IFLA WLIC 2013
SINGAPORE

Submitted on: July 4, 2013

Classification of Keywords Selected from Research Articles on Physics and Development of a Quantitative Subject Access Tool

Bidyarthi Dutta
Assistant Professor
Dept. of Library & Information Science
Vidyasagar University, Midnapore
West Bengal, India
E-mail: bidyarthi.bhaswati@gmail.com

Krishnapada Majumder
Professor
Dept. of Library & information Science
Jadavpur University
Kolkata; India

Bimal Kanti Sen
80, Shivalik Apartments
Alaknanda
New Delhi; India

Abstract:

All research articles begin with a title. Most include an abstract. Several include keywords. All three of these features describe an article's content in details. The title sends an instant reflection of the central theme of the research topic. The abstract summarizes the content. The keywords indicate the core and allied fields of concern. The researchers and indexers quickly and easily locate particular articles within their areas of interest with the aid of keywords. Keywords hold prime importance in abstracting and indexing services. Keywords play major role in information retrieval function. This paper is based on analysis of 14,221 keywords collected from 2,526 research articles published in three journals, viz. Chaos, Physics of Plasmas *and* Low Temperature Physics *since 2006 to 2012. Out of all these author-assigned keywords, the number of distinct bits obtained was 2571. After collection, the lexically close keywords are identified that form clusters. Several such clusters are found and the composition of keywords in nearly all clusters varies over the said time span.*

Four indicators have been defined on the basis of fluctuating keyword composition within clusters. The name given to these four indicators are stability index, integrated visibility index, momentary visibility index and potency index respectively. These indicators hold different values for different clusters. The value ranges of them are categorized in five groups, viz. very high, high, medium, low and very low. A new quantitative subject access tool has been proposed on the basis of these *indicators, which can predict the probable new and obsolete keywords in any subject domain. The name given to this new tool is keysaurus, i.e., keyword-based-thesaurus.*

Keywords: subject access tool; information retrieval; information retrieval thesaurus; keysaurus, knowledge classification

1 INTRODUCTION

The “Keyword” is an inseparable part of our daily life that we are constantly using either consciously, or unconsciously. The keywords generally come from a controlled vocabulary or may be freely assigned. Keywords collected from controlled vocabulary, however allow improved retrieval precision of documents on a selected topic. The selection of keywords is thus a vital measure of an information system. The indexers generally read a literature or text to locate the best terms in a thesaurus, and then assign the terms that best describe the document content. The keywords collected in this way are stored in a search index. The function of indexing actually depends on human analysis of a topic or subject. Different indexers may assign different keywords to represent same topic or subject[1]. Common words like articles (a, an, the), prepositions (by, with, for, to etc.) and conjunctions (and, or, but) are not treated as keywords because they can’t reflect any essence of the document. Almost every English-language document or site has the article "the", and so it makes no sense to search for it. The most popular search engine, Google removes stop words such as "the" and "a" from its indexes. Sometimes, nascent themes or concepts may lack appropriate keyword to be described compatibly. Suraud et al.[2] observed the non-existence of well-defined keywords in newly-emerging fields, which makes bibliographic searches difficult. The keywords may be sometimes described as “Subject descriptors”, the term, which was coined by Calvin Mooers in 1948.

The standard subject access tools, such as lists of subject headings (e.g., *Sears List of Subject Headings* or *Library of Congress Subject Headings*) or classification schedules (e.g., Dewey Decimal Classification, or Colon Classification) are based on controlled vocabulary rather than on the users' terminology. Studies of controlled vocabularies have indicated that they work well when there is an accepted common terminology describing concepts in the concerned subject area and when users are familiar with the terminology[3]. Solomon[4] stated, "Classification schemes fail too often because they are not grounded in the language and knowledge of users or in the task or situation of use." Hurt[5] suggested that it is necessary to renew and expand indexing and classification systems. Soergel et al.[6] pointed out that existing classification schemes and thesauri lack well-defined semantics and structural consistency. With the advent of electronic information and the Internet, the physical location of the material is of much less importance. This has brought forth a re-scrutinization of classification schemes with a greater emphasis placed on intellectual access. Bates et al.[7] proposed development in the structures of thesauri and in the design of online information systems. If the classification schemes be freed from the requirement of shelving of one document in one location, then the subject hierarchies can be made more flexible. There is also a greater possibility of customizing classification schemes to fit specific groups of users with particular needs. In traditional library systems users need document-title and author’s name primarily for starting any search, whereas, in electronic environment the foremost need is centered on keywords for doing so. The users from different subject areas use different keywords, and large numbers of keywords form different clusters. The 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 of science and social sciences and it, particularly document-based cluster analysis, paves the way towards automatic classification[8].

One of the major shortcomings of existing information systems is that they are silent about the behavioral aspects of the keywords, i.e., the modes of occurrences of the keywords in a database. Also, no system ever described the properties of keywords in quantitative form. However, one of the strengths of the model studied in this paper is its interpretation of the behavioral aspects of the keywords in quantitative form. Keyword clusters have been generated here through indexing of keywords. The indicators defined in this model describe quantitative aspects of the keyword clusters. In all, four quantitative indicators of trend-analysis have been defined here.

2 OBJECTIVES

The main objectives of this study include:

1. To analyze the assigned keywords from 2526 research articles published in three journals, viz. *Chaos*, *Physics of Plasmas* and *Low Temperature Physics*
2. To identify different groups of keywords, as keywords generally occur centering a common term. The name given to such groups is keyword clusters
3. To define four indicators that describe some modes of occurrences of the keyword clusters
4. To propose a quantitative subject access tool comprising of keyword clusters

3 SCOPE AND METHODOLOGY

This study has been executed after collecting author-assigned keywords from 2526 research articles in all, published in three journals, viz. *Chaos*, *Physics of Plasmas* and *Low Temperature Physics*. The number of articles taken from *Chaos* is 1037, from *Low Temperature Physics* is 769 and from *Physics of Plasmas* is 720. The time span for three journals are different, i.e., 2006 to 2012 for *Chaos*; 2006 to 2010 for *Low Temperature Physics* and 2010 to 2012 for *Physics of Plasmas*. The numbers of keywords collected from three journals are as follows in Table 1, below.

Table 1: No. of keywords collected from three journals

Journal	Total no. of keywords	No. of distinct keywords	Average frequency of each keyword
Chaos	4901	1155	4.2
Low Temperature Physics	5105	920	5.5
Physics of Plasmas	4215	496	8.5

After collection the keywords have been collated to find out different clusters, i.e., to trace groups of keywords with a common key-term. For instance the following seven keywords, i.e., *crystal defect, crystal field interaction, crystal growth, crystal microstructure, crystal orientation, crystal structure* and *crystal symmetry* form a keyword cluster where the common key-term is crystal. In such cases the name given to the cluster has been taken from the common key-term, i.e., *crystal* in this case. The variables associated with a keyword cluster have been taken under consideration to define the indicators. The corresponding representative symbols are given in the adjacent parenthesis.

1) Total number of keywords in an arbitrary cluster is “N”, say
2) Frequency of Occurrence of all keywords belonging to the same cluster k_r during the entire span ‘l’ is F
3) Occupancy of the said cluster during the time span “l” is “A”
4) Highest possible Occupancy of the same cluster is “A_{Max}”
5) Concerned Time span of occurrence of keywords is “l”

The highest possible occupancy (A_{Max}) of a cluster is equal to span of occurrence of keywords (l) multiplied by total number of keywords (N) in that cluster.

i.e., $A_{Max} = l * N$

Let us take an example from Table 2 for the keyword “Semiconductor, elemental”, which is the third keyword of this cluster. The frequency of occurrence of this keyword is 15, as it appeared in 15 different journal articles; while its occupancy is 4, as it appeared 4 times only in between 2006 and 2010. The maximum possible occupancy is 5, as it can appear maximally 5 times within the stipulated time span, i.e., 2006 to 2010. Again, if the whole cluster ‘Semiconductor’ is considered, then the total frequency of occurrence and total occupancy will be equal to 129 and 63 respectively. The total number of keywords in this cluster is 28. The numerical values of the above variables for the cluster ‘Semiconductor’ is given below in Table 3.

Table 2: The cluster ‘Semiconductor’ and its member keywords with their frequencies of occurrences over 5 years (2006-2010)

S. No.	Year Keywords	2006	2007	2008	2009	2010	Total
1	Semiconductor (cluster name)					1	1
2	semiconductor, amorphous			1			1
3	semiconductor, elemental	4	5	3	3		15
4	semiconductor, ferroelectric		1			1	2
5	semiconductor, III-V	1	7		4	1	13
6	semiconductor, III-VI					2	2
7	semiconductor, II-VI	1	6	1	11		19
8	semiconductor, IV-VI			1			1
9	semiconductor, magnetic		2		1		3
10	semiconductor, narrow band-gap	1		1			2
11	semiconductor, piezoelectric			1			1
12	semiconductor, semimagnetic		2		3	1	6
13	semiconductor, superconducting			1			1
14	semiconductor, ternary		1				1
15	semiconductor, wide band-gap	2			5	1	8
16	semiconductor-doped-glass				1		1
17	semiconductor-doping		2	1	3	1	7
18	semiconductor-epitaxial-layer			1	1		2
19	semiconductor-growth			1	1		2
20	semiconductor-heterojunction	1	3	2	3		9

21	semiconductor-laser		1				1
22	semiconductor-material		3	1	1	1	6
23	semiconductor-metal boundary		1	1			2
24	semiconductor-nanotube			1			1
25	semiconductor-quantum-dot		2	1	1		4
26	semiconductor-quantum-well	3	6	1	4	1	15
27	semiconductor-quantum-wire	1				1	2
28	semiconductor-superlattice			1			1
	All						

Table 3: Numerical values of some variables for the cluster 'Semiconductor'

Variable	Representative Notation	Numerical Value
Total number of keywords	N	28
Frequency of Occurrence	F	129
Occupancy	A	63
Highest possible Occupancy	A_{Max}	l* N = 5*28 = 140

A keyword occurs with certain frequency in any year. It may occur with a very high frequency but within a narrow time span; on the other hand, it can also come with trifle frequency but over a large time span. The phenomena of occurrence over certain time span has been termed here as 'Occupancy'. Hence, *frequency of occurrence* and *occupancy* are two vital variables associated with a keyword or keyword cluster.

These two variables indicate two fundamental dimensions of a keyword/keyword cluster. High 'Occupancy' indicates higher stability over certain time span or higher temporal stability, whereas high 'Frequency of Occurrence' is an indicator of greater coverage of a keyword cluster over journal articles. The journal-articles may be looked as the intellectual space, where the keywords exist. A higher value of 'Frequency of Occurrence' thus points out higher spatial stability.

Another important variable is number of keywords in a cluster, represented by N, which says the strength of a cluster. The three fundamental variables of a keyword/keyword cluster thus indicate three fundamental features of the same in the subject space as shown in Table 4, below.

Table 4: Three fundamental variables of a keyword/keyword cluster

Variable	Representative Notation	Feature indicated in subject space comprised by journal articles
Frequency of Occurrence	F	Stability over Space
Occupancy	A	Stability over Time
Total number of keywords	N	Energy
Maximum occupancy	A(max)	Maximum possible stability over Time

Keyword Characteristic Indicators: The following four indicators based on four variables of a keyword cluster have been identified and are defined as noted in Table 5, below.

Table 5: Keyword Characteristic Indicators

Serial No.	Indicator	Denoted by	Defined as
1	Integrated Visibility Index	v	F / N
2	Momentary Visibility Index	m	F / A
3	Potency Index	p	ln(N*F)
4	Stability Index	s	(A/A(max))*100

1) Integrated Visibility Index, denoted by v, reflects the exposure of a keyword cluster over the entire journal-article space during concerned time span. This is defined as number of journal articles covered by a single keyword over the entire time span.
2) Momentary Visibility Index, denoted by m, reflects the exposure of a keyword cluster over the entire journal-article space in a single appearance. This is defined as number of journal articles covered by a single keyword in a single appearance.
3) Potency Index, denoted by p, tells the energy of a keyword cluster and defined as the natural logarithm of product of total number of keywords and frequency.
4) Stability Index, denoted by s, tells about temporal stability of a keyword cluster and defined as ratio of actual occupancy of a cluster to the maximum occupancy of the same cluster, multiplied by 100.

In short, these four indicators define five basic properties of keyword clusters, as given below in Table 6, viz. (1) Visibility, (2) Scattering, (3) Strength, (4) Stability and (5) Density.

Table 6: Basic properties and corresponding trends indicated

Basic Properties Studied		Corresponding Indicators	Trends Indicated at high values of the indicators
Visibility	Integrated	v	High visible keyword, which may be subject-specific, subject-generic or supporting.
	Momentary	m	Highly visible, i.e., myriad but isolated. Generally keywords belonging to an area that is supportive to the central area of research fall under this category.
Potency or Strength		p	Cluster with large number of keywords and high occupancy indicating highly relevant and subject-centric keywords.
Stability		s	Ratio of actual occupancy to maximum possible occupancy, which tells average occupancy over entire time span. High value of average occupancy indicates higher stability.

Keysaurus: the proposed quantitative subject access tool: A subject access tool, in general terms, can be defined as a classification tool to assist libraries, archives or other documentation centres to manage their records and other information. This tool is designed to facilitate users to identify preferred (or authorized) terms for classifying and titling records and to provide a range of paths to reach these terms. The classification schedule, subject heading list and thesaurus are well-known examples of subject access tool. The subject access tool also facilitates strategies for retrieving documents and reduces the probability of an unsuccessful research, which results from a confusing or irrelevant retrieval. This functionality is achieved by establishing semantic relationships between keywords.

The design and development of a subject access tool is based on knowledge classification. The following aspects are chiefly considered for classifying any domain of knowledge in the universe of knowledge:

1. Classification principles for organizing information and displaying subject relationships;
2. Controlled vocabulary features, particularly the control of synonyms and homographs for the purpose of improving recall and precision;
3. Search strategies formulated and pre-stored for the purpose of optimizing search results and current awareness.

In the present study, a subject-access-tool has been proposed for information retrieval, which is based on keyword cluster analysis that is a bottom-up approach of processing and organisation of information. The name given to the proposed new tool is *Keysaurus*, (i.e., keyword + thesaurus) as it is based on keyword cluster analysis, and describes the necessary quantitative aspects of keywords. This subject access tool, i.e., a keysaurus, shows the numerical values of some quantitative parameters of the keywords. The parameters of the keyword have been termed as “Keyword cluster locus indicator (KCLI)”, as these are the predictors of the future orientations (or future path or locus) of the keyword (cluster). Four such indicators have been included to describe the state of a keyword cluster in the proposed tool, which are stability index, integrated visibility index, momentary visibility index and potency index.

The stability index describes the temporal stability of the keyword cluster within the stipulated time span. The keyword cluster that appears regularly for a long time possesses high values of stability index. The visibility index of the keyword cluster describes the appearance of the same in journal articles. The keyword clusters appear in large number of journal articles possess high visibility index. The integrated visibility index encounters single keyword (isolated or clustered) over entire occurrences, while the momentary visibility index encounters multiple keywords (single in case of isolated keyword) over single occurrence. The potency index encounters weightage of a keyword cluster. The keyword clusters that contain wide varieties of keywords possess high weightage, and hence high potency index also. The variations occur in those keyword clusters where large numbers of research projects are commencing. Those keyword clusters having high values of potency index thus describe the thrust areas of research. In the present study, it has been observed that these four keyword cluster locus indicators are independent with each other. These indicators are listed in Table 7, below.

Table 7: Indicators and related phenomenon indicated

Keyword Cluster Locus Indicator (KCLI)	**Particular phenomenon of keyword belonging to the said cluster**	**Corresponding property of the keyword cluster indicated**	**Research trend indicated**
Stability index	Persistence (transience and/or continuance) of keyword	Stability	Persistence of research
Integrated visibility index	Average number of journal-articles covered by single keyword in multiple occurrences	Integrated visibility	Research potential
Momentary visibility index	Average number of journal-articles covered by multiple keywords in single occurrence	Momentary visibility	Research intensity
Potency index	Keyword variety within a cluster	Weightage	Thrust areas of research

Table 8: Indicators and gradation of their numerical values

Keyword Cluster Locus Indicator (KCLI)	Keyword Cluster Property Indicated	GRADATION of Numerical Values of KCLI				
		++ (Very High)	+ (High)	0 (Medium)	(-) (Low)	(-)(-) (Very Low)
Stability index	Stability	Perfectly continuing	Strongly continuing	Moderately Continuing	Weakly Continuing	Transient
Integrated visibility index	Integrated visibility	Very high potential	High potential	Moderate potential	Low potential	Very low potential
Momentary visibility index	Momentary visibility	Very high intensity	High intensity	Moderate intensity	Low intensity	Very low intensity
Potency index	Weightage	Perfectly strong research area	Highly strong research area	Moderately strong research area	Weak research area	Alien

The range of numerical value of any keyword cluster has been divided in five equal zones that is obtained by dividing the difference between maximum and minimum values by five. The names given to different ranges are shown in Table 8. The very high zone is indicated by ++, high zone by + and so on.

Tables 9 through 12 bring together the three fundamental variables of a keyword/keyword cluster (see Table 4), and the four keyword characteristic indicators (see Table 5), along with indicators and gradation of their numerical values (see Table 8) to illustrate sample layouts for the *keysaurus* subject access tool based on keyword clusters collected from the study's three target journals, i.e., *Low Temperature Physics* (Tables 9 and 10), *Chaos* (Table 11), and *Physics of Plasmas* (Table 12).

Table 9: Sample layout of the subject access tool (*Keysaurus*) for the keyword clusters collected from the journal *Low Temperature Physics*

Keyword cluster-name	**N**	**F**	**A**	**A(max)**	**v = F/N**	**m = F/A**	**p = ln(N*F)**	**s = (A/A(max)) *100**
Alloy	54	168	114	270	3.11 (-)(-)	1.47 (-)(-)	9.11 (+)(+)	42.22 (-)
Antiferromagnetism	3	73	14	15	24.33 (+)(+)	5.21 (+)	5.39 (-)	93.33 (+)(+)
Compound	70	448	191	350	6.4 (-)	2.35 (-)	10.35 (+)(+)	54.57 (0)
Crystal	11	56	28	55	5.09 (-)	2 (-)	6.42 (0)	50.91 (0)
Dislocation	9	24	15	45	2.67 (-)(-)	1.6 (-)	5.38 (-)	33.33 (-)

Doping	2	24	10	10	12 (0)	2.4 (-)	3.87 (-)(-)	100 (+)(+)
Electricity	12	56	25	60	4.67 (-)(-)	2.24 (-)	6.51 (0)	41.67 (-)
Electron	34	139	72	170	4.09 (-)(-)	1.93 (-)	8.46 (+)	42.35 (-)
Exchange-interaction	2	41	6	10	20.5 (+)(+)	6.83 (+)(+)	4.41 (-)(-)	60(0)
Exciton	2	20	6	10	10 (0)	3.33 (0)	3.69 (-)(-)	60 (0)
Fermion	3	9	5	15	3 (-)(-)	1.8 (-)	3.3 (-)(-)	33.33 (-)
Ferrimagnetism	3	12	7	15	4 (-)(-)	1.71 (-)	3.58 (-)(-)	46.67 (-)
Ferroelectricity	3	10	6	15	3.33 (-)(-)	1.67 (-)	3.4 (-)(-)	40 (-)
Ferromagnetism	5	100	23	25	20 (+)(+)	4.35 (+)	6.21 (0)	92 (+)(+)
Helium	11	95	28	55	8.64 (-)	3.39 (0)	6.95 (0)	50.91 (0)
Impurity	7	55	20	35	7.86 (-)	2.75 (-)	5.95 (-)	57.14 (0)
Laser	5	7	7	25	1.4 (-)(-)	1 (-)(-)	3.56 (-)(-)	28 (-)(-)
Lattice	5	28	11	25	5.6 (-)	2.55 (-)	4.94 (-)	44 (-)
Magnetism	47	352	122	235	7.49 (-)	2.89 (-)	9.71 (+)(+)	51.91 (0)
Metal	10	29	19	50	2.9 (-)(-)	1.53 (-)	5.67 (-)	38 (-)
Nanostructured-material	12	63	26	60	5.25 (-)	2.42 (-)	6.63 (0)	43.33 (-)
Optics	14	19	17	70	1.36 (-)(-)	1.12 (-)(-)	5.58 (-)	24.29 (-)(-)
Organic-compound	2	38	7	10	19 (+)	5.43(+)(+)	4.33 (-)(-)	70 (+)
Paramagnetism	4	35	14	20	8.75 (-)	2.5 (-)	4.94 (-)	70 (+)
Phonon	8	46	18	40	5.75 (-)	2.56 (-)	5.91 (-)	45 (-)
Plasma physics	6	7	7	30	1.17 (-)(-)	1 (-)(-)	3.74 (-)(-)	23.33 (-)(-)
Plasmon	6	6	5	30	1 (-)(-)	1.2 (-)(-)	3.58 (-)(-)	16.67 (-)(-)
Quantum physics	12	40	28	60	3.33 (-)(-)	1.43 (-)(-)	6.17 (-)	46.67 (-)
Semiconductor	28	129	63	140	4.61 (-)(-)	2.05 (-)	8.19 (+)	45 (-)
Spin dynamics	16	81	40	80	5.06 (-)	2.03 (-)	7.17 (0)	50 (0)

Superconductivity	30	297	90	150	9.9 (-)	3.3 (0)	9.09 (+)(+)	60 (0)
Surface physics	9	19	14	45	2.11 (-)(-)	1.36 (-)(-)	5.14 (-)	31.11 (-)(-)
Thin film	9	62	24	45	6.89 (-)	2.58 (-)	6.32 (0)	53.33 (0)
Tunnelling	3	31	11	15	10.33 (0)	2.82 (-)	4.53 (-)(-)	73.33 (+)
X-ray	4	27	11	20	6.75 (-)	2.45 (-)	4.68 (-)(-)	55 (0)

(For explanation of the symbols given in the top-most row see Table 4 and Table 5)

Table 10: Sample layout of the subject access tool (*Keysaurus*) for the single keywords collected from the journal *Low Temperature Physics*

Single keywords	N	F	A	A (max)	v = F/N	m = F/A	p= ln(N*F)	s = (A/A(max)) *100
Bose-Einstein-condensation	1	39	5	5	39 (+)(+)	7.8 (+)(+)	3.66 (+)(+)	100.00
Specific-heat	1	30	5	5	30 (+)	6 (+)	3.4 (+)(+)	100.00
Conductivity, thermal	1	25	5	5	25 (+)	5 (+)	3.22 (+)(+)	100.00
Band-structure	1	24	5	5	24 (0)	4.8 (+)	3.18 (+)	100.00
Carbon nanotube	1	24	5	5	24 (0)	4.8 (+)	3.18 (+)	100.00
Quasiparticle	1	21	5	5	21 (0)	4.2 (0)	3.04 (+)	100.00
Argon	1	18	5	5	18 (0)	3.6 (0)	2.89 (0)	100.00
Flux-pinning	1	18	5	5	18 (0)	3.6 (0)	2.89 (0)	100.00
Cryogenics	1	16	5	5	16 (-)	3.2 (-)	2.77 (0)	100.00
Fermi-level	1	15	5	5	15 (-)	3 (-)	2.71 (0)	100.00
Fullerene	1	15	5	5	15 (-)	3 (-)	2.71 (0)	100.00
Ab-initio-calculation	1	14	5	5	14 (-)	2.8 (-)	2.64 (-)	100.00
Fermi-surface	1	11	5	5	11 (-)	2.2 (-)	2.4 (-)	100.00
Boson-system	1	9	5	5	9 (-)	1.8 (-)	2.2 (-)(-)	100.00
Fermi-liquid	1	7	4	5	7 (-)(-)	1.75 (-)	1.95 (-)(-)	80.00

(For explanation of the symbols given in the top-most row see Table 4 and Table 5)

It is interesting to note that the stability index (s) for all single keywords are 100, except one keyword Fermi liquid, which is 80. But in case of keyword clusters only two clusters got more than 90 stability index values.

Table 11: Sample layout of the subject access tool (*Keysaurus*) for the *keyword clusters* collected from the journal *Chaos*

Keyword cluster-name	**N**	**F**	**A**	**A(max)**	**v = F/N**	**m = F/A**	**p = ln(N*F)**	**s = (A/A(max))*100**
Atmospheric science	3	7	6	21	2.33 (-)(-)	1.17 (-)(-)	3.04 (-)(-)	28.57 (-)
Biomedical	6	14	11	42	2.33 (-)(-)	1.27 (-)(-)	4.43 (-)	26.19 (-)
Cellular biophysics	4	45	15	28	11.25 (-)	3 (-)	5.19 (0)	53.57 (+)
Circuit theory	7	26	15	49	3.71 (-)(-)	1.73 (-)(-)	5.2 (0)	30.61 (-)
Crystal	4	5	5	28	1.25 (-)(-)	1 (-)(-)	3 (-)(-)	17.86 (-)(-)
Image processing	7	9	9	49	1.29 (-)(-)	1 (-)(-)	4.14 (-)	18.37 (-)(-)
Laser	7	9	9	49	1.29 (-)(-)	1 (-)(-)	4.14 (-)	18.37 (-)(-)
Magnetism	4	5	5	28	1.25 (-)(-)	1 (-)(-)	3 (-)(-)	17.86 (-)(-)
Nonlinear dynamics	9	417	33	63	46.33 (+)(+)	12.64 (+)(+)	8.23 (+)(+)	52.38 (+)
Numerical analysis	2	110	9	14	55 (+)(+)	12.22 (+)(+)	5.39 (0)	64.29 (+)(+)
Optics	22	53	37	154	2.41 (-)(-)	1.43 (-)(-)	7.06 (+)	24.03 (-)(-)
Pattern formation	3	58	12	21	19.33 (-)	4.83 (-)	5.16 (0)	57.14 (+)(+)
Plasma physics	13	16	16	91	1.23 (-)(-)	1 (-)(-)	5.34 (0)	17.58 (-)(-)
Polymer	5	6	6	35	1.2 (-)(-)	1 (-)(-)	3.4 (-)(-)	17.14 (-)(-)
Quantum physics	10	20	17	70	2 (-)(-)	1.18 (-)(-)	5.3 (0)	24.29 (-)(-)
Semiconductor	5	9	6	35	1.8 (-)(-)	1.5 (-)(-)	3.81 (-)(-)	17.14 (-)(-)
Surface science	5	11	11	35	2.2 (-)(-)	1 (-)(-)	4.01 (-)	31.43 (-)
Telecommunication	6	15	10	42	2.5 (-)(-)	1.5 (-)(-)	4.5 (-)	23.81 (-)(-)

(For explanation of the symbols given in the top-most row see Table 4 and Table 5)

Table 12: Sample layout of the subject access tool (*Keysaurus*) for the *keyword clusters* collected from the journal *Physics of Plasmas*

Keyword cluster-name	**N**	**F**	**A**	**A (max)**	**v = F/N**	**m = F/A**	**p = ln(N*F)**	**s = (A/A(max))*100**
Acoustics	4	4	5	12	1 (-)(-)	0.8 (-)(-)	2.77 (-)(-)	41.67 (-)(-)
Astrophysical plasma	3	43	7	9	14.33 (0)	6.14 (-)	4.86 (-)	77.78 (+)
Cyclotron	3	6	4	9	2 (-)(-)	1.5 (-)(-)	2.89 (-)(-)	44.44 (-)(-)
Dielectric function	3	3	3	9	1 (-)(-)	1 (-)(-)	2.2 (-)(-)	33.33 (-)(-)
Dispersion	3	58	6	9	19.33 (0)	9.67 (0)	5.16 (-)	66.67 (0)
Doppler effect	4	6	5	12	1.5 (-)(-)	1.2 (-)(-)	3.18 (-)(-)	41.67 (-)(-)
Electricity	5	11	8	15	2.2 (-)(-)	1.38 (-)(-)	4.01 (-)	53.33 (-)
Electromagnetism	4	4	4	12	1 (-)(-)	1 (-)(-)	2.77 (-)(-)	33.33 (-)(-)
Electron	12	30	17	36	2.5 (-)(-)	1.76 (-)(-)	5.89 (-)	47.22 (-)
Magnetism	15	99	24	45	6.6 (-)(-)	4.13 (-)	7.3 (0)	53.33 (-)
Microwave	4	7	5	12	1.75 (-)(-)	1.4 (-)(-)	3.33 (-)(-)	41.67 (-)(-)
Numerical analysis	2	68	5	6	34 (+)(+)	13.6 (+)(+)	4.91 (-)	83.33 (+)(+)
Optics	5	7	7	15	1.4 (-)(-)	1 (-)(-)	3.56 (-)(-)	46.67 (-)
Plasma physics	63	2670	165	189	42.38 (+)(+)	16.18 (+)(+)	12.03 (+)(+)	87.3 (+)(+)

(For explanation of the symbols given in the top-most row see Table 4 and Table 5)

4 CONCLUSIONS

The result shows that the rate of cluster formation is highest in the journal, *Low Temperature Physics* compared the other two. Some single keywords have also analysed for the journal *Low Temperature Physics*, almost all of which have shown highest stability index, i.e., 100. For all three journals very few clusters possessed (+)(+) values of the indicators, i.e., very high values. The keyword clusters possessing high indicator values may be reckoned as potential keyword or content descriptor for the said subject domain. The keysaurus thus may navigate towards the right keyword or content descriptor that will enable more precise and appropriate searching supportive of accurate and relevant retrieval.

REFERENCES

1) Bertrand A, Cellier J M, Psychological approach to indexing: effects of the operator’s expertise upon indexing behaviour, *Journal of Information Science*, 21 (6) (1995) 459-472.

2) Suraud M G et al, On the significance of databases keywords for a large-scale bibliometric investigation in fundamental physics, *Scientometrics*, 33 (1) (1995) 41-63.

3) Voorbij H J, Title Keywords and Subject Descriptors: A Comparison of Subject Search Entries of Books in the Humanities and Social Sciences, *Journal of Documentation*, 54 (1998) 466-476.

4) Solomon P, Use-based methods for classification development. *Proceedings of the 2nd ASIS SIG/CR Classification Research Workshop*. Washington DC (1991).

5) Hurt C D, Classification and subject analysis: looking to the future at a distance, *Cataloguing and Classification Quarterly*, 24 (1-2) (1997) 97-112.

6) Soergel D et al, Re-engineering thesauri for new applications: the AGROVOC example, *Journal of Digital Information*, 4 (4) (2004).

7) Bates M J, Wilde D N and Siegfried S, An analysis of search terminology used by humanities scholars: the Getty Online Searching Project Report, No.1, *Library Quarterly*, 63 (1) (1993) 1-39.

8) Willett P, Recent trends in hierarchical document clustering: A critical review, *Information Processing & Management*, 24 (5) (1988) 577-597.