Library Classification Trends in the 21st Century

CHANDOS
INFORMATION PROFESSIONAL SERIES

Series Editor: Ruth Rikowski
(email: Rikowskigr@aol.com)

Chandos' new series of books is aimed at the busy information professional. They have been specially commissioned to provide the reader with an authoritative view of current thinking. They are designed to provide easy-to-read and (most importantly) practical coverage of topics that are of interest to librarians and other information professionals. If you would like a full listing of current and forthcoming titles, please visit www.chandospublishing.com or email wp@woodheadpublishing.com or telephone +44(0) 1223 499140.

New authors: we are always pleased to receive ideas for new titles; if you would like to write a book for Chandos, please contact Dr Glyn Jones on email gjones@chandospublishing.com or telephone number +44(0) 1993 848726.

Bulk orders: some organisations buy a number of copies of our books. If you are interested in doing this, we would be pleased to discuss a discount. Please contact on email wp@woodheadpublishing.com or telephone +44(0) 1223 499140.

Library Classification Trends in the 21st Century

Rajendra Kumbhar

CP
CHANDOS
PUBLISHING
Oxford Cambridge New Delhi

Chandos Publishing
Hexagon House
Avenue 4
Station Lane
Witney
Oxford OX28 4BN
UK
Tel: +44(0) 1993 848726
Email: info@chandospublishing.com
www.chandospublishing.com

Chandos Publishing is an imprint of Woodhead Publishing Limited

Woodhead Publishing Limited
80 High Street
Sawston
Cambridge CB22 3HJ
UK
Tel: +44(0) 1223 499140
Fax: +44(0) 1223 832819
www.woodheadpublishing.com

First published in 2012

ISBN: 978-1-84334-660-9 (print)
ISBN: 978-1-78063-298-8 (online)

British Library Cataloguing-in-Publication Data.
A catalogue record for this book is available from the British Library.

Typeset by RefineCatch Limited, Bungay, Suffolk.
Printed in the UK and USA.

Preface

Modern librarianship began with the evolution of the printed book. From then onwards there have been manifold developments in all aspects of librarianship. These developments span over newer forms of documents, various tools for the processing and searching of documents, human resource development and services. Further dynamism was added to modern librarianship with the emergence of information and communication technologies. Classification is one of the oldest tools developed by library and information professionals and is considered to be the foundation of librarianship due to its versatile functionality. The origin of various species of classification schemes, such as enumerative, analytic-synthetic and electronic, are some of the significant developments in the field of classification. Library professionals, having a special interest in classification, and institutions established by them have made tremendous contributions to classification. These professionals and institutions not only help develop classification but also many complementary and supplementary tools to organise knowledge. Thus, the functional journey of classification began with the efficient arrangement of printed books on shelves to the management of knowledge and to resource organisation and discovery on the Internet.

The most prominent tool of librarianship, i.e. classification, has thus witnessed many transformations in librarianship over the years. The development brought by the 21st century by way of newer formats for information communication, i.e. digital, and newer methods for the dissemination of information, i.e. networked media, has brought many challenges to classification. These challenges will definitely help it to further strengthen its capabilities so as to serve humanity in creating order out of chaos. Various developments in and around classification are reported by experts through the literature. So an effort has been made through this book to trace the developmental trends in classification as reflected in the library and information science literature published in the last decade, i.e. the first decade of the 21st century. In other words, the book presents a review of literature published on classification.

Methodology, scope and limitations

The present review is based on the abstracts published in the Library and Information Science Abstracts (LISA) (*http://www.csa.com*). LISA is an international abstracting and indexing tool designed for library professionals and other information specialists. It covers more than 440 periodicals from more than 68 countries and in more than 20 different languages.

For the purpose of this review, the LISA was searched using the term 'classification', which has been listed as a prescribed term in the LISA thesaurus. The search was carried out on 3 April 2010. The search strategy included publications from 1999 to 2009. Publications on classification from all languages were searched. The database search retrieved a total of 3,179 references, which included journals, conferences and book reviews. All of the 3,179 abstracts were read, studied, assimilated and sorted into three groups: (i) publications directly related to classification and related topics; (ii) book reviews; (iii) publications *not* related to classification.

This book reviews the publications forming part of the first group mentioned above. Based on the second group of abstracts mentioned above, a list of books on classification along with bibliographic details is given at the end of this book. The publications in the third group are not considered for this review as they do not deal with classification *per se*. The reviews presented in this book are based on abstracts of publications published in English as well as in other languages. Some entries in the list of references do not provide page numbers as that entry in the LISA did not provide any page numbers.

The book is organised into 10 well-articulated chapters. Chapter 1 serves as a preliminary chapter. It reviews literature dealing with the concept and importance of knowledge organisation and knowledge organisation system. Literature highlighting the importance of classification in the digital era is reviewed in Chapter 2. It covers literature that discusses the use of classification in activities such as reference service, in information retrieval, in classification of e-mail, etc. Construction of classification schemes is a valuable but complicated task. Literature dealing with the construction of classification schemes, including notation and translation of classification schemes, is reviewed in Chapter 3. This chapter also discusses the literature that dealt with facet analysis and its application. In Chapter 4, the literature dealing with general classification schemes is reviewed. The literature particularly deals with DDC, UDC and LC.

Chapter 5 reviews the literature dealing with special classification schemes meant for classification of special types of documents, such as patents, comics, grey literature, maps, etc. This chapter also reviews literature on classification schemes developed for classifying non-bibliographic entities, such as business stakeholders, museum objects, industries, etc.

Automatic book classification has been a dream of library and information professionals and their sustained interest and highly specialised developments in the areas of computing and artificial intelligence have been encouraging for the growth of literature on automatic book classification. Literature on automatic book classification, as well as reclassification and non-classificatory approaches to knowledge organisation, is reviewed in Chapter 6. Classification has been an inherent component of library and information science education at all levels and in all countries of the world. Chapter 7 reviews literature dealing with classification education in different countries of the world. Technological advancements, particularly in the field of programming languages, software, development of the World Wide Web, Internet-based resources, etc., has led to the development of newer knowledge organisation tools. These include taxonomies, folksonomies and ontologies. Further efforts are made to evolve interoperable systems that enable crosswalk. Literature dealing with such modern knowledge organisation systems and interoperability is reviewed in Chapter 8. Text categorisation is another modern manifestation of classification and various models and algorithms have been evolved for text categorisation. Chapter 9 reviews literature dealing with text categorisation, its uses and various algorithms such as k-NN, SVM, etc. Chapter 10 concludes the book with a review of literature on classification theories, conferences that deliberated on classification, research trends in classification and noteworthy personalities in classification.

The book presents many emerging trends and challenges that the modern world has put forth for librarianship in general and to classification and knowledge organisation activity in particular. What can be noticed through the review presented in this book is that new areas are emerging in classification. These include, for example, automatic classification; use of classification in Internet resource discovery; and newer knowledge organisation tools, such as taxonomies, folksonomies, ontologies, etc. These trends and challenges are the areas of study and research for budding, enthusiastic library professionals and professionals having a special interest in classification. This book will be a useful research and reference tool for library and information professionals in general and for classification researchers in particular.

Glossary of classification terminology

Classification web: the online resource made available by the Library of Congress for subject headings and classification documentation.

Folksonomy: people who are not subject experts assign subject tags to the websites they use for future search and retrieval of websites. Thus the tags or terminologies used to describe websites assigned by untrained, non-subject expert individuals (folks) is known as 'folksonomy'. This is also known as social tagging or tagging.

Genre: web documents defined primarily by the communicative purpose they accomplish for their producers and recipients.

Genre palette: developed to show the genre ('palette' is a term used to denote a classification, attributable to Karigren, Bretan, Dewe, Hallberg, and Wolkert, 1998).

Social tagging: one of the most widespread services of web 2.0, allows users to provide labels or notes for web contents such as photographs, web links, artwork and other content.

About the author

Dr. Rajendra Kumbhar presently works as Associate Professor at the Department of Library and Information Science, University of Pune (India). Dr. Kumbhar has taught classification, knowledge organization, vocabulary control and related topics for more than 20 years. As a part of his doctoral research he has compiled an depth classification schedule and thesaurus of LIS and has published books and articles on this topic.

The author may be contacted at the following address:

Dr. Rajendra Kumbhar
Associate Professor
Department of Library and Information Science
University of Pune,
Ganesh Khind, Pune 411007
India

E-mail: kumbharrajendra@yahoo.co.in

1

Knowledge organisation and knowledge organisation systems

Abstract: This chapter presents a review of literature dealing with the meaning of knowledge organisation and knowledge organisation systems. It also discusses literature that deals with history and trends in knowledge organisation and knowledge organisation systems. Literature that discusses the importance and functions of knowledge organisation systems is also reviewed in this chapter.

Key words: knowledge organisation; knowledge organisation systems.

Introduction

Human societies have been characterised as knowledge-sharing societies. Knowledge is the sum total of ideas, emotions, beliefs and experiences conserved by the society (Satija, 2008). Knowledge can be categorised into two categories, i.e. subjective knowledge (knowledge as a thought in the individual's mind) and objective knowledge (knowledge as an independent object). Zins (2004) distinguished between subjective and objective knowledge and explained the importance of knowledge organisation by stating that knowledge organisation plays a key role in the creation, learning and dissemination of knowledge. Knowledge organisation has its own epistemological foundation and has implications for: (i) classification research; (ii) development of classification schemes; and (iii) knowledge maps. All of these are core elements of librarianship; knowledge organisation also forms a core component of librarianship.

What is knowledge organisation? According to Hjorland (2008b), knowledge organisation consists of activities such as document description, indexing, classification performed in libraries, databases,

archives, etc. These activities are carried out by librarians, archivists, subject specialists as well as by computer algorithms. Knowledge organisation as a field of study is concerned with the nature and quality of knowledge-organising processes as well as knowledge organisation systems. Knowledge organisation systems are used to organise documents, document representations and concepts. Knowledge organisation has various theories and different historical and theoretical approaches which are related to different views of knowledge, cognition, language and social organisation. Hjorland expressed that each of these approaches tends to answer differently the question: 'What is knowledge organisation?' He finally offers an understanding of knowledge organisation explicitly based on theory of knowledge. All knowledge organisation systems basically organise concepts. Thus, there is a symbiosis between the concept theory and knowledge organisation systems (Hjorland, 2009).

Knowledge organisation is facilitated by a knowledge organisation system. As defined by Hodge (2000) and quoted by Zeng (2008), a knowledge organisation system encompasses all types of schemes for organising information and promoting knowledge management, such as classification schemes, gazetteers, lexical databases, taxonomies, thesauri and ontologies. While providing an overview of these knowledge organisation systems, Zeng (2008) added two more systems – i.e. authority files and synonym rings – and categorised the various knowledge organisation systems into two categories based on their structure (from flat to multi-dimensional) and main functions (e.g., eliminating ambiguity, controlling synonyms or equivalents, establishing explicit semantic relationships, hierarchical and associative relationships, etc.).

Knowledge organisation and knowledge organisation systems – history and trends

Knowledge organisation systems are as old as librarianship itself. Though the present review is of recent literature, scholars have written about the past of knowledge organisation systems and so also about the current trends. While writing on the occasion of the 50 years of the Institute of Information Scientists (UK), Dextre Clarke (2008) presented a historical review and discussed the current status of the developments in knowledge organisation tools, theories and principles. According to Dextre Clarke,

research in classification principles, faceted classification, theories and practices in thesaurus construction, use of technology in the design and development of knowledge organisation tools are some of the landmarks in the development of knowledge organisation tools. By 1958 (the year in which the Institute of Information Scientists was launched) the principles of classification, especially faceted classification, were well established, which provided an excellent springboard for developments in knowledge organisation systems. In the next two decades of the Institute's existence, the principles of thesaurus construction and use were worked out. Then arrived the desktop computer, soon followed by the growth of networks, providing access to an almost unimaginable quantity and variety of resources, which stimulated evolution of knowledge organisation schemes to exploit the technology available.

The changing trends in knowledge organisation are analysed by Saumure and Shiri (2008). For this the authors searched the literature on knowledge organisation in the LISTA (Library and Information Science and Technology Abstracts) database by using the term 'knowledge organisation' or 'information organisation'. The retrieved literature was analysed into pre- and post-web era to explore the current trends in the knowledge organisation. Through the analysis the authors observed that the classic knowledge organisation principles remained prominent throughout both eras but the focus of the knowledge organisation literature has shifted from indexing and abstracting to new areas, such as metadata. Nevertheless, the issues of cataloguing and classification are still dominating the landscape in the post-web era. Saumure and Shiri's findings will be useful particularly to those interested in learning about upcoming trends in the knowledge organisation. Knowledge organisation systems are further gaining momentum because the development of the Internet as a major medium of publication has stretched the capacity of search engines to cope with retrieval (Broughton and Lane, 2000). Like any other field, computers and related technologies are increasingly used for designing and developing knowledge organisation tools. For example, using the WINISIS, the Sarada Ranganathan Endowment for Library Science (SRELS) developed three knowledge organisation tools: an index to Colon Classification, edition 7; a toolkit for learning and teaching postulational approach to classifying subject; and a machine-readable biosciences thesaurus (Neelameghan, 2002c). More information about the developments and trends in knowledge organisation systems with special reference to the online and Internet era can be obtained from Chan and Zeng (2004).

Importance and functions of knowledge organisation systems

What is the real problem in organisation and retrieval of knowledge and how old is it? According to Garshol (2004), to be faced with a document collection and not be able to find the information you know exists somewhere within it is a problem as old as the existence of document collections. Information architecture (which, according to the author, is synonymous with knowledge organisation) is the discipline that deals with the modern version of this problem, i.e. how to organise websites so that users can actually find what they are looking for. Classification, thesauri and taxonomies are the various traditional tools used in the creation of information architecture/knowledge organisation and retrieval.

The initial purpose of knowledge organisation systems was to help in the organisation of printed documents in libraries. With changes in the knowledge-containing formats and increase in their volume, newer knowledge organisation systems were developed. Each of the knowledge organisation systems came into existence with new purpose and/or to add to the strength of an existing knowledge organisation system. Electronic form of documents is one of the newer forms. Nielsen (2008) opines that the different knowledge organisation tools are designed and developed for different purposes. These purposes include: indexing of information and documents; search and retrieval of information; developing semantic-based domain maps; providing conceptual basis for knowledge-based systems, e.g., automated classification systems, etc. Nielsen reported that there are still challenges involved in leveraging the full potential of the above-mentioned knowledge organisation tools for advanced digital library applications. These challenges were addressed at the 7th European Networked Knowledge Organisation Systems/Services (NKOS) workshop.

Do we find a specific trend in the adoption of the various knowledge organisation tools in a specific country? Shiri and Molberg's (2004) survey answers this curiosity. The authors conducted a survey to know which of the various knowledge organisation tools are incorporated by the Canadian digital library collections in their search interfaces. The survey was based on literature review, deep scanning of Canadian governmental and academic institutions' websites, discussions with researchers involved in knowledge organisation and with knowledge organisation people in governmental organisations. The survey revealed that a total of 33 digital library collections widely used knowledge organisation systems such as

thesauri, subject heading lists and classification schemes. It is also a known fact that Canada has long been adopting the various knowledge organisation tools produced in the US. These include the Library of Congress Classification scheme, the Library of Congress List of Subject Headings, the Dewey Decimal Classification and the Sears List of Subject Headings. This is because there are many similarities between the US and Canada. However, Holley (2008) explored that there are also differences between these two countries which require changing terminology, greater depth and precision in subject headings and classification for specifically Canadian topics. The Library and Archives Canada (LAC) is making efforts to provide suitable extensions for LCC and LCSH. Holley analysed the history and philosophy of these efforts.

A knowledge organisation system primarily performs two functions, i.e. searching and browsing. These two are complementary functions. Each of these two functions is performed by different types of knowledge organisation tools. Searching involves the use of subject headings, indexing terms from a controlled vocabulary, or natural language key words. Whereas browsing, of documents on the shelf or of information present on the web, requires the application of some kind of taxonomy or classification scheme (Currier, 2002). All knowledge organisation systems model the underlying semantic structure of a domain and provide semantics, navigation and translating through labels, definitions, relationships and properties of concepts. A knowledge organisation system when embodied as web services facilitates resource discovery and retrieval by acting as a semantic roadmap, thereby making possible a common orientation for indexers and future users, either human or machine. This function of knowledge organisation system is illustrated by Koch and Tudhope (2003, 2004) and quoted by Zeng (2008).

Topic maps are one of the knowledge organisation tools. According to Garshol (2004), topic maps go beyond the traditional solutions in the sense they provide a framework within which not only the traditional tools can be represented as they are, but also extended in ways which significantly improve information retrieval. Thus mapping can be a potential solution to the terminological problems in digital libraries as it provides improved retrieval effectiveness for users and opportunity to use multilingual schemes effectively. Along with these advantages, however, it has labour-intensive, expensive, high-maintenance requirements to keep it up to date. McCulloch, Shiri and Nicholson (2005) discussed these issues and reviewed other solutions for terminological problems. A summarised account of the functions of classification, thesauri and ontologies is provided by Soergel (1999) and Gopinath (2001b).

It could be observed from the above literature dealing with knowledge organisation and knowledge organisation systems that constant efforts are made to design, develop, implement, evaluate and redesign knowledge organisation systems and tools. Beghtol's (2008) article is probably a perfect extension of these efforts, which provided guidelines for establishing ethical policies for knowledge organisation and cataloguing. These policies could be implemented consistently using the ethics toolkit. Beghtol studied three professional codes of ethics and presented the core value of access to information and indicated how these codes govern consensual issues in knowledge organisation and classification. The codes, however, do not address non-consensual issues.

2

Classification and its uses

Abstract: This chapter reviews literature that deals with the importance of classification in the digital era and uses of classification in diverse activities of librarianship. It reviews literature that discusses the use of classification in browsing online catalogues, in the development of thesauri and taxonomy, in organising web-based information sources, in classification of web pages, in classification of search queries, in e-learning and many other areas.

Key words: uses of classification; use of classification in taxonomy development; use of classification in classification of web pages; use of classification in e-learning.

Importance in the digital era

Classification is one of the oldest and most prominent knowledge organisation tools. It is the foundation of librarianship and other knowledge organisation tools. As such, a considerable literature is published on classification. Kim (2003) searched the literature on cataloguing and classification that appeared during the period 2000–2002 and reviewed it, highlighting trends and important developments in these two correlated subject areas.

In spite of being used beneficially for years together, sometimes questions are raised about the usefulness of classification in the modern context. Even experiments are carried out to prove that there is no need for classification in libraries. Particularly after the emergence of the so-called magic technology, i.e. computer and telecommunication technology, the modern community is raising questions such as: whether computer technology can be applied to the classification and cataloguing of natural language texts; whether human classification can be

superseded by the powerful searching capabilities of computerisation; how far computers can be adopted in digital retrieval techniques, etc. In 1999 Satyanarayana answered these questions, saying that as yet there are few signs that automatic procedures are sufficiently developed to completely replace manual ones. Gilchrist (2002) argued that the new technology may have its own merits; nevertheless, the principles of classification and indexing may lead to much more accurate and targeted automatic searching and retrieval than is possible at present. He further asserted that whichever system (*knowledge organisation system*) may prevail, the role of human intellect in grouping and labelling will remain very much in demand. Thus he emphasises that classification is important and human involvement in classification is equally important. Field (1999) too has a similar observation to that of Satyanarayana, and indeed makes a very apt observation. Based on his observation he also provides a valuable piece of advice: that librarians may be developing a tendency to neglect their traditional skills in the rush to embrace the new icons of the information age, such as knowledge management. Classification is an important technique of librarianship and should not be neglected in a rush to embrace modern technologies. But then what is the reality: has library automation undermined the role of classification? Slavic (2002) answers this question negatively. According to Marcella (2002) there are a number of areas where there is scope for the application of classification. These, for example, are: organisation and exploitation of a physical collection, archives, bibliographic records (electronic and printed), Internet resources and organisations' internal information resources.

Classification is surely important in the modern electronic age; it is even more important in the ultramodern semantic web context (Asundi, 2001; Gilchrist, 2002). The role of classification in the modern context is further highlighted by Matveyeva (2002), who stated that the traditional purpose of classification for locating physical objects on the shelves loses its function in the case of remote electronic resources. The other function of classification, categorisation, is more useful in the case of electronic resources, as it promotes the role of classification as a subject-organising tool. Attempts at using library classification schemes in order to organise electronic resources have moved beyond the realm of libraries and their catalogues. Information scientists, database developers and specialists in information retrieval have explored library classification's potential in organising information on the Internet in order to improve browsing and subject searching. Several projects have proved that classification systems such as the Dewey Decimal Classification, Universal Decimal Classification

and Library of Congress Classification can be useful in describing, organising and retrieving electronic resources.

Finally, how important is classification to librarianship? Can the various knowledge organisation tools and other related discoveries and inventions in LIS, including classification, be ranked so as to show their relative importance? Yes, Tyckoson (2007) has listed and describes the top 10 innovations in the history of librarianship; one of these is classification systems. The remaining nine important developments are the printing press, dictionary catalogue, catalogue cards, Poole's index to periodicals, circulation, the telephone, microfilm, the photocopier and data processing.

However, modern technologies such as information and communication technologies (ICT) do have an impact on cataloguing and classification. ICT has its impact on classification. This is proved by the results of the survey conducted to show the impact of ICT on cataloguing and classification in Nigerian libraries. The survey of 84 participants at the Cataloguing, Classification and Indexing Group of the Nigeria Library Association workshop in 2004 showed that the workshop helped develop positive attitudes for the use of ICT for cataloguing and classification purpose (Obajemu and Ibegwam, 2006). The impact of computing technology on classification is further highlighted by Dhyani (1999), who stated that library classification has been metamorphosed by computer technology so as to sustain its relevance to librarianship. It has also proved its relevance in contemporary times by constantly responding to the multifaceted, multidimensional growth of literature and the changing demands of users (Dhyani, 1999).

Library classification schemes have become increasingly available in electronic form and undergone many enhancements that make them attractive for web knowledge organisation. In fact, library professionals have been quite successful in applying library classification to Internet-based information services in a number of projects, both small and large. Yet, many opportunities remain for improving general knowledge organisation tools and using them in new ways. Vizine-Goetz (2002) asserts that general library classification schemes have a very good potential in organisation and retrieval of Internet information sources. The author, particularly, compared the Dewey Decimal Classification (DDC) hierarchy structure with the subject trees of Internet directory services in terms of categories, hierarchies and distributions of postings. Classification schemes are also compared with respect to several general characteristics that support browsing. The author sees great prospects for developing effective DDC-based browsing structures to large collections.

The literature reviewed above makes it clear that classification was an important knowledge organisation tool in the past and it is still important and will be important in the future also. Its importance lies in the fact that the library is a highly organised information store, and this feature is attained by virtue of classification. It could also be noticed from the above literature (and in what follows) that there is revolution in classification concepts, theories and practices. Moreover, according to Mills (2004), there is the relevance of the revolution in library classification that has taken place. The importance of classification also lies in the multiplicity of its functions. Due to this feature, i.e. multiplicity of functions, classification schemes are developed by many communities of research and practice. Unfortunately, the different communities involved in the design and development of classification schemes do not have proper communication amongst themselves, which results in considerable duplication of efforts (Soergel, 1999). On the positive side, such duplication is adding new dimensions to classification.

Specific uses of classification

The literature reviewed in the section above dealt with the importance of classification in general. Classification has its very specific uses in diverse aspects of core librarianship and the literature reviewed in this section deals with specific uses of classification. The sequence of the review is: first, literature dealing with the use of classification in traditional areas of librarianship or core librarianship; and then, literature dealing with the use of classification in the modern context. The examples of the first context are use of classification in collection development, reference service, etc., and of the second context is the use of classification in Internet search, resource discovery, etc.

Use in arranging periodicals and browsing online catalogues

Whether classified or alphabetical arrangement is better for shelving print periodicals is still a matter of debate for academic libraries. Both of the arrangements have pros and cons. However, Glaser (2007) argues that the classificatory arrangement's special feature is that it strengthens patrons' browsing ability. Classification and cataloguing have always

been complementary to each other. Once upon a time the classified catalogue was considered as the most sophisticated subject access device. This was particularly true in the context of research and university libraries. The classified catalogue was appreciated most for two reasons: (i) it collocated catalogue entries on the same subject and its sub-subjects; and (ii) its arrangement reflected the arrangement of documents on the shelves. It was the classified catalogue that further intensified the link between classification and cataloguing. In spite of such a great value, the classified catalogue has almost become a historical phenomenon as its use has declined drastically. The concept of a classified catalogue, or using classification as a form of subject access, has been almost forgotten by contemporary librarians (Bland and Stoffan, 2008). While observing the virtual extinction of the classified catalogue, the same authors have cited a case which showed a ray of hope for the revival of the classified catalogue. In order to enhance the capabilities of their online catalogues, the Western North Carolina Library Network (WNCLN) developed a 'classified browse' feature for its shared online catalogue that used the Library of Congress classification. This feature offers both novice and experienced library users one more way of identifying relevant materials. Markey (2006), too, expressed that only the online catalogue's shelf-list browsing capability enables users to experience classification online. While explaining the support of classification to cataloguing, Markey further stated that cataloguers have been the beneficiaries of most of the advances in classification online and operational online systems are now able to assist them in class number assignment and shelf-listing.

Use in organisation of bibliographic entries

Bibliographical entries can be organised in chronological order by the publication date of the documents listed; or in alphabetical order by author or title and/or subject headings; or can follow a subject classification scheme from the relevant field. Each of these approaches has its own merits and demerits. Amongst these approaches, it is better to prefer the classificatory approach. A faceted classification scheme grounded in solid scientific foundations could be a very good option. For a demonstrative elaboration of such a classification scheme, refer to Zins and Guttmann (2000), who also clarify the scientific foundations of the three-phase structuring methodology that was used in developing the faceted classification for the organisation of bibliographic entries.

Use in collection development

Does classification have a role beyond 'mark and park'? Sukiasyan (2008b) asserts that it has. The author postulated that if the functions of a classifier were combined with those of a cataloguer, then classification could play a role in the process of developing a document collection. With proper work planning and manpower management, it is possible to create a productive work environment for the classifier/cataloguer. Classification does play a vital role even in the development of a collection of inter-disciplinary subjects. Some subjects are of an intense interdisciplinary nature, integrating two or more disciplines, for example psychiatry, with psychology and medicine. Interdisciplinarity poses problems in collection development, particularly the problem of duplication of reading materials. It is worth avoiding unnecessary duplication of reading materials as it leads to wastage of budget, which is continuously dwindling. According to Hiebert (2009), cross-disciplinary classification schemes help in avoiding unnecessary duplication. Going a step further in this direction, Hiebert even developed a mechanism that mapped the National Library of Medicine's classification to the Library of Congress Classification for psychology/psychiatry, which could be used as a guideline for cooperative collection development.

Use in reference service

Categorisation of questions is useful in their effective and efficient answering. Reference questions can be categorised by applying different categories, such as 'where is', 'troubleshooting', 'policy and service' and 'reference'. Fennewald (2006) reported a study that applied these categories to classify online reference questions submitted at Penn State's subject libraries. The study found that the types of question were similar across the three services, i.e. online, e-mail and chat. However, there were differences in the proportion of questions by category. The study indicated that the type of questions submitted differ between in-person and online and also between e-mail and chat.

In 2001, Warner published a skill and strategy-based classification system to classify reference questions. Skill/strategy-based classification and the resource or the time allocated per question-based classification are the two systems available for the classification of reference questions. Testing of these two systems using data from a small academic library proved that the skill/strategy-based approach of classification is

more consistent and provides more accurate reflection of reference desk activity (Henry and Neville, 2008). The same authors, i.e. Neville and Henry (2009), tested the Warner's classification system by another method, i.e. they invited reference librarians from other institutions to compare the technology-sensitive Warner's system to the traditional system of classifying reference questions suggested by William Katz. The authors found that the Warner's system is superior to the traditional resource-based classification system as far as consistency and ease of use was concerned. Another application of classification in reference service is reported by Meserve et al. (2009). Meserve and his co-authors adopted the 'Warner model' of reference question classification to develop tiered reference, adjusting staffing levels and improving service in an academic library which was merged with the main branch of a large city's public library. The authors stated that this model could be used by other libraries that want to develop effective tools for analysing reference services.

Use in information retrieval

The relationship between classification and information retrieval (IR) is explored by Araghi (2004a). The author argued that 'the classification should not be just linked with the organisation of static collection of documents; rather we should view the library as a place where people fill their information needs, and also that librarianship is not mainly about classification, but is a discipline by which we retrieve information and knowledge; we may see a dynamic relationship between classification and information retrieval'. Hjorland and Pedersen (2005), who have been supported by Jones (2005), have also elaborated on the importation of classification and classification theories in IR.

Use in thesaurus and taxonomy development

Classification schemes and thesauri could be used beneficially to develop organisational taxonomies. This is demonstrated by Wang et al. (2008), who developed taxonomy in the information studies domain for the Division of Information Studies at the Nanyang Technology University, Singapore. This taxonomy was built by using the Dewey Decimal Classification, the Information Science Taxonomy, two information systems taxonomies, and three thesauri (ASIS&T, LISA and ERIC). The classification and thesauri proved useful in creating structure and categories

related to the subject of the taxonomy. To get more details about the procedure of building taxonomy based on classification, it is worth referring to this 34-page article in the *Journal of Documentation*. Depth classification schemes structure and list almost all micro aspects of the concerned subject. Speciators are the qualifiers added to isolates so as to make them semantically rich. Kumbhar (2005) discussed the advantages of using a speciator-based faceted depth classification scheme for the construction of a thesaurus. Use of a faceted classification scheme particularly helps in deriving related terms for the thesaurus. LIS professionals have always shown interest in combining the various vocabulary control tools and/or their principles so as to produce a new, more efficient vocabulary control tool. Jean Aitchison's thesaurofacet is one such example. Realising the benefits of integration of knowledge organisation tools and as a result of the influence of the Western classification theory, especially Aitchison's Thesaurofacet, China developed the Classified Chinese Thesaurus. This is developed by integrating classification and subject headings in China. Liang (2007) discusses the history and background of this integration effort.

Use in organising web-based information sources

The Internet contains useful information, but effective organisation of this information is a serious problem. Classification and indexing (as well as the other knowledge organisation tools) prove useful in organising Internet resources (Satija, 2006). These uses may be in the form of developing subject trees, resource discovery, organising and searching digital library resources or may even be in the form of understanding users' navigation patterns. The literature in this section highlights such uses of classification. However, Markey (2006) observed that the information industry and library and information community has not made enough effort to use classification in the organisation of Internet resources, apart from the efforts of the Online Computer Library Center (OCLC). And whatever use is made, Broughton and Lane (2000) claimed, is without a full understanding of modern classification principles. The authors illustrated that the analytico-synthetic schemes have been used crudely, as in the case of the Universal Decimal Classification (UDC). The fully faceted Bliss Bibliographical Classification, second edition (BC2), with its potential as a tool for electronic resource retrieval, is virtually

unknown outside academic libraries. The authors further emphasised that there is an urgent need to apply the principles of classification to Internet and World Wide Web searching in a thorough and rational manner. The writer of this book believes that these observations of Markey as well as Broughton and Lane might hold true at this time. However, at present, there are very many examples where efforts are made to apply classification to the organisation of Internet resources, and these applications are based on a better understanding of classification principles.

Wheatley (2000) identified a new role for bibliographic classification. This new role is in the development of subject trees for Internet search. Subject trees are used for developing the search mechanism on the Internet and adapt principles of conventional bibliographic classification for structuring hierarchic browsing interfaces, thus providing easily used pathways to their selected resources.

Different book classification schemes have different features in the context of their use in organisation of information on the Internet. Dal Porto and Marchitelli (2006) analysed features of different book classification schemes for web indexing and concluded that different book classification schemes are useful in indexing different types of websites. For example, Biblioatipici, a weblog about temporary workers in libraries and Italian documentation centres, is indexed with a homemade faceted scheme; Letture, a diary about reading and books, is indexed by DDC; and Marchitelli's, a blog about digital libraries and open access, is indexed by JITA, the scheme used for indexing e-prints in different open archives. Each of these classification schemes has its own positive features for web classification. For example, the faceted scheme allows multiple accesses to users on the basis of different information requirements. The indexer is able to generate the needed classes in the indexing phase; also, with a small number of facets and foci one can obtain an exponential number of classes. The DDC is coherent and intuitive and the JITA scheme is very simple and essentially pragmatic, intuitive and coherent. A similar observation is reported by Nasir Uddin et al. (2006), who, with the help of case studies, explained that appropriate use of classification structure in the web information architecture helps in creating a user-friendly interface in information retrieval. For example, the hierarchical-enumerative structure is used in web directories, subject gateways and in cataloguing of electronic contents, whereas faceted structure is used by commercial websites to effectively organise and retrieve web documents through multidimensional taxonomies. The user-oriented classification through folksonomy is also emerging. All three of these approaches could be used in the appropriate context. Another use

of classification in online systems is presented by Jansen, Booth and Spink (2008), i.e. a system to classify user intent for web searching. This classification consists of three levels, i.e. informational, navigation and transactions intent. Based on these three levels, the author developed software which automatically classifies queries using a web search engine log. Comparison of this classification system with the manual system of classification showed better accuracy.

Do the modern search engines and other technologies make the classification irrelevant? Not necessarily. For example, Google Print does not 'change everything' regarding the need for professional cataloguing and classification of books; its limitations make cataloguing and classification even more important to researchers, argued Mann (2008). This is because Google Print has its own limitations, e.g.: (i) it fails to map the taxonomies that alert researchers to unanticipated aspects of their subjects; (ii) it also fails to retrieve literature that uses key words other than those the researcher can specify; (iii) it misses not only synonyms and variant phrases but also all relevant works in foreign languages; and (iv) it fails to retrieve desired key words in contexts segregated from the appearance of the same words in irrelevant contexts. Due to these limitations of the Google search interface, researchers cannot use Google to systematically recognise relevant books whose exact terminology they cannot specify in advance. The author argued that cataloguing and classification do provide the recognition mechanisms that scholarship requires for systematic literature retrieval in book collections and thus proves that these limitations of Google make cataloguing and classification even more important.

Using the Internet's capabilities, many institutions are organising e-learning programmes. These institutions make available their course material in digital format. However, these institutions face the challenge of providing appropriate tools for searching and retrieving suitable course material. This challenge could be faced successfully by adopting ontologies (which are useful in describing learning resources) and classificatory categories of IEEE LOM metadata standard (Sanchez-Alonso, 2009).

Yet another use of classification systems is in resource discovery and delivery on the World Wide Web. There are a number of projects that have been developed to automatically assign class numbers to information on the World Wide Web. Some such projects are: the Nordic Wide Area Information Server/WWW Project at Lund University Library in Sweden; the German Harvest Automated Retrieval and Discovery project; SCORPION: automatic classification and indexing using Dewey Decimal

Classification; and the Web Searching Using Classification project carried out by Stanford University, California and IBM. For a description of these projects, refer to Robbins (1999). According to the author there are number of advantages of using classification systems for resource discovery and delivery on the web. Lim (2000) also observed that classification and other tools such as cataloguing and indexing can help in overcoming the shortcomings of Internet search engines. Further evidence of this is that there are a number of south-east Asian subject gateways that have used traditional classification schemes to develop mechanisms for cross-searching and cross-browsing.

There is a growing need for practical solutions to provide flexible access to digital documents in a structured form on the World Wide Web. The existing library classification schemes serve as good bases for achieving this goal (Toth, 2002). Automatic classification is useful in searching digital libraries (Liu et al., 2000). With the passage of time more and more digital libraries are being developed in the world. These digital libraries adopt one or the other knowledge organisation system, such as bibliographic classification schemes, thesauri, lists of subject headings, ontologies, etc. Bhojaraju and Urs (2006) studied and compared knowledge organisation systems used by major digital libraries of the world. The study revealed that there is variety and diversity in the components of knowledge organisation systems, such as term lists, classification and categories and relationship list components in each category. In addition, there appear to be no norms in the use of knowledge organisation systems. Like digital libraries, institutional repositories also contain electronic information sources to be made available to the user communities in remote places. Mondoux and Shiri (2009) studied knowledge organisation systems used by the 27 Canadian post-secondary institutional repositories, their accessibility and searching features. The study revealed that the new institutional repositories incorporated complex knowledge organisation systems such as controlled vocabularies. Shiri and Chase-Kruszewski (2009) conducted a survey of 269 North American digital library collections with the intention of finding the knowledge organisation systems used by them. The authors found that the Library of Congress Subject Headings is the most widely used subject representation tool, followed by domain-specific thesauri. A total of 113 digital library collections made use of locally developed taxonomies. A few collections used the Dewey Decimal Classification and alphabetical indexes.

Knowledge of users' navigation patterns helps in designing and improving websites. Automatic classification of users' navigation patterns

is an effective tool to categorise users' navigation data. Zhiguo, Guishi and Liping (2009) presented the architecture of a prototype system for classifying users' navigation patterns. The proposed prototype includes: (i) primary web-log pre-processing to extract user navigation sessions from datasets; (ii) mining the representative of user navigation patterns; (iii) representing the content of every web page of user navigation pattern by N-grams; (iv) building N-gram-based user navigation pattern profiles.

Portals are another Internet-based e-resource facility provider. A portal requires a system that can perform the following functions: (i) accurate description of complex digital documents and objects for retrieval purpose; (ii) organisation of those digital entities into a user-friendly and easily navigable or browsable visible structure; (iii) creation of a knowledge structure with deep levels of hierarchy and complex interconnections which can accommodate those entities and place them accurately in the collection to enable resource discovery by a number of routes or search strategies. To perform these functions, Broughton (2002) developed a faceted classification model to organise the humanities electronic resources in the portal. Wilson (2001) presented a survey of current trends in classification of electronic titles. The author also addressed future applications of classification in classifying electronic books.

Use in classification of web pages

Challenges of classification of the web and usefulness of theoretical foundation of a bibliographic classification for the construction of a classification scheme for the web are discussed by Mai (2004b). Most automatic classifications face the problem of a fixed number of categories. To represent and retrieve information from the ever-growing Internet, a large number of categories are needed and/or there should be a facility to add new categories. To overcome this problem, Choi and Peng (2004) proposed a dynamic and hierarchical classification system that is capable of: (i) adding new categories as required; (ii) organising the web pages into a tree structure; and (iii) classifying them by searching through only one path of the tree. The authors claimed that, compared with related systems, this method improved the results by 6 per cent. Liang (2006) and his six co-authors proposed a dictionary-based text categorisation system to retrieve chemical web pages. This system uses a chemistry dictionary to extract chemistry-related information more exactly from

the web pages. After automatic segmentation of the documents to find dictionary terms for document expansion, the approach adopts latent semantic indexing (LSI) to produce the final document vectors, and the relevant categories are finally assigned to the test documents by using the k-NN text categorisation algorithm.

As the Internet provides millions of web pages for each and every search term, getting relevant and required results quickly from the web becomes very difficult. Automatic classification of web pages into relevant categories is the current research topic which helps search engines to get relevant results. Irrelevant words and stop-words reduce the performance of the classifier. Extracting or selecting representative features from the web page help to control this problem. Considering the integration of various feature selection techniques may help to resolve this problem. Devi, Rajaram and Selvakuberan (2008) combined the subset evaluator with the term frequency method, which gave minimal qualitative features enough to attain considerable accuracy.

The growth of the web and the increasing number of documents electronically available has been paralleled by the emergence of harmful web content such as pornography, violence, racism, etc. This scenario created a need for filtering systems. To fulfil this need Hammami et al. (2008) proposed a violent web content detection and filtering system, which uses textual and structural content-based analysis and a dictionary of violent key words. The article also explains the process of preparation of such a dictionary and provides a comparative study of different data-mining techniques to block violent web content. Hammami and his co-authors reported that research is under way to develop effective filtering tools for other types of harmful web pages, such as racist ones, and suggested that research is needed to develop techniques to filter harmful visual web contents. Another classification system for similar purpose is developed by Lee and Luh (2008). Their system is called an inverse chi-square-based web content classification system. The authors claimed that this system can work along with an incremental update mechanism for incremental generation of a pornographic blacklist. This system operates on the linking structure of pornographic hubs to locate newly added pornographic sites. It can classify English and Chinese web pages. The precision rate achieved is 97 per cent. The blacklist created by this system proved better than the blacklists of three public domain agencies.

Images on web pages attract the attention of the user. That is because the text contained in the visual block, also called image-block, contains significant information about the page contents. In order to estimate the

visual image-block's importance, to smooth the term weight according to the importance of the blocks in which the term is located, Fersini, Messina and Archetti (2008) proposed a new system called Inverse Term Importance Metric. This system assigns higher weights to important terms contained in important image-blocks identified by performing a visual layout analysis. The authors claimed that use of this approach with different classification algorithms proved satisfactory.

Montesi (2008) defined web genres and summarised the achievements in automatic classification of web genres. To test his hypothesis that 'a single genre label does not account for the users' perspective', Santini (2008) submitted a restricted number of web pages (25) to a large number of web users (135), asking them to assign only a single genre label to each of the web pages. Users could choose from a list of 21 genre labels, or select one of the two 'escape' options, i.e. 'Add a label' and 'I don't know'. Results showed that users largely disagree on the label to be assigned to a web page. The number and variety of information items on the Internet is constantly growing. The web pages are usually categorised on the basis of subject and/or their genre, such as blog, homepage, etc. Considering the usefulness of detecting the web page genre automatically, Kanaris and Stamatatos (2009) developed a method to automatically extract the feature set of web page style for identifying their genre. The authors used the character N-grams of variable length and HTML tags, which are language independent and can be easily extracted and even they can be adapted to the properties of the still evolving web genres and noisy environment of the web. The authors conducted an experiment based on two publicly available corpora. The result showed that the performance of the proposed approach is superior to other existing methods. The results also showed that character N-grams are better features than words when the dimensionality increases while the binary representation is more effective than the term-frequency representation for both feature types. A series of cross-check experiments further proved the robustness of this approach.

Chaojun, Yi and Xiaojiang (2009) recognised users' personal domain knowledge by observing their Internet surfing actions and background knowledge. Using this personal domain knowledge, they built a knowledge bank to guide the spider's crawling behaviour. The spider downloads the web pages, predicts each URL's relevancy to the specified theme and classifies the resources while downloading the web page using the fuzzy rule-based reasoning algorithm. In order to test their algorithm, the authors developed a basic educational resource-gathering system. The experiment showed that fuzzy rule-based reasoning

algorithm is very effective in boosting the spider's performance and the classification accuracy.

Golub and Lykke (2009) reported a study aimed to investigate: (i) the usefulness of the Engineering Index (Ei) classification scheme in browsing; and (ii) the performance of an automated classification algorithm based on the Ei classification scheme. The user study conducted by the authors indicated that the Ei classification scheme is suitable for browsing. However, it was correlated, i.e. it depended on classification correctness. The authors found that describing class captions and/or listing their subclasses from the start; allowing word search from class captions with synonym search (easily provided for Ei since the classes are mapped to thesauri terms); returning the hierarchical tree expanded around the class in which caption the search term was found, are some of the improvements suggested for successful browsing with the Ei classification scheme. It is worth finding other reasons for browsing failure. Evaluation of the classification algorithm indicated that the automatically assigned classes were on average partly correct, with some classes working better than others. Overall, this is a unique study as it is a user-based evaluation of automated subject classification in the context of browsing, not adopted so far. Most search engines have low search accuracy for full-text search based on key words. To overcome this problem Jing and Hanqing (2009) designed an intelligent search engine based on subject portal, i.e. Agricultural History Portal's Search Engine. This search engine collects web pages, indexes and classifies them automatically. The article also discusses the problems in using this search engine.

Use in classification of agricultural information

Qingfeng et al. (2009) developed a system of classification and coding for web agriculture information to meet the need of information management and sharing in agricultural websites. For developing this classification scheme, the authors: (i) analysed the characteristics of information classification in domestic representative agricultural websites in China; (ii) considered the advantages and disadvantages of the China Library Classification; (iii) collected information from *www.ccain.net* as experimental material; and (iv) considered the features of agricultural web information. The new classification system suits the network environment, helps in the effective web agriculture information management, raises the information use factor and promotes the sharing of information resources.

Use in classification of search queries

Queries submitted to search engines are inherently ambiguous. But what is the proportion of ambiguous queries? To find the answer to this question, Song et al. (2009) conducted a study that involved: (i) clarification of definition of ambiguous queries with the help of taxonomy of queries, constructed from being ambiguous to specific; (ii) manual classification of queries which helped to know that query ambiguity is to some extent predictable; (iii) automatic identification of ambiguous queries with the help of a supervised learning approach. The study found that about 16 per cent of queries in a real search log are ambiguous.

Use in classification of e-mails

E-mail has emerged as the most used Internet service. It is used by individuals as well as by organisations for speedy communication. The ever-increasing number of e-mails, particularly of spam e-mails, has created a severe challenge in the use of e-mails. What is needed to face this challenge is a classification system which can classify e-mails and identify spams. Considering this need, Yu and Zhu (2009) proposed a new e-mail classification system based on linear neural network trained by perception learning algorithm and non-linear neural network trained by back-propagation learning algorithm. An efficient semantic feature space (SFS) method was introduced in this classification. The traditional back-propagation neural network (BPNN) has slow learning speed and is prone to trap into a local minimum, so the modified back-propagation neural network (MBPNN) is presented to overcome these limitations. The e-mail classification system based on the vector space model suffers from a large number of features and ambiguity in the meaning of terms, which leads to sparse and noisy feature space. So Yu and Zhu used the SFS to convert the original sparse and noisy feature space to a semantically richer feature space, which helps to accelerate the learning speed. The experiments conducted, based on different training set size and extracted feature size, showed that the models using MBPNN outperform the traditional BPNN, and the use of SFS can greatly reduce the feature dimensionality and improve e-mail classification performance.

Spam e-mails could also be detected at the packet level (i.e. layer 3) based on classification of e-mail contents (Marsono et al., 2009). The process includes: (i) pre-classification of (pre-detestation) spam on a per-packet basis, without the need for reassembly; (ii) fast e-mail class

estimation (spam detection) at receiving e-mail servers to support more effective spam handling on both inbound and outbound e-mails; (iii) adoption of the naive Bayes classification technique to support both pre-classification and fast e-mail class estimation on a per-packet basis. The classification system was evaluated on the basis of: (a) constraints on processing byte-streams over network including packet reordering; (b) fragmentation; (c) overlapped bytes; and (d) different packet sizes. The evaluation indicated that the proposed layer 3 classification techniques gives less than 0.5 per cent false positive, which approximately equals the performance attained at layer 7. This experiment showed that classifying e-mails at the packet level could differentiate non-spam from spam with high confidence for a viable spam control implementation on middleboxes.

Yet another classification system for the classification of e-mails has been suggested by Appavu (2009) and his co-authors. The authors proposed a decision-tree-based classification method to trace e-mails that contain terrorism information. It is an incremental method and adopts user-feedback-based extension of a decision-tree induction algorithm named Ad Infinitum. The proposed classification system effectively detects threat-containing e-mails. Capable of operating quickly on large and high-dimensional databases, easy to tune and high accuracy are merits of this system when compared with other popular algorithms such as Decision Tree, Support Vector Machines and Naive Bayes.

Use in knowledge representation

Kwasnik (1999) highlights the role of classification in knowledge representation and discovery for bibliographic databases. The author states that classification schemes have properties that enable the representation of entities and relationships in structures that reflect knowledge of the domain being classified. The strengths and limitations of four classificatory approaches – i.e. hierarchies, trees, paradigms and facet analysis – are described in terms of their ability to reflect, discover and create new knowledge.

Use in the legal information profession

Based on opinions of nine new legal information professionals from both academic and law firm libraries, Sands (2002) concluded that cataloguing and classification skills are important even in the electronic

environment for legal information professionals. Sands' article is useful for knowing the rationality of this sustained importance of classification.

Use in e-learning

Buchel and Coleman (2003) explained that the classificatory structures, such as faceted thesaurus, by way of concept maps and topic maps help in developing e-learning programmes. The author demonstrated the potential of classification in e-learning by using the physical geography as a domain discipline and the test bed of the Alexandria Digital Earth Prototype project.

3

Construction of classification schemes

Abstract: This chapter presents a review of literature dealing with various approaches to the construction of classification schemes. Literature that deals with the various components of classification, such as notation, mnemonics, book numbering system, etc., is also reviewed in this chapter. Literature on software useful for construction of classification schemes and facet analysis and its application in construction of classification schemes is also reviewed.

Key words: postulational approach; software for construction of classification schemes; translation of classification schemes; facet analysis.

Methods of constructing classification schemes

Construction of a classification scheme is a complex and intellectual task. It involves many steps and also needs to apply various theories and principles. Amongst them, sameness and difference are important guiding principles (Olson, 2001). Zins (2004) explored the epistemological foundations of knowledge organisation and discussed the implications of knowledge organisation for the development of classification schemes and knowledge maps. Thomas (2004) provided a useful reading in classification, facet analysis as well as thesaurus construction, including deconstruction and reconstruction of sections of classification systems and thesauri. The literature reviewed in this section deals with various approaches and methods and rationality for constructing classification schemes. It particularly deals with the use of postulates, formation of

hierarchies, professional and naive classification schemes, design of work-centred classification schemes, use of circulation logs and speciators, planes of work, etc.

A classification scheme must have a useful structure. Morphology for the design of a classification structure is given by Gopinath (2001b), who further explained the usefulness of a lexicographic approach to the design and development of a classification scheme. According to Gopinath, while designing a universal classification scheme, one has to keep in mind the 10 qualitative requirements, i.e. simplicity, brevity, freedom from bias, ease of updating, flexibility, ability to regenerate, ability to educate, representation, portability and philosophy.

Postulational approach and hierarchy

In the classification and Ranganathanian context, postulates are the assumptions. Postulates are useful in the construction and use of a classification scheme as well as in organising and searching World Wide Web resources. When Ranganathan's postulates are compared with the Peter Pin-Sen Chen's entity relationship approach to data-modelling and analysis, it is found that both the theoretical approaches are concerned with the organisation of knowledge or information, and apply almost similar theoretical principles, concepts and techniques for the design and development of a framework for the organisation of knowledge, information or data in their respective domains. Both of the approaches are complementary and supplementary to each other. Ranganathan's postulations-based approach or analytico-synthetic approach to knowledge classification can be applied for developing efficient data retrieval systems in addition to the data analysis and modelling domain (Kashyap, 2001; 2003). The procedure of application of the postulational approach is explained by Gopinath and Mangai (2003). The procedure consists of nine steps in which the notation from any classification scheme can be applied, or subject indexes derived according to chain procedure or its variations, and it can also be used for preparing subject analytical entries. The application is illustrated by classifying more than 200 books by Dewey Decimal Classification, Universal Decimal Classification and Colon Classification. Considering the above-mentioned applications and benefits of postulates, it is worth teaching them effectively to LIS professionals in general and to classification students in particular. The toolkit developed by Neelameghan (2002c) could be used for effective teaching and learning of the postulational approach.

From where did the legacy of hierarchy originate? According to Olson (2004), it originated from the Hegelianism and Charles Cutter and Scottish common-sense philosophers. Even the legacy of hierarchy visible in the Dewey Decimal Classification's structure is influenced by these thinkers, thus revealing the connection between Melvil Dewey and Hegelianism and Charles Cutter and the Scottish common-sense philosophers. Principles of hierarchy are adopted while forming hierarchies in classification schemes and taxonomies which help in the logical organisation of a large number of concepts. Information about hierarchies represented in various disciplines, guidance for forming hierarchies and new principles for the formation of hierarchies, i.e. principle of emergent whole and principle of succession of events or processes for the organisation of hierarchies could be found in the article of Neelameghan (2002b). Scholars interested in the development of classification schemes and taxonomies will benefit from this article.

Specific approaches to construction of classification schemes

Work-centred design of classification schemes is an emerging area of research and a key challenge of these schemes is the evolving semantics of work. Albrechtsen and Pejtersen (2003) suggested a work-centred approach to the design of classification schemes, based on the framework of cognitive work analysis. In order to construct a classification scheme by this approach the authors launched a collaborative task situation as a new unit of analysis for capturing evolving semantic structures in work domain. This is further explained with an example from cognitive analysis of actors' negotiations and integration of knowledge perspectives during collaborative work at the three national film research archives.

According to Beghtol (2003), classification schemes are constructed by two different communities for two different purposes: (i) classification for information retrieval purposes, called 'professional classification' because it is developed by people who have professional interest in classification; (ii) classification for knowledge discovery purposes, called 'naive' because it is developed by people who have no particular interest in studying classification as an end in itself. However, Hjorland and Nicolaisen (2004) considered it unfortunate to use such terminology (i.e. 'professional and naive classification'). They also mentioned that such

thinking is based on deeply anchored misjudgements about the relationships between scientific and scholarly classification on the one side and US classifications on the other. Hjorland and Nicolaisen argued that even scholarly classifications are not naive as scholarly activities and research lead to classification systems which subsequently are disseminated in publications which are used in information retrieval systems.

In Chinese Decimal Classification, some highly associated classes are not grouped in the same subject hierarchies, where the users feel useful. Pu and Yang (2003) analysed two years' worth of transaction logs of users' book circulation records from Taiwan University Library and obtained highly associated classes which were not grouped in the same subject hierarchies and added these to the classification scheme. The experiment showed that by adopting this method, a classification scheme can be made more adaptable to changes of users and the uses of different library collections by analysing the circulation patterns of similar users.

Idea, verbal and notational planes are the three planes enunciated by Ranganathan as a part of classification theory. The verbal plane helps to express the idea and the natural language is one of its primary attributes. While stating the importance of linguistic approach in the development of a classification scheme, Adhikary and Nandi (2004) explained that the lexical meaning (obhidha) and the contextual meaning (according to denotative, implicational and suggestive potentiality, following principles of compatibility, expectancy and proximity) help to formulate the canons of verbal plane. The classification is closely linked with language and may be defined as a mean by which the context of information sources is made accessible within established semantic contexts. A speciator-based approach is another approach that could be applied in the development of a classification scheme. Kumbhar (2005) defined 'speciator' as an isolate idea used as a qualifier for an isolate or another speciator and demonstrated how this concept can be usefully applied in constructing a depth classification schedule.

Use of shared origin in evolution or history (phylogenesis) as a basis for classification is one of the common principles. This principle has been applied in general classification by several classificationists in identifying a series of integrative levels, each originated from the previous ones, and using them as the main classes. In special classification schemes, common origin is a key principle in many domains: examples are given from the classification of climates, of organisms and of musical instruments. However, Gnoli (2006) observes that using only the principle of common origin can produce confusing results, such as having birds as a sub-class of reptiles. The author suggests and displays how we can use a

well-balanced mix of common origin and similarity in developing a general classification scheme. The author further claimed that Charles Bennett's notion of logical depth could be a promising conceptual tool for this purpose.

Most continents and countries, excluding those from where a classification scheme originates, do not get satisfactory treatment in the classification scheme. This thought is substantiated by the writing of Ndakotsu (2006), who observed that the present major book classification schemes, such as Dewey Decimal Classification and Library of Congress Classification, do not have enough provision for the African continent. Considering this fact, the author suggested that African LIS professionals should come up with an Africa-based classification scheme. Once such a classification scheme is developed, it will help the revisers of the major schemes (such as Dewey Decimal Classification, Library of Congress Classification, etc.) to integrate the African continent properly into the main classification schemes.

The above literature provides guidance for the construction of classification schemes in particular and knowledge organisation tools in general. Mai (2008), however, observes that while the technical level of the design and construction of controlled vocabulary is prescribed in great detail, the methodological level has been somewhat neglected. While one would assume that the growing body of research on human information behaviour might help guide the development of controlled vocabularies and shed light on these contexts, unfortunately much of the research in this area is descriptive in nature and of little use for systems design.

The author demonstrates how the design of controlled vocabularies can be informed by investigations of people's interactions with information. This approach is based on the cognitive work analysis framework and outlines several dimensions of human–information interactions. Application of this approach will result is a comprehensive understanding of the contexts in which the controlled vocabulary will function and which can be used for the development of controlled vocabularies.

Scientists in general and library professionals in particular are interested in creating a logical order of the universe of main subjects. Each of the book classification schemes adopts certain logic in ordering the various subjects and sub-subjects. One more effort in this direction is the consensus map of science produced by Klavans and Boyack (2009). This map is developed by analysing 20 existing maps of science. These 20 maps occur in three basic forms: hierarchical, centric and non-centric

(or circular). Klavans and Boyack's map is circular. At the top of this map circle is mathematics, followed clockwise by physics, physical chemistry, engineering, chemistry, earth sciences, biology, biochemistry, infectious diseases, medicine, health services, brain research, psychology, humanities and, finally, computer science. Thus the link of computer science and mathematics completes the circle. This new map has a high level of correspondence with the 20 existing maps and has many advantages over the hierarchical and centric forms of maps.

Araghi (2004b) proposed a new classification scheme on two principles, i.e. hierarchism and binary theory. Therefore, he called the system Universal Binary Classification. The Universal Binary Classification is useful in preparing subject headings and in constructing a thesaurus. In addition, the terms arranged in tabular form with meaningful features help researchers, through a semantic process, to find what they need. This system is fully consistent with classification of knowledge. Universal Binary Classification codes of randomly selected books of library and information science discipline from the McLennan Library of McGill University were compared with the Library of Congress Classification and found compatible. Szostak (2008) suggested that an exhaustive and universal classification of the phenomena that scholars study and methods and theories they apply is feasible and desirable. Particularly, the author states that such a classification is useful for interdisciplinary scholarship.

Notation

Notations are an inherent part of any book classification scheme. According to Broughton (1999), there are various ways in which notation can be used to express the content of documents. For example, at the most superficial level notation can correspond to the hierarchical structure of the schedules or link to literal components. The notation of compound concepts can express the structure and composition of the compound, and systems exist in which symbols denote the functional roles of the constituent elements and the relationships between them. At the highest level notation can be used to mirror the actual structure of those entities which it represents, as in the case of mathematical systems or chemical compounds. Broughton examined methods of displaying these structures, and questioned the practicality in a documentary context, with particular reference to recent revision work on the chemistry class of the Bliss Bibliographic Classification second edition (BC2).

Mnemonics

A notation must be mnemonic. Systematic, scheduled, seminal, etc., are the various types of mnemonics. According to Neelameghan and Raghavan (2006), seminal mnemonics means identification at the near seminal level, of a similar subset or group of concepts in subject domains and deriving similar patterns of organisation of subsets in different domains. Seminal mnemonics are usually represented by the same digit or digit group to denote seminally equivalent ideas, which might have appeared in different subjects and might have been referred to by different terms. Apart from being useful in designing a subject classification scheme, seminal mnemonics are also useful in the presentation of ideas in technical writing and in discourses, as well as in designing taxonomies.

Book numbering systems

A book number takes shelf arrangement of documents to a point where classification *per se* cannot. Class number alone is not able to uniquely individualise documents on the library shelves, so the function of a book number starts from where that of the class number ends. A chronological book numbering clearly portrays the development of a subject over a given period of time. On the contrary, an author number brings together all the books of an author in one specific subject. Considering the peculiarities of Indian names, the National Library of India developed an author numbering system named 'National Library Table'. The Table covers names across religion and regions. Satija (2007) described this table and stated the need for its revision and marketing so that it will be used outside the National Library of India.

The LC Cutter Table is a book numbering system used along with the Library of Congress Classification scheme. Zhao (2004) discussed the problems in assigning book numbers (Cutter numbers) to printed materials under the Library of Congress Classification system. The existing problems have blocked or invaded the uses of some numbers and letters ruled by the LC Cutter Table. Zhao states that not following the LC Cutter Table well, confusion in using the Table and directly downloading the Library of Congress record to the local database are some of the problems in using the LC Cutter Table.

Just two articles in a decade on book numbers is probably an indication of the waning interest.

Classification terminology

Satija (2000) argued that established terminology is crucial for the development of a discipline and terminology of classification is neither well settled nor widely used by its exponents. Satija examines the terminology generated by Ranganathan and also explains many terms such as canonical classes, systems, special, environmental main classes, kinds of subject (basic, compound, complex), facets, isolates, arrays and chains, etc.

Software for construction of classification schemes

Software has proved useful in the design and development of many useful entities and models in the world. So also, there is software which can be useful in the design and development of classification schemes. Varendorff (2005) describes software useful for the creation of controlled vocabularies, thesauri and business classification schemes. The author further states that MS Excel and MS Word have some limitations in this context. Gaikaiwari (2002) described the use of a computer application program named SRR to assist classificationists and classifiers in the task of designing and developing a scheme for subject classification and using the scheme to classify documents. This program is based on a freely faceted model of classification.

The literature above indicates that efforts are constantly made to evolve new methods for the construction of classification schemes and/or for their revision and modification. Efforts are also made to design and develop special classification schemes devoted to a specific subject.

Translations of classification schemes

In a sense, translation itself is a method to develop a new classification scheme. Translating a classification scheme is a difficult and demanding task. Kwasnik and Rubin (2004) described the difficulties in translating classification systems from a source language and culture to another language and culture. For demonstrating the difficulties, the authors collected kinship terms and concepts from native speakers of 14 languages and analysed them to find differences between their terms and structures

and those used in English. Using the representations of kinship terms in the Library of Congress Classification and the Dewey Decimal Classification as examples, the authors identified the source of possible lack of mapping between the domain of kinship in the 14 languages studied and the Library of Congress Classification and Dewey Decimal Classification. Suggestions are given for making translated classifications more hospitable, linguistically and culturally.

Facet analysis and its application

Faceted classification schemes gained momentum with the special efforts of the Classification Research Group and the contribution of Ranganathan and his contemporaries from the Western world. A historical account of the development and use of faceted classification systems during 1920–69 is presented by La Barre (2007). The 1950s and 1960s seem to have been an intensive period of design, testing and evaluation of faceted classification. This was the period when the international discourse sought to augment and extend the reach of faceted classification through system implementation. The author revisits the heritage of faceted classification, beginning with an examination of the Universal Decimal Classification in its early manifestations. Facet analysis is the most important component in the design and use of faceted classification. On the basis of experience of the Bliss Bibliographic Classification second edition, Mills (2004) explained the practical steps in facet analysis. Panigrahi (2007a) described the method of identification of facets from the title of document, stated the technique of selecting respective category-names of facets to follow the facet formula in general and provided a method for defining the specific facet formula for any specific subject. Facet analysis is a powerful methodology, particularly in dealing with the problems of complex subject description, retrieval and multidimensionality (Broughton, 2001b).

Facet analysis could be used for building knowledge structures which perform a range of functions when adopted as knowledge organisation system tools such as a classification scheme, thesaurus, etc. Facet analysis has a very vital role in information retrieval. Mills (2004) demonstrated the role of classification in information retrieval and in the design of classificatory structures by the application of logical division to all forms of the content of records, subject and imaginative. The natural product of such division is a faceted classification. The latter is seen not as a particular kind of library classification but the only

viable form enabling the locating and relating of information to be optimally predictable.

Apart from the above general uses of facet analysis, it has very specific uses too. One of them is its use in developing a task classification. The nature of the task that leads a person to engage in information interaction, as well as of information-seeking and searching tasks, have been shown to influence individuals' information behaviour. Classifying tasks in a domain has been viewed as a departure point of studies on the relationship between tasks and human information behaviour. Li and Belkin (2008) observed that the previous task classification schemes either classified tasks with respect to the requirements of specific studies or merely classified a certain category of task. As these approaches do not depict a holistic picture of tasks, the authors developed a faceted classification of task, which incorporated both the work tasks and information search tasks. This classification scheme analysed facets and categorised them into generic and common attributes of tasks. The generic attributes included source of task, task doer, time, action, product and goal. The common attributes of task included task characteristics and users' perception of task. This classification scheme could be used to describe users' work tasks and information search tasks and the relationship between these two tasks. It could also be used in designing advanced interactive information retrieval systems.

There are many parallels between representation of building structures (by means of plans, drawings, etc.) and the representation of a subject structure (e.g. by a classification scheme). Neelameghan (2005) observes that due to this parallelism, Ranganathan's Generalised Facet Structure can be applied to both of the domains, particularly for the purpose of hospitality to accommodate new interdisciplinary topics and expression of the relationship between them.

Term-based querying relies upon suitable query formulation; browsing of pre-organised information allows the information to be displayed against a descriptive structure. Traditional classification schemes generally cater for certain viewpoints; however, faceted schemes concurrently describe multiple perspectives. Based on the review of literature related with faceted classification schemes, Giess, Wild and McMahon (2008) observed that the literature specifying the creation of such schemata is sparse and distributed across different domains and eras. The authors proposed means by which the underlying theory of faceted classification may be more readily applied for storage and retrieval of information of engineering designs. This proposal was brought into reality by applying faceted notions outside its 'home'

community to the 'real world' engineering documentation by Wild, Giess and McMahon (2009).

Most of the information retrieval tools (classifications, subject heading lists, thesauri, etc.) developed over the latter part of the twentieth and early 21st centuries demonstrate faceted features. This is visible more in the presentation of product information on commercial websites. There is even an independent strand of theory and documentation related to the application of faceted features (Broughton, 2006). Lists of subject headings are one of the knowledge organisation tools. The Library of Congress Subject Headings is one of the famous lists of subject headings but has three main problems: (i) inconsistent syntax rules; (ii) inability to create headings that are coextensive with the topic of a work; and (iii) lack of effective displays for long lists of subdivisions under a single subject heading in OPACs and similar electronic displays. These problems could be overcome by incorporating fully faceted syntax using the facets of a modern faceted library classification (i.e. the Bliss Bibliographic Classification, second edition) with existing headings. The methodology for accomplishing this task is suggested by Anderson and Hofmann (2006).

Which principles of facet analysis are used in the thesauri and the extent to which different thesauri apply these principles in the same way is examined by Spiteri (1999). Using a self-devised measuring instrument, the author evaluated 14 faceted information retrieval thesauri and concluded that: (i) in some cases the thesauri apply both enumerative-style classification and facet analysis to arrange their indexing terms; (ii) the thesauri do not share a common definition of facet; (iii) the number of facets used in the thesauri are not homogeneous or mutually exclusive; (iv) the principle of synthesis is used in only 50 per cent of thesauri, and there is no consistency in the facet sequence. The following remarks of Aitchison and Clarke (2004) further highlight the value of a faceted approach in thesaurus construction. 'After a period of experiment and evolution in the 1950s and 1960s, a fairly standard format for thesauri was established with the publication of the influential Thesaurus of Engineering and Scientific Terms (TEST) in 1967. This and other early thesauri relied primarily on the presentation of terms in alphabetical order. The value of a classified presentation was subsequently realised, and in particular, the technique of facet analysis has profoundly influenced thesaurus evolution.' Aitchison and Clarke have discussed the influence of facet analysis on the evolution of thesaurus and cited the Thesaurofacet and Art and Architecture Thesaurus as knowledge organisation tools that applied the facet analysis technique to display the terms.

A summary of the papers presented at the National Seminar on Classification in the Digital Environment, organised by the Sarada Ranganathan Endowment for Library Science held in August 2001 in Bangalore, India, is presented by Prasad (2001). The seminar highlighted the use of facet analysis and chain indexing in the digital environment for organising and searching web-based material as well as their use in formulating queries for digital libraries. Zins (2002) examined various classification schemes used in portals and web-classified directories. Zins identifies the inadequacies of these schemes and asserts that the inadequacies prevail because their developers fail to differentiate the various classificatory models and are unaware of their different rationales. Zins identified eight classificatory models for the classification of Internet resources and suggested alternative ways to combine them in a faceted integrated classification scheme. The eight models suggested are: subjects, objects, applications, users, locations, reference sources, media and languages. Broughton (2001b) claimed that faceted classification schemes such as the Bliss Bibliographic Classification are powerful tools for the management of electronic resources. They have the strength to deal with the problems of complex subject description and retrieval and multi-dimensionality. This is because these schemes are characterised by a rigorous analytical approach to terms and clear identification of semantic and syntactic relationships and structures. The process of building knowledge structures on facet analytical principles is explained by Broughton (2001b).

With the rapid proliferation of digital repositories and digital archives comes the need for appropriate and flexible classification schemes that can be implemented in conjunction with current technology such as object-oriented programming techniques. Fox (2005) observes that Ranganathan's theories of classification and concept of faceted approach can be very useful for this purpose. Fox considered its usefulness in the development of semantic web and suggested a theoretical framework for describing and syndicating digital repository content, which is flexible and anticipatory in nature. Ellis and Vasconcelos (1999) highlighted the continuing relevance of facet analysis as a technique for searching and organising World Wide Web-based materials and outlined two approaches underlying WWW indexing and searching, i.e. word and concept-based indexing. The authors argued that facet analysis as an *a posteori* approach to classification represents an excellent approach to searching and organising the results of WWW searches using either search engines or search directories. Facet analysis provides an established rigorous methodology for the conceptual organisation of a subject field, and the

structuring of an associated classification or controlled vocabulary. Due to these features, knowledge organisation tools based on facet analysis are increasingly used in resource discovery in an online environment. Broughton and Slavic (2007) provided an overview of principles and procedures involved in creating a faceted classification scheme for such an environment. Nasir Uddin and Janecek (2007) developed and implemented a faceted classification structure to improve web information organisation. The authors adapted facet analysis theory to use semantic web tools, especially XML (Extended Markup Language) and RDF (Resource Description Framework) store and ontology and demonstrated that classifying and organising information in multidimensional hierarchies is more accessible than simple one-dimensional taxonomic hierarchies.

Thus, it could be noticed that there are various classification schemes developed by adopting the facet analysis technique. Also, the facet analysis technique is increasingly adopted in the digital information environment for purposes such as organisation of information, resource discovery, developing and use of semantic web, etc. Based on the above review, it could be concluded that facet analysis is surging beyond the traditional book classification regime. In spite of its widespread application, there are a number of pragmatic and theoretical issues about the facet notion (Wild, Giess and McMahon, 2009). These issues are: differing interpretations of the facet notion; confusion between facet analysis and faceted classification; lack of methodological guidance; the use of simplistic domains as exemplars; description verses analysis; facet recognition, etc. In an article entitled 'The need for a faceted classification as the basis of all methods of information retrieval', Broughton (2006) compared and provided different understandings of faceted methodology. Both of the above articles indicate that greater guidance about the derivation of faceted analysis is needed and also greater realism is needed when teaching faceted approaches.

4

General classification schemes

Abstract: This chapter reviews the literature published on various general classification schemes such as DDC, CC, UDC, LC, BC, etc. The review includes aspects such as different editions, their description and evaluation, translations and criticism of various classification schemes.

Key words: DDC; CC; UDC; LC; Bliss Bibliographic Classification.

Dewey Decimal Classification

Dewey Decimal Classification (DDC) was designed and published by Melvil Dewey in 1876. Dewey as a versatile LIS professional contributed immensely to the development of library associations, library and information science (LIS) education, standardisation, etc. (Harbo, 2002). However, the most noteworthy contribution of Dewey is DDC. The origin of DDC was the most valuable development in the library and information field in the nineteenth century. Since its first publication in 1876, 22 editions of DDC have been published. At present it is used by more than 200,000 libraries in 135 different countries and is translated into more than 30 languages (Coult, 2002). DDC was primarily devised as a book classification scheme. However, over time it has proved useful for the organisation of bibliographic entries by more than 60 national bibliographies. In the electronic era it is used successfully for the organisation and retrieval of web resources. DDC is certainly a long survived, successful classification scheme and there are a number of reasons for this. Maintenance of the system with suitable, timely changes is one of the reasons. While writing about the DDC on its 125th anniversary, Mitchell (2001) highlighted key features, strengths, the

changes it has undergone, the way it is maintained and how it has survived so well. Niculescu (2009) carried out a comparative study of the various editions of the DDC and arrived at the same conclusion as that of Mitchell: updating terminology, reallocations and expanding the main and auxiliary structure has helped DDC to survive successfully.

For regular, appropriate updates, support of an active institute is required for the survival of any classification scheme because institutes, unlike people, have indefinite life. Fortunately, the Library of Congress, an institute of international reputation, provided active support to DDC. Most importantly, for a book classification scheme to survive successfully, a living environment of books is needed. Availability of such an environment and support is yet another reason for its survival (Tabb, 2001). In fact, the DDC and the Library of Congress formed a successful relationship. It is the longest-serving (more than 100 years) relationship in the world of bibliographic classification. This relationship between the DDC and the Library of Congress is explored by Tabb (2001), though there were conflicts too, e.g. Melvil Dewey refused to modify the sixth edition of the DDC in 1899 to accord with the Library of Congress's own requirements. This refusal resulted in the launch of the Library of Congress's own alphabetical classification scheme (named Library of Congress List of Subject Headings). Similarly, major changes at the LC in the 1990s led to the editorship of the DDC passing from the Library of Congress to the Online Computer Library Centre (OCLC). However, through this development too, like a blessing in disguise, came another strong institutional supporter of the DDC – the OCLC. The formal relationship between DDC and the OCLC, however, came into existence when the OCLC acquired the Forest Press and the rights to the DDC from the Lake Placid Education Foundation in 1988. Having a keen interest in research in DDC, the OCLC has continuously updated it, developed interoperable translations, created systems to map DDC with other classification schemes and even presented the DDC in electronic form (Mitchell and Vizine-Goetz, 2009). Apart from regular updating and institutional support, making itself amenable to the electronic environment is another reason for the DDC's survival.

Use of DDC in the web environment

The emergence of the electronic age has had a wide impact on all aspect of library and information science, i.e. forms of document, processing techniques and tools, processes, services, user behaviour, education, etc.

Of course, classification schemes have limitations in online systems. Nevertheless, Tinker et al. (1999) found that assigning multiple class numbers to electronic collections and the faceted nature of DDC can improve access to electronic materials. A review conducted in 2001 of selected websites that used DDC to organise web resources indicated that the full potential of the DDC scheme for this purpose has not been realised (Saeed and Chaudhry, 2001). At the same time, Saeed and Chaudhry appealed that the DDC should be further enhanced to make it more suitable for organising web resources. DDC as the most powerful tool for the organisation of documents on the shelves responded to both the above appeal and the electronic phenomena by becoming available in electronic form for the first time in 1993. Since then, various versions of the electronic/digital DDC, such as Dewey for Windows and WebDewey, appeared (Lawson, 2001). The WebDewey was made available by the OCLC in July 2003. Mitchell (2003) provides guidance for accessing the WebDewey, a facility that offers online searching and browsing of the DDC. In addition, it maps DDC to Library of Congress Subject Headings (LCSH) and links from the mapped LCSH to the corresponding LCSH authority records. The newly introduced enhancement to WebDewey makes it possible for a library to check its own catalogue while attempting to develop the correct class number for the item being catalogued. It also enables users to link directly from it to the library's own online catalogue and send a DDC number search, bypassing the need to open the catalogue and perform the search there. A detailed account of this enhancement is given by Gonzalez (2005). Another version of DDC created to make it suitable for the present electronic environment is DeweyBrowser. (*www.oclc.org/researchworks/ddc/browser.htm*), OCLC's visual user interface. It allows users to search and browse collections of library resources that have been organised using the DDC. The DeweyBrowser displays results in successive rows of 10 categories based on the three main summaries of the DDC. The categories are colour-coded so as to indicate where the matching record occurs. The system is at present deployed at the WorldCat, which contains several million records and OCLC e-book collection (Vizine-Goetz and Hickey, 2006). In another writing, Vizine-Goetz (2006) explained about a prototype developed to make the most out of the Dewey numbers assigned to library materials and explored new ways of providing access to DDC.

The following literature, reviewed in chronological order, make it even clearer that DDC is a useful classification scheme, even in the web environment.

Stickley (1999) reported the ongoing efforts to enhance the features of DDC through the use of multimedia. Then there are various projects through which the LIS profession is constantly exploring the potential of DDC for the electronic environment. These projects are: (i) Renardus, a European Union project whose goal is to provide users with integrated access to subject gateways and other Internet-based distributed services; (ii) a cooperative project between the Library of Congress and OCLC to develop a prototype for a new reference service based on the Collaborative Digital Reference Service (CDRS) pilot; (iii) the OCLC Office of Research Projects is exploring new approaches to browsing, navigation and the display of results sets (Vizine-Goetz, 2001). The Science Net used DDC for organising large files of links for K-12 science curriculum resources for Canadian schools. The Science Net is a project that aims to provide an alternative to the rigours of sorting through search engine results by providing what are in effect virtual library shelves, which support online browsing that is similar to conventional browsing of library shelves (Martin and Daniels, 2000). The CORE (Cooperative Online Resource Catalog) programme of OCLC used DDC as a metadata element, which offered a unique opportunity for organising and searching web resources (Hickey and Vizine-Goetz, 2001). The use of DDC in the CORE programme proved beneficial to both the DDC and CORE. CORE benefited from having integrated access to a rich classification system and DDC got the opportunity to test and prove its suitability for the networked environment. Scorpion is another DDC-related OCLC project. This project builds tools for automatic subject assignment and worked on the hypothesis that the DDC can be used to perform automatic subject assignment for electronic items (Shafer, 2001). The hierarchical structure of the DDC has the potential for browsing online resources. To further explore this potential of the DDC, the OCLC is conducting research for developing customisable views of DDC, for enhancing links to other thesauri, for improving links to other editions and for transforming captions into end-user language (Vizine-Goetz and Mitchell, 2001).

According to Kepner (2002), DDC's universal nature has further made it useful for classifying Internet websites. This usefulness is proved by the Library and Archives Canada (LAC), which used the DDC for organising web resources in its two projects, i.e. a service that offers links to authoritative web resources about Canada categorised according to the DDC via its website and to manage web content related to Canadian culture (Zeeman and Turner, 2006). However, Zeeman and Turner also expressed that though the DDC has proved useful to organise broad

collection, problems have been encountered in adapting it for specific subject domains. Comparison of DDC with Yahoo, The Open Directory and Looksmart showed that the DDC's subdivisions are rather evenly distributed in different hierarchical levels. This and other general features make DDC suitable for organising Internet resources (Hayati and Tajer, 2006). The present networked environment has created the potential to cross-search or cross-browse groups of networked information services at regional, national or international level. However, this potential has its own problems, too. To study the problem of subject-based retrieval from multiple networked sources using different subject schemes, the JISC UK initiated a funded project named the HILT (High-level Thesaurus) project. Nicholson, Dawson and Shiri (2006) explored DDC's role in this project as a spine in developing subject interoperability solutions.

DDC consists of seven tables, which give it its robust synthetic feature. Each table lists common/semi-special subdivisions. These are applicable to the whole schedule (for example, Tables 1 and 2) or have restricted application for a particular main class (for example, Tables 3, 4, 5, 6 and 7). Table 5 lists racial, ethnic and national groups. Provisions in this table can be used to extend subject access to works about racially mixed people beyond that provided by the rules in DDC. This is possible through the use of 083 fields in the MARC bibliographic format and the techniques developed for the DeweyBrowser beta v2.0 by the OCLC research team (Beall, 2009).

Use of DDC in a non-electronic environment

DDC is the most widely used bibliographic classification scheme worldwide. An e-mail survey carried out by Arellano and Garrido (2000) for knowing the classification schemes used by the Latin American libraries revealed that more than 50 per cent of libraries were using DDC. DDC is most known as a general book classification scheme. As such, it is considered more suitable for general libraries. However, a survey result indicated that it is useful for and so used by special libraries also. An informal survey of art libraries in the UK, conducted via ARLIS e-mail discussion list to know which classification scheme the art libraries are using to arrange their book collection, revealed that DDC was the most preferred scheme. Even some of the art libraries that earlier used other classification schemes, when they went for reclassification, preferred DDC (Currier, 2002).

Though DDC originated as a book classification scheme, it has proved useful for classification of other forms of documents, including newspapers. Should newspapers be classified? And if they are to be classified, which classification scheme should be adopted? After considering the characteristics of newspapers, assessing the various alternatives available for their classification and giving thought to the problems in their classification, Kuhn (1999) arrived at the conclusion that DDC, with few modifications, can be used optimally to classify newspaper collections. Bibliographical entries can be organised in chronological order by the date of their publication, in alphabetical order by author or title and/or subject headings; or can follow a subject classification scheme (Zins and Guttmann, 2000). When it comes to choosing a subject classification scheme for the organisation of entries in a national bibliography, it seems that most choose DDC. Landry (2006a) reported that Germany, Austria and Switzerland started using the DDC for organising entries in their national bibliographies to improve access. Nevertheless, there were certain difficulties in adopting DDC for this purpose. Another national library, i.e. the Swiss National Library, adopted the DDC scheme in 1999 for its open access collections and for organising entries in its national bibliography named Das Schweizer Buch. According to Landry (2006b), the two reasons for this choice are that the DDC is used extensively in Europe and the decision taken by the libraries in Germany to translate DDC 22 (i.e. the 22nd edition of the DDC) in the German language and to adopt it.

Revisions of DDC – editorial teams

Regular revision is the mantra for the successful survival of any classification scheme. The revision has to be in all aspects of the scheme, such as structure, notations, additions, deletions, modifications, relocations, etc. For a meaningful, useful revision the classification scheme must have an expert editorial team consisting of active, involved, intellectual professionals. It requires team/s of experts. It seems that the DDC has got all of these requirements fulfilled. There is an international Editorial Policy Committee for DDC which constitutes members from public, special, academic libraries and even from library schools and it works continuously to keep the DDC updated and relevant to the current library needs (Coult, 2002). The very specific nature of team efforts is responsible for the successful maintenance of DDC (Gauder, 2003).

Revisions of DDC – classes, sections revised, bias

The DDC is revised by way of adding new subjects; by deleting subjects which have lost literary warrant; by relocating subjects at more appropriate locations; and/or by thoroughly revising a schedule of a main class or sub-class. Such a thoroughly revised schedule is called a phoenix schedule. During its lifespan, so far, the DDC has adopted all of these modes of revision. Many a time revisions are made in response to remove bias, if present. The DDC is criticised for bias. The root cause of bias in any classification scheme is that they depict the social systems that were prevalent at the time and place of their origin. Though the DDC is the most widely used book classification scheme the world over, it is based on the shape of the nineteenth-century North American academic world and that is why it shows distinct bias, particularly in the context of religion, depth of geographical area, languages and literature.

The main class '200 Religion' is criticised for its bias against non-Christian religions. One reason for this may be that the religious principles and practices change from area to area. That is why demands are made for and/or expansions are recommended for the religion main class by religious communities of different countries. For example, considering the religious diversity in Korea, in general and in the east Asia region in particular, Dong-Geun and Ji-Suk (2001) made recommendations to adapt and expand the Religion Main Class (200) of the DDC for libraries in Korea. The authors emphasised the Korean and particularly east Asian region. In the proposed revision, only the 220 division is assigned to Christianity as against the present 220–280 divisions, the 230 division is assigned to Buddhism and 240 for other religions of Asian origin. Some of the earlier expansions in the DDC relating to Islam are discussed by S.N. Khan (2004) providing a detailed account of the Dewey class number 297.63, which represents the Sirah of Holy Prophet Muhammad in the 21st edition of the DDC. Based on the 20th edition of the LCSH, Khan has expanded and assigned subject headings for the Sirah. Khan has also provided a list of experts of Islamic studies, consulted for this expansion. Similarly, considering the inadequacy of the 297 class for Islam, the Indonesian librarians and Islamic libraries made efforts to improve the notation 297 of the various editions of DDC 1973 to 2005 and suggested for future adoptions and expansions (Sulistyo-Basuki and Mulyani, 2008).

In general, it is criticised that the main classes 400 and 800 of DDC give a very scanty treatment to non-Western languages and literature.

African language and literature is one of the examples of biased treatment in the DDC. Kua (2004) reported the efforts made at official and local levels in Africa to remove this bias. It is expected that the national libraries should contribute to the appropriate revisions of the DDC in this context.

All general classification schemes have to have provisions for representing racial, ethnic and national facets. However, while designing and redesigning a classification scheme, the designers face many challenges of having balanced and useful provision for racial facets. Whatever provisions are made by a classification scheme for racial facets needs to be evaluated for suitability, adequateness, etc. How should the provisions of racial, ethnic and national facets in a classification scheme be evaluated? Furner (2007) observed that the Critical Race Theory could be useful for this purpose. Adopting this theory, Furner evaluated race-related categories in DDC and concluded that it does not fulfil the expectations successfully. However, efforts are on to 'deracialise' the DDC and Furner believes that the adoption of Critical Race Theory will help bring success in these efforts.

Different editions: description, comparisons and translations

As mentioned earlier in this section, DDC is regularly updated and new editions are brought out to cope with the ever-changing and growing universe of knowledge. Library and information professionals have shown constant interest in knowing details of the different editions of DDC in terms of similarities, differences, additions, deletions, modifications, etc. They have a sustained interest in the comparison of various editions of DDC. This section reviews literature that describes, evaluates and compares different editions of DDC.

DDC20 and DDC21 were published in 1989 and 1996 respectively. The comparison of these two editions indicates that DDC21 has three striking features which make it more superior to DDC20. The three striking features are: (i) reduced bias towards Western material and Christianity; (ii) maximum facetisation so as to enhance its retrieval capabilities in the online environment; and (iii) expansion of Relative Index and the manual (Mandal and Sain, 1999). The DDC22 was published in 2003. This edition of DDC made noteworthy changes in the classes of Religion, Social Groups, Computer Science, Mathematics and Tables (A. Khan, 2004). A detailed description of both printed and

electronic versions of DDC22 is provided by Mitchell (2003, 2004), which includes an account of the new numbers and topics added as well as the structural changes incorporated so as to promote classifier efficiency and DDC's use in the web environment. Mitchell specifically stated that the contents of DDC22 are based on the needs and recommendations of users around the world. For further details about the organisation, new changes and the salient features of the DDC22, one can refer to the article written by Satija (2004).

As soon as a new edition of a classification scheme is published, libraries wish to shift to the new edition, which is essential and a healthy sign. However, shifting from an old edition of a classification scheme to a new edition poses several problems. A. Khan (2004) suggested different strategies for changing successfully from DDC21 to DDC22.

DDC's translation in different languages is one of the reasons for its widespread use. In order to implement DDC in the library network of German-speaking countries, a project has been undertaken to translate the DDC22 into German (Heiner-Freiling, 2006). The Italian edition of DDC immediately incorporated the changes made in the English edition of DDC22. Like the English edition of DDC, the Italian edition is a result of team efforts of different institutions, i.e. Associazione Italiana Biblioteche, Biblioteca nazionale centrale di Firenze and Editrice Bibliografica (Fagiolini, 2009).

The Relative Index is one of the important components of the DDC since its origin in 1876. Miksa (2006) examined the Relative Index of the DDC, from its first edition through to the twenty-second. The aspects covered are its character as a concept indexing system, provision of conceptual context for the terms it lists and the way in which the index intersects with special tables of categories used in the system. Striking features and critical issues such as indexing of synthesised concepts are discussed by Miksa.

Teaching the DDC

Being the most practical and popular book classification scheme, the DDC is preferred all over the world. It is taught in almost all library and information science schools at different levels as part of classification theory. It is also taught as part of classification practicals/knowledge organisation practice courses. Teaching of DDC as a part of classification practicals has almost become *de facto*. As per the purpose, the DDC is taught in practicals at certificate, diploma and bachelor-level courses.

Curriculum, teaching method and evaluation system are the important components of the teaching-learning process. Like any other subject, the DDC could be taught effectively by adopting appropriate teaching methods and tools. For example, weaving Dewey's life story into a lesson plan can help in teaching DDC innovatively and effectively (Prescott, 2001). Based on communication from his colleagues, Taylor (2006) provides the curricula, suggests teaching methodologies and discusses the problems encountered in teaching DDC. Taylor has also provided a list of sources useful for teaching DDC.

The DDC is available in two formats: the printed version and the web-based version, i.e. WebDewey. From the education point of view, it is a matter of curiosity to know which of these two versions will provide better learning performance. For accomplishment of this objective, both of these versions were made available to distance education students at Charles Stuart University (for both study and application purpose). The same support material was provided to all of the students. Results of the experiments did not show any significant difference except that the students who used the printed version performed a little better (Hider, 2004).

In a library where the physical collections are organised as per DDC, familiarising users with the DDC becomes a useful exercise in enhancing users' library literacy. Adoption of attractive user education techniques adds value to user education programmes. More particularly, while educating children about the DDC, one has to use innovative and attractive methods and tools. For example, websites could prove useful in teaching DDC to children. Fortunately, these days there are many websites that could be used in creating awareness about the DDC amongst the children (Junion-Metz, 2002).

Shortcomings of the DDC

It is very difficult to find a classification scheme that satisfies all sections of the concerned community. That means all classification schemes have some limitations in one or another context. DDC too has its own limitations. Jones (2002) enumerates shortcomings of the DDC and submits that DDC disempowers users whereas the simple subject headings help them. Jones has suggested that librarians should find how users actually search books and should make efforts for the development of user-centred classification schemes based on psychology and staff implications.

DDC – action against its illegal use

The recent editions of DDC are published by Forest Press, a division of the OCLC. As an alert organisation it keeps watch on the unauthorised use of the DDC and takes action against its illegal use. For example, the OCLC took legal action against the Library Hotel in New York. Each of the hotel's 10 floors was categorised by DDC and its travel packages incorporated DDC. The OCLC accused the Library Hotel of trademark infringement, unfair competition and false advertisement (Ardito, 2003).

To sum up, it can be observed through the above review that the DDC is growing and so is the literature on it. The growth is in terms of editions, translations, revisions and updates. The growth is also in terms of innovative applications of DDC in the electronic era as well as in terms of number of users all over the world.

Colon Classification

Colon Classification (CC) is the brain child of S.R. Ranganathan. It was first published in 1933. It is now in its seventh edition, published in 1987. CC is used in selected libraries in India. Due to the lack of updated editions, newly emerging libraries are not venturing to use it. Nevertheless, LIS professionals have sustained interest in CC. The citation analysis of Ranganathan indicates that the CC is one amongst the three most cited books of Ranganathan; the other two are *Prolegomena to Library Classification* and *Five Laws of Library Science* (Harinarayana and Raju, 2009). This indicates the continued interest of library and information professionals in the CC. Due to lack of institutional support, very few editions of the CC have been published and it has had a very limited growth. However, its revitalisation is essential and possible (Singh, 1999). For a novice scholar interested in getting primary familiarity with the CC, Singh's (1999) article is a good starting point. The author discusses history, salient features and limitations of CC.

Use of fundamental categories is an important feature of CC. Ranganathan postulated that there are only five fundamental categories, namely: personality, matter, energy, space and time. However, it is difficult to recognise some of these categories. Gopinath (2002) studied problems in the recognition of manifestation of fundamental categories on the basis of a sample of 360 specific topics from four different interdisciplinary subjects. Based on the study, he identified four problems: qualification, multi-connotation, heterogeneity and ALUP in four

different groups of subject formed out of distillation, entity cluster, problem cluster and fusion.

Common isolates is one of the features of CC. Form divisions are represented as common isolates, which are essential components of any bibliographic classification scheme and are referred to by different nomenclature in different classification schemes (Panigrahi, 2007b). Anteriorising common isolates (ACI), posteriorising (PCI), common personality, common matter property and common energy isolates are the various types of common isolates enumerated in CC6 and CC7. Even though there are different and additional common isolates in CC7 than in CC6, there is scope for their further expansion (Shah and Kumar, 2002). As a step towards automation, it is worth considering the preparation of a knowledge base of the common isolates. Selecting category names befitting facet formula and assigning notations is the appropriate method for preparing a suitable knowledge base of the anteriorising common isolates (ACI) (Panigrahi, 2007b). In an another article, Kumar and Shah (2003) examined the time isolates in CC6 and CC7 and suggested for its revision through uniform schedule of time isolates in the library classification schemes. Two expanded tables of time isolates are also part of Kumar and Shah's article.

Like all general book classification schemes, the Colon Classification too has a table listing space isolates. The space isolates in CC are based on political, geographical and administrative characteristics. The provisions of 'mother country' and 'favoured country' in the space isolate table of CC holds great potential. In the context of automation of CC, Panigrahi and Prasad (2005b) mentioned the procedure for developing a knowledge base of divisions and subdivisions of geographical areas. Panigrahi and Prasad also discussed the rules of expert systems developed using Prolog to handle the procedure of picking up class numbers from other schedules and forming the special components to build compound space isolates.

The CC7 was published in 1987 without an index, but a classification scheme without an index does not fulfil the basic feature of a book classification scheme. In the absence of an index, searching of isolates/class numbers from the classification scheme becomes very difficult. To overcome this lacuna of CC7, the Sarada Ranganathan Endowment for Library Science (SRELS) developed an index using the WINISIS software. Neelameghan (2002c) described the procedure adopted in preparing this index.

Order of main classes in the different book classification schemes has always been a matter of curiosity and deliberation among LIS

professionals and there are theoretical as well as practical views about this order. The order of main classes reflects the nature of knowledge, its origin and growth. Satija (2008) defines knowledge as the sum total of ideas, emotions, beliefs and experiences conserved by the society and mentions that its growth is characterised by specialisation, inter and multidisciplinary nature. Classification is knowledge mapping and it changes as per the changing society and time. The order of main classes or knowledge mapping, Satija says, is based on different principles such as those given by D.W. Langridge. Satija evaluated the order of Social Science main classes in CC6 and argued that it does not follow the order of increasing artificiality of social laws as claimed by Ranganathan, rather it follows the order of serial dependency as given by August, Comte and practised by the Library of Congress Classification.

Like CC, the Bibliographic Classification of Bliss is another classification scheme which did not grow much like the DDC or the UDC. Both of these classification schemes share the commonality of being faceted, though there are differences too. Chatterjee (2000) explored the similarities and differences between these two classification schemes with respect to treatment of complex subjects, notations, citation order, etc., and expressed that further study of both of these classification schemes will be of great value.

Uses of CC and its terminology

Apart from its use in classification of documents, the following three articles explain CC's use in other areas. CC is based on sound theoretical principles of facetisation and as such there is scope for its application in knowledge organisation along with artificial intelligence (Singh, 1999). CC contains various devices as isolate sharpening mechanisms. The super-imposition device, alphabetical device, geographical device and subject device are some of the devices used in the CC. Using artificial intelligence and the PROLOG language, an automatic classification system named 'Viswamitra' is developed. This system is capable of constructing class numbers for compound isolates using the various facet sharpening devices (Panigrahi and Prasad, 2005a). Classification, thesauri and taxonomies are the tools used in knowledge organisation. A case study of an online integrated use of CC, bilingual thesaurus and lexicons (through hyper linking) in the domain of Tamil classics proved that the integrated use of these tools enhances organisation and retrieval efficiency (Neelameghan, 2007). Ranganathan was a man of terminology.

He coined and used many innovative, meaningful terms in all of his writings, including the CC. No other individual has contributed terminologically to the LIS profession in general and to classification in particular than Ranganathan. Many of his terms are accepted by conferences and other writers and writings (Singh, 2000).

The above review of the literature dealing with CC indicates that most of the writing published during 1999–2009 is of Indian origin. This may be because CC originated in India and is mostly used in India.

Universal Decimal Classification

Universal Decimal Classification (UDC) originated as a classification scheme for the organisation of bibliographical entries prepared as part of a giant venture of compilation of a classified index to world literature. This project was undertaken by the then International Institute of Bibliography. This institute was later renamed as the International Institute of Documentation and finally as the International Federation of Documentation (FID). The first edition of UDC was published in French between 1904 and 1907 (*www.udcc.org/*). At present the UDC Consortium looks after the maintenance of the UDC. A detailed account of the recent history, development and policies for the revision and maintenance of UDC is provided by Slavic, Cordeiro and Riesthuis (2008). The authors particularly report on the development that took place during 1990–2006. Some of the noteworthy developments during this period include: compilation of the manual for creating the UDC standards, the rise of the UDC Consortium (UDCC), decisions about the expansion of the editorial team at international level and improving technological innovation, upgrading and structuring policies, facilitating the UDC translations and education. The UDC is now maintained in electronic form as UDC Master Reference File (UDC-MRF) (*www.udcc.org/mrf.htm*).

Use of UDC

UDC originated as a classification scheme for the organisation of entries in a classified index and so it has always been proved useful in creating and maintaining an indexing system. Slavic (2004) summarises the functionalities of UDC to be used in system design and highlights issues about the relation between the UDC schedule in electronic form, i.e.

UDC Master Reference File and classification tool (an authority file) that may be built on it. He argued that classification schemes supply an underlying structure to information systems but the specific requirements for their efficient use are poorly implemented, especially in the case of synthetic schemes such as UDC. However, a better understanding of the working of UDC may improve its implementation and reduce the cost of system maintenance. The role of classification has changed as per the evolution of the Internet subject gateways (SGs), from supporting subject organisation on the interface and automatic categorisation of resources to supporting a semantic linking, control and vocabulary mapping between different indexing systems in subject hubs and federated subject gateways. The UDC has proved useful for this new role. Slavic (2006) provides an overview of the history of use of UDC in SGs from 1993 to 2006. While comparing the UDC with the DDC for adaptability, it is suggested that these two systems should be used in conjunction with each other rather than as competing systems (Marsh, 1999). As a general classification scheme, UDC is used worldwide, though it is not favoured uniformly all over the world. A world e-mail survey of the use of UDC conducted by Slavic (2008) revealed that it is used in 208 countries. In some of these countries it is a dominant system; in some countries it is used by few libraries; and in some countries it is rarely used.

Editions/revisions of UDC

Abridged editions of general classification schemes are published to fulfil the needs of small and medium-sized libraries. The DDC has been publishing its abridged edition for many years. The UDC has also published its abridged edition. The creation of an abridged edition needs to adopt special procedures and specific policies. Robinson (1999) provides a detailed account of: (i) the procedures adopted in the creation of an abridged simplified edition of the UDC; (ii) how and which subjects were selected for the abridged edition; and (iii) the difficulties encountered in maintaining consistency between the abridged edition and the complete edition.

The terminology related to the subject 'religion' has many specialties; there are differences in the understanding of the meaning of religious terms, and there is bias towards Christianity in DDC and UDC. Considering these and other features of religion class, Broughton (2000) observed that the present classification schemes do not adequately

represent the users' needs and so the author presented a revised schedule for the religion main class of the UDC based on the users' needs.

The culture and literary history of a country are the main criteria considered while designing the schedules for literary works. The subject of the literary work could also be considered for classifying literary works as research indicates that users prefer richer subject information about literary works. However, most of the classification schemes, including the UDC, do not give much emphasis to the subjects of literary works. To overcome this shortcoming, Pogorelec and Sauperl (2006) proposed an alternative model to enhance UDC's literary class numbers by adding subject information. This alternative model consists of lists of verbal and alpha-numerical denotations for the basic groups of literary works (main genres: lyrics, drama, epics) and all other categorical criteria (language of the original literary work, literature to which the work belongs, genre, sub-genre) and half-categorical (accessibility of the content of literary works, origin within the periods of literary history, the century in which the literary work was written, the rhythm of the language). The authors believed that this new model would be more helpful, efficient and exact. UDC is published in 40 different languages in whole or in part (*www.udcc.org/*). Antoshkova (2008) describes the Russian edition of the UDC, giving details about the major changes made and further changes expected.

Users and researchers of a classification scheme expect that its publisher should keep them regularly updated about the additions, deletions, modifications made or proposed and any other development regarding the classification scheme. To fulfil this requirement of the users and researchers, the UDC Consortium (The Hague, Netherlands) produces a yearly publication in November entitled *Extensions and Corrections*. This publication informs the users and researchers of UDC about the changes and additions made in the UDC editions. It consists of three sections: (i) comments and communications (collection of articles and notes on research, developments and applications of the UDC worldwide); (ii) revised UDC tables (extensions and corrections to the UDC); and (iii) proposals (preliminary drafts of tables in the process of revision, on which UDC users are encouraged to comment and make suggestions that could affect the final result) (Williamson, 2004).

Criticism of UDC

The chain procedure as an indexing technique expects that the classification system used should be an efficient one. As the UDC uses the

decimal notational system, it inherits some limitations at the idea and verbal planes. Problems also arise in using the UDC for the chain procedure due to homonyms, telescoping, missing links and lack of precision in terminology. Kumar and Parameswaran (1999) observed the limitations of UDC for chain indexing and suggested means to overcome them. Sen (2007) criticised that during the past few years the UDC has become more complex. His specific criticisms are: more deletions than additions; riddling of verbal expressions of class numbers with superfluous words; sentences like expressions; unnecessary clusters; inappropriate, confusing, misplaced, out-of-date and not index-friendly headings. Relocations, reduction of length of class numbers and other suggestions have been made to overcome these and other lacunae.

The above review of the literature on UDC published in the decade covered by this publication indicate that Slavic is the major author who has studied and written extensively on the various aspects of the UDC. The UDC is continuously updated by an efficient editorial mechanism and organisation and the users are kept updated through a periodical announcement bulletin.

Library of Congress Classification

The Library of Congress Classification (LCC) scheme, as its name suggests, is developed and published by the Library of Congress, Washington, DC (*www.loc.gov/cds/*). Publication of the LCC began in 1901; different volumes and their revisions have been published since then as and when needed. 'The printed schedules of the [Library of Congress] Classification so far comprise nearly fifty individual volumes for the main classes, subclasses and tables' (Chan, 2004a). The LCC is essentially an enumerative classification scheme. It is used by American academic and research libraries as well as by many foreign libraries (Chan, 2004b).

The most striking feature of the LCC is the individual schedules developed by a group of experts. That is the reason, as it will be noticed from the reviews presented below, that most of the descriptions, discussions, criticisms and suggestions are individual schedule-based. The review below approximately follows the sequence of the schedule.

In the LCC, class BM (Judaism), BS (Bible), DS (History of Asia), and PJ (Oriental Language and Literature) are some of the classes of major importance to Judaica libraries. Additions and modifications made to

these classes during 1995–96 are reported by Ruderman (1999–2000). The schedule G of LCC deals with Geography, Maps, Anthropology and Recreation. Moore (1999) has described the structure of this schedule, explored its development and has provided guidelines in number construction. The author has also discussed the shortcomings in this schedule.

The University of Tennessee-Knoxville Library, USA, contained a collection of Knoxville and Knox County planning documents. Organisation of these local planning documents under HT168 class of the LCC was customised. While customising, the library decided to subdivide the class HT168 so as to place works about the same sector of the city or county together on the shelf because the majority of the documents addressed planning for a specific sector. Robertson (1999) reported the difficulties encountered in making and implementing this change.

Religious laws evolved with a definite purpose and do have history of their own development. Depicting these and other aspects, the Library of Congress published new classification schedules for the Law of the Roman Catholic Church and for the History of Canon Law. The new schedules provide new opportunity as well as having their own problems (Diamond, 2001). The Library of Congress has been developing classification schedules of law since the 1930s. The schedule of class K provides class numbers for all laws and legal systems at international, regional and jurisdictional levels and also to express the relationship between historical and religious systems. Goldberg (2003), who is credited for the development of many schedules of law (e.g. Law of the Roman Catholic Church, Islamic and Jewish law, etc.) for the Library of Congress, described the schedule of class K and discussed historical and political aspects of the development of law classification. For a long time, there was no separate provision in the LCC for the classification of documents dealing with Islamic and Jewish law. Libraries using LCC were either classifying books of Islamic and Jewish law with the help of a religion schedule or having some temporary local provision, which affected organisation and retrieval. To help solve this problem, the LC developed draft schedules of Class KBM: Jewish Law and KBP: Islamic Law. Schwartz (2001) described these schedules and praised Jolande E. Goldberg for developing the schedules with an elegant structure. A comparative study of the treatment given to national legal schemes in the schedule of the LCC reveals a number of inconsistencies. van Laer (1999) explained these inconsistencies with the example of 'hypotheque' in the French system compared with 'mortgage' in the English system and

criticised the LCC for compromising between user convenience and theoretical insights.

Classics of each country require special treatment in every bibliographic classification. However, most of the book classification schemes (naturally, though) provide provisions favouring classics of their own country, i.e. the country of the origin of the classification scheme. For example, the present LCC recommends that Chinese classics should be classified as literature. Jiang (2007) observed this provision as inappropriate and inadequate and explained the practical and philosophical reasons for this inadequacy. In order to classify the Chinese classics with accuracy and consistency, Jiang made two suggestions: (i) make provision to classify Chinese Classics in Class B (Philosophy) as against the present practice of classifying them in Class PL (Literature); and (ii) make modifications and additions to the relevant schedules.

The schedule of Class PA (Supplement) of the LCC, deals with classification of Byzantine literature. Like literature in any other country, the Byzantine literature also has its own specialties. Sitas (2001) evaluated the schedule of Class PA in the context of its handling of the Byzantine literature and Classical literature texts. Based on his evaluation the author suggested: (i) that the early Christian literature and the Fathers of the Church must be excluded from Class PA; and (ii) to have uniformity, consistency and reliability in classification of Byzantine text, they must be treated according to the method proposed for classical literature by the Library of Congress Classification Schedule PA (Supplement). The same schedule, i.e. PA (Supplement), also deals with Greek folk (laiki) literature, Modern Greek poetry and Greek folk (dimotika) songs. Evaluation of this schedule and suggestions for effective and accurate classification of Byzantine folk (dimodi) literature can also be found in Sitas's article of 1999.

Library of Congress Classification, Table L7, provides a range of Cutter numbers. However, it does not provide any guidance for the allocation of the numbers within the ranges. In the absence of any guidance, this table could not be used effectively for maintaining logical shelf arrangement in a sizable library having normal changes and with cataloguers of varying degrees of expertise. Shelton (2000) discussed the problem of applying this table to the institutional publications of Georgia State University and proposed its explicit expansion for local use as a solution. The proposed table accommodates specific administrative and academic divisions. This expansion of Table L7 could be helpful to other institutes using both Table L7 and Table L13.

LCC – uses

The library of the Institut national d'histoire (Inha) in Paris, France has a large collection dealing with art and archaeology. The authorities of this library decided to provide open access to its users. The library established a special group to choose a suitable classification scheme for this library and for this purpose. The group surveyed similar libraries in France and abroad to know which classification scheme should be used by them and concluded that there is no need to create a special classification scheme. After considering the various options available, including DDC and UDC, the group decided to use the LCC (de Cours, 2002). As the LCC's schedules are subject-based, they could be considered as a special classification scheme and as such there is no surprise that it is used by a library specialising in a particular subject such as art and archaeology.

Due to the highly specialised nature, involvement of subject experts in its making, continuous updates under a robust institutional support and a live book environment for its adoption, the LCC is growing and spreading steadily.

LC's contributions to librarianship are numerous. In the field of knowledge organisation, its noteworthy contributions are by way of support to DDC, development and maintenance of LCC, the LCSH, MARC21 and so on. Another recent contribution in the field of knowledge organisation is Classification Web. It is an online resource made available by the Library of Congress for subject headings and classification documentation. A survey of the extent of use of this resource by the working cataloguers has been presented by Ferris (2009). The survey also asked about the type of institutions that subscribe to the Classification Web; reasons for the use of this resource and what other resources are used by the cataloguers for classification/subject heading analysis.

Bliss Bibliographic Classification

The Bibliographic Classification (BC) devised by Henry Evelyn Bliss was first published in four volumes in the USA between 1940 and 1953. The new, revised edition (known as BC2) was initiated by Jack Mills and was to be produced in 22 parts, comprising one or two subjects per volume. The first volume was published in 1977 (Bliss Bibliographic Classification, edited by J. Mills and Vanda Broughton, London: Butterworth, 1977–). Publication of BC2 is now undertaken by K.G. Saur. Revisions have been made to some of the BC2 volumes in order to retain subject currency.

The updates made in the BC2 are published in the BCA (Bliss Classification Association) Bulletin (*www.blissclassification.org.uk/*).

Broughton (2001a), who happens to be co-editor of the BC2, provided a description of BC2 and explained the philosophy of it. In order to know whether BC2 suits general academic libraries, it was adopted by Fitzwilliam, Jesus, King's, Queen's and Sidney Sussex College libraries in Cambridge. All of these libraries differed from each other in their application of BC2 in the location of major subjects, in minor choices of classmark, in the length and presentation of classmarks, etc. It was found that the flexibility of BC2 supported divergence and its faceted structure helped in understanding. BC2's application in these general libraries had no reactions from users, so it was concluded that BC2 works satisfactorily in general academic libraries (Attar, 2000). The BC2 is a faceted classification scheme. It is praised for short class numbers and enhanced exactitude. At the same time it is criticised for having no institutional support. To test the validity of these appraisals and criticisms, Attar (2002) compared the BC2 class numbers with DDC, LCC and UDC class numbers for works about Shakespeare and proved that the BC2 offers a greater precision and brevity. And then, with the help of the results of a survey sent to non-Bliss Cambridge libraries, the author reported that in practical terms the criticism is substantiated. However, where it is practised, it shows the evidence of its success. Like Attar, Broughton (2001a) too appreciated that the faceted schemes in general and the BC2 in particular are appropriated tools for the management of electronic resources because they are: (i) powerful tools for the management of vocabulary; (ii) characterised by a rigorous analytical approach to terms; and (iii) clear identification of semantic and syntactic relationships.

Subject classification – Brown

The first edition of Subject Classification (SC) appeared in 1906, the second edition was published in 1914 before Brown's death, and the third was published in 1939 by Brown's nephew, J.D. Stewart. Beghtol (2004) discussed the theories and practices introduced by Brown and investigated his views on the order of main classes, the concept of 'concrete' subject and the need for synthesised notations. These ideas are traced into the future through works of Ranganathan, the Classification Research Group, and the second edition of Bliss Bibliographic Classification. A further study of Brown's work may help to shed light on the current classification theories and practices.

Other general classification schemes

The Library-Bibliographical Classification (BBK) is the Russian national classification system. It is unlike the Dewey, the Universal Decimal System or the Library of Congress System. By 2015 the materials in Russia will reflect both the BBK and the UDC codes. For a history, development and description of BBK, refer to Sukiasyan (2008a). China has a long history of bibliographic classifications, although modern classifications did not emerge until 100 years ago when the Western classifications were introduced into the country. The Classification for Chinese Libraries (CCL) is the major effort in the development of library classification in modern times. For a historical account of the development of CCL, its structure, problems and comparison with LCC, refer to Zhang (2003).

The above classification schemes have their own specialties. Satija (2000) compared different classification systems, specifically the Dewey Decimal Classification, Colon Classification, and Library of Congress. Rafferty (2001) also explored the representation of knowledge in the various library classification schemes. If you are interested in the history and origin of bibliographic classification from the perspective of pragmatist philosophy and social studies of science, refer to Hjorland and Albrechtsen (1999).

Choice of classification schemes

As could be seen from the review of classification schemes presented above, many classification schemes exist for adoption. Library and information professionals have studied and written about the choice of classification scheme, reasons thereof, problems faced and other practice-based experiences in the use of classification schemes. For example, Tadasad and Maheswarappa (2002) studied the classification practices among 500 college libraries in Karnataka State (India). The authors covered aspects such as the classification scheme used, provision of book numbers, book numbering system adopted, problems faced in classification, etc. The authors found that most of them used DDC followed by Colon Classification. They identified that due to inadequate staff, insufficient funds, lack of support from authorities and lack of interest, one-third of libraries undertook no classification work at all, and of those that classified their collection, one-third used Colon and two-thirds used

DDC. Similarly, Womack (2006) investigated the reasons for the choice of classification systems and subject heading lists by academic health sciences libraries. The author also investigated systems used by general libraries at the same academic institutions. The aim was to know if there is a desire for conformity and whether shared OPAC plays a role in this desire. About 75 per cent of libraries in the survey used the National Library of Medicine classification system and 95 per cent used MeSH (Medical Subject Headings) as they found it more suitable. The study also noticed that general libraries at the same academic institutions adopted the Library of Congress Classification scheme. It was found that the most compelling reason for using the National Library of Medicine's classification system is that it is considered the most appropriate for medical collections.

Despite the Dewey Decimal Classification becoming almost universal in British public libraries, Bowman (2005a) reports that the other systems also proved their worth and survived. The author provided a historical account of the various classification schemes used by public libraries in the UK and stated that apart from DDC, British public libraries used Quinn-Brown, Brown's Adjustable and Brown's Subject Classification, which survived into the late 1960s and is still in use in a few local studies collections.

Do different libraries assign different class numbers to the same title? To answer this question, Subrahmanyam (2006) conducted a survey of Library of Congress Classification (LCC)-based class numbers assigned to a representative sample of 200 titles in 52 American libraries. The results showed that there are more than 85 per cent of titles that have the same class number. However, the remaining titles have different class numbers. These titles were mostly multi-focus titles, titles in series, bibliographies and fiction. Use of copy cataloguing, assigning alternate class number along with preferred class number and finding a method to link alternate class numbers to preferred class numbers for enriched subject access through local and union catalogues are the suggestion made.

Like most other subjects, the development of classification is not uniform in all parts of the world. The subject originated and developed at different times at different places. The pace and dimension of its development also differed from country to country. For example, by referring to Mutula and Tsvakai (2002), we would be able to know about the development of classification in Africa, the cataloguing and classification problems faced by libraries in Africa and even solutions to these problems.

5

Special classification schemes and classification of non-bibliographic entities

Abstract: Library and information professionals have been involved in developing depth classification schemes for the classification of documents on micro-subjects, special forms of documents and for the classification of non-bibliographic entities. This chapter reviews literature that deals with special classification schemes for specific subjects, such as law, and special classification schemes for specific forms of documents, such as patents, music, comics, maps, etc. This chapter also reviews literature dealing with classification of non-bibliographic entities such as business stakeholders, industries, museum objects, etc.

Key words: special classification schemes; International Patents Classification (IPC); United States Patents Classification (USPC); classification schemes for records management; classification schemes for music; North American Industry Classification System.

Introduction

When there are well-established, comprehensive, standardised general classification schemes such as the Dewey Decimal Classification (DDC), Library of Congress Classification (LCC), Universal Decimal Classification (UDC), Colon Classification (CC), etc., that alleviate the classifier's task of creating coextensive call numbers for reading materials, why do we need special (subject-based) classification schemes? Ferrari (1999) observes that historically not all librarians found it feasible to follow these standardised classification schemes and sought to create classification

systems more appropriate to the holdings and collections of their own libraries. This approach led to the development of very many special classification schemes for special subjects, from celestial bodies to patents. These schemes are for specialised groups of entities and most often designed by subject specialists (Neelameghan, 2002a). The special classification schemes are also referred to as depth classification schemes as they provide class numbers for micro aspects of a specific subject.

There is a considerable amount of literature that deals with the construction of special classification schemes, their description and use. These schemes could be meant for documents in a specific subject, such as anthropology, humanities, information science, etc., or a specific form of literature, such as grey, maps, fiction, etc. This chapter reviews the literature dealing with special classification schemes meant for specific subjects and documents published in specific form. The review of special classification schemes for specific subjects follows the DDC's main class order, whereas the review of special classification schemes for specific forms of document is presented in alphabetical order by the name of the form of document.

Special classification schemes for specific subjects

Diverse aspects covered, vast geographical and chronological range, multi- and interdisciplinary nature and continual evolution of ideas and a conscious or unconscious bias in interpretation are some of the features of religious books. Further, most church/religious libraries have extensive collections of religious texts and have very specific and focused collection goals. Above all, there is no single classification scheme that works for all of these characteristics, making it complicated to classify and arrange theological books (Longstaff and Henry, 2009). Thus, the authors believe that the existing general book classification schemes are not very useful for this purpose and so a specialised classification scheme for religion should be developed. Kotter (2002) critically studied the many existing classification and subject access schemes in anthropology and observed that none of them was able to fulfil the requirement satisfactorily. The author expressed that this is because of the diversity of the subject and the elusive nature of terminology. So the author proposed a framework for designing a faceted subject access scheme which will fulfil the requirements of anthropology information searches.

Faceted classifications prove very useful in searching information from the Internet. Considering this fact, Broughton and Slavic (2007) developed a faceted classification for humanities. The authors provided details of the principles and procedures adopted in building a faceted classification appropriate to an online environment. This classification scheme will be a model to many stakeholders in the domain of resource discovery who are considering developing their own classification system and supporting tools.

Not that there are classification schemes for other subjects only, but there are many for our own field itself. For example, Zins (2007) documented 28 classification schemes of information science that were compiled by leading scholars in the academic community. Zins claimed that this unique collection of 28 classification schemes portrays and documents the profile of contemporary information science at the beginning of the 21st century.

Some of the benefits of law libraries' intranets include the creation of a single, seamless source; a single navigation system; trustworthy and up-to-date information; information tailored to the needs of a particular firm, or even particular staff, etc. A common legal classification system can further add value to law libraries' intranets as it can improve searches, stated Dow (2000), who appealed for an agreed common legal classification system. Should legal materials in one organisation be classified or not? This is the argument presented by Mosley (2002). If it is to be classified, is it good to use an established classification scheme or a home-made, flexible, user-friendly classification scheme? Mosley (2002) recommended the well-established Moys Classification for this purpose. This recommendation is further substantiated by Brett's (2008) survey of classification practices in law, which found that most libraries used the Moys Classification scheme, though about 40 per cent used in-house classification schemes.

Special classification schemes for documents in specific forms

General classification schemes such as DDC, LCC, UDC, CC, etc., are not very useful for specialised libraries that hold collections prevailing in special form. For example, for effective organisation and retrieval, archives must be classified using a unified and standardised classification scheme. The Chinese Archive Classification (CAC) system is one such

scheme. Zhang (2002) provides the history and development of CAC and describes its nature, organisational structure and functions in the information age. Art libraries have a special collection of dealing with art objects. Most art libraries have developed their own classification schemes to classify their specialised collections. Ferrari (1999) conducted a survey of such special classification schemes, compared them and provided further guidelines for constructing a classification scheme for art libraries.

Classification schemes for comics and fiction

Comics have their own value in the areas of education and reading for children, and the general classification schemes do have some provision for the classification of comics. Marcio Pajeu et al. (2007) studied the Brazilian classification systems such as COD and CDA for knowing the provisions for the classification of comics and found it inadequate. So based on the comics collection and provisions for comics in the existing classification schemes, the authors proposed a new classification for comics.

Fiction as a form of literature has been treated specially in all general classification schemes. For example, it has been placed in separate main classes and/or divisions. Provision is also made to represent the various literary forms such as poetry, drama, fiction, etc. However, it seems that library professionals as well as the user community are not very happy about the provision for the classification of fiction in the current general classification schemes. That is the reason there emerges a fair amount of literature on fiction classification. There exists literature that discusses the classification of literature by literary forms, subject of the document, as well as different approaches adopted in classifying fiction literature by different types of libraries such as schools, public, etc.

Fiction in public libraries can be classified by genre or by considering readers. Maker (2008) observed that classification of adult fiction by genre in public libraries creates confusion. He suggested that adult fiction should be classified by adopting a reader-centred method, which will enhance the accessibility of the collection, as the underlying principles are easier to understand. Classification of fiction in general and classification of fiction in schools in particular requires special treatment. Unfortunately, the classification of fiction in school libraries is not addressed largely in library and information science (LIS) literature. Considering that need for special treatment, which is not given by the general book classification

scheme, Shenton (2006) developed a system for fiction classification in school libraries. It is based on users' responses gathered from the query logs and the nature of books in the specific collection. That is why it is called a 'reactive classification', i.e. developed as a reaction to users' queries. This system is augmented by a cross-referenced user guide describing the individual categories, again drawing on the contents of the query log, which was also used to resolve situations where a certain book could legitimately be inserted into more than one category for shelving. A key aim was to ensure that the newly developed system of 'reactive classification' was more consistent with users' likely needs than was the past arrangement. Can this system be adopted by any other library? Shenton warns that since it was never intended to be generalised to other contexts, it would be unwise to adopt elsewhere without questioning the categories created here, but the processes undertaken are more open to transfer.

Fiction can be classified as reading for pleasure, but could also be classified as scholarly discipline. There are classification schemes for fiction as the first category, but no classification scheme for fiction as the second category. To fulfil this need, Vernitski (2007) developed an innovative model for an intertextuality-oriented classification scheme for fiction, to be used by humanities scholars studying fiction. This model consists of classes, subclasses, notation and examples. Instead of arranging young adult fiction in alphabetical order, it should be classified by the theme contained in them. The theme-based arrangement of fiction provides easier and more direct access to books and thus increases borrowing. Jouin (2008) cites examples of three libraries that classified young adult novels by theme and states that this classification method takes time to set up and requires adaptability on the part of the librarians; however, there is scope for further development of such systems.

The University of Western Australia and Murdoch University in Australia adopted two special classification schemes for literature. Szunejko (2003) described and compared the two special classification schemes by considering the collocation of subject matter, partition of subject aspects, ease of application, shelf browsability, retrievability, identification, access and navigation. Similarly, Negrini (2003) describes and compares two projects meant for the organisation of fiction literature. They are 'Finnish thesaurus for fiction' and 'thesaurus of Italian literature'. These cognitive and terminological resources could also be used by other libraries. The provisions for classification of literary work in the 21st edition of Dewey Decimal Classification (DDC21) and the Chinese Library Classification edition 4 are compared by Hu and Chen (2007).

Classification schemes for grey literature and maps

Grey literature is another form of document and is produced in almost all subjects. The terminology used to represent grey literature changes as per the subject covered. To represent the specialised terminology of grey literature in different subjects, Cuvillier (2007) developed a special classification system. This classification system covered 110 different subjects, each one of which is further subdivided. It provided seven entry points for searching subjects. They are: (i) teaching and research; (ii) literature; (iii) technological documentation and data sheets; (iv) vulgarisation and everyday life; (v) citizen and social activities; (vi) arts, sports and spare time; and (vii) bank of images.

Maps as a source of information contain very specific information. From the classification context, subject of maps is a more salient facet than geographic area. Considering these features of maps, Babik (1999) expressed the need for a uniform classification scheme and suggested 21 elements to be used in representing the knowledge contained in maps for retrieval in computerised systems.

Classification schemes for music

The present review reveals that a considerable amount of literature is published on the various aspects of classification of music and a number of music classification systems have been developed over time. Bradley (2003) presents a historical overview of the development of music classification schemes as well as a subject heading list and catalogue code for music in the USA.

Actually, the general classification schemes such as DDC and LCC do have provisions for the classification of music. Nero (2006) observed that there are problems in using these classification schemes for classifying Caribbean popular music. Revision of the pertinent sections in the respective classification schemes could be a solution. Kinney (2009) also observed that for almost half a century, music librarians, media librarians and professional associations have advocated classifying audio-visual materials using the same classification scheme that a library used for print materials. However, the music moving image materials have often received the same lack of subject classification access as sound recordings and other media resources. Like Nero, Kinney also discusses the problems

of using the general book classification schemes for classifying the music moving image materials.

Since many classification schemes are available for music classification, selection of an appropriate scheme becomes a challenge. The Leeds College of Music (UK) adopted the following process for finding an appropriate classification scheme for music: (i) deciding what is needed from a new shelf arrangement; (ii) looking at the available options; (iii) deciding to use and adapt the Alpha-Numeric System for Classification of Recordings (ANSCR). Marsh (2002) not only evaluates the system but also comments on the users' reactions. Like the Leeds College of Music (UK), the Suyematsu Music Library at Minnesota State University, Mankato, decided to reclassify its 5,000 music CDs using the ANSCR, which were initially arranged by accession number. The accession number arrangement was found unsuitable. Weber and Schomberg (2007) described the process of reclassification and its advantages. The process included design of a workflow for a graduate student and undergraduate assistants and installation of each CD in checkpoint security case. Easy access, more security and satisfaction to students for contributing to a meaningful task were the advantages observed.

It has been noticed from the literature reviewed in this book and other literature that many new special classification schemes are developed to overcome problems in existing classification schemes. The Hunter Library at Western Carolina University also developed a home-grown classification scheme for its collection of music CDs. Some of the special features of this scheme are that it does not require the classifier to have knowledge of music and is quick and easy to apply. It allows the library to organise the collection in an easily understood and user-friendly arrangement (Carstens and Vihnanek, 2000).

Classification schemes for news and newspapers

News documentation has its own characteristics from an indexing and retrieval point of view. Castillo and de la Cueva (2007) observed that: (i) the news documentation lack suitable instruments of classification and indexing, the only exception being the Subject Reference System of the International Press Telecommunications Council (IPTC), which is not yet fully developed; (ii) the library and information profession has not paid much attention to the classification and indexing of news. Against this background, Castillo and de la Cueva reviewed the most relevant

contributions in classification and indexing of news in Spanish-language media, both in Spain and internationally.

Newspaper is one of the very common forms of document possessed by almost all types of library all over the world. Most libraries face problems in deciding whether to classify newspapers or not. In reality, this decision should be based on the basic research needs of users. City of publication/geographical location and subject are the major preferences of users in searching information from newspapers. Newspapers, as general serial publications, present the challenge of multiple title change, a unique relationship among other newspapers (Kuhn, 1999). Kuhn claims that the DDC with a few modifications could be used optimally to classify a newspaper collection

Classification schemes for patents

Patents are an important form of documents that provide information about the holder of the intellectual property rights as well as the description of the specific invention. Classification of patents required special provision considering the need of the patent users. At international level two main classification schemes have been developed for the effective organisation and retrieval of patents and information about patents. These are the International Patent Classification (IPC) and the United States Patents Classification (USPC). The following paragraphs review the literature about both of the patent classification schemes.

The International Patent Classification (IPC)

The International Patent Classification (IPC), established by the Strasbourg Agreement 1971, provides for a hierarchical system of language-independent symbols for the classification of patents and utility models according to the different areas of technology to which they pertain. IPC first published in 1968 and the latest (eighth) edition was published in 2006. The literature reviewed in this book mostly deals with the seventh and eighth editions of the IPC, though passing and contextual references are made by the writers to earlier editions also. Information about the latest development of IPC can be found at *www.wipo.int/classifications/ipc/en/*. The IPC was originally designed as a shelf-location tool for paper search-files. However, the classification has certain features which made it an attractive option for subject-based patent searching of electronic databases. For background history, development and modern use of the International Patent Classification (IPC) as a storage and

retrieval tool, refer to Adams (2000), and for knowing structure and illustrated examples, refer to Sangeetha et al. (1999).

The IPC has its own advantages. The United States Patent and Trademark Office (USPTO) praised the efforts that have gone into the creation of the IPC. However, it also expressed that its future value may be a problem unless substantial changes are made to it. It sets out 18 areas where problems exist and substantial changes are necessary. Kunin (1999) discussed in detail the value of discretionary and non-discretionary indexing and proposed a new concept, the alternative or supplemental use of 'catch terms' linked to each IPC term. Similarly, Bruun (1999) identified inconvenient provisions in the IPC and, for its efficient use, suggested that the rule on 'last place' allocation of terms and other similar rules should be scrapped as they are primarily designed to limit the extent of the paper files. Adopting a number of approaches to improved indexing and multiple classifications are the other two suggestions.

Considering the lacunae in the earlier edition, the process of reforms of the IPC was initiated in 1998 and was almost completed in 2003. According to Calvert and Makarov (2001) the reasons for the reforms in the IPC are:

- better worldwide consistency in the application of the IPC to invention and invention-like published information;
- creation of a single, worldwide master classification database (MCD), based on the IPC, for patent documents and classified non-patent literature;
- all documents to be classified on the current edition of the IPC;
- the IPC to become a more readily available, readily understood and easily used system;
- the IPC is to be so designed that it could be used to search both large international patent collections and also to enable users to search smaller national collections;
- and the IPC to be fully accommodated to an electronic environment.

Makarov (2000) highlighted the creation of two new classes in the seventh edition of the IPC, B81 and B82, covering micro and nano-structures, devices and systems and the concentration of artificial intelligence material in classification subclass G06N. The restructuring of class F21 includes the abolition of some subclasses and the creation of new indexing subclasses, F21W and F21Y, for uses and applications of lighting devices and systems, and for forms of light sources. Indexing is now possible in nearly 14 per cent of subclasses, but remains non-obligatory. In 2004,

Makarov reviewed the progress of the IPC reform and described the procedures for its implementation and the structure of the reformed classification scheme. The author discussed in detail the core and advanced levels; reclassification of patent collection by using the current edition of IPC; the additional information in the electronic version; principles and rules of the reformed IPC; the Master Classification Database (MCD); and aspects of information technology support for the IPC, such as automated assistance in applying classification symbols and multilingual access. The extensive IPC reform project of the last few years has led to many major changes. Pauwels (2004) describes three vital aspects of the implementation of the reformed International Patent Classification (IPC):

- the Master Classification Database (MCD), which is based on the EPO's DOCDB system;
- CONcept of OPerationS (CONOPS) documents describe the classification and reclassification process of the reformed IPC in sufficient detail to enable all industrial property offices to implement and use the reformed classification system; it covers topics such as patent families, back files, newly published documents and revision of the advanced level; and
- the HARMONY project of the Trilateral Offices represents a further stage in the development of harmonised classification between the USPTO, EPO and JPO.

The first edition of the International Patent Classification (IPC) was developed to have a unified patent classification system. However, this aim was not achieved sufficiently by all seven editions. The eighth edition of this classification system was published in 2006. Takamasa (2006) presents an account of the failures in the past editions and changes that have been made in the eighth edition (IPC8) to overcome the lacunae of the earlier editions and to achieve the original aim. According to Makarov (2006), the IPC8 incorporates both the usual technical revisions to improve the searching in certain subject areas and also the major reforms which have been proceeding in parallel for this edition. Makarov's article also provided statistical data on the revision as well as lists new and substantially revised subclasses; it includes information about both paper and electronic forms of IPC8.

United States Patents Classification (USPC)

The United States Patents and Trademarks Office (USPTO) has its own classification system for the classification of patents known as US Patent

Classification (USPC). The key components of the USPC, along with specific examples from appropriate classes, are discussed by Falasco (2002a). The components include: classification by industry, utility classification by proximate function, utility classification by effect or product, and classification by structure or configuration. Through another article, Falasco (2002b) explored the system organisation at class and subclass level. A detailed comparison of the International Patent Classification (IPC) and the USPC is carried out by Adams (2001), who considers the relative weighting given to the inherent functionality of each invention and to its application in a specific industry. Examples are provided where this difference has led to a plurality of IPC terms corresponding to a single USPC term, and vice versa. Implications of these differences for the patent searchers are explored.

Since the USPTO registers patents for various plants, the USPC system consists of a special part for the classification of patented plants. It helps in efficient retrieval of plants' patents. Rademaker (2000b) describes this part of the USPC system under various headings, such as: purpose; classification schedule; schedule organisation; relationship to design and utility patent classification; class definition; search notes; line notes; hierarchy; and special features of plant patents, such as colouration and the effect of commercial or market practice. Rademaker (2000a) also described the provisions for the classification of ornamental designs in the United States Patent Classification Scheme. The description includes: purpose, classification schedule, schedule organisation, relationship to utility patent classification, design patent titles, class and subclass definitions, hierarchy, special features, simulative designs, and cross-reference art collections, placement rules, searching for ornamental designs, etc. In addition to the above major classification schemes for patents, there also exist Japanese classification schemes. The Japanese Patent Office has developed two classification systems for patents: the File Index (FI) and F-term classification system. Since April 2001, it has been possible to search Japanese patent literature with FI classification and F-terms in English. Schellner (2002) described concept, structure and layout of these two classification systems and provided guidance in the use of these features to enhance searches of Japanese patent documentation.

Are there any lacunae in the IPC, USPC or other patent classification systems? Is it necessary to develop more specialised patent classification systems considering the needs of specific industries? Lai and Wu (2005) observed that the traditional patents classification schemes such as IPC and USPC are too general to meet the needs of specific industries.

Also, these patent classification schemes have placed some patents under incorrect categories. Such incorrect categorisation affects research and development planning, technology positioning, patent strategy-making and technology forecasting. So the authors proposed a new approach to create a patent classification system to replace the IPC or USPC system for conducting patent analysis and management. The new approach is based on co-citation analysis of bibliometrics and the needs of a specific industry. The three steps in developing such a scheme are: (i) selecting appropriate databases to conduct patent searches according to the subject and objective of this study and then select basic patent; (ii) using the co-cited frequency of the basic patent pairs to assess their similarity; (iii) using factor analysis to establish a classification system and assessing the efficiency of the proposed approach. The main contribution of this approach is that it develops a patent classification system based on patent similarities to assist patent managers in understanding the basic patents for a specific industry and the relationships among categories of technologies and the evolution of a technology category.

Automatic classification of patents

The USPTO automated its operations in the following applications: assigning new patent applications to the group of examiners most appropriate for the area of technology for each application; classification of new documents into the search files; and reclassification of patent specifications. Smith (2002) explained the current use of this automated system and explored the potential, stating that the use and development of such systems are likely to become increasingly important in handling large volumes of patent documents. Another benefit of automation of patent classification is the retrieval of ranked listings of relevant prior art. The number of patent documents is currently rising rapidly worldwide, creating the need for an automatic categorisation system to replace time-consuming and labour-intensive manual categorisation. Accurate patent classification is crucial to search for relevant existing patents in a certain field. As patent documents are structural documents with their own characteristics distinguished from general documents, these unique traits should be considered in the patent categorisation process. Kim and Choi (2007) developed an automatic patent categorisation system for Japanese patents. The system is based on the k-NN (k-Nearest Neighbour) approach. Claim, purpose and application field of the patent are the elements considered for the categorisation as these are identified by various user-defined tags.

Classification schemes for records management

In the process of business transactions, a large number of records are created. Business classification schemes are required in order to organise and retrieve appropriate records effectively and for effective record management. Business classification schemes, also referred to as records management schemes, are essential components of a good records management practice. A business classification scheme is like the foundation of a house. All subsequent records management is built on top of it. If the foundation is weak, the house will crumble and eventually collapse. Therefore, it is necessary to have a logical and systematic approach to develop a business classification scheme (Morelli, 2007). The business classification scheme is also referred to as a file plan or records classification scheme. In such a classification scheme, business classification processes feature prominently (van der Walt, 2004). Jeffrey-Cook (2005) explained the concept of a file plan and provided guidance for the development of a business classification scheme, retention policies and appropriate degree of centralisation in the context of local government. The structure of such a classification scheme is recommended by the records management standard BS 15489. Morelli (2005) discussed critical issues in developing and implementing a business classification scheme as recommended by BS 15489. Morelli, who worked as a records management consultant in the private and public sector, observed that due to statutory obligations such as the Freedom of Information Act, there is an increased interest in the UK in records management along with business classification schemes. Such a scenario is emerging in India, also due to the enactment of the Right to Information Act and other similar legislation.

van der Walt (2004) proposed a generic classification scheme for organisation of electronic documents in a business enterprise sector in small, medium and micro enterprises (SMMEs). Data for the proposed scheme were collected from literature on information organisation, business information, competitive intelligence and information systems, as well as through an empirical study of information organisation practices in a sample of 24 small businesses in three different provinces in South Africa. The concepts gathered from folder systems for documents, email and Internet favourites, augmented by concepts derived from the literature, were analysed using the technique of facet analysis. The Local Government Classification Scheme (LGCS) is another useful record management classification scheme (Dodgson and Jeffrey-Cook, 2007). This classification scheme can work as a

model for other local authority records managers in developing a classification scheme.

Of late, a number of software packages have been developed commercially for the management of electronic records. Some of them are quite costly. Smaller organisations cannot afford costly electronic records management software. Such organisations may use 'inteRM CFS', an application that can be used in conjunction with MS Office desktop features. This application, developed by Active Classification Systems, Australia, also incorporates a business classification and thesaurus facility known as 'term tree'. More about this application can be found in the article by Plater (2005).

Subject matters, record types and forms, structures and functions of the organisation are the characteristics to be considered for the design and development of a records classification scheme. However, Foscarini (2006) observed that there is no consistency in the design of records classification schemes. Archivists and records managers (whether consciously or unconsciously) have traditionally referred to any of these elements, or to a mix of some or all of them, when developing their classification tools. Only in recent times has the concept of function become central to the theory, method and practice of records classification. The advantages and limitations of this approach (function-based approach) are elaborated upon. How this approach is interpreted in the literature of different countries is also discussed. Systems used to arrange or classify government records play a key role in a government department's ability to conduct and manage its business information overtime. In recent years, there has been growing interest within the public sector records management community in using function-based classification systems as a means to link business context, instead of classifying it into groupings reflecting ever-changing organisational structures. Function-based records classification systems are in line with the principle of levels of arrangement and description as records are organised according to hierarchical levels in order to reflect the nature of their creation. Ngoepe (2009) described the function-based classification scheme, i.e. functional subject fileplan, developed and used by the National Archives and Records Service (NARS), South Africa, for the classification of government records. Smyth (2005), with a case study in Northern Ireland, discusses issues involved in the development of functional fileplan.

Records are commonly classified by considering subjects, organisational structure, geographical location, etc. However, the Australian Records Management Standard (AS 4390) and Information and Documentation

Records Management Standard (ISO 15489) gave prominence to function-based classification of records. Both of these standards helped to promote function-based classification. Orr (2006) reported that function-based classification has been strongly promoted, particularly in Australia, Canada and the United Kingdom. However, this system has a number of issues that need to be addressed.

Classification of non-bibliographic entities

What can be noticed from the available literature on classification is that it has many useful applications in other areas apart from libraries. The literature reviewed in this section deals with such classification schemes, which have been developed for the organisation of non-documentary artefacts. For example, it deals with classification of museum objects, industries, research institutes, regulatory modes, figures of speech, etc. In fact, the literature in this section discusses the non-library applications of classification. However, this literature is also useful for library professionals because it helps in further understanding classification. This understanding could be used for more appropriate application of classification in librarianship and/or for developing more versatile classification schemes. The literature in this section is organised alphabetically by the subjects of the application of classification.

Classification of business stakeholders

As the web is used increasingly to share and disseminate information, business analysts and managers are challenged to understand stakeholder relationships. Traditional stakeholder theories and frameworks adopted a manual approach for analysis and did not scale up to accommodate the rapid growth of the web. Unfortunately, existing business intelligence (BI) tools lack analysis capability, and research on BI systems is sparse. To classify business stakeholders on the web, Chung, Chen and Reid (2009) designed a Business Stakeholder Analyzer (BSA), a business intelligence system that identified and classified stakeholders on the web. The system incorporated human knowledge and machine-learned information from web pages. The classification results demonstrated better within-class accuracies in widespread stakeholder types, such as partner/sponsor/supplier and media/reviewer, and were more efficient than human classification.

Classification scheme for evaluation purpose

A book classification scheme is usually developed for the organisation of documents and/or their records. But a classification scheme with a very different purpose, i.e. for scientometric evaluation, is proposed by Glanzel and Schubert (2003). This is a two-level hierarchic classification system of fields and subfields of the sciences, social sciences and arts and humanities. The process of application of this system and results obtained can be found in the article. Thijs and Glanzela (2009) believed that the clustering of research institutions according to their publication profile would be beneficial to arrive at a benchmark which will be useful to study the national research performance on the basis of the institutional classification. To test their idea, they used a structural analysis method to evaluate research performance of specialised and multidisciplinary institutions. A breakdown by subject field is used to characterise field-specific peculiarities of individual clusters by bibliometric indicators. This allowed comparison within the same and among different clusters.

Classification of figures of speech

In knowledge organisation tools, such as classification schemes and thesauri, a figure of speech generally displays a lateral relationship (non-hierarchic associative relationships) between two or more concepts/entities. The figures of speech themselves could be categorised on the basis of various categories. Neelameghan (2009) presented a list of Tamil figures of speech categorised as per function.

Classification of ethnic data

'Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity', formerly known as 'Statistical Policy Directive 15', is a classification system that has formed the basis of the US government's collection and presentation of data on race and ethnicity since 1977. Robbin (2000) examined the history of Statistical Policy Directive 15 from its origin through 1997. Based on the study of this classification system, Robbin puts forth a very important observation: classification systems cannot be isolated from the larger political setting.

Classification of icons

The painters of Byzantine and post-Byzantine icons used specific rules and iconographic patterns. On the basis of these rules and patterns, Tzouveli et al. (2009) developed a knowledge-based image analysis system to classify icons. Their system first detects and analyses important facial characteristics providing rich, yet still imprecise, information about the icon. Using the description logic the system then expresses the extracted information in terms of expressive terminology. Further, by using the fuzzy extension of description logic, the system produces assertional knowledge. Finally, the icons are classified by combining the terminological and assertional knowledge. Classification of 2,000 Byzantine icons demonstrated that the new classification system is effective.

Classification of industries

The US federal government classified industries using the codes given in the Standard Industrial Classification (SIC). Later, the SIC was replaced with the North American Industry Classification System (NAICS). This new classification system inherits some problems from its originator, the SIC. Boettcher (1999) explained the concerns of the researchers and information providers about this new classification system, addressed issues arising from the transition (from SIC to NAICS) and discussed its effects on companies and researchers. Sabroski (2000) described NAICS, providing details about the hierarchy, two-digit industry sectors, the relationship between NAICS and the North American Product Classification (NAPCS), and NAICS-friendly Gale online databases on DIALOG. The author has also brought out differences between the SIC and the NAICS, such as the former concentrated on what the industries do rather than whom the industries serve. Different agencies helped in the spread of use of NAICS. For example, the Gale Group included the NAICS codes in their databases and DIALOG acted as a first host to make the Gale Group databases searchable by NAICS (Tudor, 2000). Tudor's article is also useful for knowing information sources on the NAICS, which included its own website, handbooks, conversion tools and the specialist.

Classification of information databases

A classification scheme called 'The Global Data Model' has been developed by the Crystal Reference for the publishers of large information databases so as to help them organise their information. The Global Data Model classification scheme consists of 1,500 topic domain categories, which break down into 60,000 sub-elements, each of which has at least 40 key words, over 300,000 English names, words and phrases (Flood, 2002).

Classification of museum objects

Museums organise their artefact collections in the way they think convenient. Can there be a classification scheme for their logical and useful organisation? Yes! With the help of a case of a folklore museum affiliated to Mysore University, India, Rajashekara (2005) proposed a classification method for organising the museum artefacts, particularly the mementos that formed part of the collection. Characteristics such as purpose of the object and chronology could be considered for the classification of museum objects. The medium of which the objects are made could also be a feature adopted to classify museum objects. Larsen (2009) and his committee members adopted the last approach, i.e. mediums, and developed a classification scheme to classify museum objects. The categories of material considered in this classification scheme included fibre, clay, metal, wood, etc.

Classification of regulatory modes

Various types of regulation brought forth by the state and private players play a vital role in the liberalised and convergent communications sectors of developed economies worldwide. A vital feature of this transformed statehood in communications is a rising importance of independent national regulatory agencies (NRAs) and a growing reliance on alternative modes of regulation, i.e. self- and co-regulation. Most of the time the contributions of NRAs are appreciated and those of the private agencies are largely neglected. Based on the case study of the Austrian regulatory institutions active in the convergent communications sector, Latzer (2006) and his co-authors developed a new classification scheme of regulatory modes. This classification scheme makes it possible to grasp the numerous and often intertwined contributions of both state and private actors.

6

Automatic book classification, reclassification and non-classificatory approaches to knowledge organisation

Abstract: A fair amount of literature discusses automatic book classification and this chapter reviews this literature. Libraries reclassify their collections using either the new edition of their current classification scheme or, rarely, with a different classification scheme. Some libraries experiment with organising the collection by an approach other than classification. Literature on both of these subjects is also reviewed in this chapter.

Key words: automatic book classification; reclassification; non-classificatory approaches to knowledge organisation.

Automatic book classification

Automatic book classification, i.e. the construction of a call number by computer, has been one of the dreams of library professionals. After the emergence of computers and, more particularly, after the development of artificial intelligence, the hopes of success in automatic classification have reached the apex. Library professionals hope that computers and related technology-based tools and techniques should be able to identify the subject and sub-subjects of the document, search appropriate isolates and notations representing those isolates, and synthesise these notations into a coextensive call number. Initially there was doubt about the capability of computers for classification. For example, in 1999 Jaiswal argued that a computer cannot determine the subject content the way librarians can.

Therefore, automatic indexes produced by computerisation can only be considered for a similar automatic production of title indexes or keyword-enhanced indexes. Truncation and stemming may be probable solutions; however, for a robust solution, further investigation is essential. Due to technological developments, a new positive attitude along with action and effort in the direction of development of automatic book classification has emerged – which is reflected through the literature reviewed in this section.

Researchers had made numerous attempts to design a powerful automatic book classification system, but those could not bring any well-accepted results. The main problem was successful automatic analysis of titles of documents and finding out subject propositions. Because of being a purely mental process, classification demands human intelligence for analysing the document title, which is in natural language, to find out its basic subject and other facets, if any, along with its category and also synthesising those facets according to principles, postulates and canons to construct a class number. The emergence of artificial intelligence (AI) could bring the solution to this problem. The use of natural language processing (NLP) techniques would help in automatic analysis of titles and the expert system would work exactly in the same way as a classifier does to build class numbers, being guided by canons, principles and postulates. Basu, Panigrahi and Prasad (2003) suggested a methodology of building a lexicon (parser based on definite clause grammar, knowledge base based on frame-based structure and rule-based knowledge) representation model. Continuing the efforts in developing an automated book classification, Panigrahi and Prasad (2007) suggested a method to identify facets from the title of the document and demonstrated the techniques of fixing the facet sequence with regard to five fundamental categories, their respective notations, connecting symbols, etc. Techniques of selecting their respective category name to follow the facet formula in general and method for defining facet formula, based on postulates and principles, is also explained. Analysing the title of the document (i.e. in the form of natural language sentence), finding noun phrases, picking up isolate numbers, symbols, basic subject notation from the knowledge base, etc., are the steps in automatic book classification (Panigrahi, 2007a). The author suggested that the integration of expert systems and natural language processing components is useful in developing an automatic book classification system.

In another effort, Panigrahi (2000a, 2000b) developed a knowledge representational model of Analytico Synthetic Classification for Colon Classification with special reference to medical science and an expert

system for automatic classification using natural language processing techniques. Similarly, Kim and Lee (2002) developed an automatic classification system for the library science field. The system used the facet classification principles of CC. The classification system identified the subject of the document on the basis of key words in the title and created class numbers automatically.

Wang (2009) observed that library classification systems, such as the DDC, impose great obstacles on state-of-the-art text categorisation technologies, including deep hierarchy, data sparseness and skewed distribution. Therefore, he experimented with the automated assignment of DDC class numbers to bibliographic data with a supervised machine-learning approach to reshape the DDC structure into a balanced virtual tree by balancing the category distribution and flattening the hierarchy. To improve the classification effectiveness to a level acceptable to real-world applications, the author proposed an interactive classification model that is able to predict a class of any depth within a limited number of user interactions and conducted experiments on a large bibliographic collection created by the Library of Congress within the science and technology domains over 10 years. The results indicated that with no more than three interactions, a classification accuracy of nearly 90 per cent is achieved, thus providing a practical solution to the automatic bibliographic classification problem.

Reclassification

Reclassification is the process of migrating from one classification system to another. Lack of updated editions, lack of suitability of the in-use classification system and emergence of a more suitable classification system are all reasons for reclassification. The amount of literature on reclassification indicates that a fair bit of reclassification activities and thinking is taking place.

Reclassification to LCC

Reclassification was a popular trend during the 1960s and 1970s for many academic libraries wanting to change from the Dewey Decimal Classification (DDC) to the Library of Congress Classification (LCC). Nevertheless, there were libraries that did not consider changing from DDC to other classification schemes. Ahrberg, Seikel and Shorten (2005)

conducted a survey of 34 academic libraries using DDC in the United States and Canada for knowing factors which influenced the decisions of these libraries to continue using DDC, and if reclassification is something they have considered or are considering. The survey also investigated whether patrons of these DDC libraries prefer LCC and if their preference influences the library's decision to reclassify. Results from the survey indicated that the issue of reclassification is being considered by some of these libraries even though, overall, they are satisfied with DDC. The study was unable to determine if patrons' preference for a classification scheme influenced a library's decision to reclassify.

The University of Kentucky Library reclassified its collection from DDC to LCC. Reliance on automation was the unique feature of this reclassification project, which helped in reducing the number of man hours and caused minimum disturbance in the user services (Lewis and Seago, 1999). Wright and Blasé (2006) also reported a similar experience of reclassification of a collection from DDC to LCC due to the merger of two university branch libraries. The authors explain how spreadsheets were useful in manipulating and visualising data in the whole process of reclassification to reorganisation of the reading material. What is worth noting is that during the whole process of reclassification and reorganisation, the library remained open and maintained normal service levels. Another important feature of this reclassification venture is that the books and periodicals were integrated from separate shelving areas into one continuous call number sequence.

However, one cannot conclude that there is a definite trend in changing from one particular classification scheme to another (for example, from DDC to LCC), as Currier (2002) reports that the DDC is not only the most commonly used classification scheme by art libraries in the UK but is also the one that most art libraries choose when they reclassify their library. Another reclassification experience is that of the Singapore Polytechnic (SP) Library, which converted its UDC numbers to LCC. The work was carried out in-house. One staff member was seconded to the National University of Singapore Library for learning LCC and also to know how special collections are classified using LCC. Help was also provided by student assistants. At the same time, some 10,000 bibliographic records were outsourced for assigning LCC call numbers. The conversion and reshelving was implemented during the polytechnic vacation periods from February 2007 to March 2008, with the library remaining open throughout the period. This benchmarking project showed that work performance could be improved upon the conversion (Tim, 2009).

Reclassification within LCC

In 1997, the Library of Congress Classification abandoned the outdated JX schedule of Political Science and began to use JZ (International Relations) and KZ (International Law and Relations). Due to these modifications, the library of the Davis School of Law, University of California, had to reclassify 3,100 titles. The experience of this reclassification is reported by Lin and Murphy (2000). Another similar change took place at the Cataloging and Government Documents Departments of the Green Library at Florida International University (FIU) in 2001 when it embarked on a project to reclassify the Official Records of the United Nations (UN). Wartzok and Hernandez (2005) narrate the experience of reclassification of official records of the United Nations by using the JZ schedule of the LC Classification to supersede the JX and the UN document numbers. Problems encountered during this reclassification project are discussed. Addenda, corrigenda, annexes and the various subordinate bodies presented many complexities, reported the authors.

Reclassification with other classification schemes

The Narashino Media Centre of the Toho University was using the locally modified version of the DDC for classifying its collection. In 2007, the centre changed from this system to the Japan Decimal Classification system. Difficulties faced by the users and inadequacy for changes in natural sciences are some of the reasons attributed to the change of the classification system, reported Tsuneda (2009). Two important features of this system migration were: (i) the library stayed open, continued circulating materials and there was no reduction in the library usage; (ii) the change in the classification system was extended to documentation and tutorials. The author emphasised the need of additional user education to cope with reclassification.

'The contracting world of Cutter's Expansive Classification' is the title of the article written by Winke (2004). The title is very interesting, isn't it? Winke's article deals with both reclassification and staying with the current classification scheme. The author conducted a survey of libraries in the US, Canada and England that were identified as past or present users of the Expansive Classification (EC) of Cutter. Four of these libraries still used it as a primary scheme and another 23 continue to maintain some portion of their collection in EC. The survey sought answers to

questions such as reasons for abandonment, which scheme they have shifted to, whether EC was still used for part of their collection, reasons for staying with EC, etc.

Many firms have changed or are changing from Standard Industrial Classification (SIC) system to the North American Industrial Classification System (NAICS). O'Connor (2000) claims that the transition from SIC to NAICS will not be rapid, but its effects will be profound for business researchers and information professionals. Most government agencies are already in the midst of this six-year transition and, although private information producers are not compelled to switch from SIC to NAICS, most are planning to do so. The far-reaching impact of NAICS on business information will affect libraries of all types. O'Connor has discussed the challenges and costs associated with this change and made recommendations for materials and training.

A very different rationality for reclassification has been thought of by Rooney (2009). It is observed that in open-access libraries users mostly prefer to browse the actual collection than to search books through a catalogue or database. Considering this fact, it is expected that the library personnel reclassify their collection considering user preferences. On the contrary, Rooney's experience in Latin America is that librarians are reluctant to reclassify their collection, which creates problems in organisation and collection development. The author suggested that reclassification should be a routine part of classification development. The experiences about reclassification narrated in the literature reviewed above indicate that, at times, the work of reclassification is necessary and could be carried out with the help of computers speedily, economically and without much disturbance to the users.

Non-classificatory approaches to knowledge organisation

Library professionals largely believe that a library collection need necessarily be classified and arranged by using some book classification scheme. However, experiments have been carried out to test other document arrangement systems also. That is, efforts are made to arrange library collections in a non-classified manner. Most libraries of the world prefer to organise their print collections by using some book classification scheme, such as DDC. However, some libraries, particularly public libraries, feel that their collections can be organised in a more effective manner by a method other than systems such as DDC. This is because the

main areas of interest to borrowers in a suburban public library service centre are domestic and personal: their health, their homes, their holidays, their money and their leisure. While DDC is a satisfactory method of linking a catalogue record to an item on the shelf it does not facilitate browsing in the areas of most interest to public library users, observed Hopkins (2007). His observation is substantiated by an experience of Bayside (public) Library in Australia, which did not use DDC in the traditional sense to determine shelf order. The argument for not using DDC is that it does not facilitate browsing in the areas of most interest to public library users. The library organised documents in six new non-fiction collections, which combines DDC sequences within subject areas that are meaningful to a contemporary Australian audience. The system proved so successful that the authority decided to implement it in other branches. Another similar experience narrated by Lau Whelan (2007) is of the recently opened Perry Branch Library, part of the Maricopa County Library District in Arizona, which decided not to use DDC; rather, it arranged the entire collection by topic and alphabetised by author's last name. This decision was taken to make the library more customer-friendly, attracting more patrons. The library officials argued that DDC is not fail-safe and the classification system often confused and frustrated patrons. The author particularly mentioned that 'patrons could not have cared less and in fact did not seem to notice the difference' (i.e. the library collection is not organised using DDC). This further proves that the Dewey-less (which in a sense is classification-less) library does not have any adverse impact on patrons. Another 'no need of classification' ideology is expressed by McKenna (2009), who stated that records managers actually do not need a classification system (taxonomy) when computerising records. What they need is a way to index and then search for information. They need to organise their data to give ordinary end-users an easy time in finding what they need without having to be a trained, professional records manager.

However, most libraries which reported that they are not using DDC or any other classification scheme do not mean that they are not classifying their collection. They are probably not using any recognised classification scheme. They must be adopting some classification approach which suits their purpose.

In the modern age of outsourcing, apprehensions are raised about the effectiveness of in-house classification and its outsourcing. Centralised classification is another approach to library classification. Weaver (1999) and his co-authors presented a study of centralised classification carried out at the Library Management Group of the University of Huddersfield,

UK. The study concentrated on two areas: establishing a benchmark for throughput in technical services in terms of speed, cost and availability of externally supplied class numbers and analysis of differences in classification numbers and subject indexing terms between those that a human assigned and those that a system assigned. The results of the study showed considerable savings to be made in terms of both speed and cost if the classification is carried out centrally. The hit rate for records and the impact on the shelf arrangement were at an acceptable level.

Thus the literature published in the past ten years indicates that classification is indeed a very useful technique of librarianship. Classification schemes and other tools such as thesauri, taxonomies and other controlled vocabularies play important roles in the organisation and retrieval of information in many different environments (Mai, 2008). Classification has diverse applications in librarianship, even in modern aspects of librarianship such as digital library environments, and in other areas such as e-learning (Buchel and Coleman, 2003), the business world (Field, 2000), etc. Classification is a transdisciplinary activity that occurs during all human pursuits. Classificatory activity serves different purposes in different situations. In information retrieval, the primary purpose of classification is to find knowledge that already exists, but one of the purposes of classification in other fields is to discover new knowledge (Beghtol, 2003). These are some of the reasons it is listed as one of the top ten innovations in librarianship (Tyckoson, 2007) and this importance of classification will be sustained in the future.

7

Classification education

Abstract: This chapter reviews literature that deals with classification education, including its importance, curricula, teaching method and levels at which it is taught. The review also covers the literature that discusses the status of classification education in different countries.

Key words: classification education; classification curricula; teaching of classification; classification education in different countries.

Introduction

A fair amount of literature has been produced in the decade covered by this book on classification education. The literature on classification education particularly deals with courses, teaching methods and the content taught or expected to be taught, i.e. the syllabi and to some extent the evaluation system adopted or to be adopted. Classification as a foundation of librarianship is taught in all library and information science schools/departments in all countries. Curriculum, teaching methods and evaluation systems adopted in classification education does change from country to country. Given below is the review of literature that deal with the classification education in different countries of the world. The review in this section is arranged alphabetically by the name of the countries.

Like many other subjects, the library and information science subject is developing in its many forms. Library and information science educators have been updating their curricula so as to accommodate the newly emerging subjects and topics. In this process of updates of curricula, of late it is observed that the emphasis on classification in the LIS curricula is decreasing. However, classification and other information storage and retrieval tools are basic to library activities and so their importance

should not be minimised from the LIS curricula (Brunt, 2007). This view was further supported by Poulter and Brunt (2007). Both of these articles are of interest to classification educators and practitioners.

Classification is mostly taught as one of the subjects in bachelor-level LIS courses. However, Gopinath (1999) suggested that it should be taught as an advanced course in LIS and presented a curriculum structure for such a course, which focused on classification and indexing languages. Use of tutorials and project method as an effective teaching method has been recommended. Classification is taught as a theory subject as well as practical. Oyler (2002) observed that during the past century a theoretical overlay was present in the teaching of classification. The thrust of the courses in the past century was not the organisation of information but the organisation of books in catalogues and on shelves. Over the last 30 years or so the format of information has changed and so have the needs of the users of classification. Considering this change, library and information science schools should teach how to use classification for the purpose of organisation of information, whether in digital form, the WWW or databases, as well as for the organisation of printed books.

Classification education in different countries

Harvey and Reynolds (2005) presented an overview of the status of education for cataloguing and classification in Australia (considered broadly and encompassing descriptive cataloguing, subject access, classification, metadata, knowledge organisation, bibliographic control and other related areas for all formats of library resources). The overview is based on the data collected through websites, printed handbooks and informal discussions with practising cataloguers and classifiers as well as library and information science educators. Britain is rightly credited for the propagation of library and information science education in the Commonwealth countries in general and classification research (e.g. through Classification Research Group) in particular. However, after the introduction of automation in libraries, the basic skills of classification and subject indexing have been little taught in the UK, observed Broughton and Lane (2000). This observation is further confirmed by Bowman (2005a), who, based on a survey conducted in 2003, concluded that cataloguing and classification have become largely invisible in

professional education, but it appears that most courses still include something about them, though not always as a compulsory module and usually without much practical work. Both recent graduates and chief cataloguers complained that what is taught about cataloguing and classification is inadequate. These observations are based on a survey of postgraduate education and training for cataloguing and classification in the UK. Data for the survey was gathered by web content analysis and by sending e-mail requests. Considering the critical status of cataloguing and classification education in the UK, Bowman explored the possibilities of training through commercial providers.

The status quo of cataloguing and classification education in China is discussed by Si (2005). Discussions covered programmes, their curricula, degree offered, textbooks, etc. Problems related with cataloguing and classification education in China are discussed and solutions thereto are recommended. The author projects the probable improvements expected in the next five to ten years. Education of cataloguing and classification in China includes university education, continuing education, professional training and is provided at basic training, junior college, undergraduate and graduate levels. Cataloguing, classification and subject analysis are generally the core courses in the university curricula and offered with other required courses (Ma, 2005). Some of the recent changes introduced in cataloguing and classification education in China are the application of computer technology, the increase of practice, update of course contents and the improvement of teaching methods. Like Si, Ma hoped that there would be constant improvements in the teaching methods adopted for teaching cataloguing and classification.

Like China, the syllabus of classification in the Department of Information Science, University of Zagreb, Croatia, is also updated from time to time and it included (in 2001), apart from the routine contents, the topics on classification of Internet information sources and application of classification for information retrieval and discovery (Slavic, 2002). At this department the wider objective of classification curriculum was to teach content analysis and classification as a process and to demonstrate how to adapt and use classification for different purposes and different environments. Unlike the situation in the UK (Bowman, 2005a), classification has an important place in the curriculum in the Croatian library school at the Department of Information Science, University of Zagreb, because classification is the most important indexing language in Croatian libraries, documentation centres and services, and its role has not been undermined by automation as is the case elsewhere (Slavic, 2002).

In Egypt, too, all library and information science courses are inclusively related to cataloguing and classification of library materials. Abdel Hady and Shaker (2005) analysed the curricula of cataloguing and classification in Egypt and also gathered views of the LIS faculty through a questionnaire and interviews and summarised the changes made in the cataloguing and classification education. Based on the survey the authors also recorded the changes expected in cataloguing and classification education in Egypt in the next five years. They expressed that cataloguing and classification education in Egypt must add more practicals and facility should be made available for the continuing education of the faculty.

As far as India is concerned, two documents (i.e. the LIS Curriculum Development Committee's Reports) – one issued in 1965 and another in 2001 – have influenced university-level courses of knowledge organisation (Raghavan, 2005). These CDC reports formed the basis of LIS curricula in India.

The status of cataloguing and classification education in Iran can be seen by referring to Kokabi's (2005) article. The article is useful for knowing: cataloguing and classification curricula taught in Iran, classification schemes, subject heading lists and cataloguing codes taught during practicals, number of faculties teaching cataloguing and classification courses, number of students taking cataloguing and classification related coursework, etc.

Like many other countries, in Japan the library and information science programmes are offered by library and information science schools (i.e., departments in traditional universities) and by colleges and other universities. Similarly, cataloguing and classification education in Japan has its own issues and future. The specialty of Japan in this context is that there are facilities of on-the-job training and continuing education in cataloguing and classification (Taniguchi, 2005). A very important observation of Taniguchi is that the practices in librarianship influence the cataloguing and classification education. Indirectly this is an advice that while designing the curricula of cataloguing and classification its existing practices should be considered.

Continuing education is essential for updating the skills of working LIS professionals. Continuing education is provided through or accomplished by on-the-job training and self-training. Kwak (2005) conducted a survey of the current status of on-the-job training and self-training provided for cataloguing and classification librarians in 98 Korean academic libraries. All respondents emphasised the need for on-the-job training and 64.3 per cent were trained. The training was given mainly by using print media. The survey also indicated that most academic libraries provided financial

support for staff training. Observing the benefits of on-the-job training and self-training for classification and cataloguing staff, Kwak suggested that more such programmes should be organised and more support should be given for cataloguers' and classifiers' self-training.

The Mexican library and information science schools also have traditionally given considerable importance to cataloguing and classification learning, which still continues. As the recent technological developments have influenced every walk of life, it has influenced cataloguing and classification education in Mexico (Martinez Arellano, 2005). The author compared the different Mexican LIS schools and their curricula and depicted general trends in education for cataloguing and classification.

The curricula taught and teaching methods adopted in teaching cataloguing and classification in Pakistan are dominated by the practices of the 1960s and 1970s. Cataloguing and classification education in Pakistan is also characterised by a lack of new technology, non-availability of competent teachers and poor laboratory facilities. According to Haider (2006), some of the solutions to improve this situation are: (i) revise the curricula immediately; (ii) arrangements should be made for the training of cataloguing and classification teachers (for this, help should be sought from developed countries); (iii) cataloguing and classification laboratories should be updated; and (iv) facilities for the continuing education of cataloguing and classification teachers should be made available.

What can be observed from the above review is that classification education is losing attention compared with some other subjects such as information technology applications. The philosophy, psychology and practices of education are continuously developing. These developments should be tracked and adopted to provide quality classification education. The outcome of the review brings forth so many new concepts and topics which may be incorporated in the curricula. Innovative teaching methods, such as those suggested by Prescott (2001) and Hider (2004) should be developed or adapted and tested for their suitability for teaching classification. So also innovations should be implemented in the evaluation system. Research along these lines will definitely make the teaching and learning of classification more interesting and fruitful. Based on the above review, another observation is that most of the literature related to classification education treats 'cataloguing and classification' together. This joint treatment further proves co-association and strong symbiosis between cataloguing and classification. This symbiosis between these two subjects should be used beneficially and optimally in all aspects of education.

8

Modern knowledge organisation systems and interoperability

Abstract: With the increased use of information and communication technology for information processing, search and retrieval, newer knowledge organisation tools are being developed. These include taxonomies, folksonomies, ontologies, etc. This chapter reviews literature that deals with these knowledge organisation tools, covering various aspects such as their development, use and integration. The emergence of various knowledge organisation tools also created the need for developing interoperable mechanisms. As a result of emergence interoperability, literature on this topic is also increasing, which is reviewed in this chapter.

Key words: taxonomies; folksonomies; tagging; ontology; interoperability.

Introduction

Classification is one of the prime knowledge organisation tools. Thesauri, taxonomies, folksonomies, ontologies, etc., are the other useful knowledge organisation tools. Thesauri, taxonomies and ontologies are used by people working in different domains with different connotations. These terms are used primarily by information scientists, artificial intelligence practitioners and developers of the semantic web. Use of these terms in various domains has given rise to a degree of confusion in the terminology applied to information problems. Gilchrist (2003), in his etymological note, sheds light on the current use of these terms by the various related agencies and clarifies their meanings. This section deals with the literature on these other knowledge organisation tools.

Thesaurus

The thesaurus is one the most used knowledge organisation tools. Thomas's (2004) article is a good starting point for knowing about the thesaurus; it defined thesaurus, explained the procedure of thesaurus construction and the rationale for self-instruction in thesaurus-making. Thomas also discussed the availability and use of free software for thesaurus construction and means of examining individual thesauri. Gopinath (2001a) provided practical steps for applying the classificatory approach to the construction of various types of thesauri, while Kumbhar (2005) explained the advantages of using the speciator-based depth classification schedule for developing thesauri. J. Wang (2006) described an automatic thesaurus development process-based term extraction from title metadata. The three steps in the process included: (i) extracting words and phrases from the title field of the metadata; (ii) applying a method to identify and select the specific and meaningful key words based on the associated controlled vocabulary terms from the thesaurus used to catalogue the objects; and (iii) inserting selected key words into the thesaurus as new terms (most of them are in hierarchical relationships with the existing concepts), thereby updating the thesaurus with new terminology that is being used in the literature. Experiments to add terms from scientific and technical publications demonstrated this method's effectiveness with Chinese Classification Thesaurus (CCT) and bibliographic data in China MARC. Claims are made for its equal effectiveness in large-scale collections and in other languages as well.

Although there exists a sizeable amount of literature on style and structure of thesaurus and other controlled vocabularies, Aitchison and Clarke (2004) observe that the expansion of end-user access to vast networked resources is imposing further requirements on the style and structure of controlled vocabularies.

Taxonomies – meaning

Taxonomy is the classification of terms within subject domains. It retains many of the characteristics of classical library classification arrangements and can be hierarchical systems, employing controlled vocabularies and thesauri, but is not always based on the accepted standards used in traditional library settings (Miskin, 2002; Cote, 2005). Raschen (2005) traced the practices of taxonomy, which have been in use since the

nineteenth century. Z. Wang et al. (2006) expressed that the term taxonomy is of recent origin and still used loosely and obscurely; a novice reader interested in knowing about all aspects of taxonomy may refer to this article of Z. Wang et al. (2006). It is a very useful reading on taxonomy as it defines taxonomy, explains its nature, scope and role and also differentiates between taxonomies and other knowledge organisation tools. For a list of resources including websites, software and examples of taxonomies on law websites, one can refer to Miskin (2002). A glossary of taxonomy is a feature of Miskin's article.

Uses/importance of taxonomy

Due to the need to improve knowledge-sharing through classification and retrieval and the need to develop the semantic web, the importance of taxonomies has been realised much more in recent years. Their importance is growing because it: (i) helps to avoid the problems created by similar-sounding words or words with multiple meanings; (ii) gives web resources a more professional, organised look; and (iii) improves site navigation (Raschen, 2005). Unstructured knowledge is inconvenient and difficult to acquire, organise and share, but taxonomies help to organise and share this knowledge. More specifically, as stated by Gilchrist (2001), taxonomy is useful as a source of authority for tagging, it is useful as an aid in navigation, it is a tool that supports search engines, it works as a knowledge map and it is also useful as a depository of enterprise retrieval languages. Taxonomies are developed considering such multiple uses and Bertolucci (2003) discussed advantages of using taxonomies. According to Corcocan (2002), taxonomies can provide a structure for content management at an enterprise or intra-enterprise level, and are being developed to support the needs of internal and external web initiatives.

The legal profession is one of the professions which deals with interpretations of terms. Use of appropriate terms enhances the efficiency of the information storage and retrieval system. Miskin (2002) explained the usefulness of taxonomies in information retrieval, more particularly in retrieving legal information. Taxonomies are domain-based tools and law is one of the domains where there are a number of taxonomies. Similarly, Cisco and Jackson (2005) explained that taxonomies are useful in managing mass electronic records for legal, business and regulatory purposes and are required for efficient retrieval of electronic records.

E-governance has been having a wide and rapid impact, but appropriate controlled vocabulary is a prerequisite for the successful implementation of e-governance. For example, the UK government previously used the Local Government Category List (LGCL) for e-governance and this list was replaced by a more effective single taxonomy entitled Integrated Public Sector Vocabulary (IPSV). It contained 2,700 terms and this taxonomy had a wide impact on UK public sector websites. Cameron (2005) explains the nature and working of this taxonomy with the help of a case study of Leicestershire County Council's website. This taxonomy helps in easy search of information from the local authority websites by harmonising contents and classifying it consistently within each council and with every other UK council. Skinner, Han and Chang (2006) developed a taxonomy of terms related to 'information privacy' which will be useful in formulating better information privacy policies, practices and privacy-enhancing technologies.

Taxonomies are useful in identification, control and utilisation of all contents developed by a publisher over the years, which is illustrated by Levinson (2006), who is both a writer and a publisher. Levinson applied taxonomy in his publishing firm's 10-volume *Encyclopedia of World Cultures*, *Encyclopedia of Community*, *Encyclopedia of Modern Asia* and *Encyclopedia of Leadership*. Taxonomies have a role in knowledge management as well. To conduct a business successfully, it is essential to manage the knowledge related to the concerned business. Classification and taxonomy play a vital role in knowledge management. Earley (2006) highlights the importance of classification in knowledge management as it offers the benefits of providing a continuum of knowledge processes, from unstructured to highly structured. While emphasising the importance of a taxonomy covering business processes to the most minute extent, the author observed that a taxonomy developed for a wide set of processes may not be suitable for a specific set of capabilities. Taxonomies have proved useful in classifying information sources and services within intranets and portals (Milne, 2007). Taxonomies, particularly corporate taxonomies, facilitate knowledge audits and lead to a greater organisational impact (Sharma, Foo and Morales-Arroyo, 2008).

Design and development of taxonomy – procedure-oriented literature

Usually taxonomy development involves four phases, i.e. planning/analysis, design/development/testing, implementation and maintenance.

Cisco and Jackson (2005) elaborate these four phases and best practices in the development of taxonomies. Design and development of taxonomy is a demanding task and for an effective taxonomy one has to select and use appropriate software, coordinate the taxonomy development work through colleagues and test and implement it optimally (Raschen, 2005). These and other activities involved in the development of taxonomies are further illustrated by Cote (2005). One can also refer to Miskin (2002) for an elaborated account of the idea of self-help in designing and developing taxonomy. Steps in developing successful enterprise taxonomy plans are explained by Corcocan (2002). Basically, taxonomies organise information into hierarchical categories. Bertolucci (2003) asserts that category creation is an art, which is useful not only in developing taxonomies but also in understanding them. Bertolucci explained the concept of category creation with the help of the cartoon character 'Snoopy'. This is indeed a very innovative method adopted to explain the concept of categorisation. The author further stated that in developing a useful taxonomy, it is essential to discover the client's individual method of interacting with information.

Albertsen and Van Nuys (2005), who formed various relationships into component aggregates and grouped these aggregates into various classes, describe another taxonomy development procedure. This procedure was adopted in the Paradigma Project at the National Library of Norway and its work to ensure the legal deposit of all types of digital document. Cheung, Lee and Wang (2005) developed a broader, multifaceted taxonomy system for efficient management of unstructured knowledge, which not only helped to reduced human effort but also saved time and the cost of knowledge management. The system consisted of five components, i.e. multidimensional taxonomy structure, thesaurus model, automatic classification mechanism, intelligent searching and self-maintenance of taxonomy. Artificial intelligence (AI) and natural language process (NLP) technologies were also used in the development of this multifaceted taxonomy system. Cote (2005) explained the various taxonomy architectures using which taxonomies could be developed. The architectures explained by Cote include: flat taxonomy architecture; facet taxonomy architecture; hierarchical taxonomy architecture; network taxonomy architecture, etc. The insights of faceted classification and the possible relationships between terms and concepts in ontology prove helpful in taxonomy development. This was experienced while developing the Integrated Museum and Archives System taxonomy in Singapore (Chaudhry and Jiun, 2005).

Taxonomies could also be developed by considering the information and communication actions. Action-oriented service and technology development begins with the idea that people use technologies to reach their action goals. Parkkola, Saariluoma and Berki (2009) investigated family communication and mobile communication service types for families. For this the authors interviewed ten mothers to investigate the nature of their everyday information and communication needs. They also studied the various knowledge and information communication actions that were adopted in their families. Based on the qualitative analysis, the authors developed taxonomy. The authors believed that this taxonomy will help to improve individual services and family-centred design models. This method could also be adopted to develop taxonomy based on library users' everyday information communication actions. There are various types of taxonomy designs: functional, corporate-wide or uniform taxonomies are examples. Designing functional taxonomy is a complex task and, as such, there can be many shortcomings in these designs. Connelly (2008) explains the tactics useful in overcoming probable shortcomings in designing functional taxonomies.

Dutta, Majumder and Sen (2008) test another method of taxonomy development. The authors classified the key words extracted from research articles published in science journals using the INSPEC bibliographic database. The key words related to Fermi liquid, a micro topic from condensed matter physics, and were extracted on the basis of decreasing order of relevance. The authors first classified the key words into three categories, i.e. keyphrase, modulator and qualifier. The keyphrase reflects the central concept, which is usually post-coordinated by the modulator to amend the central concept in accordance with the relevant context. The qualifier comes after the modulator to describe the particular state of the central concept and/or amended concept. The key words were further classified in 16 classes on the basis of four parameters, i.e. associativeness, chronological appearance, frequency of occurrence and category. The authors claimed that this taxonomy will enable analysis of research trends of a subject and potential research topics. Availability of an appropriate design framework makes the taxonomy designing task easier. A framework for developing a corporate taxonomy and guidance in executing taxonomy development projects is provided by Sharma, Foo and Morales-Arroyo (2008).

Taxonomies consist of categories. The basis of categorisation does affect the effectiveness of the taxonomy. To prove this point Hider (2009) compared the resource description and access (RDA) taxonomies with

the end-user categorisation of content and carrier. It is known that the AACR II used general material designations (GMDs) and specific material designations (SMDs) characteristics for describing the document in bibliographic record. The RDA taxonomy provides a new list of content and carrier types to replace the GMDs and SMDs. Both the GMD and SMD are taxonomies designed to facilitate searching on content and carrier attributes of resources. The author claimed that these taxonomies are not designed by analysing the end-user categorisation nor have they been tested on end-users. Hider investigated how end-users categorise library resources. For this, he employed a free listing technique. The results indicated that the end-user categorisation of library resources might emphasise other facets such as purpose, audience and extent. These facets are in addition to the content and carrier which need to be considered in developing taxonomies.

So where should the taxonomy development process begin? According to Chaudhry and Jiun (2005), collecting terms and concepts from various external and internal sources is sufficient to kick-start the taxonomy development process. Then, taxonomies could be either developed in-house or they could be purchased from outside agencies. Cisco and Jackson (2005) discuss the advantages and disadvantages of each of these approaches. In addition, a centralised or decentralised approach could be adopted for the development of taxonomies. However, considering the merits and demerits of each of these practices, Brenner, Kolbe and Kremer (2005) warn that only a well-suited trade-off between centralised and decentralised terminology management will be sustainable.

Domain-based taxonomies

Modern-day taxonomies are the result of the emergence of an e-governance culture in the corporate world, where a huge amount of electronic information is generated, managed, searched and retrieved. Gilchrist (2001) conducted a survey to study the practices of building and using taxonomies by corporate enterprises in managing their information portals and specifically reported the efforts of four corporate enterprises in developing taxonomies: the British Broadcasting Corporation (BBC), Glaxo Wellcome, Microsoft and Unilever. A narrative account of the experience of developing a taxonomy of computer science terms is provided by Saeed and Chaudhry (2002). For developing the taxonomy of computer science, the authors combined terms from the Dewey Decimal Classification's relative index and the terms in the IEEE Web

Thesaurus with Dewey Decimal Classification hierarchies. The taxonomy was displayed using 'MyInfo' shareware and proved promising for browsing electronic information sources.

Caton, Hawkins and Larson (2003) suggested a taxonomy structure for information science. The authors provided development of a new definition of the field of information science and the creation of a 'map' of the field showing subjects central to it and their relationships to those on the periphery. The case study presented by the authors described the creation of a new classification structure (taxonomy) for the Information Science Abstracts (ISA) database, aiming to reflect and accommodate the rapid and continued technological and market changes affecting the information industry today and in the future. Based on a sample of some 3,000 abstracts from ISA, two validation experiments were conducted by a three-member team comprising a database editor, a reference librarian and an abstractor-indexer, who represent three of the major communities within the information science field. In the first experiment, the sample of abstracts was classified according to the proposed new taxonomy; after analysis of the data and revision of the taxonomy, it was revalidated and fine-tuned in a second experiment. Indexer consistency measures obtained in this study were significantly higher than those found in previous studies.

Considering that glossaries and taxonomies help in solving a wide range of terminological problems, Brenner, Kolbe and Kremer (2005) developed a procedural model for building and maintaining glossaries and taxonomies. The model is based on extensive literature review and analysis of three action research cases: (i) introduction of a glossary at an insurance company; (ii) setting up a corporate taxonomy at an international professional service firm; and (iii) combining glossary and taxonomy for document classification and retrieval. Storage, organisation and retrieval of cultural heritage information pose various problems due to its cross-cultural, cross-domain nature. However, taxonomies prove useful in organisation and retrieval of networked information about cultural heritage. Considering this benefit, Chaudhry and Jiun (2005) developed a 500-term taxonomy for a Singapore-based cultural heritage network. The authors described the methodology adopted for developing this taxonomy and discussed the problems and advantages of this methodology. Taxonomies play a vital role in managing information in a business consulting environment. So, Chaudhry and Ling (2005) developed a taxonomy system for a regional business consulting company. This taxonomy consisted of 12 main categories and approximately 500 terms. It is based on existing knowledge structure and information needs

of consultants in the selected companies. The authors claim that other companies can conveniently adopt this prototype in their efforts to develop their own taxonomy.

The concept of information privacy is fast developing, with many new definitions and terminologies to express its various aspects. At the same time the collaborative environment is also expanding. The proper organisation of terms related to information privacy will enable better understanding of the concept. Considering the usefulness of organised terminologies of information privacy in the collaborative environment, Skinner, Han and Chang (2006) developed a taxonomy of terms related to information privacy. For developing this taxonomy, terms were collected by surveying each aspect of information privacy and were classified into a number of sub-categories according to their nature and relevance. The taxonomy consisted of three high-level dimensions, each of which is further sub-classified. It is the first of its kind.

All knowledge organisation systems such as a classification scheme, subject heading lists, thesauri, taxonomies, etc., have their own merits and demerits. In order to gain from the features of different knowledge organisation systems, efforts have been made/are made to combine the meritorious features of different knowledge organisation systems to produce a new knowledge organisation system. Thesaurofacet and Classarus are examples of such efforts in the past. These two new knowledge organisation systems were developed in the past by combining the features of faceted classification with thesauri. So, is it possible to combine features of existing knowledge organisation systems with those of taxonomy? Yes: one of the earliest efforts in this direction was development of taxaurus for the UK Government Communications Headquarters (GCHQ). It combined the features of a thesaurus. The taxaurus was specially developed to improve the information retrieval on GCWeb, the departmental intranet (Foster et al., 2001). Another example of the development of such a hybrid knowledge organisation system is the taxaurus developed by the Information Services Department at Ricardo, UK, for the domain entitled 'mechanical engineering and engineering consultancy in automotive engineering and heavy duty engine engineering'. Ward (2006) described the process, rationality and constraints of developing such a terminological control tool.

Taxonomies could also be developed by using contextual classification. Based on his personal experience, literature and discussions with peers, Milne (2007) expressed that contextual classification is a useful tool for taxonomy development. More particularly, the author tested the value of contextual classification in taxonomy development by developing

taxonomy for a university portal and suggested that further exploration in this direction is needed.

As could be seen from the above two and many other examples, principles, theories and practices of librarianship are adopted in designing and developing taxonomies. This point affirms that the traditional library and information science skills and competencies are useful in the design and construction of taxonomies – and that is the reason that library and information professionals can play a vital role in the development of taxonomies (Cote, 2005). Corcocan (2002) also stated that library and information professionals worked on classification for many decades and that the principles of classification are useful in developing taxonomies. This is why library and information professionals should get involved in developing corporate taxonomies. Corcocan reiterated the emphasis by stating that taxonomies are nothing new for librarians, who have been developing and using classification schemes for years, but are now being recognised at a higher level in corporate management.

Software for taxonomies

The task of design and development of taxonomy could be carried out in an effective manner by using appropriate software. Collaborative Classifier 1 (VCC1) is one of the software packages useful for the development and maintenance of taxonomy. It is an enterprise-wide taxonomy application launched by Verity (*www.verity.com*) and provides the facility of real-time workflow editing. Using this facility a library can grant editing permissions to subject experts to help classify documents and establish rules for their classification within taxonomy (Pickering, 2004). MultiTesPro, version 2007.02.01 from Multisystems (Miami, Florida); Term Tree 2000, version 2.3 from This to That Pty Ltd. f/a A.C.S. Active Classification Solutions (Australia); and TCS-10, version 2.26 from Webchoir, Inc. (Los Angeles, California) are some other software packages that exclusively facilitate the creation and maintenance of taxonomies (Hedden, 2008).

Ontology – meaning

Ontology, in its philosophical meaning, is the discipline investigating the structure of reality, which can be structured into a series of integrative levels, such as the physical, the biological, the mental and the cultural,

and each level acts as a base for the emergence of more complex levels. The finding of ontology can be relevant to knowledge organisation and models of knowledge can, in turn, offer relevant ontological suggestions (Gnoli and Poli, 2004). The notion of ontology could be better understood with the help of the means-ends tool provided by cognitive work analysis (Park, 2008). For conceptual understanding of ontologies along with a brief overview of the topic, refer to Gokhale (2009).

Development of ontology

The library and information profession has evolved many theories and practices for developing knowledge organisation tools. The knowledge organisation tools developed by the library and information science profession could be adopted in developing ontologies. Zeng (2005), who suggested a procedure to convert the Chinese Classified Thesaurus (CCT) into ontology in OWL Lite, demonstrated this. The CCT contains a large number of compound concepts, which create problems in the development of ontology. Zeng has provided a solution for revealing the semantics behind the compound concepts. Park (2008) proves the usefulness of the context-centred approach and faceted classification in developing ontologies with the help of a case study of wine ontology. The task of ontology construction is critical and time-consuming. How computers can make this task a little easier is demonstrated by Jun and Yuhua (2009), who introduced an automatic approach for ontology-building by integrating traditional knowledge organisation resources. Jun and Yuhua first developed a primary ontology describing the classes and relationships involved in bibliographic data with OWL, then they filled the primary ontology with instances of classes and their relations extracted from the catalogue dataset and thesauri and classification schemes used in cataloguing.

Ontology could also be developed by using the user's borrowing record. This is demonstrated by the Personal Ontology REcommender (PORE) project. In this project, Liao et al. (2009) developed users' personal ontology by using the traditional cataloguing scheme, classification for Chinese libraries, as the reference ontology. The authors then transformed this reference ontology to a unique personal ontology for each user, based on mining results from library borrowing records of that user. This PORE ontology is presently implemented for Chinese collections. For English collections a new version of this ontology can be developed which may be based on the Library of Congress Classification or any other classification scheme.

Uses of ontology

Ontologies are useful in organisation of knowledge. They are particularly useful in the organisation of the vast stock of knowledge on the Internet (Gokhale, 2009). More particularly, ontologies could be useful in: (i) conceptual navigation; (ii) semantic retrieval; (iii) enhancing power of search engines; and (iv) web-based retrieval (Jun and Yuhua, 2009; Liao et al., 2009). Ontologies could also be used as a personalised recommender system. The personalised recommender becomes a very important service when library collections are growing rapidly. PORE, designed and developed by Liao et al. (2009), was developed by using the patrons' borrowing records. This ontology can be used for three purposes: to know unique user interest on specific subjects; to filter out unsuitable recommendations based only on a keyword matching method and the recommended books can be organised into the personal ontology; and provide the patron with a user-friendly interface to access library collections. The usefulness of classification schemes in ontology construction is evidenced from the writings of Zeng (2005) and Park (2008). However, classification schemes could also benefit from ontology as it provides structural levels of reality. These structural levels are considered as a source for structuring principles in bibliographic classification by CRG and by Ingetraut Dahlberg (Gnoli and Poli, 2004).

The above literature on ontology indicates that the subject of ontology is developing fast. However, Park (2008) believes that it is not sufficiently explored as yet and so there is scope for further research in this area.

Folksonomies – meaning

People who are not subject experts assign subject tags to websites and/or to their contents, such as photographs, web links, artwork and other resources. These tags are used for search and retrieval of websites. Thus the tags or terminologies assigned by untrained, non-subject expert individuals (folks) for describing web content is known as folksonomy. The process of assigning tags is called tagging. As it is assigned by people in general, thereby indicating society's participation, it is also referred to as social tagging. In the process of assigning tags, many people join each other and so it is also known as collaborative tagging. Tagging is one of the most widespread services of web 2.0.

Noruzi (2006) referred to collaborative tagging as folksonomy and defined it as a free-form tagging, a user-generated classification system of

web content. He also refers to tagging/folksonomy as concepts, categories, facets or entities. Peterson (2006) defined folksonomy as an Internet-based information retrieval methodology based on collaboratively generated, open-ended labels categorising content such as web pages, online images and web links; or as 'social bookmarking, emphasising the social networking often achieved by a collaborative effort of assigning subjects and tagging'. Tagging has its own nature and purpose (Noruzi, 2006; Macgregor and McCulloch, 2006). Hammond and his co-authors (2005) present a valuable general review of social bookmarking tools. According to the authors, server-side software aimed specifically at managing links with, crucially, a strong, social networking flavour and an open and unstructured approach to tagging, or user classification, of those links are some of the important features of these social bookmarking tools. The tags are also referred to as marginalia. Abbas (2007) in his article 'In the margins: reflections on scribbles, knowledge organisation, and access', defined marginalia or 'scribbling in the margins' (i.e. tags) as a means for readers to add a more in-depth level of granularity and subject representation to digital documents such as those present in social sharing environments like Flickr and del.icio.us.

Folksonomies are useful to both libraries and users. Their strength, particularly from the libraries and information centres' points of view, is that they allow the user to get involved in information management, thereby making the information system interactive. Folksonomies are also useful in knowledge and resource discovery (Macgregor and McCulloch, 2006; Matusiak, 2006; Noruzi, 2006; Spiteri, 2006; Sanders, 2008). Tags being users' terminology helps libraries in knowing their perspective of knowledge organisation (Abbas, 2007). For users, it allows them to tag their favourite web resources with their chosen words or phrases selected from natural language; users can express their preferences for the descriptive words (Noruzi, 2006).

According to Matusiak (2006), low precision, lack of collocation, etc., are some of the shortcomings of tagging. These shortcomings are the reasons for favouring controlled vocabulary, explained Noruzi (2006). Both user-generated tags (also called user-generated metadata) and professionally developed metadata, i.e. controlled vocabularies such as taxonomies, have their own features and advantages. Because of this, folksonomies and controlled vocabularies will co-exist and complement each other (Matusiak, 2006; Macgregor and McCulloch, 2006). Users and information professionals have their own views about folksonomy. From the information professional's point of view, folksonomies are unsystematic and unsophisticated; however, from the Internet user's

point of view, they are easy to use (as there is no complicated, hierarchically organised nomenclature to learn) and are cost-saving tools. Continuing the argument for and against folksonomies, Tebbutt (2005) stated that it has become difficult for the creators of taxonomies to keep pace with the user-contributed tags because of their ever-growing and changing nature. On the other hand, the author expressed optimism, stating that more good than ill can come from social tagging because it may enable taxonomy developers to gain access to terms that more realistically indicate the real usage of such terms in practice.

Tagging patterns

Del.icio.us has one of the largest folksonomies on the Internet in terms of number of users and tagged websites. Munk and Mork (2007) studied the procedure of tagging on del.icio.us to find out the patterns in tagging. The statistical analysis of 76,601 different key words with a total frequency of 178,215 from 500 randomly chosen taggers showed that: (i) the distribution of key words follows classic power law; (ii) there are distinct tagging communities; (iii) the most frequently used tags are situated on a general–specific axis; and (iv) there are nine distinct tagging strategies. Another effort in exploring the tagging pattern is made by Angus, Thelwall and Stuart (2008). The authors intended to know the general patterns of tag usages and determine the usefulness of the tags used within university image groups to the wider Flickr community. Flickr is one of the most popular social networking websites. It is an online image management application. The study was carried out by adopting the webometric data collection, classification and informetric analysis method and it led to the conclusion that the members of university image groups tend to tag in a manner that is of use to users of the system as a whole rather than merely for the tag creator.

Folksonomies and controlled vocabularies

While differentiating between folksonomies and taxonomies, Bianco (2009) stated that people who are not subject experts assign subject tags to the websites they use for future search and retrieval of websites. Thus, the tags or terminologies used to describe websites assigned by untrained, non-subject expert individuals (folks) are known as folksonomies. This is also known as social tagging or tagging. On the contrary, taxonomies are

developed by subject specialists using authorised terms determined by professionals. Bruce (2008) conducted a study to determine the overlap between the controlled and uncontrolled vocabulary. For this the author used the descriptors and tags from journal articles indexed in the Education Resources Information Centre (ERIC) database and the folksonomy-based website CiteUlike. The author collected metadata from 2,786 journal articles indexed in the above two datasets using Perl and MySQL. The tags used by the users, i.e. the uncontrolled vocabulary, did not match with the subject specialists' terminology, i.e. the controlled vocabulary, thus revealing that there was not much overlap between the two types of terminology.

Integration of folksonomies and controlled vocabularies

Teng and Ke (2009) present a case study of creating a tag map by combining social tagging and descriptors assigned by professionals to artworks. The study was carried out by using the case of the artworks available at the Yuyu Yang Digital Art Museum. Initially the authors used the CKIP Chinese Segmentation System to process the metadata of each artwork and extract the key words according to their weights. These key words represent professional descriptors. Then the public was allowed to freely tag different artwork. The descriptors and the tags were then combined to form hierarchical tag clusters and generate a tag map for every category of artwork. The authors stated that the map achieved the goal of folksonomy; however, it is more than that, as it reflects not only the users' perspective but also the experts'. Thus, it can be said that folksonomies can play an important role in developing user-based controlled vocabulary. However, to what extent they follow the standard guidelines recommended by institutions such as the National Information Standards Organisation (NISO) for the construction of controlled vocabulary is a matter of investigation. To find an answer to this curiosity, Spiteri (2007) collected folksonomy tags from the daily tag logs of del.icio.us, Furl and Technorati and evaluated them against the NISO guidelines for the construction of controlled vocabulary. The evaluation indicated that the folksonomy tags assigned by the users closely resembled the NISO guidelines in terms of types of concept expressed, the predominance of single terms and nouns, and the use of recognised spellings. However, the author found that there were inconsistencies in the use of count nouns, homographs, abbreviations and acronyms. The article

suggested that with the addition of guidelines for developing unambiguous tags and links to appropriate resources, more effective folksonomies could be developed.

Integration of folksonomies and ontologies

For effective retrieval, web content needs to be organised systematically. Various knowledge organisation tools are used for effective retrieval. Classification schemes, thesauri, ontologies and folksonomies are some of the tools forming part of the web information classification engineering. All of these tools have their own strengths and weaknesses. For example, flexibility, collaboration and information aggregation are the strengths of folksonomies; whereas standardisation, automated validation and interoperability are the merits of ontologies. Though the structure of ontology and folksonomies is different, their role is common. Thus, considering that the reconciliation of these two web information organisation tools may be beneficial, Dotsika (2009) proposed a framework for the same.

Application of folksonomies

As can be seen from the above references, there is criticism as well as appreciation of social tagging. As only their usefulness can prove their worth, Goh (2009) and his co-authors investigated if documents could be assigned to their associated tags. For this purpose, the authors downloaded web pages and their associated tags from del.icio.us. Content analysis was also adopted as a part of the methodology to find characteristics of effective and ineffective tags. Based on these methodologies, the author carried out two text categorisation experiments using support vector machine classifiers. In the first experiment the author used the terms from the documents as features and in the second he included both, i.e. the terms from the documents and tags. The performance of both of these sets was tested using criteria such as precision, recall, accuracy and F1 score. The results indicated that: (i) the tags did not have much impact on the performance on either side, i.e. did not improve or degrade much; (ii) all tags are not useful in resource discovery by public users, thereby confirming that there may be other dynamic reasons for tagging which only the tagger knows. These conclusions may be useful to library and information professionals for contributing to the social tagging system and using tags and tagging for development and management of

information storage and retrieval systems. Folksonomies are useful for enhancing interactivity of public library catalogues; they also help users to organise personal information spaces, supplement existing controlled vocabularies and form informal online communities of interest (Spiteri, 2006). Folksonomies are also useful for initiating user-driven readers' advisory services (Spiteri, 2007). Application of tagging by medical librarians is reported by Bianco (2009), who cites a survey of tagging by Americans, which indicated that use of tagging is not restricted by income, age or ethnic group.

Scope for research in folksonomies

Quantitative and qualitative analysis of folksonomies and users' meta-tagging behaviour are some of the areas for further research in tagging (Spiteri, 2006). The knowledge organisation community should critically examine its understanding of tags as an emerging classificatory schema and determine how best to adapt, augment and revitalise existing knowledge organisation structures (Abbas, 2007).

Interoperability – meaning, need and methods

With the increase in the volume and complexity of information, newer knowledge organisation systems are evolving. From the traditional classification-based systems to modern taxonomies, ontology is the direction of evolvement for knowledge organisation systems. The multiplicity of knowledge organisation system has created a need for interoperability amongst these systems. Language and structure of knowledge organisation systems are the core issues involved in the emergence of interoperability. Considering the importance and need, various interoperability systems have evolved in recent years. Derivation/modelling/translation/adoption, satellite and leaf mode linking, direct mapping, co-occurrence mapping, switching, linking through a temporary union list and linking through thesaurus server protocol are some of the interoperability methodologies that have been developed and tested (Chan and Zeng, 2004). The authors also reviewed trends in the development and use of interoperability and stated that the process and issues of interoperability are unavoidable in today's networked environment. Mai's (2004c) survey report of the concepts of interoperability and

switching language is a useful reading on this topic. Agreeing with Chan and Zeng, McCulloch (2004) stated that multiple terminologies create a number of obstacles in knowledge organisation and retrieval. Of course these obstacles can be overcome through interoperability, particularly terminology mapping. Semantic interoperability could also be achieved by improving compatibility between terminologies and classification schemes. Interoperability enables users to search multiple resources simultaneously. For example, Mai (2004c) expressed that switching systems helps to search multiple collections using a single retrieval language. Yet another interoperability solution to improve retrieval performance suggested by McCulloch is to use the associated terms drawn from several schemes. Various interoperability solutions tend to have their own strengths and weaknesses. One such weakness is that mapping between more indexing languages will always be an approximation (Mai, 2004c). Shiri and Molberg (2004) evaluated the various knowledge organisation system interfaces and concluded that searching, browsing and navigation facilities as well as bilingual features call for improvements.

The HILCC (Hierarchical Interface to Library of Congress Classification) project at Columbia University Library generated a structured, hierarchical menuing system for subject access to resources in the library's electronic collections by using the Library of Congress Classification numbers as provided in standard catalogue records. Developed jointly by the library's systems, cataloguing and reference staff, this classification mapping table linked each Library of Congress Classification range with entry vocabulary in a three-level subject tree. Classification numbers and other metadata elements were extracted from catalogue records in the OPAC on a weekly basis and matched against the HILCC mapping table and then used to create browsable subject category menus to guide users to electronic resource subject content (Davis, 2002). Frank and Paynter (2004) provide another example of an interoperability system. They developed an interoperability system which automatically assigned Library of Congress class numbers to a work given its set of Library of Congress subject headings. The system organised the Library of Congress Classification in tree structure, i.e. the root node comprised all possible topics and the leaf node comprised the most specialised topics. This interoperability system used machine learning techniques and training data from a large library catalogue to learn a model that maps from sets of Library of Congress List of Subject Headings to classification from the Library of Congress Classification tree. Mapping is one of the methods of interoperability. Jianbo and Hanqing (2005)

analysed the feasibility mapping of enumerative classification, discussed the mode of mapping and the principles of an automatic mapping system. The authors constructed four semantic mapping relationships based on the differences of class meaning between the Chinese Library Classification and the Dewey Decimal Classification. When one has to choose between a general classification and special classification as a switching language, one should prefer general classifications as they provide access to collections nationally and internationally (Mai, 2004c).

9

Text categorisation

Abstract: This penultimate chapter reviews literature that discusses the meaning and uses of text categorisation, as well as the literature that describes the application of various text categorisation algorithms, such as support vector machine (SVM), k-NN, Naive Bayes, etc.

Key words: text categorisation; k-NN; k-Nearest Neighbours; Naive Bayes; SVM; support vector machine; machine learning; uses of text categorisation.

Introduction

There has been a dramatic increase in web information. Information on the web is unstructured, non-indexed and unorganised but automatic text categorisation techniques help to organise and retrieve information effectively from the web. For a detailed discussion along with a description and features of the various tools and techniques adopted in text categorisation, one can refer to the article written by Farmer (2006): 'Automatic categorisation: what's it all about?' Farmer discusses the issue of human–machine interaction required for an effective categorisation. Xiaoge (2009) provides a brief review of the literature dealing with the various aspects of text categorisation. The importance of automatic classification is discussed by Liu (2008) and Humphrey et al. (2009). Automatic classification involves definite procedures, e.g. feature extraction, feature selection and building classifier (Ishida, 2006), and has its own advantages and disadvantages (Lubbes, 2001; Schewe, 2002; Golub, 2006; Waters, 2007). Toth (2002) explained the value of practical solutions, including automatic text classification, for accessing digital documents on the Internet and presented a brief review of various

methods applied in automatic classification. Realising the importance of automatic classification, various organisations have undertaken projects to explore its potential. Some such projects are: Nordic WAIS/WWW; DESIRE2; Engineering Electronic Library System (EELS); GERHARD; and SCORPION. Toth provides details of the activities of these projects.

Text categorisation – different nomenclatures

The term automatic classification has a magnitude of meanings based on different contexts. These contexts could be grouped into two broad categories: (i) automatic classification of electronic/digital text – this includes classification of websites and text available in electronic format; (ii) automatic classification of books, i.e. a system that helps to construct class numbers using the electronic version of a book classification scheme. This is referred to in this book as 'automatic book classification'. The automatic classification in the second sense is more useful for managing print collections of books and other similar forms of document. Unfortunately, though, the nature and quantity of literature available shows that automatic classification in this second sense is yet to get the desired momentum. Nevertheless, research on automatic book classification is in progress and literature on this topic does exist (reviewed in Chapter 6 of this book). Of course, there is literature that discusses the applicability of the principles and practices of bibliographic classification to automatic text classification. For example, Yi (2007) reviewed the tools, methods and models developed for automatic text classification, which are based on bibliographic classification schemes, and discussed the issues and challenges in the adoption of bibliographic classification schemes in automatic text classification.

There is sizeable literature representing the first context, i.e. categorisation of electronic text, websites, etc. This section reviews literature related to automatic categorisation/classification of text. This is mostly referred to as automatic classification and/or automatic text categorisation or simply text categorisation. Nevertheless, there remain structural and semantic differences between classification and categorisation, which, according to Jacob (2004), affect the functional activities of an information environment. Another similar term is clustering. Clustering (also called cluster analysis) is the assignment of a set of observations into a sub-set. Though broadly similar, there are differences between clustering

and text categorisation. Clustering is a way to group text according to its character; text categorisation classifies text according to predefined categories (Ishida, 2006). The literature reviewed here adopts a general approach and, as such, treats all of these terms synonymously and uses them interchangeably.

Considerable work is being carried out on automatic classification. Analysis of this work indicates that there exist three different approaches to automatic classification. These are based on the three types of research communities: machine learning, information retrieval and library science. All of these approaches have two similarities: document pre-processing and utilisation of web-specific document characteristics. They have differences in algorithms applied, employment or not of the vector space model and controlled vocabularies. Golub (2006) described, compared, evaluated and discussed the applications and problems of these approaches. Probably this is the first attempt to discuss more than one community's approaches towards automatic classification in an integrated manner. As such, it is a useful document for library and information professionals, students, practitioners as well as for computer science professionals.

Text categorisation of non-English text

Most of the research in text categorisation carried out so far has concentrated on the categorisation of English-language text. Most of the algorithms developed so far are capable of classifying text in English. Now the professionals are exploring the potential of categorising texts in other languages. One such effort was reported by Kanaan et al. (2009). The authors applied k-NN, Rocchio and Naive Bayes techniques to classify Arabic text. By classifying 1,445 Arabic text documents belonging to nine categories, the authors concluded that the Naive Bayes performed better, followed by k-NN and Rocchio. Another similar effort was made by Duwairi (2006), who proposed a distance-based classifier for categorising Arabic text, which consisted of: (i) representing each category as a vector of words in an m-dimensional space; (ii) classification of documents on the basis of their closeness to feature vectors of categories; (iii) extraction of features of categories that capture inherent category-specific properties by scanning the set of training document – this is done by the classifier at the learning stage; (iv) categorisation of unclassified documents by using the previously determined category-specific features. On the basis of several categorisation tasks on

an in-house collected Arabic corpus, the author reported that the proposed classifier is very accurate and robust. Xin and Renren (2008) reported a similar effort made to develop an algorithm to categorise Chinese text.

Tools, models, methods and algorithms used in text categorisation

There are various techniques, methods and/or algorithms developed for automatic classification. A review of various methods used for automatic classification could be found in the article written by Toth (2002). Sen et al. (2008) provide a historical account of the classification techniques used in classifying networked data along with a comparison of various algorithms. Since there are many algorithms, they themselves could be categorised into strictly supervised and weakly supervised algorithms (Zhang, 2008). Zhang submits that a variety of weakly supervised learning algorithms can be applied to take advantage of large amounts of unlabelled data when labelling is expensive.

The following paragraphs review the literature dealing with the various algorithms developed and tested for automatic classification. The literature in these paragraphs is reviewed mostly in chronological order and, within a chronology, in alphabetical (by author name) order. It is reviewed in chronological order because it indicates the developmental trends. Of course thematic association overrules the above arrangement.

A language-independent document classification system that provides easy access through the web was developed by Liu (2000) and his co-authors. The system is capable of displaying the classification results using graphical user interface and is useful for retrieving and analysing sets of documents from public websites. Panigrahi (2000a) reports on the usefulness of artificial intelligence and natural language processing in developing an expert system for automatic classification. Due to the enormous dynamic range of documents and high computational cost, existing approaches were seen as inadequate in classifying documents reliably. To reduce the computational complexity, Siva et al. (2000) suggested a new classification approach involving a preliminary data reduction step, grey-scale projections, transform and neural networking techniques. Toth (2002) further highlighted the potential of artificial intelligence and artificial neural networks in automatic classification, which were applied by the US Department of Education to classify their electronic documents (Schewe, 2002).

Discourse provides an opportunity to understand and use context in classification. Does contextual information improve the results of automatic classification? This was the hypothesis investigated by Geneva et al. (2003). To test the hypothesis, Geneva and his co-authors used a functional framework for information analysis in an automated environment where the n-grams (filtering) and the k-means and Chen's classification algorithms were tested against sub-collections of documents based on the following discourse variables: 'genre', 'register', 'domain terminology' and 'document structure'. The results showed that: (i) n-grams does not appear to have a clear dependence on discourse variables; (ii) k-means algorithm has dependence on discourse variables, but only on domain terminology and document structure; and (iii) Chen's algorithm has a clear dependence on all of the discourse variables.

A new algorithm of text categorisation using feature projection (TCFP) technique was developed by Ko and Seo (2004). In this technique, training data were represented as the projections of training documents on each feature. This technique is based on a normalised voting method. The authors claimed that this method outperforms k-NN, Rocchio and Naive Bayes. It is a hundred times faster than k-NN in the newsgroups data set, it is robust from noisy data and it is very simple to implement. Park and Zhang (2004) observed that most document classification systems considered only the distribution of content words of the document and ignored the semantic information. To overcome this lacuna, the authors suggested a new approach which incorporated both the lexical and syntactic information of the documents. To demonstrate this approach, the authors used co-training algorithm, a partially supervised learning algorithm and a small amount of labelled data augmented by a large amount of unlabelled data. The results of this algorithm were encouraging.

Cai et al. (2004) studied noise in the knowledge base used for automatic classification and suggested an algorithm to remove multi-category noise based on the statistics of frequency. Binary classification usually uses the one-against-rest method. In this method, if a document belongs to a particular category, it is regarded as a positive example of that category; if not, it is regarded as a negative category. The problem with this method is that only the documents of a positive set are labelled manually, whereas the negative data set is not labelled manually and so it may contain noisy data. To solve this problem Han, Ko and Seo (2007) developed a sliding window technique and revised the EM (expectation maximisation) algorithm. The first technique, i.e. the sliding window technique, extracts potentially noisy documents from the

negative data set and the EM algorithm removes actually noisy documents. The traditional text-based document classifiers mostly perform poorly on web classification. This is because text in web documents is mostly noisy and often does not contain enough information to determine the topic.

To overcome this drawback, Calado et al. (2006) used the hyperlink structure of the web to determine the measure of similarity appropriate for document classification. The authors experimented with five different similarity measures and determined their adequacy for predicting the topic of the web page. Results of the experiments showed that use of link information alone provided 86 per cent precision and, when combined with a traditional text-based classifier, the accuracy increased to 90 per cent.

Paradis and Nie (2007) suggested that 'noisy' documents can be classified by using bigrams and named entities. The proposed approach combines conventional feature selection with a contextual approach to filter out passages around the selected features. The system was originally designed to classify call tender documents; however, the authors claimed that it is useful for other web collections also. Classification of web pages becomes more difficult than pure-text classification due to the noisy information contained in them. Removing noise can improve their classification. Apart from the above-mentioned techniques, the noise could also be removed through summarisation. In an experiment towards this end, Shen, Yang and Chen (2007) developed a web page summarisation algorithm based on the web page layout and evaluated it along with several other state-of-the-art summarisation algorithms on the LookSmart web directory. Augmentation of this technique with NB or SVM improved the classification performance by 5 per cent when compared with pure-text-based classification. The authors further improved the classification results by more than 12 per cent with an ensemble method that combined the different summarisation algorithms.

Gama (2000) presented a method for combining classifiers by means of constructive induction. It consists of two parts: a method for building multivariate trees and a method for combining classifiers. Loosely coupling and tightly coupling classifiers are the two schemas suggested for the combination of algorithms. Experimental tests of the new classifier proved that it is efficient in terms of error rate, learning time and interpretability. Automatic classification is based on a set of rules and each rule in the set is a disjunct. A small disjunct is a rule covering a small number of examples, but its prediction is unreliable and error-prone. To overcome this problem, Carvalho and Freitas (2005) investigated six candidate solutions (algorithms). The algorithms investigated belonged to different machine learning paradigms and their hybrid combinations.

For example: two versions of a decision-tree (DT) induction algorithm; two versions of a hybrid DT/genetic algorithm (GA) method; one GA; and one hybrid DT/instance-based learning (IBL) algorithm. Evaluative experiments of these algorithms indicated the hybrid DT/IBL maximises only predictive accuracy whereas a hybrid DT/GA provides both maximum predictive accuracy and rule-set simplicity. The authors also reported the results of a meta-learning experiment, which produced meta-rules predicting which algorithm will tend to perform best for a given data set.

With the help of a decision tree one can obtain a proper set of rules from a large number of instances. However, it is not efficient in obtaining the relationship between continuous-valued data points. The decision tree also leads to overgeneralisation and/or overspecialisation. The Naive Bayes node helps to solve this problem. In order to reap the benefits of the strengths of both of these algorithms, L. Wang et al. (2006) proposed the Self-adaptive NBTree algorithm. This is a hybrid of Naive Bayes and decision tree algorithms. The Bayes measure is used to construct the decision; the tree handles continuous attributes and automatically finds the most appropriate boundaries for discretisation and the number of intervals.

Like Wang, Xiaoyue and Rujiang (2006) presented a hybrid classifier called Rough-ANN (artificial neural network). This time it is a combination of rough set theory and BP neural network. The three steps in the algorithm are: (i) documents are denoted by vector space model; (ii) feature vector is reduced by using rough sets; and then (iii) the documents are classed by BP neural network. Experiments showed that the Rough-ANN classifier has better precision, stability and fault-tolerance compared with Bayesian classifiers SVM and k-NN. The authors claimed that this classifier performs better for the complex classification problems with many feature vectors.

Yet another hybrid categorisation technique was proposed by Yu, Zheng'ou and Mingchun (2005), who suggested a web text categorisation rule extraction based on CHI value theory, rough set and decision tree. The decision tree is used as it has high efficiency of data analysis and easily abstracts the understandable categorisation rules. However, the decision tree could hardly be applied to thousands of dimensions of features. To overcome this shortcoming, CHI value is proposed for feature selection. For reducing the attributes, a further rough set is used. The authors claimed that by applying this algorithm, an understandable categorisation rule can be easily extracted and better categorisation accuracy could be attained.

The decision tree induction algorithm fails to handle: (i) the situation in which the majority voting makes incorrect decisions (generating two different types of rule for the same data); and (ii) in case of dimensionality reduction by decision tree induction algorithms, the determination of the appropriate attribute at a node where two or more attributes have equal highest information gain. As a solution to these, Appavu and Rajaram (2009) presented a knowledge-based text classification algorithm named ID6NB. This algorithm is used for extending the decision tree induced by Quinlan's non-incremental ID3 algorithm. Results of the new algorithm on a number of real and synthetic databases showed that it performs better than other methods based on the decision tree algorithms.

Janssens et al. (2009) developed a hybrid text/citation-based method for the clustering. Janssens and his co-authors clustered journals from the Web of Science database listed during 2002–06. The authors aimed to validate and if possible to improve an existing journal-based subject-classification scheme. The method included: (i) cross-citation links determined on an item-by-paper procedure for individual papers assigned to the corresponding journal; (ii) text mining for the textual component based on the same principle; and (iii) textual characteristics of individual papers attributed to the journals in which they have been published. In the first step of this classification venture, the authors evaluated and visualised the 22-field subject classification scheme of the Essential Science Indicators. In the second step, the hybrid clustering method was applied to classify about 8,300 journals meeting the selection criteria concerning continuity, size and impact. The study found that the hybrid method is superior to its two components when applied separately. The authors particularly appreciated the role of each component of the hybrid, i.e. the textual component allowed labelling the clusters using cognitive characteristics and the citation component allowed visualising the cross-citation relationship.

Another novel text categorisation technique based on a reduction in the original data information using numerical conversion of a symbolic expression and an onion-layer algorithm was proposed by Chrissikopoulos, Papavlasopoulos and Poulos (2005). Five texts from each of the three different semantic categories were categorised applying this technique, which gave highly accurate results.

A feature (or a word) may be able to represent the characteristic of a document in one application context but may not reflect its nature in another. Considering this fact, Chen and Lu (2006) presented a feature selection method with the consideration of the application context for the documents to be classified. This method is called PBMCD. The results

of testing this new method indicated that this algorithm filters out irrelevant features before the classification process and also increases accuracy in classification. A multi-objective genetic algorithm could be used for developing highly predictive and comprehensible classification rules from a large database. However, the accuracy and comprehensibility of the rules often conflict with each other. This affects optimisation. To solve this problem, Dehuri and Mall (2006) proposed a multi-objective evolutionary algorithm called Improved Niched Pareto Genetic Algorithm (INPGA). Comparative evaluation indicated that this algorithm performed better than Simple Genetic Algorithm (SGA) and Niched Pareto Genetic Algorithm (NPGA). Many feature selection algorithms emphasise only on the reduction of high dimensionality of the feature space. However, in the case of highly redundant features, use of more complex dependence models such as Bayesian becomes essential. To resolve this situation, Lee and Lee (2006) proposed an information gain and divergence-based feature selection method for text categorisation, which does not rely on more complex dependence models. This feature selection method strives to reduce redundancy between features while maintaining information gain in selecting appropriate features for text categorisation. Data mining aims to understand data. Mostly rule-based methods are used for understanding data in data mining. Most of the rule-based classifiers use a small number of rules and a default prediction for making mining results comprehensible. But this reduces the explanatory ability of the rule-based classifier. Li and Jones (2006) proposed multiple and negative target rules to improve the explanatory ability of rule-based classifiers. The result of the experiment showed that the understandability increases without affecting accuracy. Feature selection is an important aspect of automatic classification and various methods are available for feature selection. Tan et al. (2006) studied the comparative efficiency of six commonly used feature selection algorithms, i.e. document frequency, information gain, expected cross entropy, chi (super 2) statistical, the weight of text and mutual information. Amongst these, the mutual information algorithm has lowest performance, which could be improved by removing a single word.

Document keyphrases help to get the summary of a document's contents. Keyphrases have multiple applications in knowledge organisation and management. For example, keyphrases can be used in automatic text summarisation, development of search engines, document clustering, document classification, thesaurus construction, etc. All authors do not assign keyphrases and assigning keyphrases manually is a time-consuming and costly venture. Considering the cost-effectiveness of

a machine for this purpose, Wu et al. (2006) developed a keyphrase identification program (KIP). KIP extracts document keyphrases by using prior positive samples of human-identified keyphrases to assign weights to the candidate keyphrases. Realising the value of keyphrases in classification, clustering, data mining, etc., El-Beltagy and Rafea (2009) also developed an algorithm called KP-Miner for automatic keyphrase extraction. This algorithm is capable of extracting keyphrases from English and Arabic language documents. According to the authors, two advantages of this system are: (i) it does not need to be trained on a particular document set; and (ii) it is configurable as the rules and heuristics adopted by the system are related to the general nature of documents and keyphrases.

Most web documents are classified by their subjects. However, web documents could be classified by their genre (style) also. Considering this fact, Kim, Lee and Lim (2005) suggest a set of features extracted from URL and HTML tags, useful in automatic genre classification of web documents. A number of text classification systems have been developed which focus on the topic of the text, but most of them neglect the style aspect of the text. The style aspect includes information about its author, its purpose, feelings it is meant to evoke, text's sentiment (positive/negative), etc. Argamon et al. (2007) developed a new type of lexical feature for use in stylistics text classification. This system is based on the taxonomies of various semantic functions of certain choice words or phrases. The authors claimed that: (i) this classification system is useful for determining author identity and nationality, the gender of literary characters, text's sentiment and rhetorical character of scientific journal articles; (ii) the functional features help in gaining insight about the stylistic differences among different kinds of texts. Similarly, Fujino, Ueda and Saito (2007) developed a classifier for text data samples consisting of main text and additional components such as titles, links, authors, etc. The classification algorithm developed is based on a hybrid composed of probabilistic generative and discriminative approaches. This classification system is more useful for classification of multiclass and single-labelled text.

Text analysis is one of the primary tasks in document classification and techniques of text mining help in text analysis. Text segmentation, summary extraction, feature selection, term association, cluster generation, topic identification and information mapping, etc., are some text mining techniques. A descriptive and evaluative account of these techniques can be found in the article written by Tseng, Lin and Lin (2007). This article also explains how these techniques are useful in

automatic text classification. Pons-Porrata, Berlanga-Llavori and Ruiz-Schulcloper (2007) also discussed the use of text mining in topic discovery and automatic text classification. The authors proposed a new incremental hierarchical clustering algorithm, which combines both partitional and agglomerative approaches, taking the main benefits from them and a new summarisation method based on testor theory.

In practical text classification tasks, the ability to interpret the classification result is as important as the ability to classify exactly. Associative classifiers are excellent at interpretation as well as having good classification accuracy and rapid training. However, when applied to text classification, the associative classifiers also have their limitations. Some of their limitations are: (i) target text collection generally has a very high dimension, which prolongs the training process; and (ii) the training process of the associative classifiers produces a huge amount of classification rules, which makes the prediction with a new document ineffective. Yoon and Lee (2007) discussed these limitations and to overcome them they proposed: (i) a feature selection based on the mutual information between the word and class variables to reduce the space dimension; and (ii) a new efficient method for storing and pruning classification rules.

Numbers of classification algorithms are developed for classification of actual, complete data sets. However, often the actual data sets are incomplete, for many reasons. To classify incomplete data sets, Chen et al. (2008) presented a selective Bayes Classifier with a simpler formula for computing gain ratio. The experiments of using this algorithm on 12 benchmark incomplete data sets found that this algorithm greatly improved the accuracy of classification and simplified the data sets and classifiers by sharply reducing the number of attributes.

The amount of user-generated content in the form of wikis, blogs, forums, etc., is increasing fast. Cosh, Burns and Daniel (2008) explored the use of content clouds to classify such ad hoc contents. Cosh and his co-authors applied natural language process (NLP) tools to automatically extract contents of some text, visualising the results in a content cloud. The authors observed that the content clouds share the visual simplicity of a tag cloud, but display the details of an article at a different level of abstraction, providing a complimentary classification. According to the authors, further research is needed to refine the process.

Identification of the information sender of web documents constitutes an important part of 'information credibility analysis'. Kato et al. (2008) developed an algorithm to classify information senders of web documents. The algorithm is based on support vector machine (SVM). The title and

URL of the document and information on the top page were the features used for developing this algorithm. The authors found that due to the subjective nature of the categories, some of the information sender categories were found more difficult to classify, so they suggested that there should be further refinement of categories. An ideal goal of automatic text classification is to achieve high-quality text classification by accepting almost all documents that should be accepted (i.e. high recall) and rejecting almost all documents that should be rejected (i.e., high precision). In reality, many a time classifier's erroneous decision incurs high cost and/or serious problems. Consulting users to confirm the classifier's decisions can help avoid potential errors. However, this may incur a heavy cognitive load on users, so controlling the number of confirmations can be the main challenge. Liu (2008) analysed this problem; to resolve it, he developed an intelligent and classifier-independent confirmation strategy called ICCOM. The test of ICCOM showed that it helps various kinds of classifiers to achieve very high precision and recall by conducting fewer confirmations.

Digital image analysis technique is used for automatic classification. Recognition and extraction of object/s of interest from an image are the two important tasks in digital image analysis. Manual image analysis becomes more time-consuming, expensive and error prone. Against this, automated recognition and extraction of visual information brings more consistency, efficiency and accuracy in image analysis. Anouncia and Joseph (2009) studied and discussed the various automated approaches available for recognition and extraction of objects of interest (OOI) from an image in various scientific and engineering applications. The authors suggested that categorisation could be made based on four principal factors: input, object, feature and attention. The authors claimed that all the approaches studied and discussed by them proved to work efficiently in a real environment.

Duwairi et al. (2009) presented a comparative account of three feature reduction techniques applied to Arabic text. These techniques are: (i) stemming, which reduces words to their stems; (ii) light stemming, which removes common affixes from words without reducing them to their stems; and (iii) word clusters, which groups synonymous words into clusters and each cluster is represented by a single word. The authors analysed effects of these techniques on the k-NN classifier to reduce the size of document vectors without affecting the accuracy of the classifiers. Size of document vectors, classification time and accuracy (in terms of precision and recall) were the aspects considered for comparison. Several

experiments were carried out by using four different representations of the same corpus covering sports, economics and political science subjects. The first version used stem vectors, the second used light stem vectors, the third used word clusters and the fourth used the original words (without any transformation) as representatives of documents. The result indicated that in terms of vector size and classification time, the stemmed vectors consumed the smallest size and the least time to classify. In terms of classification accuracy, the light stemmed vectors superseded the other three representations.

Most text classifiers use flat classifiers. Flat classifiers treat each concept as independent, even when the concept space is hierarchically structured. Against this, hierarchical text classifiers use the structural relationship between concepts. Gauch et al. (2009) explored the effectiveness of hierarchical classification for a large concept hierarchy. For testing the efficiency of this algorithm, the authors used: (i) documents selected from sub-concepts to address the sparseness of training data for the top-level classifiers; and (ii) document relationship to identify the most representative training documents. The hierarchical classifiers proved superior for large, three-level concepts in terms of accuracy over the flat classifier.

Humphrey et al. (2009) described and compared two systems of automatic categorisation of documents in biomedical literature, i.e. CISMeF and JDI. The CISMeF is a system that uses rules based on human indexing of documents by the Medical Subject Headings (MeSH) controlled vocabulary in order to assign metaterms (MTs). The JDI – journal descriptor indexing – is a system based on human categorisation of about 4,000 journals and statistical associations between journal descriptors (JDs) and textwords in the documents. The authors evaluated and compared the performance of these systems against the humanly assigned categories for 100 MEDLINE documents and concluded that the results favour JDI, given the significantly greater intellectual overhead involved in human indexing. The authors suggested that it would be useful to investigate whether these two systems could be combined to see if they are complementary to each other.

Classification model, similarity measure and document representation are the three factors involved in text classification. Keikha et al. (2009) claimed that the choice of document representation has a deep impact on the quality of classifier. To substantiate their claim, the authors compared four different types of document representation: N-gram, single term, phrases and RDR, a logic-based document repre sentation. The authors used the centroid-based text classifier, a simple and sound text

classification scheme and concluded that the RDR is more effective than other document representation systems and it is also more complex. Phrase approach was the next best approach.

In many applications, a huge amount of unlabelled data is available at little cost. Lee (2007) analysed the role of unlabelled data in the context of Naive Bayesian learning. An experiment including unlabelled data as a part of training data resulted in significant classification accuracy. Most automatic classification algorithms are tested on a set of manually labelled training documents. However, Hung and Chien (2007) claimed that sometimes the manually labelled documents are unavailable or are of poor quality. So the authors developed a self-learned approach to extract high-quality training documents from the web. The test result showed that it performs satisfactorily. Supervised learning is learning based on a general inductive process, which automatically builds a text classifier. However, for accurate learning it requires a large number of labelled training documents. Labelled documents are generated by human developers and so have difficulty in generation. Against this, unlabelled documents are easily collected and are available in plenty. To take advantage of the unlabelled documents, Ko and Seo (2009) proposed a text classification system based on unsupervised or semi-supervised learning. This system launches text classification tasks with only unlabelled documents and the title word of each category for learning, and then it automatically learns text classifier by using bootstrapping and feature projection techniques. On the basis of the experiments, the authors claimed that this system is faster, economical and performed reasonably better than the supervised system.

A text comprises a number of passages. Sometimes they can be hidden within the text. In such a case their search, retrieval and transfer becomes difficult. Passage retrieval is the search for passages relevant to a user's query. Passage detection is classification of passages. For detection purposes, the passages are labelled with one or more categories from a set of predetermined categories. Mengle and Goharian (2009b) presented a keyboard-based dynamic passage (KDP) approach for passage detection. Effects of feature selection, passage length, ambiguous passages and training-data category distribution on passage detection were evaluated and found favourable. According to Mengle and Goharian (2009a), one of the major problems in text classification is the high dimensionality of feature space. The authors presented the ambiguity measure (AM) feature-selection algorithm, which selects the most unambiguous features from the feature set. Unambiguous features are those features whose presence in a document indicates a strong degree of confidence that a

document belongs to only one specific category. Application of the AM feature selection on a Naive Bayes text classifier, using five datasets, outperformed eight existing feature selection methods. Use of this feature selection method on SVM classifier reduced the training time by more than 50 per cent. The Hidden Markov Model (HMM) is used in a wide range of applications in information processing. However, Yi and Beheshti (2009) claimed that it is not used in text classification. In order to test its application in text classification, the authors developed a prototype for HMM-based text classifier and implemented it to categorise medical documents into MeSH and concluded that its performance was comparable to other text classification methods.

Use of k-NN in text categorisation

k-NN (k-Nearest Neighbours) is one of the algorithms used for text categorisation. It classifies objects based on closest training examples. It is an instance-based learning algorithm. It classifies by majority vote of its neighbours with the object being assigned to the class most common amongst its k nearest neighbours. In the k-NN, k is a positive integer. A large amount of literature does exist on the k-NN and other algorithms. Reviewing them is beyond the scope of this book, which reviews those articles that have been listed in the LISA. Thus, only those articles that discuss the application of k-NN in automatic text categorisation are reviewed.

Kwon and Lee (2003) suggested a classification algorithm for websites based on the k-NN approach. The algorithm first chooses several representative web pages using connectivity analysis. Next, the k-NN classifier classifies each of the selected web pages. Then, the classified web pages are extended to classify the entire website. Finally, a feature selection method is used to improve the performance of k-NN and markup tags are used to improve the term weighting scheme. Experiments on a Korean commercial web directory showed that this system worked efficiently compared with a system that uses only the home page.

For improving the performance of k-NN, Wang and Yang (2004) developed an interactive k-NN based on k-NN and automatic retrieval. The algorithm works on a small-scale database. An important character of this system is that if enough similar documents are not found, part of the document subjects are magnified to find more similar documents by Internet retrieval. Document text as well as its metadata is useful in machine learning and in automated classification.

However, Macfarlane and Richter (2005) observed that many classifiers use document text and ignore metadata (such as authors, publication date and author affiliation). The authors experimented with classifying patents using k-NN algorithm with two classification systems. One used metadata such as name of the inventor, applicant's name, International Patent Classification code, etc., and another ignored it. The result of use of metadata was improved accuracy; however, quality deteriorated, so the authors suggested it must not be used indiscriminately.

Organisation of text documents on the Internet in manageable and easy to understand categories is one of the ways to control these documents. For this purpose, Tan (2005) suggested the binary k-NN (BKNN) classifier for automatic text categorisation. This algorithm is superior to the traditional k-NN as far as speed and accuracy are concerned. k-NN is a simple, yet effective classifier. However, in case of ambiguity, i.e. having more than one candidate category to which a document can be assigned, its precision decreases.

To overcome this drawback, Bang, Yang and Yang (2006) proposed a new algorithm that incorporates relationship of concept-based thesauri into the document categorisation using the k-NN classifier. This new system entails structuring categories into hierarchies. The k-NN uses various relationships in thesaurus corresponding to the structured categories. Use of concept-based thesauri along with k-NN provided improved performance.

Laguia and Castro (2008) introduced a method known as local distance-based classification for text categorisation. In this method, every training point learns what is happening in its neighbourhood. For this purpose, a hyperlane is learned and associated to each point. The hyperlane can be used to define the bands distance, a distance measure that brings points closer or moves them away depending on their classes. Tests of this algorithm on 18 well-known repository datasets, one private dataset and 49 ad hoc synthetic datasets proved that this system is accurate. More particularly, the bands distance has obtained the best overall results with 1-NN and k-NN classifiers when compared with other distances.

Many text classifiers have been developed in the recent past and amongst them k-NN is the most widely used. However, this algorithm suffers from the presumption that the training data are evenly distributed among all the categories and it is sensitive to the parameter k. To overcome this drawback, Xiaopeng and Feicheng (2008) suggested an unsupervised strategy (UKNN) for the k-NN classifier, which is not

sensitive to the parameter k and performs well on imbalanced corpora. This strategy adopts sum-of-squared-error criterion to adaptively select the contributing part from these neighbours and classifies the input document in terms of the disturbance degree, which it brings to the kernel densities of these selected neighbours.

Use of SVM in text categorisation

SVMs (support vector machines) are a set of related supervised learning methods used for classification. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. A large amount of literature exists on the SVM and other algorithms. Reviewing them is beyond the scope of this book. It only reviews those articles which have been listed in the LISA, thereby to indicate that only those articles are reviewed which are discussing the application of SVM in automatic text categorisation.

Choosing a kernel and specific parameters for that kernel is a problem in using the support vector machine (SVM) algorithm. To overcome this problem, Howley and Madden (2005) proposed the genetic kernel SVM (GK SVM). This technique uses genetic programming to evolve a kernel for an SVM classifier. The article provides a descriptive and evaluative account of this new technique. A dendogram-based support vector machine (DSVM) algorithm for treating the multi-class problem in automatic text classification was proposed by Benabdeslem and Bennani (2006). Steps in this approach are: (i) building a taxonomy of classes based on ascendant hierarchical clustering method; (ii) separating the two subsets of the current node, SVM is injected at each internal node of the taxonomy; (iii) presenting the pattern query to the 'root' SVM for classification. As per the nature of the output, the pattern is presented to one of the two SVMs of the subsets, and so on through the 'leaf' nodes. Thus, the classification procedure works in a descendant way in the taxonomy from the root through the end level, which represents the class. Wang and Chiang (2007) presented the OAA-FSVM text categorisation system to solve the multi-class categorisation problem. It consists of two modules: (i) the processing module uses ICF and Uni, which extract relevant terms; and (ii) the classifying module, i.e. OAA-SVM, uses the fuzzy set theory. Evaluation of the system indicated that the OAA-FSVM handled the multi-class text categorisation in a better way than OAA-SVM.

Does morphological complexity affect text classification and, if it does, how and to what extent? This was the topic of enquiry of Melenica et al. (2008). For this purpose, the authors chose Croatian and English languages, as they had radically different morphological complexity. For the experiment purpose, the authors obtained a large-scale dataset of different feature subset sizes by using different feature selection methods. The authors then carried out parallel experiments on both the languages using SVM classifier and different levels of morphological normalisation to quantify, compare and test the statistical effect of morphological normalisation on the SVM classifier. Based on the experiments, the authors concluded that the improvements in SVM classifier performance is statistically significant; improvements are greater for a small and medium number of features, especially for Croatian language, whereas for a large number of features the improvements are rather small and may be negligible in practice for both languages.

Text representation is to look for appropriate terms to transfer documents into numerical vectors and is an important task in text mining. In the recent past, many efforts have been made to enrich text representation using SVM to improve performance of text classification. Zhang, Yoshida and Tang (2008) investigated the effectiveness of using multi-words for text representation on the performance of text classification. The authors adopted the following process to obtain an answer to their query: (i) multi-words were extracted from documents based on the syntactical structure; (ii) using the extracted multi-words, documents were represented as general concept representation and sub-topic representation; (iii) the dynamic k-mismatch was used to determine the presence of a long multi-word which is a sub-topic of the content of a document; (iv) using the representation of multi-words, a series of experiments were carried out on classifying the Reuters-21578 documents (Reuters-21578 is the most widely used test collection for text categorisation research) – performance of representation of individual words was used as the baseline; (v) linear kernel and non-linear polynomial kernel in SVM were compared to investigate the effect of kernel type on their performances; (vi) index terms with low information gain (IG) were removed from the feature set at different percentages to observe the robustness of each classification method. Through the experiment, the authors found that: (a) in multi-word representation, sub-topic representation outperforms the general concept representation and the linear kernel outperforms the non-linear kernel of SVM in classifying the Reuters data; (b) the effect of applying different

representation strategies is greater than the effect of applying the different SVM kernels on classification performance; (c) the representation using individual words outperforms any representation using multi-words. The results obtained were consistent with the major opinions concerning the role of linguistic pre-processing on documents' features when using SVM for text classification.

Mehler and Waltinger (2009) developed a topic classification model using the DDC as the target scheme. For this the authors: (i) obtained the document snippets, as minimum document representation using the metadata provided by the Open Access Initiative (OAI); (ii) selected and extended features by using the social ontologies and related web-based lexical resources – the ontologies and other web-based lexical resources were used to provide reliable topic-related classification while circumventing the problem of data sparseness; (iii) evaluated the model by means of two language-specific corpora. The model is helpful to apply computational linguistics to thematic classification in digital libraries. Further investigation is essential to develop SVM-based Dewey Decimal classifiers by using larger training datasets, possibly of more than two languages. Zhang (2008) evaluated a number of learning algorithms on a standard dataset and concluded that a supervised SVM classifier using various lexical and syntactic features can achieve competitive results.

Hierarchical classification system for text categorisation

Yoon, Lee and Lee (2006) suggested a method for developing a hierarchical classification system for text categorisation. The system is based on internal node classifiers. It is useful to build a hierarchical classification system that performs well with large collections of practical data. A hierarchical text classification aims to classify each incoming document into zero, one or several categories in the text hierarchy. Liu (2009) presented the CRHTC (context recognition for hierarchical text classification) technique of hierarchical text classification. It classifies text by recognising the context of discussion (COD) of each category. The COD of a category is governed by its ancestor categories, whose contents indicate contextual backgrounds of the category. A document is classified into a category if its content matches the category's COD. The author claimed that the CRHTC is portable and easier to implement than other methods (as it does not require any trials to manually set parameters) and

its performance is better and more stable than several hierarchical and non-hierarchical text-classification methodologies.

Different performance measures are used to evaluate the performance of hierarchical classification experiments. Lim, Ng and Sun (2003) observed that most performance measures often assume independence between categories and do not consider documents misclassified into categories that are similar or not far from the correct categories in the category tree. Therefore, the authors proposed a new performance measure that consists of category similarity measures and distance-based measures that consider misclassified documents. A comparative experiment on SVM and Binary Naive Bayes classifiers showed that SVM classifiers perform better than Naive Bayes on the Reuters-21578 collection according to the extended measures. Lim and his co-authors also define a classifier-centric measure called blocking measure to examine sub-tree classifiers in a top–down level-based hierarchical classification method. Many a time the performance of various classification algorithms are compared by using real datasets. However, Jamain and Hand (2008) believed that such comparisons often yield inconclusive or limited results and a broader approach combining these studies is necessary to have trustworthy results. Believing that the current state of literature hardly allows large-scale investigations, the authors created a dataset of 5,807 classification results. Based on this data, the authors suggested methods for assessment of the classification algorithms along with their merits and demerits.

Uses of text categorisation

Spam e-mails are one of the disturbing elements of e-mail facility. With the growth of the Internet, the number of spam e-mails is also increasing. Proper classification of spam can be of great help and various algorithms with varying performance are available for spam classification. Yu and Xu (2008) evaluated the performance of four machine-learning algorithms for classification of spam e-mails. The four algorithms evaluated by the authors were: Naive Bayesian (NB), neural network (NN), support vector machine (SVM) and relevance vector machine (RVM). The purpose was to evaluate the relative efficiency of each of these algorithms. Efficiency of these algorithms was tested on different training set size and extracted feature size. The study revealed that: (i) NN classifier was unsuitable for using alone as a spam rejection tool; (ii) SVM and RVM proved superior to NB; (iii) SVM and RVM showed similar results with less relevance

vectors and much faster testing time; (iv) RVM was slower in learning but proved more suitable than SVM in terms of the applications that required low complexity.

The aviation industry consolidates large amounts of data. This data may be related to various activities, such as data reported by pilots, data about maintenance, data about accidents, data about delays, etc. Gurbuz, Ozbaki and Yapici (2009) conducted a study to know the effective attributes so as to reduce the number of fatalities in incidents. The experiment included: (i) application of data mining technique to the accidents/incidents data of the US Federal Aviation Administration; (ii) use of rough sets concepts to reduce the attributes of datasets; (iii) various categorisation tools; and (iv) decision tree. The resultant categorisation was useful in achieving the expected results.

The Internet is increasingly used by all sections of society as an expert system. The Internet's use for obtaining medical information through websites such as 'ask the doctor' is continuously growing. Himmel, Reincke and Michelmann (2009) developed a system to automatically classify the lay requests put to the web-based medical expert forums. According to the authors, the generation of indicator variables based on the chi-square analysis and Cramer's V proved to be the best approach for automatic classification of the requests.

The amount of course material, including syllabi, made available on the Internet is increasing day by day. In order to make available such material at a single web location and to increase its accessibility and availability, Ireland established a national repository of course syllabi. Joorabchi and Mahdi (2009) reported the development of this repository, including the development of a fully unsupervised document classification system that automatically collected training documents with the help of a search engine. This classification system classified syllabus documents based on an extended version of the International Standard Classification of Education. Apart from getting information about this innovative repository and its text classification system, this article is also useful for knowing the importance of classifying resources in a syllabus digital library.

The above literature dealing with automatic classification indicates the following trends: (i) the amount of research on automatic classification is growing and so is its depth; (ii) text, images, type of websites and even the nature of information such as spams, harmful e-mails, etc., are the diverse characteristics considered for classification; (iii) varieties of algorithms are developed; (iv) accuracy, speed and cost are some of the major criteria considered; (v) k-NN and SVM are the most used classifiers;

(vi) efforts to test hybrid systems are increasing. Chronological analysis of publications on automatic classification indicates that the literature on the topic is increasing, which means there is more research on the topic and more benefits of this research.

10

Classification: theories, research trends and personalities

Abstract: This chapter reviews literature published on classification theories, conferences and personalities that have contributed to the development of classification. This chapter also reviews literature that depicts trends in classification research.

Key words: classification theories; classification research trends; conferences on classification; classification personalities; George W. Bordner; Henry Evelyn Bliss; Ingetraut Dahlberg; Melvil Dewey; Julia Pettee; Ranganathan.

Classification theories

Theories/principles are brief, precise research-based statements that explain patterns of phenomena. Ranganathan, Sayers, Hulmes and others have formulated theories of classification. Mai (2004a) has discussed the various theories of classification and Marcella (2002) examined areas where classification theory usefully contributes. The concept of categories had been successfully applied in bibliographic classification schemes for grouping isolates.

Aranalde (2009) discussed the theories of categories formulated by Aristotle, Kant and Ranganathan and concluded that these three thinkers thought about the categories differently. Aristotle developed categories for expressing the ways of being, while Kant thought about categories as ways of thinking. On the other hand, Ranganathan developed categories as ways of classifying.

Textual analysis is a basic step in classification and various approaches/theories can be applied for textual analysis. These include grounded theory, content analysis, discourse analysis, conversation analysis, etc.

All of these approaches form part of classification theory. Dilevko and Gottlieb (2009) demonstrate the applicability of classification theory to various text analysis approaches. The applicability is based on three factors: (i) extant and elicited texts can be broken down into categories; (ii) extant texts are themselves explicit or implicit classification systems; and (iii) classificatory frameworks can be applied to extant and elicited texts 'in order to clarify their contribution to processes of meaning-making'. Considering the relevance of classification theory to textual analysis, Dilevko and Gottlieb recommended that classification theory should be incorporated in the teaching of textual analytic approaches in university-level research methods courses, especially in the field of library and information science. Hjorland (2008a) outlined main points in a core theory of classification and stressed that theoretical classification is important for the future development of the field.

Conferences related to classification

The classification literature published during the period covered by this book has the following four references to conferences and seminars related to classification.

The National Seminar on Classification in the Digital Environment, organised by the Sarada Ranganathan Endowment for Library Science, was held in August 2001 in Bangalore, India. It discussed the use of facet analysis and chain indexing in the digital environment for organising and searching web-based material as well as their use in formulating queries for searching reading material from digital libraries (Prasad, 2001).

Deliberations at the 25th Annual Conference of Cataloguing, Classification and Indexing of the Nigerian Library Association held in October 2005 were reported by Onohwakpor (2006). The conference concluded that the increased role of technology in the cataloguing, classification and indexing functions of technical services remains challenging in many libraries. Considering this fact, the conference resolved that special and continuing staff training is necessary to promote new procedures and efficiencies. This article is of particular value to library and information professionals interested in African librarianship.

Williamson (2009) reported on the three conferences held in 2007 which discussed classification issues. These conferences and their themes were: (i) at the IFLA Conference in Durban, South Africa, one of the papers discussed the 'Library of Congress controlled vocabularies, the

Virtual International Authority File, and their application to the semantic web'; (ii) the UDC Consortium organised a two-day seminar on Universal Decimal Classification at The Hague, Netherlands, on 4–5 June 2007, focusing on the theme 'Information Access for the Global Community'; (iii) the inaugural conference of the ISKO North American Chapter took place at the University of Toronto in Ontario on 14–15 June 2007, and included a symposium on 'Knowledge Organization Research in North America: What have we done, what are we doing, and where do we go from here?'

In 2008, the Cataloguing and Indexing Group of the Chartered Institute of Library and Information Professionals (CILIP) organised a conference with the theme 'Classification and Subject Retrieval in the 21st Century'. The conference examined a variety of issues, including traditional classification schemes, classifying and ordering information, social tagging, semantic web, and so on (Taylor, 2009).

Research trends in classification

Progress in any discipline or subject depends on research. Similarly, research in classification has also helped the subject to develop and grow in multiple aspects. It could be said that classification research began informally from its very origin. Formal research in classification, however, got its impetus with the establishment of the CRG (Classification Research Group) in 1952 and the FID/CR. McIlwaine (2003) traced the origin of CRG and discussed its contributions in the development of facet analysis, indexing principles and systems.

So what topics/aspects are researched in classification? A bibliometric analysis of the *Knowledge Organization* journal, covering the period from 1988 to 2000, revealed that classification is one of the major topics covered by this international journal. The other topics included are knowledge organisation, indexing, thesauri, vocabulary control, conceptual analysis, categorisation, etc. (Rekha and Parameswaran, 2002). Hjorland and Albrechtsen (1999), in their article entitled 'An analysis of some trends in classification research', observed: (i) academic disciplines as the main structural principle; (ii) the fiction/non-fiction distinction; and (iii) the appropriate unit of analysis in online retrieval systems are the three prevailing themes in knowledge organisation *research*. Hjorland and Albrechtsen also suggested that a change in classification research is needed, founded on a more historical and social understanding of knowledge.

So, have the trends in classification research changed? Answering affirmatively, Mai (2004a) asserted that that the focus of classification research has shifted from 'establishing a scientific foundation for classification research' to 'using contextual information as the guide for the design and construction of classification schemes'. Richardson et al. (2000) have also set forth a millennium agenda for classification research. The authors posed a set of questions to researchers and practitioners in the field of cataloguing and classification and asked them to address what they consider the key questions in the field.

Classification personalities

Some library and information professionals have a special interest in classification. These professionals have accordingly contributed extensively to the field of classification in particular and knowledge organisation in general. The following section reviews literature that deals with the life and contributions of the great personalities of classification. The review in this section is arranged alphabetically by surname.

George W. Bordner compiled a classification scheme for Pennsylvania State Publications during his tenure as head of technical services from 1961–1980. The classification scheme was published by the State Library in 1975. What made Bordner compile this classification scheme, as well as his other contributions, is reported by Spila (2008).

Henry Evelyn Bliss, the developer of Bibliographic Classification, contributed immensely to the theory and practice of bibliographic classification. Broughton (2008) presents a retrospective review of the work of Bliss as a classificationist and classification theorist. He examined the major features of Bliss's writings and philosophy and evaluated them for the originality of their contribution to the corpus of knowledge in the discipline, analysed reactions to Bliss's work, compared him with Ranganathan and assessed the influence of Bliss on the classification theory of the twentieth century. The author opines that Bliss has not been given the credit he deserves.

Ingetraut Dahlberg is one of the authorities on knowledge organisation domain. Dahlberg is the founder of the International Society for Knowledge Organization (ISKO) and the *Knowledge Organization* journal (formerly published as *International Classification*). Issue 2–3, Volume 35 of *Knowledge Organization* presents an interview with Ingetraut Dahlberg (2008). The interview provides details of her educational background, how she got interested in the field of knowledge

organisation, her collaboration with colleagues outside Germany, what prompted her to start the journal *International Classification*, what made her shift from classification to knowledge organisation, how knowledge organisation can gain widespread acceptance as a scientific discipline, etc. All in all, this is a useful document to learn about Dahlberg, an influential personality in the field of knowledge organisation.

Melvil Dewey, as the library and information world knows, was a multifaceted personality. An account of the life and ideas of Melvil Dewey is presented by Prescott (2001).

Julia Pettee was a pioneer in the field of cataloguing codes, classification and subject theory. She devised the Union Theological Seminary classification system. She recast the universe of knowledge from the point of view of the theologian for a classification system adopted by more than 50 libraries; she was a strong advocate of specialised tools for special collections. Her writings encouraged other librarians to consider whether the large-scale systems then in the process of development (Cutter Expansive, Dewey, Bliss and the Library of Congress system) really suited the needs of their patron communities. These and other valuable contributions of Julia Pettee are discussed by Walker and Copeland (2009). They examined her important writings and speeches, identified the principles of cataloguing and classification she championed, and surveyed contemporary standards to trace her continuing influence in current practice and theory.

Kabir (2003) praised Ranganathan as 'Universal Librarian' for his world-famous contributions in the form of Colon Classification. These classification personalities deserve much more appreciation and remembrance with many more articles should be written about them and their contributions to the development of classification.

Appendix
Books on classification published during 1999–2009

Andrew, P.G. and Larsgaard, M.L. (1999). *Maps and related cartographic materials cataloging, classification, and bibliographic control*. New York: Haworth Information Press.

Batley, S. (2005). *Classification in theory and practice*. Oxford: Chandos Press.

Beall, J. and Couture-Lafleur, R. (eds). *Dewey Decimal Classification: Francophone perspectives*. Dublin, OH: OCLC, Forest Press.

Beghtol, C. and Williamson, N.J. (2004). *Knowledge organization and classification in international information retrieval*. New York: Haworth Information Press.

Bock, H.H., Jajuga, K. and Sokolowski, A. (2002). *Classification, clustering, and data analysis: recent advances and applications*. Berlin-Heidelberg: Springer-Verlag.

Bowker, G.C. and Star, S.L. (1999). *Sorting things out: classification and its consequences*. Cambridge, MA: MIT Press.

British Standards Institute (1999). *Universal decimal classification – pocket edition*. London: British Standards Institute.

Broughton, V. (2004). *Essential classification*. London: Facet Publishing.

Chan, L.M. (2007). *Cataloging and classification: an introduction*. 3rd edition. Lanham, MD: Scarecrow Press.

Decker, R. and Gaul, W. (2000). *Classification and information processing at the turn of the Millennium*. Berlin: Springer-Verlag.

Ellis-Barrett, L. (2004). *Essential Dewey*. London: Facet Publishing.

Gnoli, C. (2008). *Axiomathes* (Vol. 18, No. 2): Special Issue on Facet Analysis. Guest editor. Netherlands: Springer.

Hill, J.S. (2002). *Education for cataloguing and the organisation of information: pitfalls and the pendulum*. 4th edition. New York: Haworth Information Press.

Hunter, E.J. (2002). *Classification made simple*. 2nd edition. Aldershot: Ashgate.

Hussain, S. (2004). *Dewey Decimal Classification: A complete survey of twenty-two editions*. Delhi: B.R. Publishing.

Joachim, M.D. (2003). *Historical aspects of cataloging and classification*. Binghampton, NY: Haworth Information Press.

Kao, M.L. (2001). *Cataloguing and classification for library technicians*. Binghampton, NY: Haworth Information Press.

Marcella, R. and Maltby, A. (eds) (2000). *The future of classification*. Aldershot: Gower.

Mitchell, J.S. (ed.) (2003). *Dewey Decimal Classification and relative index*. 4 volumes. 21st edition. Dublin, OH: OCLC.

Mitchell, J.S. and Vizine-Goetz, D. (eds) (2006). *Moving beyond the presentation layer: content and context in the Dewey Decimal Classification (DDC) System*. New York: Haworth Information Press.

Moys, E.A. et al. (2001). *Moys classification and thesaurus for legal materials*. Munich: Saur.

Neelameghan, A. and Prasad, K.N. (2001). *National seminar on classification in the digital environment*. Papers contributed to the national seminar on classification in the digital environment, Bangalore, 9–11 August 2001. Bangalore: Sarada Ranganathan Endowment for Library Science.

Raghavan, K.S. and Prasad, K.N. (eds) (2006). *Knowledge organization, information systems and other essays: Professor A. Neelameghan Festschrift*. New Delhi, India: Ess Ess Publications for the Ranganathan Centre for Information Studies.

Roe, S.K. and Thomas, A.R. (2004) *The thesaurus: review, renaissance, and revision*. Binghampton, NY: Haworth Information Press.

Satija, M.P. (2007). *The theory and practice of the Dewey Decimal Classification System*. Oxford: Chandos Publishing.

Satija, M.P. (2002). *Manual of practical colon classification*. 4th revised edition. New Delhi, India: Concept Book Publishing.

Scott, Mona L. (2005). *Dewey Decimal Classification: A study manual and number building guide*. 2nd edition. Westport, CT: Libraries Unlimited.

Sen, B.K. (2008). *DDC readymade: a treasury to 15,000 readymade DDC class numbers relating to Indian subjects, English language and literature*. New Delhi, India: Ess Ess Publications.

Smiraglia, R.P. (2008). *Shelflisting music: guidelines for use with the Library of Congress Classification: M*. 2nd edition. Lanham, MD: Scarecrow and the Music Library Association.

Stone, A.T. (2000). *The LCSH century: one hundred years with the Library of Congress subject headings system*. New York: Haworth Information Press.

Taylor, A.G. (2006). *Introduction to cataloging and classification*. 10th edition. Westport, CT: Libraries Unlimited.

Taylor, A.G. (2000). *Wynar's introduction to cataloging and classification*. 9th edition. Engelwood, Colorado: Libraries Unlimited.

Wellisch, H.J. (2001). *Glossary of terminology in abstracting, classification, indexing and thesaurus construction*. Medford, NJ: American Society of Indexers.

References

Abbas, J. (2007). In the margins: reflections on scribbles, knowledge organization, and access. *Knowledge Organization*, 34(2), 72–77.

Abdel Hady, M.F. and Shaker, A.K. (2005). Cataloging and classification education in Egypt: stressing the fundamentals while approaching toward automated applications. *Cataloging & Classification Quarterly*, 41(3/4).

Adams, S. (2001). Comparing the IPC and the US classification systems for the patent searcher. *World Patent Information*, 23(1), 15–23.

Adams, S. (2000). Using the International Patent Classification in an online environment. *World Patent Information*, 22(4), 291–300.

Adhikary, M.M. and Nandi, A. (2004). Ranganathan's promulgation of Verbal plane: a linguistic approach. *SRELS Journal of Information Management*, 41(3), 275–281.

Ahrberg, J.H., Seikel, M. and Shorten, J. (2005). Why do you still use Dewey? Academic libraries that continue with Dewey decimal classification. *Library Resources & Technical Services,* 49(2), 123–136.

Aitchison, J. and Clarke, S.D. (2004). The thesaurus: a historical viewpoint, with a look to the future. *Cataloging and Classification Quarterly*, 37(3/4), 5–21.

Albertsen, K. and Van Nuys, C. (2005). Paradigma: FRBR and digital documents *Cataloging & Classification Quarterly,* 39(3/4), 125–149.

Albrechtsen, H. and Pejtersen, A.M. (2003). Cognitive work analysis and work centered design of classification schemes. *Knowledge Organization*, 30(3/4), 213–227.

Anderson, J.D. and Hofmann, M.A. (2006). A fully faceted syntax for Library of Congress Subject Headings. *Cataloging & Classification Quarterly*, 43(1).

Angus, E., Thelwall, M. and Stuart, D. (2008). General patterns of tag usage among university groups in Flickr. *Online Information Review*, 32(1), 89–101.

Anouncia, S.M. and Joseph, J.G. (2009). Approaches for Automated Object Recognition and Extraction from Images – A Study. *CIT – Journal of Computing and Information Technology*, 17(4), 359–370.

Antoshkova, O.A. (2008). Overview of UDC Print Editions. *Nauchno-tekhnicheskaya informatsiya*, 8, 1–4.

Appavu, S. and Rajaram, R. (2009). Knowledge-based system for text classification using ID6NB algorithm. *Knowledge-based systems*, 22(1), 1–7.

Appavu, S. et al. (2009). Data mining based intelligent analysis of threatening E-Mail. *Knowledge-Based Systems*, 22(5), 392–393.

Araghi, G.F. (2004a). A dynamic look toward classification and retrieval. *Cataloging and Classification Quarterly*, 38(1), 43–53.

Araghi, G.F. (2004b). A new scheme for library classification. *Cataloging & Classification Quarterly*, 38(2), 75–99.

Aranalde, M.M. (2009) Reflections on the categorical systems of Aristotle, Kant and Ranganathan. *Ciencia da Informacao*, 38(1), 86–108.

Ardito, S.C. (2003). OCLC v. The Library Hotel. *Information Today*, 20(11), 17–20.

Arellano, F.F.M. and Garrido, O.A.Y. (2000). Classification systems used in Latin American libraries. *Cataloging and Classification Quarterly*, 30(1), 123–135.

Argamon, S. et al. (2007). Stylistic text classification using functional lexical features. *Journal of the American Society for Information Science and Technology*, vol. 58, no. 6, pp. 802–822.

Asundi, A.Y. (2001). Reorienting concept classification to electronic information organisation: an exploratory study. *International Information Communication and Education*, 20(2), 179–185.

Attar, K.E. (2002). The practice of Bliss. *Cataloging and Classification Quarterly*, 34(4), 47–65.

Attar, K.E. (2000). The application of the Bliss Bibliographic Classification in Cambridge College libraries. *New Review of Academic Librarianship*, 6, 35–49.

Babik, W. (1999). Knowledge representation in map collections for information retrieval systems. *Liber Quarterly: the Journal of European Research Libraries*, 9(2), 172–179.

Bang, S.L., Yang, J.D. and Yang, H.J. (2006). Hierarchical document categorization with k-NN and concept-based thesauri. *Information Processing & Management,* 42(2), 387–406.

Basu, A., Panigrahi, P. and Prasad, A.R.D. (2003). NLP based automatic classification system for analytico synthetic scheme. *SRELS Journal of Information Management*, 40(4), 289–312.

Beall, J. (2009). Racially Mixed People, DDC Table 5 Ethnic and National Groups, and MARC 21 Bibliographic Format Field 083. *Cataloging & Classification Quarterly*, 47(7), 657–670.

Beghtol, C. (2008). Professional values and ethics in knowledge organization and cataloging. *Journal of Information Ethics*, 17(1), 12–19.

Beghtol, C. (2004). Exploring new approaches to the organization of knowledge: the subject classification of James Duff Brown. *Library Trends*, 52(4), 702–718.

Beghtol, C. (2003). Classification for information retrieval and classification for knowledge discovery: relationships between 'professional' and 'naive' classifications. *Knowledge Organization*, 30(2), 64–73.

Benabdeslem, K. and Bennani, Y. (2006). Dendogram-based SVM for multi-class classification. *CIT – Journal of Computing and Information Technology*, 14(4), 283–289.

Bertolucci, K. (2003). Happiness is taxonomy: four structures for Snoopy. *Information Outlook*, 7(3), 36–44.

Bhojaraju, G. and Urs, S.R. (2006). Knowledge organisation systems in digital libraries: a comparative study. *Information Studies*, 12(4), 229–257.

Bianco, C.E. (2009). Medical librarians' uses and perceptions of social tagging. *Journal of the Medical Library Association (JMLA),* 97(2), 136–139.

Bland, R.N. and Stoffan, M.A. (2008). Returning classification to the catalog. *Information Technology and Libraries*, 27(3), 55–60.

Boettcher, J. (1999). Challenges and opportunities presented by NAICS. *Journal of Business and Finance Librarianship*, 5(2), 3–13.

Bowman, J.H. (2005a). Classification in British public libraries: a historical perspective. *Library History*, 21(3), 143–173.

Bowman, J.H. (2005b). Education and training for cataloguing and classification in the British Isles. *Cataloging & Classification Quarterly*, vol. 41(3/4), 309–333.

Bradley, C.J. (2003). Classifying and cataloguing music in American libraries: a historical overview. *Cataloging and Classification Quarterly*, 35(3/4), 467–481.

Brenner, W., Kolbe, L.M. and Kremer, S. (2005). Towards a procedure model in terminology management. *Journal of Documentation*, 61(2), 281–295.

Brett, R. (2008). Classification practice in law libraries: a brief survey. *Legal Information Management*, 8(1), 61–63.

Broughton, V. (2008). Henry Evelyn Bliss – the other immortal, or a prophet without honour? *Journal of Librarianship and Information Science*, 40(1), 45–58.

Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings*, 58(1), 49–72.

Broughton, V. (2002). Organizing a national humanities portal; a model for the classification and subject management of digital resources. *Information Research Watch International*, 2–4.

Broughton, V. (2001a). Current research on the Bliss Bibliographic Classification 2nd edition. *Information Research Watch International,* 2–3.

Broughton, V. (2001b). Faceted classification as a basis for knowledge organization in a digital environment; the Bliss Bibliographic Classification and the creation of multi-dimensional knowledge structures, *New Review of Hypermedia and Multimedia*, 7, 67–102.

Broughton, V. (2000). A new classification for the literature of religion. *International Cataloguing and Bibliographic Control*, 29(4), 59–61.

Broughton, V. (1999). Notational expressivity; the case for and against the representation of internal subject structure in notational coding. *Knowledge Organization,* 26(3), 140–148.

Broughton, V. and Lane, L. (2000). Classification schemes revisited: applications to Web indexing and searching. *Journal of Internet Cataloging*, 2(3/4), 143–155.

Broughton, V. and Slavic, A. (2007). Building a faceted classification for the humanities: principles and procedures. *Journal of Documentation*, 63(5), 727–754.

Bruce, R. (2008). Descriptor and Folksonomy Concurrence in Education Related Scholarly Research. *Webology*, 5(3).

Brunt, R. (2007). Information storage and retrieval in the professional curriculum: spring, 1980. *Library Review*, 56(7), 552–556.

Bruun, A (1999). Development of the IPC as a search tool. *World Patent Information*, 21(2), 97–100.

Buchel, O. and Coleman, A. (2003). How can classificatory structures be used to improve science education? *Library Resources and Technical Services*, 47(1), 4–15.

Cai, W. et al. (2004). An algorithm to remove multi-category noise based on the statistics of frequency. *Journal of the China Society for Scientific and Technical Information*, 23(4), 399–403.

Calado, P. et al. (2006). Link-based similarity measures for the classification of web documents. *Journal of the American Society for Information Science and Technology,* 57(2), 208–221.

Calvert, J. and Makarov, M. (2001). The reform of the IPC. *World Patent Information*, 23(2), 133–136.

Cameron, G. (2005). Revising government Web content structure: what is IPSV? *Information Management and Technology,* 38(4), 162–164.

Carstens, T.B. and Vihnanek, E.M. (2000). Classification of a small collection of music compact discs: a practical approach for the non-musician. *Technical Services Quarterly*, 18(1), 25–32.

Carvalho, D.R. and Freitas, A.A. (2005). Evaluating six candidate solutions for the small-disjunct problem and choosing the best solution via meta-learning. *Artificial Intelligence Review,* 24(1), 61–98.

Castillo, L. and de la Cueva, A. (2007). Evolution and use of controlled languages in news documentation. *Profesional de la Informacion*, 16(6), 617–626.

Caton, B.Q., Hawkins, D.T. and Larson, S.E. (2003). Information science abstracts: tracking the literature of information science. II. A new taxonomy for information science. *Journal of the American Society for Information Science and Technology*, 54(8), 771–781.

Chan, L.M. (2004a). A guide to the Library of Congress Classification. 5th ed. Englewood, Libraries Unlimited, pp. 14 and 28

Chan, L.M. (2004b). A guide to the Library of Congress Classification. 5th ed. Englewood, Libraries Unlimited, p. 1

Chan, L.M. and Zeng, M.L. (2004). Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology*, 55(5), 377–395.

Chaojun, X., Yi, L. and Xiaojiang, Y. (2009). Design and implementation of theme based Resource Gathering System using fuzzy rule reasoning algorithm. *Journal of the China Society for Scientific and Technical Information,* 28(6), 815–820.

Chatterjee, A. (2000). Treatment of complex subjects in documentary classification with special reference to CC and BC. A comparative study. *Herald of Library Science*, 39(1–2), 41–49.

Chaudhry, A.S. and Jiun, T.P. (2005). Enhancing access to digital information resources on heritage: a case of development of a taxonomy at the Integrated Museum and Archives System in Singapore. *Journal of Documentation*, 61(6), 751–776.

Chaudhry, A.S. and Ling, G.H. (2005). Building taxonomies using organizational resources: a case of business consulting environment. *Knowledge Organization*, 32(1), 25–46.

Chen, Z. and Lu, K. (2006). A preprocess algorithm of filtering irrelevant information based on the minimum class difference. *Knowledge-Based Systems*, 19(6), 422–429.

Chen, J. et al. (2008). A selective Bayes classifier for classifying incomplete data based on gain ratio. *Knowledge-Based Systems*, 21(7), 530–534.

Cheung, C.F., Lee, W.B. and Wang, Y. (2005). A multi-facet taxonomy system with applications in unstructured knowledge management. *Journal of Knowledge Management*, 9(6), 76–91.

Choi, B. and Peng, X. (2004). Dynamic and hierarchical classification of Web pages. *Online Information Review*, 28(2), 139–147.

Chrissikopoulos, V., Papavlasopoulos, S. and Poulos, M. (2005). A text categorization technique based on a numerical conversion of a symbolic expression and an onion layers algorithm. *Journal of Digital Information,* 6(1) Retrieved on 10 April 2010, from: http://journals.tdl.org/jodi/article/view/63.

Chung, W., Chen, H. and Reid, E. (2009). Business Stakeholder Analyzer: an experiment of classifying stakeholders on the Web. *Journal of the American Society for Information Science and Technology*, 60(1), 59–74.

Cisco, S.L. and Jackson, W.K. (2005). Creating order out of chaos with taxonomies. *Information Management Journal,* 39(3), 44–46.

Connelly, J. (2008). Functional taxonomies: myth or magic? *Records Management Bulletin*, 142, 10–15.

Corcocan, M. (2002). Taxonomies: hope or hype? *Online*, 26(5), 76–78.

Cosh, K.J., Burns, R. and Daniel, T. (2008). Content clouds: classifying content in Web 2.0. *Library Review*, 57(9), 722–729.

Cote, J.A. (2005). Knowledge taxonomies: what's the role for information professionals? *Information Outlook,* 9(6), 45–46, 48, 51–52.

Coult, G.R. (2002). Taxonomy. *Managing Information*, 9(4), 28–30.

Currier, S. (2002). Classification schemes in art libraries in the United Kingdom. *Art Libraries Journal*, 27(1), 18–22.

Cuvillier, J. (2007). Indexing grey resources: considering the usual behaviour of library users and the use of Dublin Core Metadata using a database of specialised vocabulary. *The Grey Journal*, 3(2), 169–175.

Dal Porto, S. and Marchitelli, A. (2006). The functionality and flexibility of traditional classification schemes applied to a Content Management System (CMS): Facets, *DDC*, *JITA*. *Knowledge Organization*, 33(1), 35–44.

Davis, S.P. (2002). HILCC: a Hierarchical Interface to Library of Congress Classification. *Journal of Internet Cataloging*, 5(4), 19–49.

de Cours, I. (2002). Choosing a classification scheme for the Inha Library in Paris. *Art Libraries Journal*, 27(1), 23–26.

Dehuri, S. and Mall, R. (2006). Predictive and comprehensible rule discovery using a multi-objective genetic algorithm. *Knowledge-Based Systems*, 19(6), 413–421.

Devi, M.I., Rajaram, R. and Selvakuberan, K. (2008). Generating best features for web page classification. *Webology*, 5(1) Retrieved on 10 April 2010, from: http://www.webology.org/2008/v5n1/a52.html.

Dextre Clarke, S.G. (2008). The last 50 years of knowledge organization: a journey through my personal archives. *Journal of Information Science*, 34(4), 427–437.

Dhyani, P. (1999). Library classification in [the] computer age. *DESIDOC Bulletin of Information Technology,* 19(3), 5–14.

Diamond, L. (2001). Library of Congress religious law classification schedules: elegant, flexible and user-friendly. *International Journal of Legal Information*, 29(2), 488–496.

Dilevko, J. and Gottlieb, L. (2009). The relevance of classification theory to textual analysis. *Library & Information Science Research*, 31(2), 92–100.

Dodgson, P. and Jeffrey-Cook, R. (2007). Developing the local government classification scheme. *Records Management Bulletin*, 140, 16–18.

Dong-Geun, O. and Ji-Suk, Y. (2001). Suggesting an option for DDC class Religion (200) for nations in which religious diversity predominates. *Knowledge Organization*, 28(2), 75–84.

Dotsika, F. (2009). Uniting formal and informal descriptive power: reconciling ontologies with folksonomies. *International Journal of Information Management*, 29(5), 407–415.

Dow, R. (2000). Legal intranets: the new law library. *Law Librarian*, 31(3), 147–151.

Dutta, B., Majumder, K. and Sen, B.K. (2008). Classification of key words extracted from research articles published in science journals. *Annals of Library and Information Studies*, 55(4), 317–333.

Duwairi, R.M. (2006). Machine learning for Arabic text categorization. *Journal of the American Society for Information Science and Technology*, 57(8), 1005–1010.

Duwairi, R. et al. (2009). Feature reduction techniques for Arabic text categorization. *Journal of the American Society for Information Science and Technology*, 60(11), 2347–2352.

Earley, S. (2006). To keep KM current, pay attention to context changes. *Information Outlook*, 10(4), 31–32.

El-Beltagy, S.R. and Rafea, A. (2009). KP-Miner: a keyphrase extraction system for English and Arabic Documents. *Information Systems*, 34(1), 132–144.

Ellis, D. and Vasconcelos, A. (1999). Ranganathan and the Net: using facet analysis to search and organise the World Wide Web. *Aslib Proceedings*, 51(1), 3–10.

Fagiolini, A. (2009). CDD 22. Italian edition: the evolution of a project, the main innovations *Biblioteche Oggi*, 27(4), 5–12.

Falasco, L. (2002a). Bases of the United States Patent Classification. *World Patent Information*, 24(1), 31–33.

Falasco, L. (2002b). United States patent classification: system organization. *World Patent Information*, 24(2), 111–117.

Farmer, L. (2006). Automatic categorization: what's it all about? *Serials Librarian*, 51(2), 91–101.

Fennewald, J. (2006). Same questions, different venue: an analysis of in-person and online questions. *Reference Librarian*, 46(95/96), 21–35.

Ferrari, R.C. (1999). The art of classification: alternate classification systems in art libraries. *Cataloging and Classification Quarterly*, 28(2), 73–98.

Ferris, A.M. (2009). They will use it, if you buy it! Results of an expanded survey on the use of classification Web. *Cataloging & Classification Quarterly*, 47(5), 427–451.

Fersini, E., Messina, E. and Archetti, F. (2008). Enhancing web page classification through image-block importance analysis. *Information Processing and Management*, 44(4), 1431–1447.
Field, M. (2000). The art of classification: making use of an old new business tool. *Law Librarian*, 31(3), 187–189.
Field, M. (1999). Why classification is sexy. *Library Association Record*, 101(10), 580–582.
Flood, G. (2002). Crystal classification shines for big databases. *Information World Review*, 181, 4.
Foscarini, F. (2006). Records classification and functions: an archival perspective. *Knowledge Organization*, 33(4), 188–198.
Foster, J. et al. (2001). When is a thesaurus not a thesaurus: the GCHQ story continues. *Catalogue & Index*, 140, 12.
Fox, R. (2005). Cataloging our information architecture. *OCLC Systems and Services,* 21(1), 23–29.
Frank, E. and Paynter, G.W. (2004). Predicting Library of Congress classifications from Library of Congress subject headings. *Journal of the American Society for Information Science and Technology*, 55(3), 214–227.
Fujino, A., Ueda, N. and Saito, K. (2007). A hybrid generative/discriminative approach to text classification with additional information. *Information Processing & Management*, 43(2), 379–392.
Furner, J. (2007). Dewey deracialized: a critical race-theoretic perspective. *Knowledge Organization*, 34(3), 144–168.
Gaikaiwari, M. (2002). An interactive application for faceted classification systems. *Information Studies*, 8(1), 43–56.
Gama, J. (2000). Combining classification algorithms. *AI Communications*, 13(2), 135–136.
Garshol, L.M. (2004). Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science*, 30(4), 378–391.
Gauch, S., Chandramouli, A. and Ranganathan, Shankar (2009). Training a hierarchical classifier using interdocument relationships. *Journal of the American Society for Information Science and Technology*, 60(1), 47–58.
Gauder, B. (2003). Team efforts keep the DDC nimble. *OCLC Newsletter,* (261), 15.
Geneva, G. et al. (2003). Experiments in discourse analysis impact on information classification and retrieval algorithms. *Information Processing and Management*, 39(6), 825–851.
Giess, M.D., Wild, P.J. and McMahon, C.A. (2008). The generation of faceted classification schemes for use in the Organisation of Engineering Design Documents. *International Journal of Information Management*, 28(5), 379–390.
Gilchrist, A. (2003). Thesauri, taxonomies and ontologies: an etymological note. *Journal of Documentation*, 59(1), 7–18.
Gilchrist, A. (2002). From Aristotle to the 'semantic web'. *Library Association Record*, 104(1), 40–42.
Gilchrist, A. (2001). Corporate taxonomies: report on a survey of current practice. *Online Information Review*, 25(2), 94–102.

Glanzel and Schubert (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56(3), 357–367.

Glaser, R. (2007). To classify or alphabetize: the arrangement of print periodicals in academic libraries. *Serials Review*, 33(2), 91–96.

Gnoli, C. (2006). Phylogenetic classification. *Knowledge Organization*, 33(3), 138–152.

Gnoli, C. and Poli, R. (2004). Levels of reality and levels of representation. *Knowledge Organization*, 31(3), 151–160.

Goh, D.H. et al. (2009). Resource discovery through social tagging: a classification and content analytic approach. *Online Information Review*, 33(3), 568–583.

Gokhale, P. (2009). Ontology: a tool for organization of knowledge. *Information Studies*, 15(4), 233–242.

Goldberg, J.E. (2003). Development of a Universal Law Classification: a retrospective on Library of Congress Class K. *Cataloging and Classification Quarterly*, 35(3/4), 355–436.

Golub, K. (2006). Automated subject classification of textual web documents. *Journal of Documentation*, 62(3), 350–371.

Golub, K. and Lykke, M. (2009). Automated classification of Web pages in hierarchical browsing. *Journal of Documentation*, 65(6), 901–925.

Gonzalez, L. (2005). WebDewey change streamlines workflow. *Action for Libraries*, 31(11).

Gopinath, M.A. (2002). Study of causal factors for problems in the recognition of manifestations of fundamental categories. *SRELS Journal of Information Management*, 39(3), 235–244.

Gopinath, M.A. (2001a). Lexicographic approach to classification and thesaurus construction. *International Information Communication and Education*, 20(1), 13–24.

Gopinath, M.A. (2001b). Universal classification system for knowledge fields. *Information Studies*, 7(2), 71–73.

Gopinath, M.A. (1999). Curricula in teaching classification and indexing in advanced courses in library and information science: a case study. *Library Science with a Slant to Documentation and Information Studies*, 36(1), 3–9.

Gopinath, M.A. and Mangai, A. (2003). Postulational approach to classifying subjects embodied in documents in any medium. *Information Studies*, 9(1/2), 31–42.

Gurbuz, F., Ozbaki, L. and Yapici, H. (2009). Classification rule discovery for the aviation incidents resulted in fatality. *Knowledge-Based Systems*, 22(8), 622–632.

Haider, S.J. (2006). Teaching of cataloging and classification in Pakistan. *Cataloging & Classification Quarterly*, 43(1).

Hammami, M. et al. (2008). Automatic violent content Web filtering approach based on the KDD process. *International Journal of Web Information Systems*, 4(4), 441–464.

Hammond, T. et al. (2005). Social bookmarking tools (1): a general review. *D-Lib Magazine,* 11(4) Retrieved on 10 April 2010, from: http://www.dlib.org/dlib/april05/hammond/04hammond.html.

Han, H., Ko, Y. and Seo, J. (2007). Using the revised EM algorithm to remove noisy data for improving the one-against-the-rest method in binary text classification. *Information Processing & Management*, 43(5), 1281–1293.

Harbo, O. (2002). Melvil Dewey, rationalitetens apostel. Melvil Dewey, apostle of rationalism. *Synopsis*, 33(6), 311–316.

Harinarayana, N.S. and Raju, N.V. (2009). Citation analysis of SRR's works: a look through the window of Google Scholar. *Information Studies*, 15(3), 165–178.

Harvey, R. and Reynolds, S. (2005). MARCup to markup: education for cataloguing and classification in Australia. *Cataloging & Classification Quarterly*, 41(3/4).

Hayati, Z. and Tajer, P. (2006). The comparison of hierarchical structure of subject directories Yahoo, The Open Directory & Looksmart with Dewey Decimal Classification in 10 selective subjects. *Iranian Journal of Information Science and Technology*, 4(2).

Hedden, H. (2008). Taxonomy tool roundup. *EContent*, 31(3), 40–44.

Heiner-Freiling, M. (2006). DDC German – the project, the aims, the methods: new ideas for a well-established traditional classification system. *Cataloging & Classification Quarterly*, 42(3/4), 147–162.

Henry, D.B. and Neville, T.M. (2008). Testing classification systems for reference questions. *Reference & User Services Quarterly*, 47(4), 364–373.

Hickey, T.B. and Vizine-Goetz, D. (2001). The role of classification in CORC. *Journal of Library Administration*, 34(3/4), 421–430.

Hider, P. (2009). A Comparison between the RDA taxonomies and end-user categorizations of content and carrier. *Cataloging & Classification Quarterly*, 47(6), 544–560.

Hider, P. (2004). Learning to classify: online versus printed Dewey. *Malaysian Journal of Library and Information Science*, 9(1), 15–25.

Hiebert, J.T. (2009). Beyond Mark and Park: classification mapping as a collection development tool for psychiatry/psychology. *Collection Management*, 34(3), 182–193.

Himmel, W., Reincke, U. and Michelmann, H.W. (2009). Text mining and natural language processing approaches for automatic categorization of lay requests to Web-based expert forums. *Journal of Medical Internet Research*, 11(3).

Hjorland, B (2009). Concept theory. *Journal of the American Society for Information Science and Technology*, 60(8), 1519–1536.

Hjorland, B. (2008a). Core classification theory: a reply to Szostak. *Journal of Documentation*, 64(3), 333–342.

Hjorland, B. (2008b). What is Knowledge Organization (KO)? *Knowledge Organization*, 35(2–3), 86–101.

Hjorland, B. and Albrechtsen, H. (1999). An analysis of some trends in classification research. *Knowledge Organization*, 26(3), 131–139.

Hjorland, B. and Nicolaisen, J. (2004). Scientific and scholarly classifications are not 'naive': a comment. *Knowledge Organization*, 31(1), 55–61.

Hjorland, B. and Pedersen, K.N. (2005). A substantive theory of classification for information retrieval. *Journal of Documentation*, 61(5), 582–597.

Holley, R.P. (2008). Subject access tools in English for Canadian topics: Canadian extensions to U.S. subject access tools. *Library Resources & Technical Services*, 52(2), 29–43.

Hopkins, S. (2007). Decimating Dewey: introducing a bookshop arrangement for shelving the nonfiction collection. *Australasian Public Libraries and Information Services*, 20(1), 8–13.

Howley, T. and Madden, M.G. (2005). The genetic kernel support vector machine: description and evaluation. *Artificial Intelligence Review,* 24(3–4), 379–395.

Hu, Y. and Chen, Y. (2007). Differences between the DDC and the CLC in classifying works of literature. *Illinois Libraries*, 86(4), 5–10.

Humphrey, S.M. et al. (2009). Comparing a rule-based versus statistical system for automatic categorization of MEDLINE documents according to biomedical specialty. *Journal of the American Society for Information Science and Technology*, 60(12), 2530–2539.

Hung, C. and Chien, L. (2007). Web-based text classification in the absence of manually labeled training documents. *Journal of the American Society for Information Science and Technology*, 58(1), 88–96.

Interview with Ingetraut Dahlberg. *Knowledge Organization*, 35(2–3), 82–85.

Ishida, E. (2006). An overview of text categorization. *Journal of Information Science and Technology Association (Joho no Kagaku to Gijutsu)*, 56(10), 469–474.

Jacob, E.K. (2004). Classification and categorization: a difference that makes a difference. *Library Trends*, 52(3), 515–540.

Jaiswal, B. (1999). Automatic document classification. *DESIDOC Bulletin of Information Technology*, 19(3), 23–28.

Jamain, A. and Hand, D.J. (2008). Mining supervised classification performance studies: a meta-analytic investigation. *Journal of Classification*, 25(1), 87–112.

Jansen, B.J., Booth, D.L. and Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing and Management*, 44(3), 1251–1266.

Janssens, F. et al. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing and Management*, 45(6), 683–702.

Jeffrey-Cook, R. (2005). Developing a fileplan for local government. *Records Management Bulletin*, 125, 3–5.

Jianbo, D. and Hanqing, H. (2005). Principle of the automatic mapping system of library classifications, taking CLC and DDC as the example. *Journal of the China Society for Scientific and Technical Information,* 24(3), 299–303.

Jiang, S.Y. (2007). Lost in translation: the treatment of Chinese Classics in the Library of Congress Classification. *Cataloging & Classification Quarterly*, 45(1), 3–14.

Jing, L. and Hanqing, H. (2009). Design and construction of intelligent search engine based on subject portal: the Experiment of Agricultural History portal's search engine. *Journal of the China Society for Scientific and Technical Information*, 28(1), 114–120.

Jones, K.S. (2005). Revisiting classification for retrieval. *Journal of Documentation,* 61(5), 598–601.

Jones, S. (2002). Classification: are we overdoing it? *Library + Information Update*, 1(1), 24.

Joorabchi, A. and Mahdi, A.E. (2009). An automated syllabus digital library system for higher education in Ireland. *The Electronic Library*, 27(4), 640–658.

Jouin, S. (2008). 'We need to talk about problem novels.' A new way of classifying novels for a young adult readership. *Bulletin des Bibliotheques de France*, 6, 76–80.

Jun, W. and Yuhua, C. (2009). An automatic approach to ontology building by integrating traditional knowledge organization resources. *Journal of the China Society for Scientific and Technical Information,* 28(5), 651–657.

Junion-Metz, G. (2002). Dewey made fun. *School Library Journal*, 48(8), 31.

Kabir, A.M.F. (2003). Ranganathan: a universal librarian. *Journal of Educational Media and Library Sciences*, 40(4), 453–459.

Kanaan, G. et al. (2009). A comparison of text-classification techniques applied to Arabic text. *Journal of the American Society for Information Science and Technology*, 60(9), 1836–1844.

Kanaris, I. and Stamatatos, E. (2009). Learning to recognize webpage genres. *Information Processing and Management,* 45(5), 499–512.

Kashyap, M.M. (2003). Likeness between Ranganathan's postulations based approach to knowledge classification and entity relationship data modeling approach. *Knowledge Organization*, 30(1), 1–19.

Kashyap, M.M. (2001). Similarity between ranganathan's postulates for designing a scheme for library classification and Peter Pin-Sen Chen's entity relationship approach to data modelling and analysis. *DESIDOC Bulletin of Information Technology*, 21(3), 3–16.

Kato, Y. et al. (2008). Classifying information sender of web documents. *Internet Research*, 18(2), 191–203.

Keikha, M., Khonsari, A. and Oroumchian, R. (2009). Rich document representation and classification. *Knowledge-Based Systems*, 22(1), 67–71.

Kepner, L.T. (2002). Workable solutions with Dewey classification. *Technicalities*, 22(4), 9, 12–13.

Khan, A. (2004). Dewey Decimal Classification 22 and beyond. A study of the new edition of Dewey Decimal Classification. *Pakistan Library & Information Science Journal*, 35(3), 9–17.

Khan, S.N. (2004). Expansion of DDC 21st edition number 297.63 for Sirah. *Pakistan Library & Information Science Journal,* 35(4), 16–25.

Kim, K. (2003). Recent work in cataloging and classification, 2000–2002. *Library Resources and Technical Services*, 47(3), 96–108.

Kim, J. and Choi, K. (2007). Patent document categorization based on semantic structural information. *Information Processing & Management*, 43(5), 1200–1215.

Kim, J. and Lee, K. (2002). Designing a knowledge base for automatic book classification. *Electronic Library*, 20(6), 488–495.

Kim, G.C., Lee, K.J. and Lim, C.S. (2005). Multiple sets of features for automatic genre classification of web documents. *Information Processing & Management*, 41(5), 1263–1276.

Kinney, D.W. (2009). The classification of music moving image materials: historical perspectives, problems, and practical solutions. *Cataloging & Classification Quarterly*, 47(1), 2–22.

Klavans, R. and Boyack, K.W. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology*, 60(3), 455–476.

Ko, Y. and Seo, J. (2009). Text classification from unlabeled documents with bootstrapping and feature projection techniques. *Information Processing and Management*, 45(1), 70–83.

Ko, Y. and Seo, J. (2004). Using the feature projection technique based on a normalized voting method for text classification. *Information Processing and Management*, 40(2), 191–208.

Kokabi, M. (2005). An account of cataloging and classification education in Iranian universities. *Cataloging & Classification Quarterly*, 41(3/4), 431–441.

Kotter, W.R. (2002). Improving subject access in anthropology. *Behavioral and Social Sciences Librarian*, 20(2), 1–15.

Kua, E. (2004). Non-Western languages and literatures in the Dewey decimal classification scheme. *Libri,* 54(4), 256–265.

Kuhn, T.J. (1999). Notes on operations: classifying newspapers using Dewey Decimal Classification. *Library Resources and Technical Services*, 43(2), 106–113.

Kumar, T.V.R. and Parameswaran, M. (1999). Chain indexing and UDC. *Annals of Library Science and Documentation*, 46(2), 53–58.

Kumar, S. and Shah, L. (2003). Expansion of schedules of time isolates in colon classification edition 6 and edition 7. *SRELS Journal of Information Management,* 40(2), 105–122.

Kumbhar, R. (2005). Speciator based faceted depth classification's application in thesaurus construction. *Annals of Library and Information Studies,* 52(1), 15–24.

Kunin (1999). IPC and the new millennium: challenge and opportunity. *World Patent Information*, 21(2), 101–108.

Kwak, C. (2005). A study on the job training and self-training of the cataloging and classification librarians working in academic libraries (Korea). *Cataloging & Classification Quarterly*, 41(2), 135–147.

Kwasnik, B.H. (1999). The role of classification in knowledge representation and discovery. *Library Trends*, 48(1), 22–47.

Kwasnik, B.H. and Rubin, V.L. (2004). Stretching conceptual structures in classifications across languages and cultures. *Cataloging and Classification Quarterly*, 37(1/2), 33–47.

Kwon, O.W. and Lee, J.H. (2003). Text categorization based on k-nearest neighbor approach for Web site classification. *Information Processing and Management*, 39(1), 25–44.

La Barre, K. (2007). The heritage of early FC in document reference retrieval systems, 1920–1969. *Library History*, 23(2), 129–149.

Laguia, M. and Castro, J.L. (2008). Local distance-based classification. *Knowledge-Based Systems*, 21(7), 692–703.

Lai, K. and Wu, S. (2005). Using the patent co-citation approach to establish a new patent classification system. *Information Processing and Management,* 41(2), 313–330.

Landry, P. (2006a). The use of the Dewey Decimal Classification (DDC) for the Organisation of National Bibliographies: Switzerland, Germany and Austria. *International Cataloguing and Bibliographic Control*, 35(3), 59–61.

Landry, P. (2006b). Dewey Decimal Classification (DDC) at the Swiss National Library. *Cataloging & Classification Quarterly*, 42(3/4).
Larsen, J.L. (2009). The Longhouse Proposal for objects classified by mediums. *Journal of the American Society for Information Science and Technology,* 60(9), 1907–1914.
Latzer, M. et al. (2006). Institutional variety in communications regulation. Classification scheme and empirical evidence from Austria. *Telecommunications Policy*, 30(3–4), 152–170.
Lau Whelan, D. (2007). Arizona Library ditches Dewey: while librarians debate the pros and cons, patrons don't seem to notice. *School Library Journal*, 53(7), 14.
Lawson, D. (2001). You've come a long way, Dewey! *OCLC Newsletter*, 254, 34–35.
Lee, C. (2007). Improving classification performance using unlabeled data: naïve Bayesian case. *Knowledge-Based Systems*, 20(3), 220–224.
Lee, C. and Lee, G.G. (2006). Information gain and divergence-based feature selection for machine learning-based text categorization. *Information Processing & Management,* 42(1), 155–165.
Lee, L. and Luh, C. (2008). Generation of Pornographic Blacklist and its incremental update using an inverse chi-square based method. *Information Processing and Management*, 44(5), 1698–1706.
Levinson, D. (2006). Anthropology, taxonomies, and publishing. *Online*, 30(4), 28–30.
Lewis, N. and Seago, K. (1999). An automated reclassification project at the University of Kentucky. *Cataloging and Classification Quarterly*, 28(4), 117–134.
Li, Y. and Belkin, N.J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management*, 44(6), 1822–1837.
Li, J. and Jones, J. (2006). Using multiple and negative target rules to make classifiers more understandable. *Knowledge-Based Systems*, 19(6), 438–444.
Liang, G. (2007). A history and backgrounds of the integration of classification and subject headings in China Library. *Information and Media Studies*, 5(2), 1–8.
Liang, C. et al. (2006). Dictionary-based text categorization of chemical web pages. *Information Processing & Management*, 42(4), 1017–1029.
Liao, S. et al. (2009). PORE: A Personal Ontology Recommender system for digital libraries. *The Electronic Library*, 27(3), 496–508.
Lim, E. (2000). Southeast Asian subject gateways: an examination of their classification practices. *International Cataloguing and Bibliographic Control*, 29(3), 45–48.
Lim, E., Ng, W. and Sun, A. (2003). Performance measurement framework for hierarchical text classification. *Journal of the American Society for Information Science and Technology*, 54(11), 1014–1028.
Lin, K.R. and Murphy, E. (2000). Reflections on a JX reclassification project. *Law Library Journal*, 92(4), 459–466.
Liu, R. (2009). Context recognition for hierarchical text classification. *Journal of the American Society for Information Science and Technology*, 60(4), 803–813.
Liu, R. (2008). Interactive high-quality text classification. *Information Processing and Management*, 44(3), 1062–1075.

Liu, Y. et al. (2000). Visualizing document classification: a search aid for the digital library. *Journal of the American Society for Information Science*, 51(3), 216–227.

Longstaff, R. and Henry, E.C. (2009). To B or not to B: deliberation on the classification of religious books. *Catholic Library World*, 79(3), 209–212.

Lubbes, R.K. (2001). Automatic categorization: how it works, related issues, and impacts on records management. *Information Management Journal*, 35(4), 38–43.

Ma, Z. (2005). The education of cataloging and classification in China. *Cataloging & Classification Quarterly*, 41(2).

Macfarlane, A. and Richter, G. (2005).The impact of metadata on the accuracy of automated patent classification. *World Patent Information,* 27(1), 13–26.

Macgregor, G. and McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, 55(5), 291–300.

Mai, J. (2008). Actors, domains, and constraints in the design and construction of controlled vocabularies. *Knowledge Organization*, 35(1), 16–29.

Mai, J. (2004a). Classification in context: relativity, reality, and representation. *Knowledge Organization*, 31(1), 39–48.

Mai, J. (2004b). Classification of the web: challenges and inquiries. *Knowledge Organization*, 31(2), 92–97.

Mai, J. (2004c).The future of general classification. *Cataloging and Classification Quarterly*, 37(1/2), 3–12.

Makarov, M. (2006). The eighth edition of the IPC. *World Patent Information*, 28(2), 122–126.

Makarov, M. (2004). The process of reforming the International Patent Classification. *World Patent Information*, 26(2), 137–141.

Makarov, M. (2000). The seventh edition of the IPC. *World Patent Information*, 22(1–2), 53–58.

Maker, R. (2008). Reader centered classification of adult fiction in public libraries. *Australasian Public Libraries and Information Services (APLIS),* 21(4), 168–171.

Mandal, A. and Sain, C. (1999). A critical observation of DDC-21 over DDC-20. *Herald of Library Science*, 38(3–4), 171–175.

Mann, T. (2008). Will Google's keyword searching eliminate the need for LC cataloging and classification? *Journal of Library Metadata*, 8(2), 159–168.

Marcella, R. (2002). The role and future of subject classification: the exploitation of resources. *Legal Information Management*, 2(2), 3–6.

Marcio Pajeu, H. et al. (2007). A new proposal of classification for comics. *Biblionline*, 3(2) Retrieved on 10 April 2010, from: http.www.ies.ufpb.br/ojs2/index.php/biblio/article/viewFile/1920/1689.

Markey, K. (2006). Forty years of classification online: final chapter or future unlimited? *Cataloging & Classification Quarterly*, 42(3/4), 1–63.

Marsh, C. (2002). The 'ANSCR' to CD classification at Leeds College of Music. *Brio*, 39(1), 33–48.

Marsh, E. (1999). Improving communication and classification in the next century. *OCLC Newsletter*, 237, 29–31.

Marsono, M.N. et al. (2009). Targeting spam control on middleboxes: spam detection based on layer-3 e-mail content classification. *Computer Networks*, 53(6), 835–848.

Martin, C. and Daniels, W. (2000). Dewey applications for the simple arrangement of a link library: a case of Science Net. *Journal of Internet Cataloging*, 3(1), 67–77.

Martinez Arellano, F.F. (2005). Education for cataloging and classification in Mexico. *Cataloging & Classification Quarterly*, 41(3/4), 353–388.

Matusiak, K.K. (2006). Towards user-centered indexing in digital image collections. *OCLC Systems and Services*, 22(4), 283–298.

Matveyeva, S.J. (2002). A role for classification: the organization of resources on the Internet, *MLA Forum*, 1(2) Retrieved from 10 April 2010, from: http://www.mlaforum.org/volumeJ/issue2/roleClassification.html.

McCulloch, E. (2004). Multiple terminologies: an obstacle to information retrieval. *Library Review*, 53(6), 297–300.

McCulloch, E., Shiri, A. and Nicholson, D. (2005). Challenges and issues in terminology mapping: a digital library perspective. *Electronic Library,* 23(6), 671–677.

McIlwaine, I.C. (2003). Indexing and the Classification Research Group. *Indexer*, 23(4), 204–208.

McKenna, F. (2009). Do you really need a taxonomy/classification scheme with a records management system? Why are we still doing it the old way? *Records Management Bulletin*, 152, 12–17.

Mehler, A. and Waltinger, U. (2009). Enhancing document modeling by means of open topic models. *Library Hi Tech*, 27(4), 520–539.

Melenica, M. et al. (2008). Language morphology offset: text classification on a Croatian-English parallel corpus. *Information Processing and Management*, 44(1), 325–339.

Mengle, S. and Goharian, N. (2009a). Ambiguity measure feature-selection algorithm. *Journal of the American Society for Information Science and Technology,* 60(5), 1037–1050.

Mengle, S. and Goharian, N. (2009b). Passage detection using text classification. *Journal of the American Society for Information Science and Technology*, 60(4), 814–825.

Meserve, H.C. et al. (2009). Developing a model for reference research statistics: applying the 'Warner Model' of reference question classification to streamline research services. *Reference & User Services Quarterly*, 48(3), 247–258.

Miksa, F. (2006). The DDC relative index. *Cataloging & Classification Quarterly*, 42(3/4), 65–95.

Mills, J. (2004). Faceted classification and logical division in information retrieval. *Library Trends*, 52(3), 541–570.

Milne, C. (2007). Taxonomy development: assessing the merits of contextual classification. *Records Management Journal*, 17(1), 7–16.

Miskin, C. (2002). Taxonomies. *Legal Information Management*, 2(1), 16–23.

Mitchell, J.S. (2004). DDC 22: an introduction. *International Cataloguing and Bibliographic Control*, 33(2), 27–31.

Mitchell, J.S. (2003). DDC 22 offers many updates to Dewey users worldwide. *OCLC Newsletter*, (261), 16–19.

Mitchell, J.S. (2001). Dewey Decimal Classification: 125 and still growing. *OCLC Newsletter*, (254), 27–29.

Mitchell, J.S. and Vizine-Goetz, D. (2009). The DDC and OCLC. *Journal of Library Administration,* 49(6), 657–667.

Mondoux, J. and Shiri, A. (2009). Institutional repositories in Canadian post-secondary institutions: user interface features and knowledge organization systems. *Aslib Proceedings: New Information Perspectives*, 61(5), 436–458.
Montesi, M. (2008). Web genres: research trends. *Profesional de la Informacion*, 17(5), 551–557.
Moore, S. (1999). Navigating the G schedule. *Cataloging and Classification Quarterly*, 27(3/4), 375–384.
Morelli, J. (2007). Don't build your house on sand! *Records Management Bulletin*, 140, 10–12.
Morelli, J. (2005). Business classification schemes: issues and options. *Records Management Bulletin*, (124)15, 17–21.
Mosley, G. (2002). To classify or not to classify, that is the question? *Legal Information Management*, 2(2), 8–10.
Munk, T.B. and Mork, K. (2007). Folksonomies, tagging communities, and tagging strategies – an empirical study. *Knowledge Organization*, 34(3), 115–127.
Mutula, S.M. and Tsvakai, M. (2002). Historical perspectives of cataloguing and classification in Africa. *Cataloging and Classification Quarterly*, 35(1/2), 61–77.
Nasir Uddin, M. and Janecek, P. (2007). Faceted classification in web information architecture: a framework for using semantic web tools. *Electronic Library*, 25(2), 219–233.
Nasir Uddin, M. et al. (2006). Information description and discovery method using classification structures in web. *Malaysian Journal of Library & Information Science*, 11(2), 1–20.
Ndakotsu, T.M. (2006). Classifying Africa. *African Research and Documentation*, 101, 39–42.
Neelameghan, A. (2009). Tamil figures of speech: categorization. *Information Studies*, 15(3), 141–164.
Neelameghan, A. (2007). Online integrated use of bilingual thesaurus + Colon Classification + lexicon: a case study in the domain of Tamil classics. *Information Studies*, 13(4), 207–218.
Neelameghan, A. (2005). Subject architecture model and building construction model: a system perspective. *Information Studies,* 11(4), 207–212.
Neelameghan, A. (2002a). Classification in the digital environment. *Information Studies*, 8(1), 1–8.
Neelameghan, A. (2002b). Hierarchy, hierarchical relation and hierarchical arrangement. *Information Studies*, 8(1), 9–23.
Neelameghan, A. (2002c). WINISIS based information handling tools from SRELS (1). *Information Studies,* 8(2), 71–82.
Neelameghan, A. and Raghavan, K.S. (2006). Seminal mnemonics in knowledge organization: ancient traditions and modern practices. *Information Studies*, 12(1), 5–26.
Negrini, G. (2003). Organizing the literary work of art – a common aim for Finland and Italy. *Signum*, 36(4), 91–95.
Nero, L.M. (2006). Classifying the popular music of Trinidad and Tobago. *Cataloging & Classification Quarterly*, 42(3/4).
Neville, T.M. and Henry, D.B. (2009). Reference classification – is it time to make some changes? *Reference & User Services Quarterly*, 48(4), 372–383.

Ngoepe, M. (2009). Organising public records to achieve service delivery: the role of the National Archives and Records Service of South Africa's Functional Subject File Plan in Government departments. *ESARBICA Journal: Journal of the Eastern and Southern Africa Regional Branch of the International Council on Archives*. Vol. 28.

Nicholson, D.M., Dawson, A. and Shiri, A. (2006). HILT: a pilot terminology mapping service with a DDC spine. *Cataloging & Classification Quarterly*, 42(3/4).

Niculescu, Z. (2009). Dewey Decimal Classification Editions. *Studii de Biblioteconomie si Siinta Informarii/Library and Information Science Research*, 13, 33–50.

Nielsen, M.L. (2008). Networked Knowledge Organization Systems/Services (NKOS): ECDL 2008 Conference Report. *D-Lib Magazine*, 14(11–12).

Noruzi, A. (2006). Folksonomies: (un)controlled vocabulary? *Knowledge Organization*, 33(4), 199–203.

Obajemu, A.S. and Ibegwam, A. (2006). A survey of librarians' attitudes to training programmes on ICT application to cataloguing and classification workshops in Nigeria. *African Journal of Library, Archives and Information Science*, 16(1), 19–27.

O'Connor, L. (2000). Approaching the challenges and costs of the North American Industrial Classification System (NAICS). *Bottom Line*, 13(2), 83–89.

Olson, H.A. (2004). The ubiquitous hierarchy: an army to overcome the threat of a mob. *Library Trends*, 52(3), 604–616.

Olson, H.A. (2001). Sameness and difference: a cultural foundation of classification. *Library Resources and Technical Services*, 45(3), 115–122.

Onohwakpor, J.E. (2006). Annual Conference of the Cataloguing, Classification and Indexing Section of the Nigerian Library Association. *Library Hi Tech News*, 23(9), 6–8.

Orr, S. (2006). Functions-based classification of records: is it functional? *Archives & Manuscripts*, 34(1), 44–96.

Oyler, P.G. (2002). Teaching classification in the 21st century. *International Cataloguing and Bibliographic Control*, 31(1), 16–17.

Panigrahi, P. (2007a). Facet analysis techniques for automatic classification systems: a study for colon classification. *SRELS Journal of Information Management*, 44(4), 309–332.

Panigrahi, P. (2007b). Implementation of the Anteriorising Common Isolates in automatic classification system. *SRELS Journal of Information Management*, 44(3), 225–236.

Panigrahi, P.K. (2000a). An artificial intelligence approach towards automatic classification (Part 1). *IASLIC Bulletin*, 45(2), 73–81.

Panigrahi, P.K. (2000b). An artificial intelligence approach towards automatic classification (Part 2). *IASLIC Bulletin*, 45(3), 104–118.

Panigrahi, P. and Prasad, A.R.D. (2007). Facet sequence in analytico synthetic scheme: a study for developing an AI based automatic classification system. *Annals of Library and Information Studies*, 54(1), 37–43.

Panigrahi, P. and Prasad, A.R.D. (2005a). Inference engine for devices of Colon Classification in AI based automated classification system. *Annals of Library and Information Studies,* 52(4), 124–134.

Panigrahi, P. and Prasad, A.R.D. (2005b). Inference engine for space isolates of colon classification for AI based automated classification system. *SRELS Journal of Information Management*, 42(4), 383–406.

Paradis, F. and Nie, J. (2007). Contextual feature selection for text classification. *Information Processing & Management*, 43(2), 344–352.

Park, O. (2008). Opening ontology design: a study of the implications of knowledge organization for ontology design. *Knowledge Organization*, 35(4), 209–221.

Park, S. and Zhang, B. (2004). Co-trained support vector machines for large scale unstructured document classification using unlabeled data and syntactic information. *Information Processing and Management*, 40(3), 421–439.

Parkkola, H., Saariluoma, P. and Berki, E. (2009). Action-oriented classification of families' information and communication actions: exploring mothers' viewpoints. *Behavior & Information Technology,* 28(6), 525–536.

Pauwels, H. (2004). Pillars of the IPC reform. *World Patent Information*, 26(3), 205–208.

Peterson, E. (2006). Beneath the metadata: some philosophical problems with folksonomy. *D-Lib Magazine*, 12(10).

Pickering, B. (2004). Collaborative taxonomy tool unveiled by Verity. *Information World Review*, (204), 7.

Plater, B. (2005). Records management for the masses? *Records Management Bulletin,* (124), 27–29.

Pogorelec, A. and Sauperl, A. (2006). The alternative model of classification of belles-lettres in libraries. *Knowledge Organization*, 33(4), 204–214.

Pons-Porrata, A., Berlanga-Llavori, R. and Ruiz-Schulcloper, J. (2007). Topic discovery based on text mining techniques. *Information Processing & Management*, 43(3), 752–768.

Poulter, A. and Brunt, R. (2007). On reading 'Information storage and retrieval in the professional curriculum' by Rodney Brunt. *Library Review*, 56(7), 557–562.

Prasad, K.N. (2001). Classification in the digital environment. *Information Studies*, 7(3–4), 137–138.

Prescott, S. (2001). If you knew Dewey. *School Library Journal*, 47(8), 50–53.

Pu, H. and Yang, C. (2003). Enriching user-oriented class associations for library classification schemes. *Electronic Library*, 21(2), 130–141.

Qingfeng, W. et al. (2009). Study on classification and coding for Web agricultural information based on China Library Classification. *Journal of the China Society for Scientific and Technical Information*, 28(1), 28–33.

Rademaker, C.A. (2000a). The classification of ornamental designs in the United States Patent Classification Scheme. *World Patent Information*, 22(3), 123–133.

Rademaker, C.A. (2000b). The classification of plants in the United States Patent Classification system. *World Patent Information*, 22(4), 301–307.

Rafferty, P. (2001). The representation of knowledge in library classification schemes. *Knowledge Organization,* 28(4), 180–191.

Raghavan, K.S. (2005). Education for knowledge organization: the Indian scene. *Cataloging & Classification Quarterly,* 41(2).

Rajashekara, N. (2005). Information/knowledge organization in a folk museum. *Information Studies,* 11(4), 265–277.
Raschen, B. (2005). A resilient, evolving resource: how to create a taxonomy. *Business Information Review,* 22(3), 199–204.
Rekha, G. and Parameswaran, M. (2002). 'Knowledge Organization', 1988–1999: a bibliometric analysis. *SRELS Journal of Information Management*, 39(4), 355–362.
Richardson, J.V. et al. (2000). Millennium project research agenda: cataloging and classification. *Library Quarterly,* 70(2), ix–xx.
Robbin, A. (2000). Classifying racial and ethnic group data in the United States: the politics of negotiation and accommodation. *Journal of Government Information*, 27(2), 129–156.
Robbins, F. (1999). An exploration of the application of classification systems as a method of resource delivery on the World Wide Web. *Cataloguing Australia*, 25(1/4), 60–65.
Robertson, M.M. (1999). Customizing the organization of local planning documents. *Technicalities*, 19(7), 6–7.
Robinson, G. (1999). Abridging the UDC: the compiling of the pocket edition. *Knowledge Organization*, 26(3), 149–156.
Rooney, E.M. (2009). Classification development: a continuing problem. *Catholic Library World*, 79(3), 204–208.
Ruderman, E. (1999–2000). Library of Congress Classification for Judaica: recent changes (1995–1996). *Judaica Librarianship*, 10(1–2), 31–40.
Sabroski (2000). NAICS Codes: a new classification system for a new economy. *Searcher*, 8(10), 18, 20, 22, 24, 26, 28, 30.
Saeed, H. and Chaudhry, A.S. (2002). Using Dewey Decimal Classification scheme (DDC) for building taxonomies for knowledge organisation. *Journal of Documentation*, 58(5), 575–583.
Saeed, H. and Chaudhry, A.S. (2001). Potential of bibliographic tools to organize knowledge on the Internet: the use of Dewey Decimal Classification Scheme for organizing Web based information resources. *Knowledge Organization*, 28(1), 17–26.
Sanchez-Alonso, S. (2009). Enhancing availability of learning resources on organic agriculture and agroecology. *The Electronic Library*, 27(5), 792–813.
Sanders, D. (2008). Tag – you're it! User tagging lets patrons set their own terms. *American Libraries*, 39(11), 52–54.
Sands, G. (2002). Cat and class: what use are these skills to the new legal information professional? *Legal Information Management*, 2(2), 19–22.
Sangeetha, M.A. et al. (1999). Tools for retrieving patents, Part 1: the International Patent Classification. *IASLIC Bulletin*, 44(2), 67–70.
Santini, M. (2008). Zero, single, or multi? Genre of web pages through the users' perspective. *Information Processing and Management*, 44(2), 702–737.
Satija, M.P. (2008). Mapping of Social Sciences in the Colon Classification. *Annals of Library and Information Studies*, 55(3), 204–211.
Satija, M.P. (2007). Book numbers in India with special reference to the author table for Indian names designed and used by the National Library of India. *Knowledge Organization*, 34(1), 34–40.

Satija, M.P. (2006). Use of classification and indexing in the Internet organization and search. *SRELS Journal of Information Management,* 43(2), 123–136.

Satija, M.P. (2004). Salient features of the DDC 22 (2003). *SRELS Journal of Information Management,* 41(4), 309–316.

Satija, M.P. (2000). Library classification: an essay in terminology. *Knowledge Organization*, 27(4), 221–229.

Satyanarayana, N.R. (1999). Role of classification and cataloguing in the information age. *DESIDOC Bulletin of Information Technology*, 19(3), 3–4.

Saumure, K. and Shiri, A. (2008). Knowledge organization trends in library and information studies: a preliminary comparison of the pre- and post-Web eras. *Journal of Information Science*, 34(5), 651–666.

Schellner, I. (2002). Japanese File Index classification and F-terms. *World Patent Information*, 24(3), 197–201.

Schewe, D.B. (2002). Classifying electronic documents: a new paradigm. *Information Management Journal*, 36(2), 54, 56–59.

Schwartz, E. (2001). Library of Congress Islamic and Jewish law classification schedule. *International Journal of Legal Information*, 29(2), 497–500.

Sen, B.K. (2007). UDC as it could be made. *Annals of Library and Information Studies*, 54(1), 44–63.

Sen, P. et al. (2008). Collective Classification in Network Data. *AI Magazine*, 29(3), 93–106.

Shafer, K.E. (2001). Evaluating Scorpion results. *Journal of Library Administration*, 34(3/4), 237–244.

Shah, L. and Kumar, S. (2002). Expansion of schedules of common isolates in CC6 and CC7. *SRELS Journal of Information Management*, 39(4), 387–394.

Sharma, R.S., Foo, S. and Morales-Arroyo, M. (2008). Developing corporate taxonomies for knowledge auditability: a framework for good practices. *Knowledge Organization*, 35(1), 30–46.

Shelton, J.M. (2000). Library of Congress' Class L: Education, Table L7: an expansion for local use. *Cataloging and Classification Quarterly*, 31(1), 31–42.

Shen, D., Yang, Q. and Chen, Z. (2007). Noise reduction through summarization for Web-page classification. *Information Processing and Management*, 43(6), 1735–1747.

Shenton, A.K. (2006). The role of 'reactive classification' in relation to fiction collections in school libraries. *New Review of Children's Literature and Librarianship*, 12(2), 127–146.

Shiri, A. and Chase-Kruszewski, S. (2009). Knowledge organisation systems in North American digital library collections. *Program: Electronic Library and Information Systems*, 43(2), 121–139.

Shiri, A. and Molberg, K. (2004). Interfaces to knowledge organization systems in Canadian digital library collections. *Online Information Review*, 28(5), 604–620.

Si, L. (2005). The status quo and future development of cataloging and classification education in China. *Cataloging & Classification Quarterly*, 41(2).

Singh, S. (2000). Development of terminology in Colon Classification: a researcher's viewpoint. *Herald of Library Science*, 39(3–4), 177–184.

Singh, S. (1999). Potentialities, limitations and the future of the Colon Classification. *Herald of Library Science*, 38(3–4), 186–194.

Sitas, A. (2001).The classification of Byzantine literature in the Library of Congress Classification. *Knowledge Organization*, 28(2), 85–94.
Sitas, A. (1999). Greek folk literature, poetry, folk songs and the Library of Congress PA (Supplement) Schedule. *Cataloging and Classification Quarterly*, 28(1), 53–68.
Siva, S. et al. (2000). Document identification and classification using transform coding of gray scale projections and neural tree network. *Information Studies*, 6(2), 101–109.
Skinner, G., Han, S. and Chang, E. (2006). An information privacy taxonomy for collaborative environments. *Information Management and Computer Security*, 14(4), 382–394.
Slavic, A, (2008). Use of the Universal Decimal Classification: a world-wide survey. *Journal of Documentation*, 64(2), 211–228.
Slavic, A. (2006). UDC in subject gateways: experiment or opportunity? *Knowledge Organization*, 33(2), 67–85.
Slavic, A. (2004). UDC implementation: from library shelves to a structured indexing language. *International Cataloguing and Bibliographic Control*, 33(3), 60–65.
Slavic, A. (2002). Teaching classification to fit a modern and sustainable LIS curriculum: the case of Croatia. *International Cataloguing and Bibliographic Control*, 31(1), 13–15.
Slavic, A., Cordeiro, M.I. and Riesthuis, G. (2008). Prepared and carried out changes in the management of the universal decimal classification. *Kniznica*, 9(10), 17–22.
Smith, H. (2002). Automation of patent classification. *World Patent Information*, 24(4), 269–271.
Smyth, Z.A. (2005). Adopting a functional classification of business processes in Northern Ireland. *Journal of the Society of Archivists,* 26(2), 233–242.
Soergel, D. (1999). The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Scienc*e, 50(12), 1119–1120.
Song, R. et al. (2009). Identification of ambiguous queries in Web search. *Information Processing and Management*, 45(2), 216–229.
Spila, M. (2008). George W. Bordner and Pennsylvania documents: a Remembrance. *DttP: Documents to the People*, 36(4), 34–35.
Spiteri, L.F. (2007). The structure and form of folksonomy tags: the road to the public library catalog. *Information Technology and Libraries*, 26(3), 13–25.
Spiteri, L.F. (2006). The use of folksonomies in public library catalogues. *Serials Librarian*, 51(2), 75–89.
Spiteri, L.F. (1999). The essential elements of faceted thesauri. *Cataloging and Classification Quarterly*, 28(4), 31–52.
Stickley, L. (1999). Karen Drabenstott envisions a multimedia Dewey Decimal Classification. *OCLC Newsletter*, (240), 37.
Subrahmanyam, B. (2006). Library of Congress classification numbers: issues of consistency and their implications for union catalogs. *Library Resources & Technical Services*, 50(2), 110–119.
Sukiasyan, E.R. (2008a). The (BBK) Library-Bibliographical Classification. The Social Characteristics. Basic Functions and Long-Term Development. *Nauchno-tekhnicheskaya informatsiya*, 2, 25–30.

Sukiasyan, E.R. (2008b). The Organization of Work of a Classifier. *Nauchno-tekhnicheskaya informatsiya*, 8, 14–22.

Sulistyo-Basuki, L. and Mulyani, A.S. (2008). Indonesian Librarians' Efforts to Adapt and Revise the Dewey Decimal Classification (DDC)'s Notation 297 on Islam. *Malaysian Journal of Library & Information Science*, 13(2), 89–101.

Szostak, R. (2008). Classification, interdisciplinarity, and the study of science. *Journal of Documentation*, 64(3), 319–332.

Szunejko, M. (2003). Literature classification schemes at two west Australian university libraries: Murdoch University and the University of Western Australia. *Cataloging and Classification Quarterly*, 36(2), 45–57.

Tabb, W. (2001). The Library of Congress and the DDC. *OCLC Newsletter*, (254), 32–33.

Tadasad, P.G. and Maheswarappa, B.S. (2002). Classification practices among college libraries in Karnataka State. *SRELS Journal of Information Management*, 39(2), 211–218.

Takamasa, M. (2006). International patent classification from the eighth edition onward. *Journal of Information Processing and Management*, 49(7), 359–370.

Tan, S. (2005). Binary k-nearest neighbor for text categorization. *Online Information Review,* 29(4), 391–399.

Tan, J. et al. (2006). An improved mutual information algorithm for feature selection. *Journal of the China Society for Scientific and Technical Information*, 25(6), 651–656.

Taniguchi, S. (2005). Current status of cataloging and classification education in Japan. *Cataloging & Classification Quarterly*, 41(2).

Taylor, A.G. (2006). Teaching the Dewey Decimal Classification system. *Cataloging & Classification Quarterly*, 42(3/4).

Taylor, W. (2009). The Hot Debates in Cataloguing. *Library + Information Update*, 27.

Tebbutt, D. (2005). Categorisers are second guessing future trends. *Information World Review,* 214, 17.

Teng, Y. and Ke, H. (2009). Social tagging on digital archives: a case study of Yuyu Yang Digital Art Museum. *Journal of Librarianship and Information Studies,* 1(68), 80–107.

Thijs, B. and Glanzela, W. (2009). A structural analysis of benchmarks on different bibliometrical indicators for european research institutes based on their research profile. *Scientometrics*, 79(2), 377–388.

Thomas, A.R. (2004). Teach yourself thesaurus: exercises, readings, resources. *Cataloging and Classification Quarterly*, 37 (3/4), 23–34.

Tim, W.L. (2009). LCC from UDC: managing the conversion. *Library Collections, Acquisitions, & Technical Services*, 33(2–3), 73–79.

Tinker, A.J. et al. (1999). The Dewey Decimal Classification and the transition from physical to electronic knowledge organization. *Knowledge Organization*, 26(2), 80–96.

Toth, E. (2002). Innovative solutions in automatic classification: a brief summary. *Libri*, 52(1), 48–53.

Tseng, Y., Lin, C. and Lin, Y. (2007). Text mining techniques for patent analysis. *Information Processing & Management*, 43(5), 1216–1247.

Tsuneda, K. (2009). Changing classification systems at Toho University's Narashino Media Center – aiming at user-oriented service. *Yakugaku Toshokan/Pharmaceutical Library Bulletin*, 54(1), 14–19.
Tudor, J.D. (2000). Nearer to NAICS? *EContent*, 23(2), 68–72.
Tyckoson, D.A. (2007). Top ten innovations in library history. *Against the Grain*, 18(6), 42–43.
Tzouveli, P. et al. (2009). Semantic classification of Byzantine icons. *IEEE Intelligent Systems*, 24(2), 35–43.
van der Walt, M.S. (2004). A classification scheme for the organization of electronic documents in small, medium and micro enterprises (SMMEs). *Knowledge Organization*, 31(1), 26–38.
van Laer, C.J.P. (1999). A comparative lawyer's review of the LC Classification System. *Law Library Journal*, 91(2), 305–311.
Varendorff, L. (2005). Misconceptions in the records management fraternity. *Records Management Bulletin*, (129), 6–7.
Vernitski, A. (2007). Developing an intertextuality-oriented fiction classification. *Journal of Librarianship and Information Science*, 39(1), 41–52.
Vizine-Goetz, D. (2006). DeweyBrowser. *Cataloging & Classification Quarterly*, 42(3/4).
Vizine-Goetz, D. (2002). Classification schemes for Internet resources revisited. *Journal of Internet Cataloging*, 5(4), 5–18.
Vizine-Goetz, D. (2001). Dewey research: new uses for the DDC. *OCLC Newsletter*, 254, 24–26.
Vizine-Goetz, D. and Hickey, T. (2006). Getting visual with DeweyBrowser. *NextSpace*, 1, 18–19.
Vizine-Goetz, D. and Mitchell, J.S. (2001). Dewey 2000. *Journal of Library Administration*, 34(1/2), 103–109.
Walker, C.H. and Copeland, A. (2009). The eye prophetic: Julia Pettee. *Libraries & the Cultural Record*, 44(2), 162–182.
Wang, J. (2009). An extensive study on automated Dewey Decimal Classification. *Journal of the American Society for Information Science and Technology*, 60(11), 2269–2286.
Wang, J. (2006). Automatic thesaurus development: term extraction from title metadata. *Journal of the American Society for Information Science and Technology*, 57(7), 907–920.
Wang, T. and Chiang, H. (2007). Fuzzy support vector machine for multi-class text categorization. *Information Processing & Management*, 43(4), 914–929.
Wang, Y. and Yang, J. (2004). Application of iterative-KNN based on KNN and automatic retrieval in automatic categorization. *Journal of the China Society for Scientific and Technical Information*, 23(2), 137–141.
Wang, L. et al. (2006). Combining decision tree and Naive Bayes for classification. *Knowledge-Based Systems*, 19(7), 511–515.
Wang, Z. et al. (2008). Using classification schemes and thesauri to build an organizational taxonomy for Organizing Content and Aiding Navigation. *Journal of Documentation*, 64(6), 842–876.
Wang, Z. et al. (2006). Potential and prospects of taxonomies for content organization. *Knowledge Organization*, 33(3), 160–169.

Ward, M. (2006). A taxaurus of engineering terms. *Managing Information*, 13(2), 42–47.

Wartzok, S. and Hernandez, M. (2005). Reclassifying the official records of the United Nations: a library's united effort. *Technical Services Quarterly*, 23(1), 53–69.

Waters, J. (2007). Classification versus search – which is more important? *Records Management Bulletin*, 140, 14–15.

Weaver, M. et al. (1999). Centralised classification of library materials – a benchmarking study. *Library and Information Research News*, 23(74), 23–32.

Weber, L. and Schomberg, J. (2007). Access plus security: compact disc classification and packaging at Minnesota State University, Mankato. *Music Reference Services Quarterly*, 9(2), 49–54.

Wheatley, A. (2000). Subject trees on the Internet: a new role for bibliographic classification? *Journal of Internet Cataloging*, 2(3/4), 115–141.

Wild, P.J., Giess, M.D. and McMahon, C.A. (2009). Describing engineering documents with faceted approaches: observations and reflections. *Journal of Documentation*, 65(3), 420–445.

Williamson, N.J. (2009). Classification issues in 2007. *Knowledge Organization*, 36(1), 66–74.

Williamson, N. (2004). Classification research issues. *Knowledge Organization*, 31(4), 252–254.

Wilson, M.D. (2001). Flying first class or economy? Classification of electronic titles in ARL libraries. *Portal: Libraries and the Academy*, 1(3), 225–240.

Winke, R.C. (2004). The contracting world of Cutter's Expansive Classification. *Library Resources and Technical Services*, 48(2), 122–129.

Womack, K.R. (2006). Conformity for conformity's sake? – The choice of a classification system and a subject heading system in academic health sciences libraries. *Cataloging & Classification Quarterly*, 42(1).

Wright, S. and Blasé, N. (2006). Using spreadsheets to map a library reclassification, reorganization, and merger. *Issues in Science and Technology Librarianship*, no. 45.

Wu, Y.B. et al. (2006). Finding nuggets in documents: a machine learning approach. *Journal of the American Society for Information Science and Technology*, 57(6), 740–752.

Xiaoge, P. (2009). A literature review on Web automated text categorization technology. *Journal of the China Society for Scientific and Technical Information*, 28(2), 233–241.

Xiaopeng, Y. and Feicheng, M. (2008). An unsupervised text classification algorithm based on k-nearest neighbors. *Journal of the China Society for Scientific and Technical Information*, 27(4), 550–555.

Xiaoyue, W. and Rujiang, B. (2006). A hybrid classifier based on the rough sets and neural networks. *Journal of the China Society for Scientific and Technical Information*, 25(4), 475–480.

Xin, L. and Renren, L. (2008). A fast text categorization technology for Chinese based on subject terms list. *Journal of the China Society for Scientific and Technical Information*, 27(3), 323–327.

Yi, K. (2007). Automated text classification using library classification schemes: trends, issues, and challenges. *International Cataloguing and Bibliographic Control*, 36(4), 78–82.

Yi, K. and Beheshti, J. (2009). A hidden Markov model-based text classification of medical documents. *Journal of Information Science*, 35(1), 67–81.

Yoon, Y. and Lee, G.G. (2007). Efficient implementation of associative classifiers for document classification. *Information Processing & Management*, 43(2), 393–405.

Yoon, Y., Lee, C. and Lee, G. (2006). An effective procedure for constructing a hierarchical text classification system. *Journal of the American Society for Information Science and Technology*, 57(3), 431–442.

Yu, B. and Xu, Z. (2008). A comparative study for content-based dynamic spam classification using four machine learning algorithms. *Knowledge-Based Systems*, 21(4), 355–362.

Yu, W., Zheng'ou, W. and Mingchun, W. (2005). WEB text categorization rule extraction based on rough set and decision tree. *Journal of the China Society for Scientific and Technical Information*, 24(6), 674–678.

Yu, B. and Zhu, D. (2009). Combining neural networks and semantic feature space for email classification. *Knowledge-Based Systems*, 22(5), 376–381.

Zeeman, D. and Turner, G. (2006). Resource discovery in the Government of Canada using the Dewey Decimal Classification. *Cataloging & Classification Quarterly*, 42(3/4).

Zeng, M.L. (2008). Knowledge Organization Systems (KOS). *Knowledge Organization*, 35(2–3), 160–182.

Zeng, X. (2005). Research on representation of Chinese classified thesaurus in OWL and its implied semantic reveal. *Journal of the China Society for Scientific and Technical Information,* 24(2), 151–160.

Zhang, Z. (2008). Mining relational data from text: from strictly supervised to weakly supervised learning. *Information Systems*, 33(3), 300–314.

Zhang, W. (2003). Classification for Chinese Libraries (CCL): histories, accomplishments, problems and its comparisons. *Journal of Educational Media and Library Sciences*, 41(1), 1–22.

Zhang, W. (2002). The development and the nature of the Chinese Archive Classification. *Journal of Educational Media and Library Sciences*, 39(3), 235–250.

Zhang, W., Yoshida, T. and Tang, X. (2008). Text classification based on Multi-Word with Support Vector Machine. *Knowledge-Based Systems*, 21(8), 879–886.

Zhao, L. (2004). Save space for 'newcomers' – analyzing problems in book number assignment under the LCC system. *Cataloging and Classification Quarterly*, 38(1), 105–119.

Zhiguo, Z., Guishi, D. and Liping, K. (2009). Algorithm research on classifying Web users navigation patterns based on N-Gram. *Journal of the China Society for Scientific and Technical Information*, 28(3), 389–394.

Zins, C. (2007). Classification schemes of Information Science: twenty-eight scholars map the field. *Journal of the American Society for Information Science and Technology*, 58(5), 645–672.

Zins, C. (2004). Knowledge organization: an epistemological perspective. *Knowledge Organization*, 31(1), 49–63.
Zins, C. (2002). Models for classifying Internet resources. *Knowledge Organization*, 29(1), 20–28.
Zins, C. and Guttmann, D. (2000). Structuring Web bibliographic resources: an exemplary subject classification scheme. *Knowledge Organization*, 27(3), 143–159.

Index