

ORGANIZING INFORMATION
from the shelf to the web

ORGANIZING INFORMATION

from the shelf to the web

G G Chowdhury and Sudatta Chowdhury



facet publishing

© G G Chowdhury and Sudatta Chowdhury 2007

Published by
Facet Publishing
7 Ridgmount Street
London WC1E 7AE
www.facetpublishing.co.uk

Facet Publishing is wholly owned by CILIP: the Chartered Institute of Library and Information Professionals.

G G Chowdhury and Sudatta Chowdhury have asserted their right under the Copyright, Designs and Patents Act 1988 to be identified as authors of this work.

Except as otherwise permitted under the Copyright, Designs and Patents Act 1988 this publication may only be reproduced, stored or transmitted in any form or by any means, with the prior permission of the publisher, or, in the case of reprographic reproduction, in accordance with the terms of a licence issued by The Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to Facet Publishing, 7 Ridgmount Street, London WC1E 7AE.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.

ISBN 978-1-85604-578-0

First published 2007

Typeset from author's disk in 11/14 pt Bergamo and URW Grotesk by Facet Publishing.

Printed and made in Great Britain by Cromwell Press Ltd, Trowbridge, Wiltshire.

To our two wonderful sons

Avirup and Anubhav

Contents

Preface xiii

Acknowledgements xv

Glossary xvii

1 Organizing information: what it means 1

Introduction 1

Information services in today's world 1

Organization of information: what and why? 3

Classification 6

Cataloguing 8

Library approaches to organizing information 9

Summary 14

Review questions 15

References 15

2 Information organization in non-library environments 17

Introduction 17

Organization of electronic information 17

The online database approach 18

The DBMS approach 20

The expert systems approach 22

Organization of information on the internet 23

New approaches to organizing information on intranets
and the web 24

The semantic web 25

Summary 26

Review questions 26

References 27

3 Cataloguing 29

- Introduction 29
- A brief history of library catalogues 29
- What is a library catalogue? 30
- AACR2 32
- The process of cataloguing 32
- Subject access to catalogues 34
- Implications of basic cataloguing rules for OPACs 34
- Cataloguing of internet resources 37
- FRBR (Functional Requirements of Bibliographic Records) 38
- Entities and their relationships 40
- Importance of the FRBR model 42
- RDA: Resource Description and Access 42
- Summary 44
- Review questions 45
- References 45

4 Bibliographic formats: MARC 21 and others 47

- Introduction 47
- Bibliographic formats 47
- ISO 2709: Format for Bibliographic Information Interchange 49
- MARC format 53
- MARC 21 54
- UNIMARC format 60
- CCF 66
- Summary 68
- Review questions 69
- References 69

5 Library classification 71

- Introduction 71
- Library classification: what and why? 71
- Classification schemes 72
- Types of bibliographic classification schemes 75
- Major library classification schemes 78
- Classification of electronic resources 105
- Summary 108

Review questions 109
References 109

6 Subject heading lists and thesauri in information organization 111

Introduction 111
Vocabulary control tools 111
Subject heading lists and thesauri 112
Subject heading lists and thesauri in the organization
of internet resources 122
Summary 126
Review questions 128
References 128

7 Organization of internet information resources 131

Introduction 131
Characteristics of the web 131
Metadata: why not bibliographic formats? 133
Bibliographic classification schemes: are they suitable for
organizing the web? 134
Subject heading lists: can they be used for indexing
internet resources? 135
New tools and standards for managing internet
information 136
The semantic web 136
Summary 137
Review questions 137
References 137

8 Metadata 139

Introduction 139
Metadata: what? 139
Metadata: why? 141
Metadata: types 142
Metadata standards 144
Metadata management 151
Summary 152

Review questions 153
References 153

9 Markup languages 157

Introduction 157
SGML 157
HTML 159
XML 161
Summary 168
Review questions 168
References 168

10 Ontology 171

Introduction 171
Ontology: origin and meaning 172
Ontology, taxonomy and thesauri 173
Some common examples of ontology 175
Ontologies: what do they do? 176
Building an ontology: guidelines and methods 176
Tools for building an ontology 178
Ontology languages: DAML+OIL and OWL 179
Ontology: role in information organization and
management 182
Summary 183
Review questions 184
References 184

11 Information architecture 187

Introduction 187
What is IA? 187
Why do we need an IA? 188
What does IA involve? 189
How does one build an IA? 190
Building an IA: approaches and stages 191
Outcome of an IA exercise 193
Summary 194
Review questions 194
References 195

12 The semantic web 197

Introduction 197

What is the semantic web? 198

How does the semantic web differ from the conventional web? 199

Semantic web technologies 200

URI 201

RDF 202

Semantic web applications 206

Semantic web and information access 207

Summary 210

Review questions 210

References 210

13 Information organization: issues and trends 213

Introduction 213

Cataloguing: FRBR and semantic catalogue networks 214

Metadata 215

Classification in the digital age 217

Ontologies 218

Semantic portals and ontologies 219

Semantic web technologies and digital libraries 220

User-driven classification of web resources 220

Conclusion 221

References 222

Index 225

Preface

Over the past few years, LIS (library and information science) departments have, for various reasons, reduced their teaching of classification and cataloguing – the essential skills, and the corresponding tools and standards, required for organizing information. Today, very few LIS schools and departments in the UK or abroad offer a full and compulsory course module on information organization. The main argument supporting this move was that libraries could easily get ready-made class numbers and catalogue records for bibliographic resources and hence there was no need to emphasize classification and cataloguing. This is not a very strong or appropriate justification. Given the current state of developments in terms of the internet, the web and digital libraries, the need for organizing information has increased significantly.

In a digital library, more specifically in a hybrid library (a library providing both traditional and digital library services), information professionals need to be prepared for organizing both printed and digital information resources, and in many cases ready-made class numbers and catalogue records are not available (particularly for digital information resources). In fact, quite often LIS professionals have to try to find the most effective means of organizing such information resources themselves. In addition, every organization/institution now produces a great deal of digital information, and their day-to-day activities heavily depend on access to, and use of, the appropriate digital information resources. This reinforces the need for appropriate organization of institutional information resources for better access and management. Indeed, many new approaches, tools and techniques have emerged over the past few years for such activities. It is therefore suggested that tomorrow's LIS professionals should be conversant with these, and with the skills for organizing information in all the different domains – from the traditional to the web and digital library shelf.

This book aims to cover the broad spectrum of information

organization in different environments – from print libraries to the internet, intranet and web. It discusses the tools and techniques and the advantages and disadvantages of various approaches to organizing information. Covering every aspect of these is not possible within a single book. However, this book aims to provide a good coverage of the entire spectrum of information organization which will educate tomorrow's information professionals about the entire field, and will lead interested readers to further studies and research by pointing them to the appropriate references.

Chapter 1 provides a broad overview of the various approaches to organizing bibliographic information. Chapter 2 discusses the range of non-library approaches to organizing information, from the database to the expert systems approach, and various approaches to organizing information on the internet and web. Chapters 3 to 6 discuss the traditional tools and techniques for organizing bibliographic information: Chapter 3 discusses cataloguing, Chapter 4 bibliographic formats, Chapter 5 classification and Chapter 6 subject heading lists and thesauri. Although these tools were primarily designed for organizing bibliographic resources, they have been used by many researchers, subject gateways and digital libraries to organize digital information resources. Examples of some such applications have been provided in each of these chapters.

The major drawbacks of the tools and techniques for bibliographic organization (cataloguing, classification and vocabulary control), which make them rather unsuitable or inappropriate for the organization of digital resources, are discussed in Chapter 7. The new tools, techniques and standards that have emerged over the past few years for organizing digital information, especially the internet and web information resources, are introduced in Chapter 7 and discussed in detail in the subsequent chapters.

Chapter 8 introduces metadata and metadata standards such as Dublin Core and eGMS, and discusses the various management issues related to metadata. Chapter 9 considers markup languages and the evolution of markup languages from SGML to HTML, XML and XHTML. It also discusses the role played by these markup languages in the organization and management of digital information.

Chapter 10 discusses ontology, the new tool designed to facilitate semantic access and retrieval and the management of digital information on the web. The importance of ontologies in the context of content management and the semantic web are also considered in this chapter. Information architecture, the field of study concerned with the organization of web and

intranet resources to facilitate the easy access and management of information, is considered in Chapter 11. Chapter 12 discusses the basic concepts of the semantic web – the new vision of the web that will allow computers to understand, retrieve and process information from a variety of sources on the web, primarily on the basis of the semantics, or the meaning, of the information resources concerned. Finally, some major research issues and trends in relation to various aspects of organizing information – traditionally and digitally – are discussed in Chapter 13.

Students in various information science departments – not necessarily only in library and information science programmes but also in a range of other programmes such as information science, information management, information technologies, etc. – will benefit from this book. It will introduce them to the various information organization principles, tools and techniques.

The book can also be used by practising LIS professionals who would like to obtain a complete picture and understanding of how to organize information. If they are interested in any particular area or aspect of information organization, they can follow up the appropriate references that appear at the end of each chapter.

Acknowledgements

In order to illustrate the content of this book, we have included a number of screenshots of different information systems and services, which are acknowledged below.

Figure 5.1: ‘Dewey Services’ <http://connexion.oclc.org> © OCLC
 Online Computer Library Center 11 April 2007
<http://connexion.oclc.org/WebZ/html/corc/corcframe.html?sessionId=cnx07.prod.oclc.org-55890-f0dm5gqm-8kaymd:entitydbname=DeweyDB?>

Figure 5.2: ‘Dewey Services’ <http://connexion.oclc.org> © OCLC
 Online Computer Library Center 11 April 2007
<http://connexion.oclc.org/WebZ/html/corc/corcframe.html?sessionId=cnx07.prod.oclc.org-55890-f0dm5gqm-8kaymd:entitydbname=DeweyDB>

Figure 5.3: ‘Dewey Services’ <http://connexion.oclc.org> © OCLC
 Online Computer Library Center 11 April 2007

<http://connexion.oclc.org/WebZ/DeweyCorcQuery?sessionId=cnx07.prod.oclc.org-55890-f0dm5gqm-8kaymd>

Figure 5.4: 'Dewey Services' <http://connexion.oclc.org> © OCLC Online Computer Library Center 11 April 2007

[http://connexion.oclc.org/WebZ/DeweyCorcQuery?sessionId=cnx07.prod.oclc.org-55890-f0dm5gqm-8kaymd?](http://connexion.oclc.org/WebZ/DeweyCorcQuery?sessionId=cnx07.prod.oclc.org-55890-f0dm5gqm-8kaymd)

Figure 5.5: 'BUBL Information Service' © Centre for Digital Library Research, Strathclyde University, Glasgow G1 1XH, Scotland

Figure 6.5: 'Intute: Health & Life Sciences'

www.intute.ac.uk/healthandlifesciences/browse.html © The Intute Consortium

Figure 6.6: 'Intute: Health & Life Sciences'

www.intute.ac.uk/healthandlifesciences/medicine/mesh_a.html
© The Intute Consortium

Figure 6.7: 'Intute: Social Sciences' www.intute.ac.uk/socialsciences/
© The Intute Consortium

Figure 6.8: 'Intute: Social Sciences' www.intute.ac.uk/socialsciences/cgi-bin/browse.pl?id=120274 © The Intute Consortium

Figures 12.1 and 12.2: RDF model reproduced from the RDF Primer available at www.w3.org/TR/rdf-primer/#conceptsummary

Many individuals and institutions have directly or indirectly helped us in preparing this book. Those whose resources have been used as illustrations are acknowledged with thanks. We are also grateful to the reviewers who made useful comments on earlier drafts of the book.

We are indebted to several members of our family – with special thanks to our two sons Avirup and Anubhav – who have provided encouragement and inspiration throughout this project. Finally we would like to express our gratitude to the staff of Facet Publishing, especially to Helen, Lin and Kathryn for their kind help and support at every stage of our work on the book.

G. G. Chowdhury and
Sudatta Chowdhury

Glossary

AACR2 *Anglo-American Cataloguing Rules*, 2nd edition. This publication provides comprehensive guidelines for the preparation of catalogue entries and is used as a *de facto* standard in the context of library **catalogues**.

analytico-synthetic classification A classification scheme that to some extent is **enumerative** and at the same time makes provision for some sort of synthesis to build the class number. Example:

Universal Decimal Classification (UDC).

array A set of co-ordinate classes (or sub-classes) all derived by applying the same principle of division to the superordinate class.

associative relationship In the context of **vocabulary control**, this denotes the relationship between a pair of terms that are neither **hierarchical** nor **equivalent**, and yet they are associated to such an extent that their relationship should be made explicit.

BC Bibliographic Classification. This classification scheme was originally devised by Henry Evelyn Bliss; the first four volumes (Bliss Bibliographic Classification) were published in the USA between 1940 and 1953. The revised edition (BC2) was initiated by Jack Mills and was produced in 22 parts; further revisions have been made to some of the volumes in order to retain subject currency, and the publication is now undertaken by Bowker.

bibliographic classification A system of organizing information where the subject divisions are assigned a **notation** or a code that indicates subject content. Examples: **Dewey Decimal Classification (DDC)**, **Universal Decimal Classification (UDC)**, **Colon Classification (CC)**, etc.

bibliographic format A standard format that prescribes the various data elements to be used to create a record for a bibliographic item, and thus it governs the process of record creation and exchange. Example: **MARC 21**, **CCF**, etc.

- bibliography** A list of books and other information resources – by or on a given author, in a given subject, etc. – arranged in a specific order.
- BT** Broader term. In the context of the **hierarchical** structure of a **vocabulary control tool** like a **thesaurus** or a **subject heading list**, a broader term denotes a superordinate term, or a term of higher order in the hierarchy in comparison to another term.
- catalogue** A list of, and index to, the information resources in a library or a collection. A catalogue lets the user know details of a library's collection, i.e. it indicates what items the library holds on a given subject, by a given author, etc., and where to locate them.
- catalogue codes** Tools devised to provide rules and guidelines for preparing library catalogues to meet the objectives of **cataloguing**. Example: **AACR2**.
- cataloguing** The act of creating **catalogue** records; sometimes used synonymously (albeit loosely) with **indexing**.
- CC** Colon Classification. Devised by S. R. Ranganathan, this is an example of a **faceted classification** scheme. The seventh edition of Colon Classification, which came out in 1987, provides extreme flexibility in building class numbers of subjects through the principles of **facet** analysis.
- CCF** Common Communications Format. A **bibliographic format** that was developed in order to facilitate the exchange of bibliographic data between organizations; it was first published by UNESCO in 1984, and then in 1988 and 1992.
- class number (or classmark)** A **notational** code, designed to represent the subject content of an information resource, and used to specify its location in a library's collection.
- classification** The process of organizing things in a systematic order. In the context of libraries, classification is a process whose purpose is to (1) specify the location of every bibliographic item on the library's shelves, and (2) display the subject relationships among various bibliographic items in a library's collection.
- classification scheme** A tool for systematic organization of information resources. A classification scheme helps a classifier to represent the subject content of every document in a collection by appropriate **notations**.
- complex subject** A subject formed by the combination of two

concepts, usually from different main classes: e.g. ‘Statistics for librarians’, ‘The influence of the internet on child behaviour’, etc.

compound subject A subject formed by the combination of two or more concepts from the same main class, e.g. ‘Rural economics’, ‘Inorganic acids’, etc.

DAML+OIL DARPA Agent Markup Language and Ontology Interchange Language. DAML+OIL is a joint name for the US DAML-ONT (DAML Ontology) and the European language (OIL). It was taken as the starting point for the W3C Working Group on Web Ontology in defining **OWL**, the standard and broadly accepted **ontology** language of the **semantic web**.

database A collection of logically organized and related data that can be accessed by multiple users to meet their information needs.

DBMS Database management system. A software application system that is used to create, maintain and provide controlled access to databases.

DCMI The Dublin Core Metadata Initiative.

DDC Dewey Decimal Classification. Devised by Melvyl Dewey and first published in 1876, it is currently in its 22nd edition published by **OCLC** in both print and web versions. It is the most widely used library **classification** system in the world.

digital library An information service based on the assemblage of digital content and the appropriate tools, techniques and standards, including the computing, storage and communications machinery and software needed to reproduce, emulate and extend the services provided by conventional print-based libraries.

DTD document type definition. An **SGML** or **XML** application that is used to define the structure of a particular type of document.

Dublin Core A format for **metadata** that defines a core set of 15 data elements that can be used as containers for metadata. The **Dublin Core Metadata Initiative (DCMI)** began in 1995; the first meeting took place at Dublin, Ohio, giving rise to a metadata format called the Dublin Core.

EAD Encoded Archival Description. A **metadata** standard that defines the structural elements and designates the content of descriptive guides to archival and manuscript holdings.

eGMS e-Government Metadata Standard. This lays down the elements, refinements and encoding schemes to be used by UK

government staff when creating **metadata** for information resources.

enumerative classification A **classification** scheme in which all the possible classes are enumerated according to certain characteristics.

Example: **Library of Congress Classification (LC)**.

equivalence relationship In the context of **vocabulary control**, this is a relationship between a pair of terms where one is a preferred term (i.e. an index entry can be made under that term) and another a non-preferred term (i.e. an index entry cannot be made under that term; instead a reference entry is made to direct the user to the preferred term).

expert systems Computer systems that embody knowledge about a specific problem domain and can solve problems from the domain using its knowledge with a degree of expertise that is comparable to that of a human expert.

facet A generic term used to denote a component of a subject.

faceted classification A **classification** scheme that, instead of listing all the classes and the corresponding numbers, lists the various **facets** of every subject or main class and provides a set of rules for constructing class numbers through facet analysis.

folksonomy Also known as ethnoclassification. An internet-based information organization method consisting of collaboratively generated, open-ended labels that categorize content such as web pages, online photographs and web links.

FRBR Functional Requirements of Bibliographic Records. Developed under the auspices of **IFLA**, this is a framework for relating the data elements in bibliographic records to the needs of the users of those records.

fundamental categories Proposed by S. R. Ranganathan, these are the building blocks of **facet** analysis. According to Ranganathan, the constituent components of any compound subject can be the manifestation of one of the five fundamental categories: Personality, Matter, Energy, Space and Time, famously known as PMEST.

hierarchical relationship In the context of **vocabulary control** this is a relationship where a pair of terms are represented in their superordinate-subordinate status to show whole-part, thing-kind, genus-species or item-instance relationships.

hospitality In the context of **classification**, this is a property of

notation that denotes the capacity of a classification scheme to accommodate new subjects, classes or sub-classes.

HTML HyperText Markup Language. HTML is a formal recommendation by the **World Wide Web Consortium (W3C)** and is generally adhered to by the major browsers like Internet Explorer and Netscape Navigator. HTML markup tells the web browser how to display a web page's content – text, image, etc. – for the user.

HTTP HyperText Transfer Protocol.

IA Information architecture. An area of study that is concerned with the organization and labelling of websites, intranets, online communities and software to support usability and findability.

ICCP International Conference on Cataloguing Principles. Held in Paris, 9–18 October 1961, and also known as the **Paris Principles**.

IFLA International Federation of Library Associations and Institutions.

indexing The process of creating an index, i.e. access points for information resources; sometimes used synonymously (albeit loosely) with cataloguing.

information retrieval Covers all activities related to the organization, processing and access to information in all forms and formats.

ISBD International Standard Bibliographic Description.

LC Library of Congress Classification. This was mainly constructed for the Library of Congress (LC) from 1901 onwards, but is also used widely in other libraries throughout the world. It is an **enumerative** classification and provides a long list of all classes of subjects in the universe.

LCSH Library of Congress Subject Headings. This was originally designed as a controlled **vocabulary** for representing the subject and form of the books and serials in the Library of Congress collection, with the objective of providing subject access points to the bibliographic records contained in the library's catalogues. It is now the most widely used tool for assigning **subject headings** to bibliographic information resources.

library catalogue An organized set of records of the bibliographic information resources in a library's collection.

library classification *see* **Bibliographic classification**.

LIS Library and information science.

literary warrant In the context of information organization, this is a concept that suggests that a new class or heading will be introduced

in a tool to represent a topic only when there is a demand, i.e. when an information resource exists that embodies the topic concerned.

main class In the context of **classification**, this is one of the primary divisions produced by a **classification scheme**.

main entry In the context of **cataloguing**, this is a **catalogue** record that contains the complete catalogue entry for the bibliographic item concerned.

MARC MACHine-Readable Catalogue.

MARC 21 A **bibliographic format** that provides a set of codes and content designators and guidelines for managing and formatting electronic records of information resources.

metadata Structured data about information resources (mainly in digital form) that can be used to help support a wide range of operations, such as resource description and discovery, and management of information resources and their long-term preservation.

metadata standard Metadata formats designed for creating **metadata** for materials in a specific domain or for materials of a specific kind and format, etc.

mnemonics In the context of **classification** this is a property of **notation** that acts as an aid to memory; a mnemonic notation should be easy to remember.

notation A symbol that is used to represent a subject or a class in a **classification** scheme.

NT Narrower term. In the context of the **hierarchical** structure of a **vocabulary control tool** like a **thesaurus** or a **subject heading list**, a narrower term denotes a subordinate term, or a term of lower order in the hierarchy in comparison to another term.

OAI Open Archives Initiative.

OCLC Online Computer Library Center.

online databases Traditional online databases and online searching denote information searches that are conducted by means of a local computer that communicates with a remote computer system containing bibliographic and/or full-text databases. It is also used as a generic term to denote any **databases** of information resources.

ontology A formal, explicit specification of a shared conceptualization. A formal ontology is a controlled **vocabulary** expressed in an ontology representation language that has a grammar and formal constraints specifying how the vocabulary terms can be used.

OPAC Online public access catalogue.

OWL Web Ontology Language.

page layouts Also called wireframes, blueprints or screen details; they define page level navigation, content types and functional elements and are useful in conveying the general page structure and content requirements for individual pages on a website.

page templates In the context of **information architecture**, these are templates that are used to define large-scale websites and intranets; they define the layout of common page elements such as navigation and content.

Paris Principles These define the objectives of a library **catalogue**. The objectives of **cataloguing** proposed by C. A. Cutter in 1876 were modified by Lubetzky and were accepted internationally at the International Conference on Cataloguing Principles held in Paris in 1961; they are known as the **ICCP** Paris Principles.

personas In the context of **information architecture**, these are hypothetical archetypes for actual users created through a series of interviews of the users of a system.

phase relations Relationships between the components of a **complex subject**.

post-coordinate indexing An **indexing** system where one entry is prepared for each keyword selected to represent the subject of a given document, and all the entries are organized in a file; when a user submits a query, individual query terms are matched against the file and documents with matching terms are retrieved.

pre-coordinate indexing An **indexing** system where candidate keywords for every document are chosen and arranged (coordinated) in a specific order to create the index; when a user submits a query, the query terms are matched with the pre-coordinated index strings for matching and retrieval. **Subject heading lists** are examples of pre-coordinate indexing systems.

prototypes In the context of **information architecture**, these are quick designs to elicit user feedback and identify problems, if any exist.

RDBMS Relational database management system. This is a **DBMS** that stores data in the form of tables where the rows represent database objects (records) and columns the characteristics or attributes of those objects. Examples: Access, Oracle.

RDF Resource description framework.

RDF schema This provides a model for organizing web objects into **hierarchies** by supplying properties, subproperty and subclass relationships, and domain and range restrictions.

RT Related term. In the context of the **hierarchical** structure of a **vocabulary control tool** like a **thesaurus** or a **subject heading list**, a related term denotes a co-ordinate term, or a term of same order in the hierarchy in comparison to another term.

search engine An **information retrieval** system designed to provide access to electronic information in standalone or networked computer systems, especially on the web.

semantic web A term coined by the creator of the web, Tim Berners-Lee, to describe a process whereby data on the web will be defined and linked in such a way that it can be used by machines not only for display purposes, but for the processing, integration and re-use of data across various applications.

semantic web portal A web portal based on **semantic web** technology, especially **ontology**.

SGML Standard Generalized Markup Language. This is a standard markup language for marking text documents so that they can be processed by computers independently of software and hardware differences.

site maps In the context of **information architecture**, these are high level diagrams showing the **hierarchy** of a system that reflects the information structure of an organization.

storyboards In the context of **information architecture**, these are sketches showing how a user would interact with a system to complete a common task.

subject heading list An alphabetical list of terms and phrases, with appropriate cross-references and notes, that can be used as a source of headings in order to represent the subject content of an information resource. Example: **Library of Congress Subject Headings (LCSH)**.

subject indexing The process of creating an **index** based on the subject content of information resources.

taxonomy A collection of controlled **vocabulary** terms organized into a **hierarchical** structure where each term is in one or more parent-child relationships to other terms in the collection.

- TEI** Text Encoding Initiative. TEI guidelines for Electronic Text Encoding and Interchange were first published in 1994; they specify a set of tags which may be inserted to mark the textual structure and other features of interest in the text.
- thesaurus** An organized set of controlled **vocabulary** terms showing synonyms, and **hierarchical** and other relationships and dependencies that are used for the indexing and retrieval of information.
- UDC** Universal Decimal Classification. First developed in 1905, this **analytico-synthetic classification** scheme is now one of the most widely used **classification** schemes in the world.
- UNIMARC** UNIversal MARC. In 1977 IFLA brought out the UNIMARC format, whose primary purpose is to facilitate the international exchange of data in machine-readable form between national bibliographic agencies.
- URI** Uniform resource identifier. This is a generic term for all types of names and addresses that refer to information objects on the world wide web.
- URL** Uniform resource locator; simply speaking, an address of a web resource.
- URN** URN, or uniform resource name, is a URI that identifies a resource by a name. As opposed to URL, it simply gives the name of the resource – but does not say how to locate or obtain it.
- VLE** Virtual learning environment.
- vocabulary control tool** An organized list of terms and phrases that can be used to index and search a collection by subject terms and phrases. **Subject heading lists**, e.g. the **Library of Congress Subject Headings (LCSH)**, and **thesauri**, e.g. the UNESCO Thesaurus, are examples of **vocabulary control** tools.
- W3C** World Wide Web Consortium.
- WebDewey** The web version of the **Dewey Decimal Classification (DDC)** scheme.
- XML** eXtensible Markup Language.
- XML schema** A language for describing the structure and constraints to which **XML** documents must conform.

