ISSN (Print): 0972-2467
ISSN (Online): 0976-2477
SRELS Journal of Information Management, Vol 53(3), June 2016, p. 177-185
DOI: 10.17821/srels/2016/v53i3/86330

Introducing Facetometrics: A New Keyword-Based Measure for Locating Research Areas of a Subject

Kalipada Jana[1,2] and Bidyarthi Dutta[3*]

[1]Library Assistant, B.C. Roy Memorial Library, Indian Institute of Management Calcutta, Joka, Kolkata - 700 104, West Bengal, India; klpdjn@gmail.com
[2]Research Scholar, Department of Library and Information Science, Vidyasagar University, Midnapore - 721102, West Bengal, India
[3]Assistant Professor, Department of Library and Information Science, Vidyasagar University, Midnapore - 721 102, West Bengal, India; bidyarthi.bhaswati@gmail.com

Abstract

This paper has introduced a technique to find out research areas in a subject by the analysis of keywords assigned to articles. Keywords assigned to 1227 research papers published between 2004 and 2013 in a specific subject area "Hawking radiation" have been collected and analyzed. The research articles were retrieved from Web of Science using the search term "Hawking Radiation". The assigned keywords occurred with different frequencies. The keywords formed the clusters of words. The names given to the clusters were in accordance with the names of the most frequently occurring word in a keyword clusters. The clusters have been classed into three groups by size, i.e. small cluster, medium cluster and large cluster. As fairly large number of keywords formed large clusters, it has been assumed that the potential facets are represented by such clusters. Three basic parameters associated with the keyword clusters were identified, viz. no. of keywords in a cluster, frequency of occurrence and occupancy. Four indicators, viz., stability index, integrated visibility index, momentary visibility index and potency index have been defined on the basis of these three parameters and their fluctuations over the study period have been noticed. These indicators hold different values for different clusters and facets. The value ranges of these are categorized in five groups, viz. very high, high, medium, low and very low. Each indicator describes a particular aspect of a facet. These indicators may be used to measure various aspects of facets.

Keywords: Facetometrics, Keyword Cluster, Hawking Radiation, Research Areas, Web of Science, Astrophysics Research Trend, Research Trend, Facet Analysis

1. Introduction

Identifying research areas is a basic requirement for study of a subject. As pointed out by Seetharama (1997), the study of development of a subject has three major components: 1. Landmarks in the development of the subject, i.e., 2. The research trends collected from review documents, indexing and abstracting periodicals, state-of-the-art report, trend report etc. Also attempts are to be made to ascertain in general (a) the growth of literature (documents) on the subject concerned and (b) the degree of seepage and scatter of information in the subject concerned. 3. Trends in education in the subject concerned at different levels. The research trend analysis basically turns into focusing the facets that are the thrust areas in the concerned subject. This paper attempts to develop a quantitative method of finding facets constituting thrust research areas of the subject "Hawking Radiation". This quantitative method involves selection of keywords from the research articles followed by frequency analysis. The frequency analysis marks keyword clusters indicating active research topics of the subject. The name given to this study is *Facetometrics* - quantitative study of facets. This study points out the growing and decaying facets of

*Author for correspondence

a subject domain.

2. Related Works

Price (1963, 1975) investigated the growth in the number of scientists, scientific journals, and papers over the past two centuries. He found that the numbers doubled every 15 years. Literature and information are assumed to grow exponentially, but in individual disciplines the growth could be linear. The concept of ageing or obsolescence is intimately linked with the growth of science. Some authors considered growth and obsolescence as inversely related, suggests that faster the growth of literature in a field, the faster it ages and the literature becomes obsolete in a shorter time (Shodhganga, 2015). This can also be done by analyzing growth and obsolescence of keywords in a subject field. There are studies on growth and obsolescence of literature using citation analysis also. Content analysis is another research method, which is extensively used in library and information science to study the state-of-the-art of a subject. Content analysis, first applied in mass communication in the 1950s, is a systematic and rigorous approach to analyze research papers generated in the course of research (Berelson, 1952). The fundamental ingredients of the content of a subject are keywords. And content analysis by keywords is based upon the assumption that a paper's keywords constitute an adequate representation of its content. Two different keywords co-occurring within the same paper are an indication of a link between the topics to which they refer (Cambrosio, 1993). Coulter, Monarch and Konda (1998) selected keywords chosen by professional indexers. They believed that it is useful to study a fixed system that imposes a common nomenclature. Professional indexers' experiences assure standard application of that taxonomy. Looze and Lemarie (1997) conducted co-word study based on the keywords proposed by the experts. Courtial, Cahlik and Callon (1994) downloaded keywords from online databases to study social interaction. Law and Whittaker (1992) mapped acidification research by co-word analysis. Noyons and Van Raan (1998) presented the adjustments as implemented in the mapping procedure of science and technology. The improvements concerned the implementation of graphical user interfaces, and the addition of 'map-external' information. This interface enabled the users of the maps to focus on their specific areas of interest and to determine the position of actors in the field. Noyons and Van Raan (1998) mapped the overall structure of neural networks using co-occurrence of classification codes. Van Raan and Tijssen (1993) discussed the limits and potentials of bibliometric mapping based on co-word analysis. Van Raan (1997) analyzed the major research areas covered by the journal Scientometrics by co-word analysis. Courtial (1994) computed the interaction network within all kinds of authors by co-word analysis for the journal Scientometrics. To analyze temporal variation of content of a subject, a Keyword Associated Content Variation Analysis or KACOVA model was developed, which involves a detailed abstraction and analysis of keywords (Dutta, 2008 and Dutta, Majumder, Sen, 2010). The keywords in the subject Fermi Liquid were categorized as: Keyphrase; Modulator and Qualifier and then analyzed (Dutta, Majumder, Sen, 2008). To analyze research-trend and to identify potential research-areas of a subject, new taxonomy of keyword was proposed (Dutta, Majumder, Sen, 2010).

Either consciously, or unconsciously we always use "Keyword" in our every-day life. The term 'Keyword-in-Context' was introduced by Luhn in 1960 (Luhn, 1960). An early use of keywords was found in 1975 in the Journal of Applied Behaviour Analysis (Hartley and Kostoff, 2003). . The keywords represent the content, and keyword cluster analysis turns into content analysis of a subject. As subject becomes more specialized, the only words which will describe a subject are those by which the workers in that field refer to it. This is, of course, what Luhn calls the 'Keyword' or significant word (Black, 1962).

Cluster analysis of keywords is an effective method for examining the user's view of information space with the goal of producing flexible and customizable classification scheme. This is based on statistical analysis of different characteristics of keywords. Cluster analysis is used in a wide range of applications in all major disciplines and it, particularly document-based cluster analysis, paves way for automatic classification (Willett, 1988).

A major shortcoming of existing subject access tools is that they are silent about the behavioral aspects of the keywords, i.e., the modes of occurrences of the keywords in a database, either full-text or bibliographic, whatever it may be. One of the strengths of the model proposed here is the quantitative interpretation of the behavioral aspects of keywords to examine the growth and obsolescence of different domains of a subject.

3. Purpose, Methodology and Limitations

The main objective of this paper is to develop a method for identifying research areas of a subject from primary sources of information which reflect the actual state-of-the-art of the subject. In all, 1227 research articles on "Hawking radiation" published between 2004 and 2013 have been retrieved from Web of Science. It is assumed that the search term "Hawking Radiation" indicates a specific subject domain under the broad area astrophysics. Then 8592 keywords were selected from titles and abstracts of these 1227 articles, i.e. an average of seven keywords per article (Table 2).

The year-wise variations of articles and keywords are presented in Figures 1 and 2. All these keywords were then separated into two groups, viz. "Clustered keywords" and "Unclustered keywords". To identify clustered keywords, at first keywords with some common word(s) were identified and a group of keywords with a particular common word(s) therein were deemed a cluster. The clusters were categorized by size, i.e. number of keywords in the cluster. The clusters having 3 to 5 keywords have been classed as "Small clusters"; the clusters with 6 to 10 have been classed as "Medium clusters" and the clusters having more than 10 have been classed as "Large clusters". Clustering keywords means grouping the same into a common facet. The term "Facet" here indicates a keyword cluster with a common word. Each cluster is named by the common word. Some examples of keyword clusters are shown in Table 1.

Total number of selected keywords over ten years (Table 2) is 3509, which includes repetitions of same keywords in different years. Any particular keyword may occur in different years. The number of distinct keywords over ten years, i.e. without repetitions was 1630. Hence, on an average each keyword was repeated (8592/1630 = 5.3) nearly five times over ten years. On an average, three keywords were selected from each article over ten years, (3509/1227 = 2.9) and there are 1.3 (1630/1227 = 1.3) distinct keywords per article on average. Out of total 1630 distinct keywords, 796 keywords occurred without repetitions. These 796 keywords formed no cluster and name given to such keywords is "Unclustered keywords". The average frequency of occurrence of unclustered keywords is 3.7 (2965/796 = 3.7). The remaining 834 keywords occurred 5627 times and formed 84 clusters. Of these 84 clusters, 48 clusters are small, 25 clusters are medium

Table 1. Some examples of small, medium and large keyword clusters

Size of the cluster	Name of the cluster and no. of keywords	Keywords of the cluster
Small	Accretion (4)	Accretion; Accretion disc; Accretion efficiency; Advection-dominated accretion
Small	Dark energy (4)	Dark energy; Dark energy theory; Holographic dark energy; New agegraphic dark energy
Medium	Brane (6)	Brane; D-Brane; D-Brane approach; Brane production; Brane world; Brane-world scenario
Medium	Quark (6)	Strange quark matter; Heavy quark physics; Quark production; Quark star; Quark gluon plasma; Quark jet
Medium	Thermodynamics (9)	Thermodynamic geometry; Thermodynamic property; Thermodynamic quantity; Thermodynamics; Generalized law of thermodynamics; First law of thermodynamics; Thermodynamics of FRW universe; Generalized second law of thermodynamics; Thermodynamics of generalized uncertainty principle
Large	Entropy (27)	Entropic force; Entropic acoustic instability; Entropy; Entropy and area quantization; Entropy bound; Entropy conservation; Entropy correction; Bekenstein-Hawking entropy correction; Entropy density; Entropy of horizon; Entropy quantization; Entropy spectrum; Bekenstein-Hawking entropy; Hawking entropy; Canonical entropy; Corrected entropy; Correctional entropy; Entanglement entropy; Fermionic entropy; Geometric entropy; Gravitational entropy; Hole entropy; Logarithmic Entropy; Logarithmic correction entropy; Modified entropy; Planck absolute entropy; Semiclassical entropy

Table 2. Sample for the study

Year	No. of articles (A)	No. of selected keywords (B)	Total no. of occurrences of all keywords (C)	Average frequency (rounded) of occurrence per keyword (C/B)	No. of selected (rounded) keywords per article (B/A)	No. of occurrence of keywords (rounded) per article (C/A)
2004	57	200	374	2	4	7
2005	62	195	356	2	3	6
2006	109	338	706	2	3	6
2007	145	370	934	3	3	6
2008	170	406	1145	3	2	7
2009	165	433	1252	3	3	8
2010	151	431	1182	3	3	8
2011	127	403	912	2	3	7
2012	124	381	911	2	3	7
2013	117	352	820	2	3	7
Total	**1227**	**3509**	**8592**	**2**	**3**	**7**

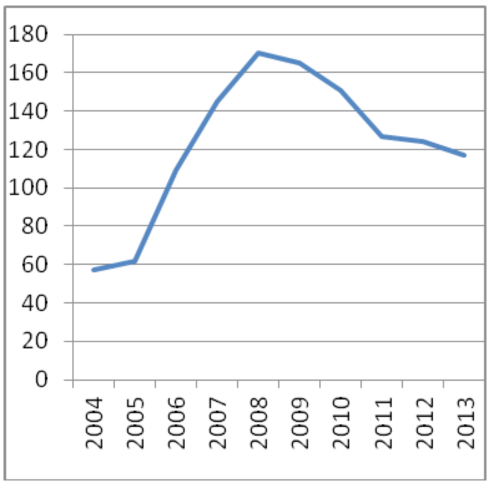

Figure 1. Number of articles over ten years.

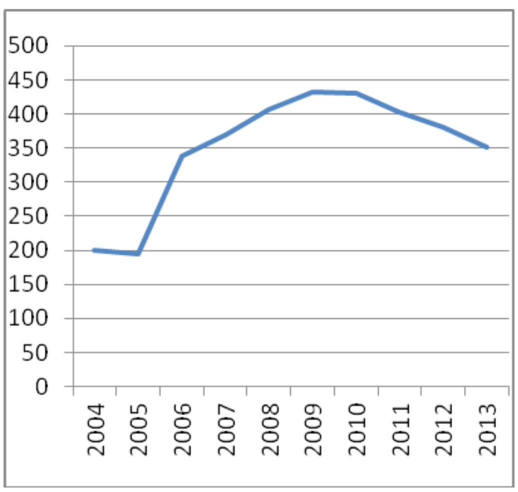

Figure 2. Number of selected keywords over ten years.

and 11 clusters are large in size. As large clusters contain higher number of relevant keywords, they may be reckoned as "Core" descriptors or core ideas of the subject. Similarly the medium-sized clusters and small-sized clusters may be reckoned as "Allied" descriptors and "Alien" descriptors respectively. Interestingly the number of articles show a growth till 2008 (170) and then started to decrease. The variation pattern of selected keywords was almost identical with that of the number of articles.

The numerical values of four characteristic indicators (CI) for each large and medium cluster has been calculated and presented in Table 3. These four characteristic indicators (CI) are: Integrated Visibility Index, Momentary Visibility Index, Stability Index and Potency Index, which are defined as follows (Dutta, Majumder and Sen, 2010):

Integrated Visibility Index, i.e. $v = F/N$;

Momentary Visibility Index, i.e. $m = F/A$;

Potency Index, i.e. $p = \ln(N*F)$ and

Stability Index, i.e. $s = (A/A_{max})*100$;

where,

N ~ No. of keywords in a cluster

F ~ Frequency of occurrence of all keywords of a cluster

A ~ Occupancy of all keywords in a cluster, i.e. Number of appearance of a keyword over the stipulated time span.

A_{max} ~ Maximum occupancy of all keywords in the said cluster, i.e. number of keywords of a cluster multiplied by number of years in the said time span.

It is to be noted that frequency of occurrence is different from number of appearance of a keyword. Say, a keyword is found in a database in three years, say 2011, 2012 and 2014 with frequencies 12, 14 and 15 (say). Here the total frequency of occurrence of this keyword in three years is 12 + 14 + 15 = 41, but its number of appearance is 3, i.e. number of appearance or occupancy just counts

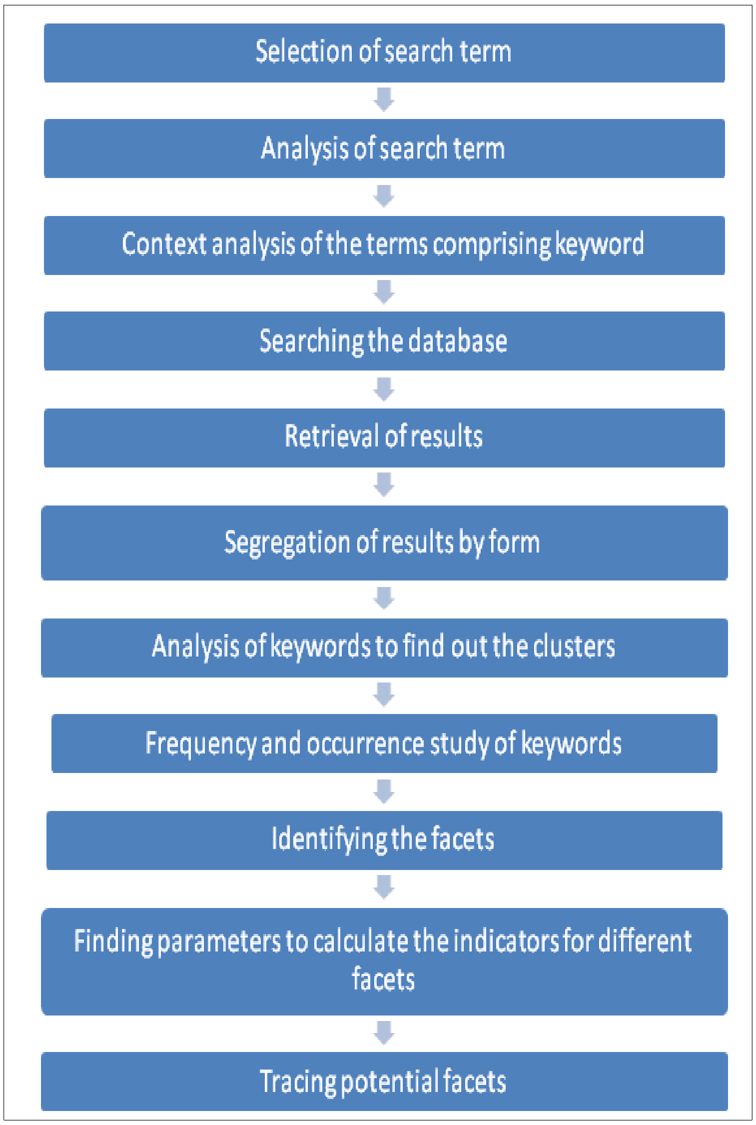

Figure 3. The steps involved in Facetometrics.

in how many years it occurred. The maximum occupancy indicates the highest appearance of a keyword within the time span. For instance, since 2001 to 2015, if a keyword is found only in the years 2005, 2007 and 2010, then its occupancy (A) would be 3, while maximum occupancy (A_{max}) would be 15, i.e. the length of the total time span.

4. Schematic Presentation

The steps involved in the function of *facetometrics* are schematically presented (Figure 3):

5. Results and Analysis

Four Characteristic Indicators for core and allied facets have been found in this study. The Integrated Visibility Index (v) of a facet or cluster is defined as average frequency of occurrence per keyword. The Momentary Visibility Index (m) of a facet or cluster is defined as average frequency of occurrence per keyword for unit appearance or occupancy, i.e. how frequently on average a keyword is assigned in research papers in a year. Potency Index (p) is defined as the natural logarithm of product of total number of keywords and frequency. It is thus an indicator of both number of keywords and their frequencies of a facet that may be reckoned as the strength of the facet. Stability index is defined as the ratio of actual occupancy to the maximum occupancy of a facet. It indicates the temporal stability or the stability over time span of a facet. It indicates whether the keywords of a facet appear uniformly over the entire time span. In all, 11 core descriptors, 25 allied descriptors and 48 alien descriptors

Table 3. Characteristic Indicators (CI) for core descriptors

Large Keyword clusters (Core facets)	(No. of keywords in the cluster, N)	(Frequency of occurrence, F)	(Occupancy, A)	[Maximum occupancy, A(max)]	Integrated Visibility Index, v = F/N	Momentary Visibility Index, m = F/A	Potency Index, P = ln(N*F)	Stability Index, S = (A/A(max))*100
Black hole	154	832	301	1540	5.40	2.76	11.76	19.55
Cosmology	24	85	50	240	3.54	1.70	7.62	20.83
Einstein's theory	11	19	17	110	1.73	1.12	5.34	15.45
Entropy	27	321	64	270	11.89	5.02	9.07	23.70
Gravity	51	300	130	510	5.88	2.31	9.64	25.49
Quantum physics	61	276	136	610	4.52	2.03	9.73	22.30
Radiation	11	110	32	110	10.00	3.44	7.10	29.09
Scalar field	11	36	27	110	3.27	1.33	5.98	24.55
Space-time	36	150	75	360	4.17	2.00	8.59	20.83
Tunneling	22	158	63	220	7.18	2.51	8.15	28.64
Universe	13	75	46	130	5.77	1.63	6.88	35.38

Table 4. Ranking of core descriptors

CIs Core Facets' Rank	Ranking of core facets in accordance with values of the four CIs			
	Integrated Visibility Index (v)	Momentary Visibility Index (m)	Potency Index (p)	Stability Index (s)
1	Entropy	Entropy	Black hole	Universe
2	Radiation	Radiation	Quantum physics	Radiation
3	Tunneling	Black hole	Gravity	Tunneling
4	Gravity	Tunneling	Entropy	Gravity
5	Universe	Gravity	Space-Time	Scalar Field
6	Black hole	Quantum physics	Tunneling	Entropy
7	Quantum physics	Space-time	Cosmology	Quantum physics
8	Space-time	Cosmology	Radiation	Space-time
9	Cosmology	Universe	Universe	Cosmology
10	Scalar field	Scalar field	Scalar field	Black hole
11	Einstein's theory	Einstein's theory	Einstein's theory	Einstein's theory

have been observed. The ratio of the number of descriptors in three zones is, 11:25:48; or 11(1:2.3:4.4). This ratio nearly tallies with Bradford's pattern, i.e. $k(1:n:(n)^2)$, where k is the Bradford's multiplier. Here $k = 11$, $n = 2.3$ and $(n)^2 = 4.4$ (close to actual value, i.e. 5.3). The values of Characteristic Indicators for core and allied descriptors are shown in Tables 3 and 5 respectively. All descriptors are ranked according to numerical values of each of the four indicators. The ranking of core descriptors and allied descriptors are presented in Tables 4 and 6 respectively. To compare the four rankings, Spearman's Rank Correlation Coefficient was used. and the data is presented in Table 7. It has been noticed that all coefficients for allied descriptors are strong positive correlation, while only three coefficients for core descriptors are strong positive correlation, two coefficients are weak positive correlation and one coefficient is weak negative correlation. The ranking of core descriptors (Table 4) reveals that the following five descriptors, viz. Black hole, Entropy, Quantum physics, Radiation and Universe constitute the central facets of the subject. However, rankings in accordance with stability index and potency index in case of core descriptors

Table 5. Characteristic Indicators (CI) for allied descriptors

Medium Keyword clusters (Allied facets)	(No. of keywords in the cluster, N)	(Frequency of occurrence, F)	(Occupancy, A)	[Maximum occupancy, A(max)]	Integrated Visibility Index, v = F/N	Momentary Visibility Index, m = F/A	Potency Index, P = ln(N*F)	Stability Index, S=(A/A(max))*100
Anomaly	7	138	27	70	19.71	5.11	6.87	38.57
Area spectrum	6	44	15	60	7.33	2.93	5.58	25.00
Brane theory	6	70	24	60	11.67	2.92	6.04	40.00
Conservation theory	6	7	7	60	1.17	1.00	3.74	11.67
Co-ordinate transformation	8	13	12	80	1.63	1.08	4.64	15.00
Dirac's theory	7	124	26	70	17.71	4.77	6.77	37.14
Electromagnetic field	8	10	11	80	1.25	0.91	4.38	13.75
Energy	10	84	30	100	8.40	2.80	6.73	30.00
Gauge theory	9	78	22	90	8.67	3.55	6.55	24.44
Horizon	9	146	18	90	16.22	8.11	7.18	20.00
Information	9	59	29	90	6.56	2.03	6.27	32.22
Particle physics	9	210	39	90	23.33	5.38	7.54	43.33
Perturbation	7	39	9	70	5.57	4.33	5.61	12.86
Planck's distribution	6	21	13	60	3.50	1.62	4.84	21.67
Quark	6	7	6	60	1.17	1.17	3.74	10.00
Quasi function	6	49	16	60	8.17	3.06	5.68	26.67
Relativity	6	28	16	60	4.67	1.75	5.12	26.67
Semiclassical approach	6	9	8	60	1.50	1.13	3.99	13.33
Space	8	95	26	80	11.88	3.65	6.63	32.50
Statistical approach	6	24	15	60	4.00	1.60	4.97	25.00
String theory	6	77	26	60	1.28	2.96	8.44	43.33
Thermodynamics	9	177	20	90	19.67	8.85	7.37	22.22
Uncertainty principle	9	59	20	90	6.56	2.95	6.27	22.22
Wave	7	17	13	70	2.43	1.31	4.78	18.57
WKB approximation	6	17	14	60	2.83	1.21	4.62	23.33

are not in consonance, but nearly reverse. It implies that in core areas the potential facets seem to be less stable over time, and also the temporally stable descriptors hold comparatively less number of keywords. A glance at Table 4 clears this feature, for instance, 'Black Hole' and 'Quantum Physics' are ranked as '1' and '2' by potency index, whereas these two descriptors are ranked as '10' and '7' by stability index. The ranking by all indicators for allied descriptors gives nearly identical results.

6. Conclusion

In this study, a method has been proposed to quantify the characteristics of the facets of a subject. As the behavioral

Table 6. Ranking of allied descriptors

CIs Allied Facets' Rank	Ranking of allied facets in accordance with values of the four CIs			
	Integrated Visibility Index (v)	**Momentary Visibility Index (m)**	**Potency Index (p)**	**Stability Index (s)**
1	Particle physics	Thermodynamics	String theory	String theory
2	Anomaly	Horizon	Particle physics	Particle physics
3	Thermodynamics	Particle physics	Thermodynamics	Brane theory
4	Dirac's theory	Anomaly	Horizon	Anomaly
5	Horizon	Dirac's theory	Anomaly	Dirac's theory
6	Space	Perturbation	Dirac's theory	Space
7	Brane theory	Space	Energy	Information
8	Gauge theory	Gauge theory	Space	Energy
9	Energy	Quasi function	Gauge theory	Quasi function
10	Quasi function	String theory	Uncertainty principle	Relativity
11	Area spectrum	Uncertainty principle	Information	Area spectrum
12	Information	Area spectrum	Brane theory	Statistical approach
13	Uncertainty principle	Brane theory	Quasi function	Gauge theory
14	Perturbation	Energy	Perturbation	WKB approximation
15	Relativity	Information	Area spectrum	Thermodynamics
16	Statistical approach	Relativity	Relativity	Uncertainty principle
17	Planck's distribution	Planck's distribution	Statistical approach	Planck's distribution
18	WKB approximation	Statistical approach	Planck's distribution	Horizon
19	Wave	Wave	Wave	Wave
20	Co-ordinate transformation	WKB approximation	Co-ordinate transformation	Co-ordinate transformation
21	Semiclassical approach	Quark	WKB approximation	Electromagnetic field
22	String theory	Semiclassical approach	Electromagnetic field	Semiclassical approach
23	Electromagnetic field	Co-ordinate transformation	Semiclassical approach	Perturbation
24	Conservation theory	Conservation theory	Quark	Conservation theory
25	Quark	Electromagnetic field	Conservation theory	Quark

Table 7. Rank Correlation Coefficients between different CIs for core and allied descriptors

Rank Correlation Coefficients (r) between the CIs	**Values of 'r' for Core Clusters**	**Nature of correlation**	**Values of 'r' for Allied Clusters**	**Nature of correlation**
r_{vm}	0.864	Strong +ve	0.867	Strong +ve
r_{vp}	0.391	Weak +ve	0.791	Strong +ve
r_{vs}	0.627	Strong +ve	0.623	Strong +ve
r_{mp}	0.645	Strong +ve	0.892	Strong +ve
r_{ms}	0.236	Weak +ve	0.529	Strong +ve
r_{ps}	-0.173	Weak -ve	0.729	Strong +ve

Nature of correlation: $1 \geq r \geq 0.5$ indicates Strong +ve Correlation;
$0.5 > r > 0$ indicates Weak +ve Correlation;
$r = 0$ indicates No Correlation;
$0 > r \geq -0.5$ indicates Weak -ve Correlation;
$-0.5 > r \geq -1$ indicates Strong -ve Correlation;

aspects of the facets are measured in terms of numerical figures, the name 'Facetometrics' is given to this method. The Characteristic Indicators imply different properties of the facets, which in turn define the specific behaviour of the subject. For instance, it is found for 'Hawking Radiation' that the facets formed by temporally stable descriptors consist of less number of keywords with low frequency, that might be a special behaviour of this subject. However, this would be clear after applying this method to other subjects. This study may be extended to other areas of physical sciences, and also to other broad subjects like bio-sciences, medical science, agricultural science, social science, engineering science, cognitive science, behavioral science etc.

7. References

1. Berelson, B. 1952. *Content Analysis in Communications Research*. New York: Free Press.
2. Black, J.D. 1962. The Keyword: It's use in abstracting, indexing and retrieving information. *Aslib Proceedings*. 14(10): 313–321.
3. Cambrosio, A., et al. 1993. Historical scientometrics? Mapping over 70 years of biological safety research with Co-word analysis. *Scientometrics*. 27(2): 119–143.
4. Coulter, N., Monarch, I. & Konda, S. 1998. Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science*. 49(13): 1206–1223.
5. Courtial, J.P. 1994. (A) Co-word analysis of scientometrics. *Scientometrics*. 31(3): 251–260.
6. Courtial, J.P., Cahlik, T. & Callon, M. 1994. (A) Model for social interaction between cognition and action through a key-word simulation of knowledge growth. *Scientometrics*. 31(2): 173–192.
7. Dutta, B. 2008. An analytical investigation into the characteristics of a subject - A case study. PhD thesis. Jadavpur University.
8. Dutta, B., Majumder, K.P. & Sen, B.K. 2008. Classification of keywords extracted from research articles published in science journals. *Annals of Library and Information Studies*. 55(4): 317–333.
9. Dutta, B., Majumder, K.P., Sen, B.K. 2010. Content management by keywords: an analytical study. *SRELS Journal of Information Management*. 47(6): 599–620.
10. Hartley, J. & Kostoff, R.N. 2003. How useful are 'keywords' in scientific journals? *Journal of Information Science*. 29(5): 433–438.
11. Index Term. Retrieved on 16 Jan 2016 from https://en.wikipedia.org/wiki/Index_term
12. Retrieved on 5 Mar 2015 from http://shodhganga.inflibnet.ac.in/bitstream/10603/5109/10/10_chapter%201.pdf
13. Law, J. & Whittaker, J. 1992. Mapping acidification research: A test of the co-word method. *Scientometrics*. 23(3): 417–461.
14. Looze, M.D. & Lemarie, J. 1997. Corpus relevance through co-word analysis: An application to plant proteins. *Scientometrics*. 39(3): 267–280.
15. Luhn, H.P. 1960. Keyword-in-Context for technical literature (KWIC Index). *American Documentation*. 11.
16. Noyons, E.C.M. & van Raan, A.F.J. 1998. Advanced mapping of science and technology. *Scientometrics*. 41(1–2): 61–67.
17. Noyons, E.C.M. & van Raan, A.F.J. 1998. Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research. *Journal of the American Society for Information Science*. 49(1): 68–81.
18. de Solla Price, D.J. 1963. *Little science, big science*. New York: Columbia University Press.
19. de Solla Price, D.J. 1975. *Science since Babylon*. Enlarged ed. New Haven: Yale University Press; p. 165–186.
20. Seetharama, S. 1997. *Information consolidation and repackaging: framework, methodology, planning*. New Delhi: Ess Ess Pub.
21. Van Raan, A.F.J., & Tijssen, R.J.W. 1993. The neural net of neural network research: An exercise in bibliometric mapping. *Scientometrics*. 26(1): 169–192.
22. Van Raan, A.F.J. 1997. Scientometrics: State-of-the-art. *Scientometrics*. 38(1): 205–218.
23. Willett, P. 1988. Recent trends in hierarchical document clustering: A critical review. *Information Processing and Management*. 24(5): 577–597.