Content organization of a subject in a classification scheme and organizing knowledge in the digital environment

Chudamani Kuppahalli

JRD Tata Memorial Library, IISc, Bangalore 560012, India. ksc@library.iisc.ernet.in

Abstract

Physics abstract is one of the most important services available to researchers for literature search. It has organized and reorganized its subject content several times to enable incorporation of new subjects. This study has identified the Physics classification scheme used by Physics Abstract to study Superconductivity development as it profiles the subject in a sucinct and explicit manner. The subject affiliation of superconductivity has changed over a period of time. Hence, an attempt is made in brief here to profile its changing contexts. It was Cammerlingh Onnes who identified the phenomena in 1911. In 1966, in Physics Abstracts, superconductivity was first included under class 15: Solid state physics. The terminology of the Physics special classification scheme is as exhaustive as possible. The revision of schedules has been done many times. Also, the broad structure of Physics Abstracts classification scheme has changed often and only stabilized in the last two or three revisions. One fact to be noted here is that the evolution of a classification scheme is based on the literary warrant observed in the earlier classification. From the previously classified literature new subjects get distilled or fused out. Hence it should be possible to map the previous edition of a classification scheme on to a new edition. Thus a chain of mapping gets developed. Here such a mapping has been attempted for superconductivity between 1999 classification scheme and 1991 classification scheme of Physics Abstracts. A nascent article appears in a journal or a conference. It reaches the information managers desktop. He recognizes a new insight that a new subject has arrived. Such a subject can be posted in a separate file. The classifier classifies the document into known categories. Further, these categories give insight to the relationships of a new subject. For this a metadata approach can be followed. For this purpose all the required metadata and associated information, except journal information, has been entered in Protégé 2000, a knowledge management software. This is owing to the fact that the terms were taken from books and tested for effectiveness. But only limitation of Protégé is that it does not behave like an information retrieval system but can be used by someone knowledgeable about the software. But the advantages of Protégé are that it is a knowledge-based system with high capability of

handling a large set of terms in any field. A separate retrieval mechanism should be built into it. For each document the context hierarchies need to be represened through the metadata like context subjects, keywords, etc. Further indexes cross linked with other metadata fields need to be created to make it an information retrieval system.

Keywords: Classification scheme, Content analysis, Indexing, Metadata, Physics Abstract, Protégé 2000, Superconductivity.

Resumen

Physics Abstract es uno de los servicios más importantes disponibles para los investigadores del campo de la Física. Ha organizado y reorganizado su contenido varias veces con el fin de permitir la incorporación de nuevos temas. Este estudio ha identificado el esquema de clasificación de la Física usado por Physics Abstract para estudiar el desarrollo de la Superconductividad. La ubicación de dicha materia ha cambiado a lo largo del tiempo. Se realiza una aproximación a su evolución. Fue Cammerlingh Onnes quien identificó estos fenómenos en 1911. En 1966, en Physics Abstract, la Superconductividad se incluyó en la clase 15: Física en estado sólido. El esquema de la clasificación es exhaustivo y su estructura ha cambiado con frecuencia, habiéndose observado cierta estabilidad en las dos o tres últimas revisiones. Cada una de ellas se fundamenta en la anterior, produciéndose divisiones y/o unificaciones. En este trabajo se analiza la evolución de la Superconductividad entre el esquema de clasificación de 1999 y de 1991 del Physics Abstract. Asimismo, se aplican metadatos a los nuevos conceptos aparecidos en la literatura actual sobre Superconductividad transpasándolos a Protégé 2000, software de gestión del conocimiento. Se observa que es necesario añadir a tal proceso un sistema de recuperación de información.

Palabras clave: Análisis de contenido, Esquema de clasificación, Indización, Metadatos, Physics Abstract, Protégé, Superconductividad.

1 Introduction

Physics abstract is one of the most important indexing services available to researchers for Physics literature search. It has organized and reorganized its subject content several times to enable incorporation of new subjects. Over a period of time it has reorganized its subjects, the notation used for representing subjects etc. depending on the need. One of the subjects which has seen major revisions is superconductivity. In this paper development of superconductivity classification will be examined to understand the dynamics of development of content and its organization.

Development of superconductivity can be studied from the analysis of the contents of sources of information such as, reference sources, textbooks, etc. However, it is a lengthy and tedious process. This study has identified yet another source, the Physics classification scheme used by Physics Abstract as it profiles the subject in a succinct and explicit manner.

2 Physics Abstract and its classification scheme

Physics Abstract which was Pt A of Science Abstract from 1903 – 1942 used a classification scheme which had primarily six broad classes, namely, general Physics, light,

heat, sound, electricity and magnetism and chemical physics. Under these divisions subjects were listed alphabetically. An alphabetical index to these subjects is provided at the end of the volume.

The classification incepted in 1898 has changed from time to time whenever there emerged a shift in perspective view. With this in view the profile of the development of the subject will be presented here.

In 1942, Physics abstracts was separated from Science Abstracts and named as Physics Abstracts. The Universal Decimal Classification Scheme was introduced for arranging the Abstracts. The schedule had the following major divisions in its listing: 51 Mathematics, 52 Astronomy, 53 Physics, 54 Chemistry, 55 Geology, 57/59 Biological sciences, 61 Medical sciences, 66/68 Industrial Technology, 77 Photography.

Under these divisions, further subdivisions required to classify the abstracts were listed (see Appendix H of the scheme). The UDC Numbers were combined using the colon (:) symbol to indicate compound subjects.

In 1961, the new version contained broad divisions listed along with subdivisions enumerated under them. Initially no notation was used to signify this arrangement. But, in 1962, contents page was introduced. Later in 1964, new scheme of classification was introduced to accommodate the expanding universe of Physics knowledge. This scheme has undergone 4 revisions since its inception. The next section deals with the scope of the subject mainly in relation to its main divisions.

2.1 Scope of the subject

The scope of the subject as seen from the initial classification scheme points to inclusion of only traditional classes deemed to belong to physics, namely, general physics, light, heat magnetism and chemical physics. When this was revised using UDC in 1942, mathematics, chemistry, Astronomy, geology (geophysics), biology, medicine, industrial technology, photography were added to the main divisions. Thus, the scope got expanded. Then, later in 1964, when the new classification scheme for physics was introduced, new subjects like mathematical physics, nuclear physics, atomic and molecular physics, physical chemistry were added to the main divisions. There was a major revision to the Scheme in 1973. Further, there have been 4 revisions carried out and the latest is the revision of 1999. From studying these revisions of schedules after 1973 the following facts can be noticed.

1. There are no noticeable changes in the first order divisions except for some relocations; ex. geophysics and astronomy and astrophysics.
2. Sometimes a second order subdivision or a third order subdivision gets elevated to the first order or second order division. Though superconductivity was initially placed as a third order subdivision under electrical properties has been elevated to second order division in 1978 because of literary warrant and understanding that it is a solid state phenomena.
3. Physics Abstracts uses a special classification scheme used for organizing articles in physics abstract. Dewey Decimal Classification is a general classification scheme employed by libraries to organize books.
4. The terminology of the special classification scheme for Physics is as exhaustive as possible. The revision of schedules has been done many times. Also, the broad

structure of Physics Abstracts classification scheme has changed often and only stabilized in the last two or three revisions.

5. One new feature noticeable in the 1999 edition is that a correspondence has been drawn between the detailed headings of the scheme used by Journal of the Acoustical Society of America and the Physics classification scheme. aeroacoustics and atmospheric sound, etc subdivisions carry the same notation as that used in the acoustic classification; viz. 43.25, 43.28, etc. This makes it possible to search the two schemes using the same broad classes.
6. Thus, as pointed out in a paper by Chudamani (1994), it is possible to update a general scheme using a special scheme if some correspondence can be established between the subdivisions enumerated as illustrated in the case of acoustics from the International Classification scheme for Physics and Astronomy 1992, it is possible to update a special scheme from another special scheme; Ex. Acoustics has been updated using the Journal of the Acoustical Society of America classification scheme. This can bring about compatibility between classification schemes2.
7. The structure of knowledge did not change for nearly 40 years. During the period only few subdivisions to these main divisions were added. Later, when UDC was employed a thorough revision has been attempted – i.e. 8 main classes which are totally different from physics as explained earlier. In 1961, when a new classification was started, acoustics, nuclear physics, atomic and molecular physics, solid-state physics, physical chemistry all emerged out as main classes enjoying the status of other traditional divisions. All these developments are based on literary warrant and emergence of new subjects by research.
8. Whenever new subjects emerge by any one process, they are given first order, second order, and etc., status in the verbal plane depending on its formation level. However, if it has to be fit into a notation then a number of problems are encountered as can be seen from the revision of the notation in physics abstract. A drastic revision of the notational plane makes it difficult to incorporate the changes in a system. Reorganization on the verbal plane however can be easily effected.

Now, let us observe the development of superconductivity

3 Superconductivity and its development

The subject affiliation of superconductivity has changed over a period of time. Hence an attempt is made in brief here to profile its changing contexts. To repeat, it was Cammerlingh Onnes who identified the phenomena in 1911. However, in 1966, in Physics Abstracts classification scheme, superconductivity was first included under class 15: Solid state physics. In 1968, after rearrangement of contents of Physics Abstracts it was listed under electrical properties of solids (class 17.6) as superconducting materials (17.63) and superconducting materials and devices (17.64). In 1969, it was again shifted to 17.24 under solid state electronics and magnetic properties.

In 1973-74, with the revision of the whole classification scheme, it was assigned a second order subdivision status with class number 8.4 under condensed matter: electrical, magnetic and optical properties (Class 8). When major revisions were undertaken in 1977-78 retaining its second order status, a new class no. 74 under "Condensed Matter Physics" was assigned. Under 74, 22 subdivisions have been enumerated. After this revision, in recent years until 1999 only new subclasses have been added and are listed as 12 further subdivisions. Again in

1999, after a further revision, which is not highly affecting the classification scheme has been implemented (Table of Transformation- Table 1).

4 Evolution of content in a classification scheme

One fact to be noted here is that the evolution of a classification scheme is based on the literary warrant observed in the earlier classification. From the previously classified literature new subjects get distilled or fused out. Hence it should be possible to map the previous edition of a classification scheme on to a new edition. Thus a chain of mapping gets developed. Here such a mapping has been attempted for superconductivity between 1999 classification scheme and 1991 classification scheme of Physics Abstracts.

Table 1. Table of Concordance(transformation) from 1999 to 1991 for Superconductivity

Subject	**Class no in 1999**	**Class no in 1991**
Occurrence	74.10 +v	74.10
Theories and models of sctg state	74.20 –z	74.20
Phenomenological theories (two-fluid, Ginzburg – London)	74.20.De	74.20
BCS Theory and its development	74.20.Fg	74.20F
Non conventional mechanism (spin fluctuation, polaron and bipolarons, resonating valence bond model, anyon mechanism, marginal fermi liquid, Luttinges liquid,etc)	74.20.Mn	74.20
General properties; correlation between physical properties in normal and superconducting states	74.25 –q	74.30
Thermodynamic properties	74.25.Bt	74.30E
Superconducting phase diagram	74.25.Dw	
Transport properties (electric and thermal conductivity, thermoelectric effects, etc)	74.25.Fy	74.30E
Optical properties	74.25-Gz	74.30 ?
Magnetic properties	74.25. Ha	74.30 ?
Electronic structure	74.25.Jb	74.30 ?
Phonons	74.25.Kc	74.30 ?
SO ON		

From the above exercise the following facts have emerged out:

1. Some classes of the earlier edition easily map on to the equivalent classes of the later edition.
2. Some classes distill out of the earlier classes and can be recognized from scope notes if provided.
3. For some classes, though may be distilled out of other classes some literature search has to be undertaken before mapping it.

This can be explained as follows. Usually classification schemes such as Dewey Decimal Classification scheme meant for use in libraries provide for detailed information on a revised schedule. They indicate the revised structure in comparison with the old structure. Then it is very easy to map the classification schedules. Otherwise the index for the different years has to be consulted and a map generated. In libraries it is easy to consult indexes. Then a literature search can be carried out or an expert or a textbook will have to be consulted.

5 Representing subject relationship in Physics Abstract

Initially Physics literature, as noted earlier was organized under broad subjects. Its contents were not much interdisciplinary in nature as viewed from that day's point of view. But today it has expanded so much and a number of subjects have emerged out gaining equivalent status as that of physics; namely, geophysics, astrophysics, etc. Also, electrical engineering has emerged out from electricity and magnetism and so on. Articles were distributed under these headings and an index was appended at the end. As the subject developed, until 1960, UDC, Universal Decimal Classification was fairly suitable for organizing knowledge. It was becoming increasingly inadequate to organize the growing micro thought and finally abandoned it in 1960. This can be mainly attributed to the following facts. 1. The schedules are created mainly to classify macro thoughts and not micro thoughts. 2. When articles are to be classified they are done at micro level. 3. Hence, a detailed classification is necessary to organize the micro thought. The Physics Abstract classification of 1973 has enumerated 5 subdivisions; revision carried out in 1977-8 has enumerated 22 subdivisions. So, articles on all these subdivisions will be classified under 537.31262 in UDC except facility for compounding it with geophysics, chemistry, etc. By using a special classification scheme as the one used in Physics Abstract, it is possible to get detailed schedule of hierarchically related subjects.

Associative subject relationship can be handled by creating a knowledge base of subject relationship as proposed in a paper by Chudamani and Asundi (2002).

6 Content management in the digital environment

A nascent article appears in a journal or a conference. It reaches the information managers desktop. He recognizes a new insight that a new subject has arrived. Such a subject can be posted in a separate file. The classifier classifies the document into known categories. Further, these categories give insight to the relationships of a new subject. For this a metadata approach can be followed.

The metadata can be as follows: term, journal, subjects, BT, NT, RT, cited journals. Context terms, thesterm, classificationcode, use, used, frequency.

Except journal information, all other information has been entered in Protégé 2000, a knowledge management software. This is owing to the fact that the terms were taken from books and tested for effectiveness. But only limitation of Protégé is that it does not behave like a information retrieval system but can be used by someone knowledgeable about the software. A separate retrieval mechanism should be built into it. But the advantages of Protégé are that it is a knowledge-based system with high capability of handling a large set of terms in any field.

On the other hand, citation analysis can also be carried out to know all the related terms, journal articles and books. For example, in the citation analysis of superconductivity conference proceedings (refer paper by author submitted to SIS 2003), it has been noticed that papers on a subject first appear in broad subject areas. A second set of journals which may be non core to the subject will be cited. This will be a first step in managing a new subject if new articles arrive they also can be posted to the database and citation analysis and content analysis continued. Finally, a bibliometric analysis and citation analysis of data can be used to

identify the literary warrant, placement, coding, identifying related terms, narrower terms etc. using a Protégé type database. Finally, a committee can be set up to examine the concept and naming it, if necessary. The multiple placements of the term may also be decided and added to the content managers desk. Once this decision is made, the associated subject lists can be listed in a new metadata element set where the occurrence is recorded in an article indexing system for retrieval. For each document the context hierarchies need to be Chudamani Kuppahalli
represented through the metadata like context subjects, keywords, etc. Further indexes cross linked with other metadata fields need to be created to make it an information retrieval system.

7 Conclusions

This dynamics of superconductivity shows that the development of a subject is a dynamic process and needs to be evolved with the development of the subject based on literary warrant. Tables of development and concordance (or transformation) need to be created based on metadata to have continuity of content organization and management. Also, bibliometric analysis of literary warrant and citation analysis need to be carried out to identify its status, placement etc., in order to keep pace with the developments.

References

CHUDAMANI, K. S. *Citation analysis of conference proceedings, books etc. as a tool for information management in SIS 2003*. Kharagpur: IIT, 2003, n. 9-11 apr., p. 545-550.

CHUDAMANI, K. S. Development of a subject and classification schedules: a historical study of superconductivity. In: *DRTC Annual seminar*, 1994, Paper H.

CHUDAMANI, K. S.; ASUNDI, A. Y. Use of a knowledge base of subject relationship for enhancing retrieval effectiveness in the digital environment. In: *DRTC Workshop on Information Resource Management (Bangalore, India: 2002),* Paper DG, p. 1-7.

CHUDAMANI, K. S.; NAGARATHNA, H. C. *Citation analysis as a tool for user oriented content management of an emerging subject* [electronic resource]. Vizhakhapattanam: SIS, 2005, n. 27-29 <http://sis-india.netfirms.com/sis7.htm.>

Physics Abstract. In: *INSPEC, 1909, 1931, 1942, 1961, 1964, 1973, 1978, 1991,1999.*

Physics and astronomy classification scheme. New York: AIP, 1999.

Protégé [electronic resource]. <http://protege.stanford.edu/>.

RANGANATHAN, S. R. *Prolegomena to Library classification.* Bombay: Asia, 1967.

SAYERS, W. C. B. *Manual of Classification for Librarians and Bibliographers*. London: Grafton, 1955.