and systematized body of ideas. It may consist of one idea or a combination of several..." (Gopinath, 1976, p. 51)".

Ranganathan's definition of "subject" is based on Colon Classification system. The Colon system is based on the combination of single elements from facets to subject designation. This aspect of the theory has been discussed by Metcalfe (1973, p. 318). In psychology and philosophy, the concept "subject" has been used in older literature, but it is almost absent in recent literature. "Subject" has, for example, been used in the sense of "intentional object" in phenomenology. The concept is also used in linguistics. A subject (as noun) can thus only be expressed using familiar nouns or general terms, not indefinite specific terms. Some information is turned to subjects; other information is placed in focus. The subject of a sentence is not identical with the content of that sentence. For instance, the sentence: "The colour of Indian flag is saffron, white and green" has subjects, for example, the Indian flag, colors and the colours of Indian flag. Its content is, however, that Indian flag's colours are saffron, white and green. The subject is thus a categorical determination of content.

The concept of subject in Library and Information Science was given by Patrick Wilson. Wilson (1968) examined by thought experiment the suitability of different methods of examining the subject of a document. The methods described are:

1. To identify the author's purpose for writing the document
2. To assign weightage to the relative dominance and subordination of different elements in the picture, which the reading imposes on the reader.
3. To group or count the documents used of concepts and references
4. To deduce a set of rules for selecting the elements which are necessary as opposed to unnecessary for the work as a whole

Patrick Wilson concluded that each of these methods is insufficient to determine the subject of a document and remarked: "The notion of the subject of a writing is indeterminate" (Wilson, P; p. 89). He also said that authors of documents often use terms in ambiguous ways. Even if the librarian could personally develop a

very precise understanding of a concept, he would be unable to use it in his classification, because none of the documents use the term in the same precise way. Based on this argumentation Wilson made conclusion: "If people write on what are for them ill-defined phenomena, a correct description of their subjects must reflect the ill-definedness".

Wilson's concept of subject was discussed by Hjørland (1992) who found that it is problematic to give up the precise understanding of such a basic term in Library and Information Science. Wilson's arguments led him to an agnostic position which Hjørland found unacceptable and unnecessary. Concerning the authors' use of ambiguous terms, the role of the subject analysis is to determine which documents would be useful for users to identify whether or not the documents use one or another term or whether a given term in a document is used in one or another meaning. Clear and relevant concepts and distinctions in classification systems and controlled vocabularies may be fruitful even if they are applied to documents with ambiguous terminology.

A number of researchers in LIS have tried to escape the difficulties as to the concept of subject by preferring to use the concept "aboutness" as an alternative. A justification for this decision is given by Hutchins (1975, p. 115):

"From this account of indexing one thing should now be clear, namely, that the notion of the "subject" of a document is peculiarly vague. We may mean the "extensional aboutness" or the "Intentional aboutness", as given by the author in his title or as given by the abstractor or by the indexer; we may mean the NL [natural language; BH] phrase expressing the Topic or we may mean the DL [documentary language; BH] expression denoting the document content. There are clearly so many variables involved that whenever we talk of the "subject" of a document we ought always to say what kind of subject we are intending.

.....

As we have seen, judgments of subject content (by authors, readers and indexers) are influenced by so many factors that any particular statement of a document's content should never be regarded as anything other than just one of many possible such statements. In other contexts and from other perspectives the same document may have other, quite different "subjects."

The concept of aboutness is thus introduced in order to solve the problem relating to the concept of "subject". Hjørland (1992, 1997) found that any practice of subject determination as well as any theory of subject analysis is necessarily based on epistemological views. Those views are, however, seldom explicit, and often unknown because of lack of epistemological knowledge in Library and Information Science. Each approach to subject analysis and information retrieval is more or less based on specific epistemological assumptions. Facet analysis, IR-approaches, user-oriented approaches, bibliometric approaches etc. are basically related to different epistemological views which implies different conceptions of what subjects are. Based on this analysis, Hjørland (1992) developed a new understanding of subjects as "informative potentials" (first formulated as "epistemological potentials"), i.e. the subjects of a document are its informative potentials. The basic idea is simple to explain. Rather than seeking the subject of a document, for example, in some inherent objectives and facts about that document, the indexer should ask: "What is this document useful for"? In other words, the subject assignment is seen as a human act, which aims at supporting some activities of the users (comparable with request oriented indexing). The subject determination that is most successful in accomplishing this goal is the most correct one. Consequently subject determinations are situational and context-dependent. The subject of a document is also theory-dependent. Just as one could not describe the potentials of uranium as an energy source before the development of physical theories of radioactivity, the potentials of documents are changing when theories change. This is best understood by considering the citation patterns and reception history of documents. Although uranium could not be described as an energy source before the development of

theories on radioactivity, uranium nonetheless contained the potentials all the time. The same is the case with documents. Their potentials may be unrecognized for a long time, but nevertheless they exist.

The concept of *meaning production* in cultures and societies is an important background for a proper understanding of what subjects are. This way we are back to Cutter's view of concept as something shaped by social processes. Stam (2000) is critical about subjects as basis for groupings. However, he was concerned with that aspect of subject matter, which is usually called topic or topicality. He stated that subject matter is the weakest criterion for generic groupings because it fails to take into account how the subject is treated (Stam, 2000, p. 14). There is nothing to prevent the use of, for example, terms related to method and genre as a part of the subject description of a document. The ambiguities in the concept of subject along with different logical aspects were discussed by Hjørland & Nicolaisen (Hjørland, B. & Nicolaisen, J. 2005) from the viewpoint of Bradford's phenomenon of scattering.

Different scientists describe the concept of a subject from different views. No single description leaves any complete picture of the concept of subject, but an overall study of all theoretical formalism draws a comprehensive layout of the criteria. The notable feature is that, in all theories so long discussed the concept of "subject" is based on the epistemological formalism. The subjects are considered as the conceptual entity associated with any discipline.

The present study is based on the assumption that a subject can be represented with a set of keywords. The number of keywords in a set ranges from one to many. Keywords are considered as fundamental building block or quanta of a subject. The keywords arranged in logical order describe the content of a subject with relevance and accuracy, and portray the complete subject from the microscopic point of view.

## Subject Formation

The results of experiment, observation and thinking may reveal themselves in a number of ways, forms and styles; which may be a theory, some process, phenomena, application, or else. The final outcome of an experiment, observation and thinking if forms any logical entity in some well-defined shape in human mind, then a subject is incepted in concept space of human brain. A subject is born usually in the form of a research paper, a short communication, or conference proceedings etc. at a particular point of time. At that time, it is not known whether the subject will grow further or not. If it has a promise more researchers start working on it, and the subject starts growing and the literature on the subject starts appearing in a scattered way in different journals, conference proceedings, and so on. At a latter stage, the scattered literature is gleaned, examined, and a review paper or a book is written where the ideas pertaining to the subject are organized and systematized and in many cases the subject is given a name. Hence in the very beginning, the ideas of a nascent subject is usually not organized or systematized. From the last two definitions it appears that a subject is an entity existing independently of the universe of knowledge.

In view of what has been said above, a new definition of 'subject' is being given here (Sen, 2007). A subject is a segment of the universe of knowledge and possesses all the characteristics a segment possesses. A subject can be composed of single, combined, mixed, overlapping, clustered and other types of segments. A subject is usually identified by a name, a notation or notations, a symbol or symbols, etc. The name can be composed of a single keyword like physics, or a set of keywords like India: History: British period. The name of a subject sometimes undergoes change due to various reasons. For example, the name of our profession has evolved like this: Library economy ---- Librarianship -- -- Library science ---- Library and information science. Now, in many cases, it is referred only as Information science/s. With the advent of space age the subject Aeronautical engineering has changed as Aerospace engineering.

In the context of library and information science, the organisation of knowledge is the organisation of documented messages in which knowledge or information is represented. Knowledge is accumulated centering a particular entity that is the nucleus of a subject. An organized set of ideas may or may not represent an individual or a cluster of subjects. Hence, a continuum or universe of knowledge needs to be divided in different segments, which is the process of knowledge classification. This process creates different subjects. According to *The Concise Oxford Dictionary* a subject is 'theme of or of discussion or description or representation, matter (to be treated of or dealt with)'. An organized set of well-defined ideas and concepts form a subject. As defined by Neelameghan "A subject is an organized or systematic account of an idea or body of ideas whose extension and intension are likely to fall within the intellectual competence and field of inevitable specialization of a normal person." (Chakrabarti, B). Different specialists have recognized different types of relations between the components of a subject or the modes of subject formation in the universe of knowledge. According to Ranganathan subjects in the universe of knowledge could be formed by five methods: Loose Assemblage, Lamination, Dissection, Denudation and Superimposition. Later his school of thought developed this view point. In 1975, M.A.Gopinath and S.Seetharama put forward the following modes of formation of subjects: Loose Assemblage, Lamination, Fission, Fusion, Distillation, Agglomeration and Cluster (Chakrabarti, B).

## Subject Proliferation

The phenomenon of subject proliferation is a direct consequence of information explosion and information overload, as well. The milestone of this phenomenon was laid in nineteenth century when natural philosophy split into three components, physics, chemistry and biology. As information on a subject increases, the subject gradually gets overloaded with information, and starts

dividing into several components; while each component appears as a separate subject-entity. Ranganathan described various ways in which the universe of subjects going with a basic subject can get proliferated. For example, a complex subject in that universe of subjects may be formed by attaching some other basic subject as phase or by coupling two or more distinct subjects on the basis of some relationship between them. A compound subject may be formed by attaching one or more isolate ideas to the basic subject. Such isolate ideas can themselves proliferate by various ways. The compound subject is composed of basic subject and isolates drawn from two or more facets of that subject. The proliferation results acute documentary chaos as described by Hood and Wilson (Hood, W and Wilson, C.S. 2001). The documents are getting scattered over databases belonging to different subject domains. The number of databases is, in turn increasing, and bibliographic control of databases is crucially required for retrieval of information.

## Subject: Fundamental Components

A subject is an ever-changing entity. The changes are governed by many factors, viz. 1) Time 2) Interaction with other subjects and 3) Applicability in real life. The dynamic locus of a subject is best traced out by the R & D works executed on that area, the information of which are available from the research papers published in various scholarly journals on the concerned area. Hence, the trends of growth and development of a subject are clearly reflected from the journal articles. The subject of documents however may contain any number of facets. If it contains only one facet then it must be a basic subject such as physics, chemistry. They may be called simple subjects. They imply that documents deal with the whole field named and does not identify any particular phenomenon for attention. A subject has following fundamental components:

- Content: actual matter dealt therein
- Context: relevant framework in which the subject is discussed

- Concept: the main theme playing behind the subject

All these three components affected by above-mentioned three factors, i.e. time, interaction with other subjects and applicability in real life. The present study describes these components with the aid of keywords and the effects of changes are also studied.

### Subject: Principal Characteristics

In this study the following characteristics of a subject are investigated and analyzed with the aid of keywords:

- Growth of literature and content
- Obsolescence of literature
- Stability of research
- Strength of research
- Thrust areas of research
- Potential keywords for searching and retrieval with higher precision
- Disciplinarity of the subject (whether inter, single or multi-disciplinary)

### Subject: Research Trend

Research is an essential function for human development in all-possible orientations. Research is the dynamic indicator of a subject, as large number of research projects comes from a dynamic subject only. The research trend of some specific subjects is analyzed in this study with the help of keywords. The research trend also reflects the latest state-of-the-art of a subject.

## Subject Analysis: Theoretical Approach

Subject analysis of an item is done for classification and generating subject headings. Classification serves the general purpose of mapping a general area and/or keeping documents on subject together in the collection. Subject headings serve the same purpose, except that they keep entries providing access to the items together in an alphabetical arrangement. The subject heading serves another important purpose, which is the retrieval function for common users. The first step in subject analysis of an item is to eliminate peripheral material from consideration. Subject analysis of an item should be as specific as the work warrants. Specificity of treatment does not depend on the holdings of material in a particular area. Generally, the classification for an item should not be narrower than the primary subject heading. Subject analysis of a work is made based on usefulness and content, not aesthetics. For materials where judgment of subject content is especially difficult, publisher's special emphasis is given in classification. If the author's or publisher's statement of the purpose or contents of an item is actually misleading, the statement of content (title, etc.) may be ignored, and the work classed without regard to the stated publisher's intent. Each and every well-defined entity belonging to a subject must have a subject heading. An entity if found absent from the concerned subject area receives no subject headings. There is no limit imposed on the number of subject headings that can be given to a particular item for full-fledged cataloging, provided they are properly applied. A general entity may receive a specific subject heading when discussed in the context of some other specific subject. For instance, 'Liquid' is a general entity, but it receives subject headings like Fermi liquid or Luttinger liquid when discussed in the context of Condensed matter physics. On the other hand, a broad subject receives the subject heading that corresponds to the content of the item as a whole. Only rarely one component of the broad heading is singled out for tracing. All subject headings traced must be justified by something that appears in the record. The justification may appear in the call number, title, series, notes, etc. Three major steps of subject analysis are:

- Determining the focus of the intellectual content of a work—“what is this item about?”
- Translating what it is about into the framework of the classification and subject heading systems used
- Then finally translating the concept into a classification number and string of headings

## Subject Analysis: Tools and Techniques

The basic tools for subject analysis are as follows:

- Subject heading lists
- Thesauri
- Classification schemes

The major subject analysis methods are as follows:

### ➤ **Purposive method**

This is an attempt to understand what the author is attempting to describe, prove, show, narrate, explain, question or report. There may be a statement of purpose. To try to determine what objective(s) the author aimed for.

### ➤ **Figure-ground method**

This method includes determination of the core theme that stands out from the background detail. From this impression of dominance or by referring to what stands out one can determine the subject

Disadvantage—what stands out for one person doesn't mean it stands out for another

### ➤ **Objective method**

This method includes counting references to names or topics, including instances where the name or topic is not explicitly noted.

Disadvantage—sometimes referenced items reflect background rather than central topics

➤ **Appeal to unity or to rules of selection and rejection**

This method includes determination of what unity and completeness are there that makes the work cohesive

What is said (selection) and not said (rejection)?

Disadvantage—varies from person to person, and one must know a lot about the topic to make the determination with confidence.

Once the main subject of a work is chosen, then the subject must be translated into the conceptual framework of the classification system after determining disciplines, subdisciplines, subtopics, place, time and form (is it fiction or a dictionary?). The subject is broken into components from general to specific and finally the class number from the classification scheme is assigned. After that, the subject headings from the subject heading lists (Sears, MeSH, and LCSH) are determined. A subject heading is an access point to a bibliographic record. A subject heading consists of a word or a phrase which designates the subject of the work or works contained in the bibliographic record. Subject headings identify pertinent material on a given subject. Subject headings also enable the user to find material on related subjects.

## Subject Indexing

The two principal steps involved in subject indexing are conceptual analysis and translation. These are intellectually quite separate, although they are not always clearly distinguished and may actually occur simultaneously. Conceptual analysis involves deciding what a document is about, i.e. what it covers. Effective subject indexing involves deciding not only, what a document is about but also why it is likely to be of interest to a particular group of users. In other words, there is no one unique set of index terms for any item. The same publication could be indexed rather differently in different information centres and should be indexed

differently if the groups of users are interested in the item for different reasons. The indexer, then, must ask several questions about an item:

- What is it about?
- Why has it been added to our collection?
- What aspects our users are interested in?

Beghtol (1990, 1994, 1995, and 2001) and Hutchins (1975 and 1977) both draw upon text linguistics in discussing the subject. Maron (1977) adopts a probabilistic approach, and Swift et al., (1978) pointed out that “aboutness” in indexing may not coincide with the “aboutness” that searchers of information are concerned with. Bernier (1980) discussed on major steps that should be remembered for subject indexing. Frohmann (1990) has used the work of Wittgenstein for criticizing some approaches to indexing theory. Wilson (1968) goes so far as to imply that subject indexing faces “intractable” problems because it is so difficult to decide what a document is about. The other word is “Concept” that some scientists like to philosophize around. Conceptual analysis means nothing more than identifying the topics discussed in a document. Preschel (1972) takes “Concept” to mean “Indexable matter” and defines “Conceptual analysis” as “Indexer perception of indexable matter”. The process of translation involves the representation of the conceptual analysis by means of a term or terms drawn from some vocabulary.

## Keyword Analysis

All research articles begin with a title. Most include an abstract. Several include ‘Keywords’. All three of these features describe an article’s contents with varying degrees of detail and abstraction. The title is designed to stimulate the reader’s interest. The abstract summarizes the content. The keywords indicate main concepts and fields of concern. The keywords are also known as ‘Descriptors’ or ‘Subject headings’; the terminology ‘Keyword’ appears in common usage, while ‘Descriptor’ should strictly speaking be called ‘Keywords and phrases’, since multiword phrases can be used as descriptors in most publications. The five

essential parameters of an article, viz, title, author, abstract, keywords and location are required in any bibliographic database designed to aid electronic information retrieval. The earliest use of keywords was found in 1975 in the *Journal of Applied Behaviour Analysis*.

Keywords allow readers to decide whether or not an article contains material relevant to their interest. Keywords provide readers with suitable terms to use in web-based searches to locate other materials on similar topics. Keywords also enable indexers and editors to group together related materials. Any subject, whether very specific or broad one, always undergoes through incessant changes, which are very picturesquely reflected by the keywords pertaining to the concerned field. Keywords indicate growth of a subject in different orientations but the notable feature is that selection of keywords is a vital aspect. If the keywords are not selected precisely, they get scattered over irrelevant peripheral areas then they will mislead both the information professional and clientele. Keywords occur within titles, abstracts, series names, content notes, subject headings and index terms supplied by the authors. The keywords may also be used as effective subject access points. It is to be noted that every word in titles, abstract or content-notes are not keywords. Articles, prepositions and conjunctions are too common to serve as keywords.

## References

Beghtol, Clare. 1990. Access to Fiction: A Problem in Classification Theory and Practice: Part 1. *International Classification* 17 (1): 21-27.

Beghtol, Clare. 1994. *The Classification of Fiction: The Development of a System Based on Theoretical Principles*. New Jersey: The Scarecrow Press.

Beghtol, Clare. 1995. Domain Analysis, Literary Warrant, and Consensus: The Case of Fiction Studies. *Journal of the American Society for Information Science* 46: 30-44.

- Beghtol, Clare. 2001. The Concept of Genre and its Characteristics. *Bulletin of the America Society for Information Science and Technology* 27 (2): 1-5.
- Bernier, C.L. 1980. Subject Indexes. *Encyclopedia of Library and Information Science*: Kent, A.; Lancour, H. & Daily, J.E. (ed.). Volume 29. New York: Marcel Dekker Inc.: 191-205.
- Chakrabarti, B. 1994. *Library Classification Theory*. Calcutta: World Press.
- Cutter, C.A. 1904. *Rules for a Dictionary Catalog*. Washington, DC: Government Printing Office.
- Drake, C. L. 1960. What is a subject? *Australian Library Journal* 9: 34-41.
- Frohmann, B. 1990. Rules of indexing: A critique of mentalism in information retrieval theory. *Journal of Documentation* 46 (2): 81-101.
- Frohmann, B. 1994. The Social Construction of Knowledge Organization: The Case of Melvin Dewey. *Advances in Knowledge Organization* 4: 109-117.
- Gopinath, M. A. 1976. *Colon Classification*. (I: Classification in the 1970's. A second look. Revised edition. Ed. by Arthur Maltby. London, Clive Bingly, 51-80).
- Herring, C. 1968. Distill or drown: the need for reviews. *Physics Today* 21 (9): 27-33.
- Hjørland, B. & Sejer Christensen, F. 2002. Work tasks and socio-cognitive relevance: A specific example. *Journal of the American Society for Information Science and Technology* 53 (11): 960-965.
- Hjørland, B. 1992. The concept of "subject" in Information Science. *Journal of Documentation* 48 (2): 172-200.

Hjørland, B. 1997: *Information Seeking and Subject Representation. An Activity-theoretical approach to Information Science*. Westport, CT & London, UK: Greenwood Press.

Hjørland, B. 2001. Towards a theory of aboutness, subject, topicality, theme, domain, field, content . . . and relevance. *Journal of the American Society for Information Science and Technology* 52 (9): 774–778.

Hjørland, B. 2002. Epistemology and the socio-cognitive perspective in Information Science. *Journal of the American Society for Information Science and Technology* 53 (4): 257-270.

Hjørland, B. & Nicolaisen, J. 2005. Bradford's Law of Scattering: Ambiguities in the Concept of "Subject". IN: Crestani, F. & Ruthven, I. (Eds.): CoLIS 2005, LNCS 3507, pp. 96 – 106. Berlin: Springer-Verlag. [http://www.db.dk/bh/Bradford\\_Colis5.pdf](http://www.db.dk/bh/Bradford_Colis5.pdf)

Hood, W. & Wilson, C.S. 2001. The scatter of documents over databases in different subject domains: How many databases are needed? *Journal of the American Society for Information Science and Technology* 52 (14): 1242-1254.

Hutchins, W. J. 1975. *Languages of Indexing and Classification: A Linguistic Study of Structures and Functions*. London, UK: Peter Peregrinus.

Hutchins, W.J. 1977. On the problem of “aboutness” in document analysis. *Journal of Informatics* 1: 17-35.

Hutchins, W.J. 1978. The concept of “aboutness” in subject indexing. *Aslib Proceedings* 30: 172-181.

Maron, M. E. 1977. On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science* 28: 38-43.

Metcalf, J. 1973. When is a Subject not a Subject? IN: *Towards a theory of Librarianship*. Ed. by Conrad H. Rawski. New York: Scarecrow Press.

Miksa, F. 1983. *The Subject in the Dictionary Catalog from Cutter to the Present*. Chicago, IL: American Library Association.

Miksa, F. 1998. *The DDC, the Universe of Knowledge, and the Post-Modern Library*. Albany, NY: Forest Press.

Molina, M. P. 1994. Interdisciplinary Approaches to the Concept and Practice of Written Text Documentary Content Analysis. *Journal of Documentation* 50 (2): 111-133.

Preschel, B.M. 1972. *Indexer consistency in perception of concept and in choice of terminology*. New York: Columbia University.

Ranganathan, S. R. 1963. *Documentation and its facets*. London: Asian Publishing House.

Ranganathan, S. R. 1967. *Prolegomena to Library Classification*. London: Asia Publishing House.

Sen, B.K. 2007. *Universe of Knowledge: Some Random Thoughts*. (To be published)

Soergel, D. 1985. *Organizing Information: Principles of Data Base and Retrieval Systems*. Orlando, FL: Academic Press.

Stam, R. 2000. *Film Theory: An Introduction*. Oxford: Blackwell.

Swift, D. F., Winn, V. & Bramer, D. 1978. "Aboutness" as a strategy for retrieval in the social sciences. *Aslib Proceedings* 30: 182-187.

Wilson, P. 1968. *Two Kinds of Power. An Essay on Bibliographical Control*. Berkeley: University of California Press.