

## An analytical model to interpret continuous expansion of universe of subjects

Bidyarthi Dutta<sup>1</sup> and B.K. Sen<sup>2</sup>

<sup>1</sup>Librarian, St. Xavier's College, 30, Park Street, Kolkata 700016, Email: bidyarthi.bhaswati@gmail.com

<sup>2</sup>80, Shivalik Apartments, Alaknanda, New Delhi 110019, Email: bkzen@airtelmail.in

Tries to find the pattern of change of keywords of an arbitrary science subject with its age. The mode of temporal evolution of a subject is formulated on the basis of some fundamental assumptions. It has been mathematically derived that the number of keywords of a subject will always tend to attain a saturation value with the age of the subject. When a subject gets overloaded with keywords, then it tries to share the same with some peripheral subject areas and consequently new subjects are formed.

### Introduction

Different branches of sciences are involved in the study of the physical world and its phenomena. A branch of science involved in studying phenomena or laws of the physical world, is commonly known as natural science, which is a general term of physics, chemistry, biology, astronomy, and so on. In science, the term natural science thus refers to a naturalistic approach to the study of the universe, which is understood as obeying rules or laws of natural origin. Natural sciences basically deal with nature and/or natural phenomena. The term natural science is also used to distinguish those fields which use the scientific method to study human behavior and society. In ancient and medieval times, the objective study of nature was known as natural philosophy. In late medieval and early modern times, a philosophical interpretation of nature was gradually replaced by a scientific approach using inductive methodology. The works of Ibn al-Haytham and Sir Francis Bacon popularized this approach, thereby helping to forge the scientific revolution<sup>1</sup>. By the 19th century the study of science had come into the purview of professionals and institutions, and in so doing it gradually acquired the more modern name of natural science. The term scientist was coined by William Whewell in an 1834 review of Mary Somerville's *On the Connexion of the Sciences*<sup>2</sup>. However the word did not enter general use until nearly the end of the same century.

Before the inception of experimental science, knowledge was solely based on ideas and concepts

only. The various streams of natural science were basically built upon some fundamental laws or axioms. Probably Archimedes gave birth to the first axiom in natural science around 260 BC. One of such universal axiom was the law of conservation of energy, which states that energy cannot be created or destroyed, but can change its form. The total quantity of matter and energy available in the universe is thus a fixed amount and never more or less.

The fundamental objective of this study is to search out how a subject grows and develops with its age or time. In order to search for an answer of this quest, we have taken help of some fundamental axioms that are discussed in the next section.

### Basic axioms

1. The facts and phenomena involved in any subject are normally recognized as data
2. The data are represented by some keywords, which may be either alphabetic, numeric or alphanumeric
3. A subject is born borrowing the keywords from other subjects. For instance, when the first article on Raman Effect was published, the article contained keywords from physics, chemistry, and so on. Later on the term 'Raman Effect' was coined<sup>3</sup>. Today, if somebody compiles a vocabulary on Raman Effect, then it will be seen that keywords from a number of subjects are figuring there apart from its own. To exemplify this fact mathematically, let us consider a set of

- subjects, where the number of subjects are 5 (say), which are denoted by  $s(1)$ ,  $s(2)$ ,  $s(3)$ ,  $s(4)$  and  $s(5)$ . All these subjects have optimal stock of keywords. At a particular point of time, suppose, another new subject  $s(6)$  is born from these five existing subjects. Now, initially all keywords held by the subject  $s(6)$  are only from the five parent subjects. As time goes on and the age of  $s(6)$  increases, its own keyword stock will gradually develop, and the percentage of keywords from parent subjects will gradually decrease.
4. The primary sources of information will gradually develop from these keywords in course of time
  5. The secondary sources of information will gradually develop from these primary sources in course of time. Thus it can be stated that keywords start to evolve to generate primary literature, and eventually consolidated in secondary literature.
  6. Any subject has a core area, which entirely belongs to the study and research pertains exclusively to that area only. The core area of a pure subject (e.g. Particle physics) is absolutely self-disciplinary, but not inter-disciplinary. Besides this, the subject has an allied zone which is feebly self-disciplinary, but strongly inter-disciplinary. The study and research belonging to the allied zone do not cover the concerned subject solely, but embraces the peripheral subject areas also. Beyond the allied zone, the alien area starts that is almost completely interdisciplinary. The alien area has trifle touch with the core area of the subject
  7. The core area is the own area of the subject concerned, therefore almost no interaction with other subject(s) occur in this zone. The keywords belonging to this core area are known as core keywords of the subject concerned and very feebly interactive with keywords from other subject areas. The core keywords, thus grow with the growth of the subject but eventually reach a saturation due to poorly interactive nature
  8. The allied and alien keywords are strongly interactive with keywords from other areas and continuously grow with the growth of the subject concerned and hardly reach a saturation point on account of strong interactive nature
  9. The core keywords represent the corpus of primary literature on the subject. Eventually they figure in the secondary literature
  10. Since the corpus of core keywords saturates with the age of the subject concerned, therefore the core stage is attained at the last (saturation) phase of the growth of the subject (Figure 1)
  11. The sum of core keywords, core primary and secondary literature at any instant during the last phase (saturation phase) of growth of the subject (Figure 1) remains constant; otherwise saturation would not be attainable.
  12. The keywords from a parent subject gradually decrease in a daughter subject as their ages increase. The daughter subject gradually develops its own keyword stock with the age
  13. The quantum of primary literature is directly proportional to temporal rate of change of keywords and the age of the subject
  14. The quantum of secondary literature is directly proportional to temporal rate of change of primary literature and the age of the subject
  15. The age of a subject is represented by time function here, because age of anything is a representative of time.
- Subject: some fundamental concepts**
- A subject may be defined in various ways. According to S. R. Ranganathan<sup>4</sup>, a subject is 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". According to M.A. Gopinath<sup>5</sup>, a subject is an organized and systematized body of ideas. It may consist of one idea or a combination of several". According to Neelameghan and Seetharama, 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<sup>6</sup>.
- A subject is something inherent in the documents or corpus of information. Contributors to the theory of

subject analysis include Cutter<sup>7</sup>, Drake<sup>8</sup>, Wilson<sup>9</sup> (1968), Hutchins<sup>10,11,12</sup>, Maron<sup>13</sup>, Miksa<sup>14</sup>, Soergel<sup>15</sup>, Hjørland<sup>16,17,18,19,20</sup> and Molina<sup>21</sup>. 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. 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. 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." 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 in deciding principles of subject analysis? According to Sen<sup>22</sup>, a subject is composed of a segment or segments of the universe of knowledge, which is the totality of knowledge that has survived till date. The knowledge that is being generated now and will be generated in future will also be a part of the universe of knowledge.

All these interpretations of subject are more or less subjective, but not objective. An objective insight of subject interpretation encounters some sort of quantification of a subject through some measurable parameters. At this point a subject may be looked upon as a collection of well-defined keywords, primary literature and secondary literature or a set of keywords, primary literature and secondary literature, which are relevant to the subject concerned<sup>23</sup>. Different characteristics of keywords, primary and secondary literature may be considered as the aforesaid measurable parameters. The growth and development of the subject may be looked upon in terms of the growth and decay of number of keywords. Considering this objective point of view, any subject generally undergoes through different phases during its lifetime as depicted in Figure 1.

In the above Figure 1, X-axis represents age of subject and Y-axis represents volume of subject. The Formation phase is the initial phase, where the core keywords are born and the subject grows gradually.

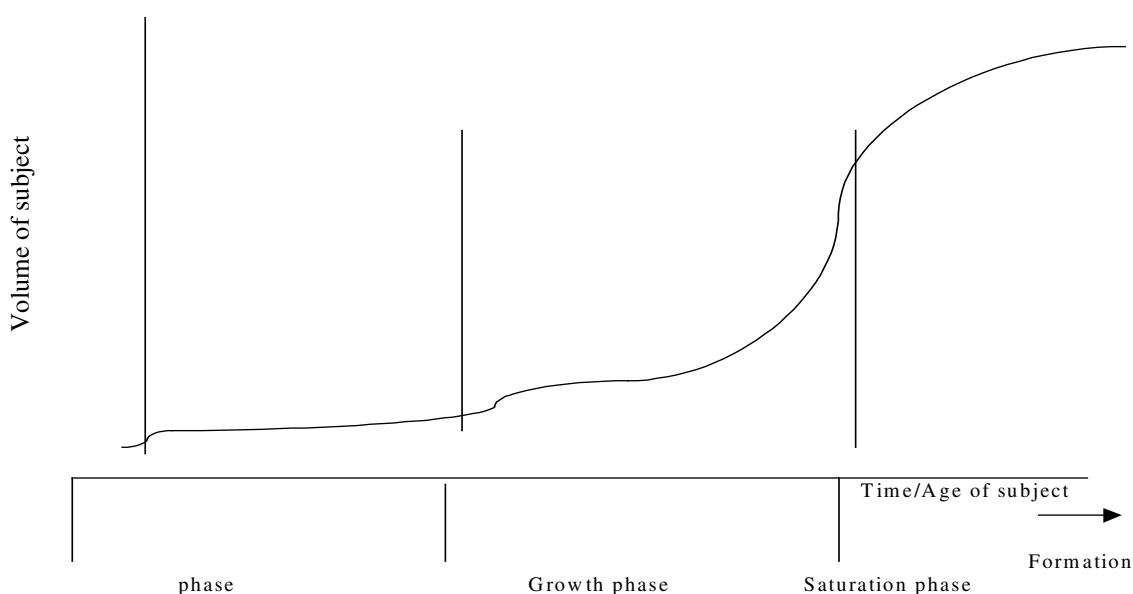


Figure 1—Different phases of a subject

The intermediate phase is the Growth phase, where huge number of core keywords comes out as a result of continuously growing number of research and studies that result information explosion within the domain of the subject, and accelerates growth of the subject in no time. The final phase is the saturation phase, where the number of core keywords starts to be saturated and the subject gradually approaches towards an asymptotic infiltration.

### Notations used

The notations used for the derivation are furnished below (The symbol  $\rightarrow$  represents 'Stands for'):

k	$\rightarrow$	number of core keywords belonging to the concerned subject
p	$\rightarrow$	quantum of core primary literature belonging to the concerned subject
s	$\rightarrow$	quantum of core secondary literature belonging to the concerned subject
t	$\rightarrow$	time
dk/dt	$\rightarrow$	temporal rate of change of core keywords or differential coefficient of k with respect to time 't'
$d^2k/dt^2$	$\rightarrow$	temporal rate of change of dk/dt or differential coefficient of dk/dt with respect to time 't'
dp/dt	$\rightarrow$	temporal rate of change of core primary literature or differential coefficient of p with respect to time 't'

### Analytical Formalism

$$k + p + s = C_0, \quad \dots(1) \text{ (by axiom-11)}$$

where  $C_0$  is a constant

$$\text{Now, } p \propto t \quad \dots (2a) \text{ (by axiom 13)}$$

and

$$p \propto (-dk/dt) \quad \dots (2b) \text{ (by axiom 13)}$$

The minus (-) sign is due to axiom 12.

Here 't' and (dk/dt) provide time and space components respectively for information expansion. The symbol  $\propto$  stands for 'proportional to'.

Combining equations (2a) & (2b), we get equation (2),

$$p = -A*t*(dk/dt) \quad \dots(2)$$

where A is constant of proportionality

$$\text{Also, } s \propto t \quad \dots(3a) \text{ (By axiom 14)}$$

&

$$s \propto (dp/dt) \quad \dots(3b) \text{ (By axiom 14)}$$

Combining equations (3a) & (3b), we get equation (3),

$$s = B*t*(dp/dt) \quad \dots(3)$$

where B is constant of proportionality

Since A and B are constants of proportionality, therefore it is possible to make their values equal to one, after choosing suitable units, i.e. defining 'A' as unit bit of information generated per second for unit temporal rate of processing of data; and defining 'B' as unit bit of knowledge generated per second for unit temporal rate of cognition of information. The equations (2) and (3) thus may be rewritten as:

$$p = -t*(dk/dt) \quad \dots(2a)$$

$$s = t*(dp/dt) \quad \dots(3a)$$

Substituting equation (2a) in equation (3a) we get equation (4),

$$s = -t*(dk/dt) - t^2*(d^2k/dt^2) \quad \dots(4)$$

Substituting equations (2a) and (4) in equation (1), we get equation (5),

$$-t^2*(d^2k/dt^2) - 2*t*(dk/dt) + k = C_0 \quad \dots(5)$$

Let us now find the value of the constant  $C_0$  from initial condition. Say, at an initial time when the quantity of data belonging to the subject under consideration started to be gradually saturated, the value of time t has been arbitrarily taken as zero. Substituting  $t = 0$  in equation (5) we get,

$$C_0 = k_0 \quad \dots(6),$$

where  $k_0$  indicates the value of k at  $t = 0$ , that is to say,  $k_0$  is the number of keywords pertaining to the parent subject when the same started to be saturated.

Substituting equation (6) in equation (5) we get equation (7),

$$-t^2 \cdot (d^2k/dt^2) - 2 \cdot t \cdot (dk/dt) + (k - k_0) = 0 \quad \dots(7)$$

In equation (7), k is a variable, i.e. the number of keywords at any time t, and k<sub>0</sub> indicates the value of k at a particular time t = 0, i.e. when the parent subject starts to be saturated and daughter subject was born. Therefore, k<sub>0</sub> is a constant.

Let us say  $(k - k_0) = k_d \quad \dots(8)$ ,

where k<sub>d</sub> is another variable that indicates the number of keywords at any instant minus the initial number of keywords. Let us say 'k<sub>d</sub>' as *dynamic keywords*. Since k<sub>0</sub> is constant, therefore

$$dk/dt = dk_d/dt \quad \dots(9a)$$

$$\text{and } d^2k/dt^2 = d^2 k_d/dt^2 \quad \dots(9b)$$

Substituting equations (8), (9a) and (9b) in equation (7), we get the equation (9),

$$t^2 \cdot (d^2k_d/dt^2) + 2 \cdot t \cdot (dk_d/dt) - k_d = 0 \quad \dots(9)$$

Equation (9) represents standard form of the second order Euler-Cauchy equation. Let us assume a trial solution of this equation as:

$$k_d = (t)^m \quad \dots(10)$$

Differentiating we get,  $dk_d/dt = m \cdot (t)^{(m-1)} \quad \dots(10a)$

and  $d^2k_d/dt^2 = m \cdot (m-1) \cdot (t)^{(m-2)} \quad \dots(10b)$

where m is constant. Substituting equation (10), (10a) and (10b) in equation (9), we get,

$$t^2 \cdot m \cdot (m-1) \cdot (t)^{(m-2)} + 2 \cdot t \cdot m \cdot (t)^{(m-1)} - (t)^m = 0$$

After rearrangement, we get:

$$m^2 + m - 1 = 0 \quad \dots(11)$$

Equation (11) is a quadratic equation with two distinct roots, viz. 0.62 and -1.62. The solution of equation (9) will thus become as follows:

$$k_d = C_1 \cdot (t)^{0.62} + C_2 \cdot (t)^{-1.62} \quad \dots(12)$$

where, C<sub>1</sub> and C<sub>2</sub> represent constants of integration.

Equation (12) is graphically presented in Figure 2. It is observed that as time or subject-age (t) changes from the value '1' to '300', the stock of dynamic keywords (k<sub>d</sub>) changes from the value '2' to '34' only, which clearly indicates the tendency of saturation of dynamic keywords with the age of the subject or time, whatever it may be. The graphical presentation is made taking the values of the constants C<sub>1</sub> and C<sub>2</sub> arbitrarily as equal to unity. Though it is obvious whatever be their values, the stock of dynamic keywords always tends towards saturation with time.

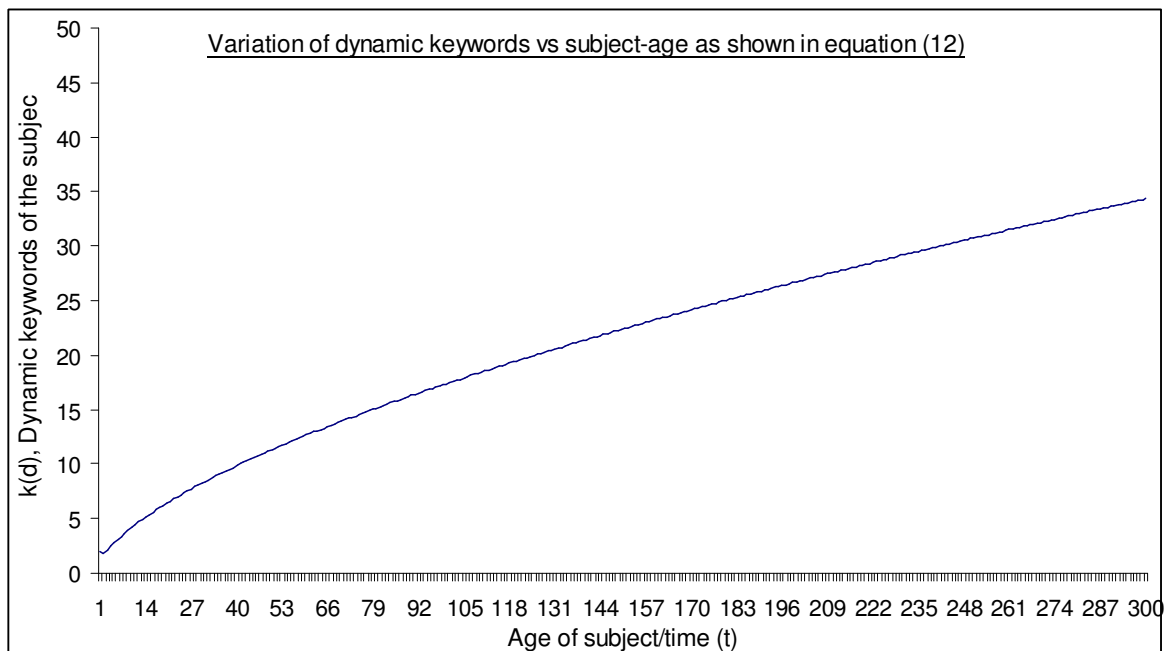


Figure 2—Variation of number of keywords with age of the subject

For high values of constants, i.e. for highly dynamic subjects the slope of variation would be higher and consecutively, the saturation value of the keyword stock would be high.

The equation (8) shows that the stock of keywords of a subject consists of two parts,  $k_0$ , which is the static keywords and represents entire stock of keywords of a subject at an instant of its inception, and  $k_d$ , which is the dynamic keywords and represents the active keywords of a subject during its voyage through active and updated courses of studies and researches. The static keywords of a subject actually belong to its parent subjects, and it remains more or less constant over time or the age of the subject; whereas the dynamic keywords belong to the subject itself and continuously changes over time or age of the subject. The above equation, equation (12), shows the changing pattern of dynamic counterpart of the keywords with time. It is clear from equation (12), that as 'time' or age of the subject tends to infinitely large value, the number of dynamic keywords or  $k_d$  still remains finite. As  $t \rightarrow \infty$ , the second part of equation (12), i.e.  $C_2^*(t)^{[-1.62]} \rightarrow 0$  and the first part of the same, i.e.  $C_1^*(t)^{0.62} \rightarrow$  a finite value. Thus the store of dynamic keywords of a subject cannot grow infinitely with time, as clear from equation (12). Therefore it can also be inferred from equation (12) that no subject can grow infinitely. Every subject has to reach saturation at a certain point of time. Since the point of saturation the rate of growth of subject decelerates, but it generates its daughter subjects by sharing its keywords with other parallel subject domains.

## Conclusion

In this paper, a theoretical expression regarding temporal variational pattern of number of dynamic keywords of a subject has been derived. It has been mathematically observed that 'Time' always limits the growth of any subject. Any subject meets saturation after an initial hike and intermediate growth, which is depicted in figure 1. The transition of keyword mass from intermediate growth phase to saturation phase limits indefinite growth of any subject.

## References

1. <http://en.wikipedia.org>
2. <http://en.wikipedia.org/wiki/Scientist>
3. Raman C V and Krishnan K S, (A) New type of secondary radiation, *Nature*, 121(3048) (1928) 501.
4. Ranganathan S R, *Prolegomena to Library Classification*, (Asia Publishing House; London), 1967.
5. Gopinath M A, *Colon Classification, (I: Classification in the 1970's. A second look. Revised edition. Ed. by Arthur Maltby. Clive Bingley; London), 1976*
6. Chakrabarti B, *Library Classification Theory*, (World Press; Calcutta), 1994.
7. Cutter C A, *Rules for a Dictionary Catalog*, (Government Printing Office; Washington, DC), 1904.
8. Drake C L, What is a subject? *Australian Library Journal*, 9 (1960) 34-41.
9. Wilson P, *Two Kinds of Power. An Essay on Bibliographical Control*, (University of California Press; Berkeley), 1968.
10. Hutchins W J, *Languages of Indexing and Classification: A Linguistic Study of Structures and Functions*, (Peter Peregrinus; London ), 1975.
11. Hutchins W J, On the problem of "aboutness" in document analysis, *Journal of Informatics*, 1 (1977) 17-35.
12. Hutchins W J, The concept of "aboutness" in subject indexing, *Aslib Proceedings*, 30 (1978) 172-181.
13. Maron M E, On indexing, retrieval and the meaning of about, *Journal of the American Society for Information Science*, 28 (1977) 38-43.
14. Miksa F, *The Subject in the Dictionary Catalog from Cutter to the Present*. Chicago, IL: American Library Association, 1983.
15. Soergel D, *Organizing Information: Principles of Data Base and Retrieval Systems*, (Academic Press; Orlando, FL), 1985.
16. Hjørland B, The concept of "subject" in Information Science, *Journal of Documentation*, 48 (2) (1992) 172-200.
17. Hjørland B, *Information Seeking and Subject Representation. An Activity-theoretical approach to Information Science*. Westport, CT & London, UK: Greenwood Press, 1997.
18. Hjørland B, 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) (2001) 774-778.
19. Hjørland B, Epistemology and the socio-cognitive perspective in Information Science, *Journal of the American Society for Information Science and Technology*, 53 (4) (2002) 257-270.
20. Hjørland B and Nicolaisen J, Bradford's Law of Scattering: Ambiguities in the Concept of "Subject". IN: Crestani, F. & Ruthven, I. (Eds.): *CoLIS 2005, LNCS 3507*, (96 – 106). Berlin: Springer-Verlag, 2005. ([http://www.db.dk/bh/Bradford\\_Colis5.pdf](http://www.db.dk/bh/Bradford_Colis5.pdf))
21. Molina M P, Interdisciplinary approaches to the concept and practice of written text documentary content analysis, *Journal of Documentation*, 50 (2) (1994) 111-133.
22. Sen B K, Universe of knowledge from a new angle, *Annals of Library and Information Studies*, 56(1) (2009) 7-12.
23. Dutta B, An analytical investigation into the characteristics of a subject; a case study. PhD thesis submitted to Jadavpur University, Kolkata, 2008.