DOI: 10.5958/0976-2469.2021.00016.6

LIBRARY HERALD
Vol 59 No 3
September 2021

From Library Classification to Knowledge Management : Developments in Knowledge Organization

PROF. K S RAGHAVAN*

1 WHAT IS KNOWLEDGE ORGANIZATION?

'Knowledge organization' is a branch of library and information science (LIS) / Information Studies concerned with activities such as document description, indexing and classification performed in libraries, databases, archives, etc. The term *'knowledge organization systems'* (KOS) encompasses all types of schemes for organizing information and promoting knowledge management. Thus, in its broadest sense it includes taxonomies, library classification schemes, clustering and categorization schemes that organize materials at a general level, subject headings, authority files that control variant forms of names such as place names, highly structured vocabularies, including thesauri, ontologies, and coding schemes, and less traditional tools, such as semantic networks and word nets. Viewed as a discipline, knowledge organization is the domain in which the heuristics of ordering knowledge are studied.

The roots of knowledge organisation can be found in philosophy. For philosophers, knowledge organisation was primarily an exercise in logically ordering / mapping different branches of knowledge. In the context of libraries, however, classification has been an application-oriented area of interest at least since the 19th century as libraries were confronted with two major tasks.

- Logical and helpful arrangement of printed books, pamphlets, and similar knowledge resources on the shelves of a library to support browsing and retrieval.

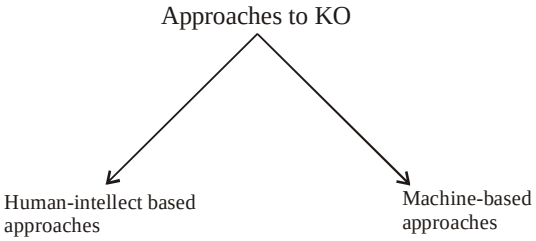
---

* *Based on the 20th Dr. S. R. Ranganathan Memorial Lecture delivered under the auspices of Delhi Library Association on 12th August 2021. FormerProfessor DRTC, Bengaluru, India.*

- Designing and building catalogues/data bases to facilitate search and retrieval of information resources.

2 APPROACHES TO KNOWLEDGE ORGANIZATION

Modern library classification can be said to have begun with the publication of the first edition of Dewey's Decimal Classification in 1876 under the title, *A Classification and Subject Index for Cataloguing and Arranging the Books andPamphlets of a Library*. It was Henry E. Bliss who used the term 'organisation of knowledge' in his famous book, '*Organization of Knowledge in Libraries*'.I. Dahlberg popularized the use of the term *Knowledge Organization*. She founded the International Society for Knowledge Organization (ISKO) and its journal *Knowledge Organization* (Earlier known as *International Classification*). Approaches to KO may be divided into human-based approaches versus machine-based approaches.



Of course, there are other ways of categorizing approaches to knowledge organization:

- Practical and Intuitive approaches: Dewey's scheme; the classification developed by ISI for its databases of journals;
- Consensus-based approach: Henry E. Bliss
- Facet-analytic approach: Ranganathan, CRG, etc.
- Domain-analytic epistemological approaches – I Dahlberg (Ontical Structures); BirgerHjorland; Richard Smiraglia
- Bibliometric approaches: based on co-citation – Garfield and his colleagues at ISI
- Information retrieval approaches: mainly by computer scientists beginning with H.P.Luhn, Gerard Salton, Karen Spark Jones, Van rijsbergen, etc.

There have also been initiatives to make authors / publishers provide metadata (e.g., pre-natal cataloguing / cataloging-in-publication, Dublin Core,

author assigned key words in journal / conference publications, etc.) to enable service providers to identify the class(es) to which a resource belongs. In more recent times, folksonomies - a user-generated classification that allows users to tag resources with their chosen words or phrases selected from natural language (a set of flat, non-explicitly related and uncontrolled terms) - have been used to tackle document organization problems (E.g., ACM Computing Classification System, the del.icio.us folksonomy and the DMOZ open directory).

3 THE CHANGED SCENARIO AND THE ROLE OF KOS

Following the Cranfield Experiments, the IR approach became dominant. Research on traditional library classification schemes lost influence. '*Information Explosion*' – both print-on-paper resources and web-based e-resources –led to the realization that human-based approaches to knowledge organization do not scale-up. Knowledge organization work grows in complexity and costs in direct proportion to the size and granularity of the domain to be organized. The IR approach is generally skeptical of all forms of human interpretation, indexing and classification. Promoted mainly by computer scientists beginning with H.P. Luhnat IBM in the mid-20th Century the research was carried forward mainly by computer scientists including Gerard Salton (Cornell University), Karen Spark Jones (Cambridge Computer Lab), Van Rijsbergen, etc. In more recent years TREC (Text Retrieval Conferences) have emerged as a major forum for reporting and communicating research in this area (https://trec.nist.gov). The techniques coming under this broad category are being continuously refined to improve the quality of indexing/ metadata extraction using technological developments in a range of areas including natural language processing and a variety of statistical techniques. This has largely been necessitated by *Information* explosion and the growth in web-based e-resources, which have made it nearly impossible to manually classify and index the huge volume of information resources being generated. To get an idea of the magnitude and scale of the problem, let us just look at some facts and figures. New York Times (NYTimes.com) alone publishes about 150 articles a day (250 articles on Sundays) and 65 blog posts per day. The screenshot from Reuters' website gives an indication of the magnitude of the problem if all these resources are to be manually indexed to support search and retrieval (Figure 1).

Figure 1: Screenshot from Reuters' Website

The TRC2 corpus (Thomson Reuters Text Research Collection) made available for experimentation to TREC covering the period from January 2008 to end of February 2009 had more than 1.8 million news stories (2,871,075,221 bytes) (Reuters Corpora @ NIST). To make the problem more complex the content types include text, video, audio, to mention a few. Manual indexing simply does not scale when we need to be able to index collections of size in the order of billions of words. Thus, to clearly understand the requirements of future KOS it is necessary to review and re-state the role and purpose of KOS in the present-day context. In other words, the approaches to knowledge organization till the middle of the 20th Century had to largely consider structured / semi-structured data in formal channels of communication such as books, journal articles, theses, etc. The major points of departure from the pre-web era are that knowledge organization should now be to overcome information overload especially in the form of unstructured data, and to represent information in a way that is searchable, retrievable, and reusable. K.O. systems today should accommodate:

- The scale at which they need to operate
- The range of material / resources from which to recall needed information – scholarly material, E-mails, balance sheets, hospital patient records, crime records, images, multimedia, etc. A very high proportion of enterprise content (Over 80%) is unstructured but has immense value for the enterprise.

- The operational space of IR - *institutional, domain-specific, or genre-specific search*

We now restate the purpose of knowledge organization systems in today's context. They are required:

- To be able to categorize information objects automatically without any human intervention.
- To be able to categorize and label information objects in such a way as to achieve acceptable levels of recall, precision, and response time.
- To be able to rank the information objects retrieved in response to a query in an order of decreasing relevance to the information need.

4 NEW AND INNOVATIVE APPROACHES TO KNOWLEDGE ORGANIZATION

Knowledge must be organized, because otherwise it cannot be retrieved and reused. The history of knowledge organization especially after the mid-20th century is largely a fluctuation between the unrealistic ideal of a complete ontology of the universe of knowledge and the more pragmatic approaches. PRECIS, commissioned by the British National Bibliography (now part of the British Library), is an example of a system that did not last long and was abandoned as it proved to be too expensive to apply. Even in the first few years after the arrival of Internet, hierarchical access was the rule,e.g., FTP services, bulletin boards or Gopher services, were hierarchical. The then most popular search service, the Yahoo categories was a classificatory one. All these relied on human work and the huge growth in the number of web pages made it impossible for such approaches to survive. It was realized that while traditional knowledge organization systems such as library classification schemes, lists of subject headings may be suitable for organizing a library collection, they fail when applied to resources such as the ones mentioned in the preceding section. Enhancing and improving the quality of automated approaches was seen as the only way forward to handle the challenges posed by the digital era. In the last couple of decades and more automatic extraction of category labels has been attempted with different degrees of success. In the following paragraphs we only mention a few of such innovative approaches to knowledge organization (It is only illustrative and by no means exhaustive).

41 SOCIAL INDEXING

Google sought to overcome the problems of poor retrieval by adding social indexing, the number of incoming links, clearly borrowed from Garfield's classical citation used for measuring impact. Google has been in corporating a

semantic dimension to its search model. With Page Rank, Google's approach was way ahead of their competitors and has a lion's share of global market.

42 LATENT SEMANTIC INDEXING

As early as in 2004, Google improved its Latent Semantic Indexing (LSI) model expanding its ability to understand synonyms. Its LSI algorithm was developed to use the character strings in a document to establish its semantic relevance to the query (search term).

43 KNOWLEDGE GRAPH

In 2012 Google began the implementation of its "Knowledge Graph" that codifies people, places and things, and the relations among them (Google Knowledge Graph - Wikipedia) to enhance its search engine's capability.

44 KNOWLEDGE MAPS

Some large corporations have employed *Knowledge Maps*, a visual representation of an organization's intellectual capital to manage enterprise knowledge. Open knowledge Maps, e.g., is a visual search engine knowledge (https://openknowledgemaps.org).

Figure 2: Screenshot of Open Knowledge Maps Partial Output (Query: Brain Tumor)

45 NATURAL LANGUAGE INDEXING

Another interesting application is the Natural Language Understanding developed by IBM, which takes a text as input, analyses it using NLP (See Figure 3 below).

Figure 3: Natural Language Understanding

This set of APIs can identify and extract metadata such as concepts, entities, keywords, from a natural language text such as, e.g., a newspaper article.

5 CONCLUDING REMARKS

Clearly extracting metadata from natural language texts has made significant strides since H.P. Luhn's keyword indexing in the middle of the 20th Century. There has been considerable progress in identifying and recognizing named entities in a natural language text. For some of the languages (especially European languages), Part of Speech (POS) Tagging tools with a high level of accuracy have been developed. POS tagging can enhance the quality of selecting appropriate categories / terms from a text, eliminating adjectives and verbs, and selecting nouns and noun phrases as more appropriate class labels. The effect of NL processing techniques on automatic classification and category creation is still not clear; but work is in progress in commercial organizations as also in academic institutions. Does this mean the end of traditional classification and indexing? Probably not. Traditional KOS (library Classification, Thesauri, etc.) will continue to be used in libraries. However, user generated metadata (Folksonomies) / author assigned metadata (Dublin Core; author keywords) will be increasingly relied upon for indexing. Categories and metadata extracted automatically from texts will be extensively relied upon for indexing and retrieval. More specialized IR situations, e.g., such as those obtaining in a corporate environment, health care delivery, etc. requiring high precision will most likely adopt new technologies such as ontology for knowledge representation and management. Knowledge organization for the sake of knowledge organization is a myth. Knowledge organization has always been purpose-oriented and to the librarian and information professional it will continue to be a specific purpose-oriented activity.

DOI: 10.5958/0976-2469.2021.00017.8

LIBRARY HERALD
Vol 59 No 3
September 2021

Ranganathan, As I Knew Him

DR. P JAYARAJAN*

I was with Dr. Ranganathan for almost three and half years, joined him as his Reseach Assistant immediately after completing my BLibSc from the Kerala University, and continued with him till a few days before his demise. Most regretting, that I was not with him when he died.

1 OFFER TO WORK WITH THE GREAT RANGANATHAN

I joined for my B Lib Sc in 1968. Started hearing about Ranganathan from the very first day in the class. Thereafter teacher after teacher in every class talked about Ranganathan at least a few times. Almost all the books they recommended were too of Ranganathan. By the time I finished my degree course, I had made a highly positive impression about Ranganathan. Then most unexpectedly came an offer to work with Ranganathan as his Research Assistant. I jumped at it, and travelled to Bangalore within a few days.

2 RANGANATHAN'S WORK STYLE

Ranganathan's office was within his residence. That was my place of work. My stay was arranged in DRTC hostel. which was only about 5 minutes walk from Ranganathan's residence. The arrangement was that I reach Ranganathan by 8.30 after my breakfast. When I reach him, he was always ready for the day, waiting for me. We start from the visitors room, then move to his bedroom and end up the day again in the visitors room, by about 7 pm, sometime even by 9 pm. He often took a nap in the afternoon. Otherwise he would take his lunch, afternoon snack and early dinner while working. He used to get fully engrossed in his work with no distractions what so ever. He hardly got any relatives or friends visiting him, which suited him! His schedule was very busy: used to get several letters everyday, from within India and

* Former Research Assistant to Dr. Ranganathan and Former Head, British Council Libraries in India.

outside. All letters were read to him same day, either by his secretary or me. He was particular that all letter were replied the same day.

3 WORK, WORK AND WORK!

I worked everyday, no Sunday or holidays. Ranganathan also worked like that , so I had no choice. That was his way of working, even everyone at DRTC also used to work like that. Though this style of Ranganathan and the association with Ranganathan benefitted everyone immensely, many did not like or enjoy the style of his working. Working closely with Ranganathan did not give any room to pursue any other interest, professional or personal.

It is a known fact most of the library leaders of today or yesterday were associated with Ranganathan. I can confidentially say it was my close association with him that gave a firm footing for me and helped throughout my long career. A mere mention of "former Research Assistant to Dr. Ranganathan" did wonders, whenever I met anyone.

He had his own likes and dislikes. Some people whom he trusted, later turned out to be out of his good books. Once he made an opinion about a person, it was very difficult to change that, even if he was wrong. I have seen many cases like that!

4 SIMPLE AND FRUGAL LIFE STYLE

Ranganathan followed simple and frugal lifestyle. He and his wife lived in a small rented house. He did not own a car or any other luxuries in life. If at all he or his wife had to travel, he used taxi. And he made sure that the taxiwallah did not cheat or overcharge! During my 3 year plus stay with him, he was mostly bed ridden, and his outings were mostly to hospitals. He even at that late stage in his life, used to work for more about 10 hours a day, everyday of every week, no holidays whatsoever. He did not get many visitors, personal or professional, which he liked. He thought such visitors would be an interruption to his work schedule.

Ranganathan's total concentration was on "Librarianship". He was not interested in anything else! He liked and even encouraged young professionals to meet him, but to discuss professional matters and not to " pay respects". I had seen him avoiding such casual visitors even coming from far off places.

VIPs and prominent personalities used to visit him, including governors, ministers, heads of institutions and so on. He never made any special arrangements to receive them or meet them. He will be in his usual attire, that is white dhoti and banian, often torn and mended!

5 HEIGHT OF PERFECTION

Ranganathan was a perfectionist. He ensured that every word he used in his writings is the most perfect one. He used to keep a dictionary by his side and often referred to it. He dictated replies to all letters received to his secretary and often made corrections to the typed drafts .

His " medical file" had a full description of his day to day health condition. His doctor had no choice but to read it fully and respond to every question! He used to communicate with DRTC officials regularly. To avoid any confusion, he maintained a "communication" notebook and everything was written in that, including replies.

6 PUNCTUALITY

Punctuality was another quality which he followed strictly and expected others also to follow. He accepted invitations to write about various professional matters from various professional agencies for different occasions. Once accepted, he never went back, and ensured that the write up was delivered on time.

7 FULL FAITH IN NEW GENERATION

He never took shortcuts, and did not like anyone doing so. Within months after joining, Ranganathan suggested that I write a paper on " patents" for a conference. For that he asked me first visit Indian Institute of Science library and see a few patents and study the structure. That made a lot of sense!

Ranganathan always encouraged youngsters. I was a beneficiary. I was introduced to writing within months of joining him. He even helped me to get a paper published in the ALA journal, Library Resources and Technical Services. When he was revising one of his books, he put me as a joint author and I was thrilled to see me as " assisted by P. Jayarajan" next to his name.It was he who encouraged me to join the the DRTC Associateship in Documentation while working with him, in fact he deputed me for the course.

He used to get excited while meeting and talking to young professionals. During a IASLIC Conference held in Indian Institute of Science, the organisers requested Ranganathan to address the delegates. Because of his ill-health he could not go to Indian Institute of Science and instead suggested that the session be held at DRTC. Noticing the response from the young audience and their enthusiasm , Ranganathan went on and on and session went on for a very long time, by the end of which he more or less collapsed and had to be brought down from the stage with help.

During my stay with him, I was totally involved with the preparation of Edition 7 of Colon Classification. Unfortunately he could not complete it, mainly because of his desire to ensure perfection. In fact, the final typescript was sent to the Press, but was withdrawn twice as Ranganathan was keen to make some changes. Any change in the schedule meant, changes in index, illustrations and so on. By that time his health was deteriorating and he had no option but to keep it aside! Pity that Ranganathan could not complete it and sadly we don't have a new edition of Colon Classification.