BRITISH STANDARD

BS 8723-1:2005

Structured vocabularies for information retrieval — Guide —

Part 1: Definitions, symbols and abbreviations

ICS 01.140.20

BSI British Standards

Committees responsible for this British Standard

The preparation of this British Standard was entrusted by Technical Committee, IDT/2, Information and documentation, to Subcommittee IDT/2/2, Indexes, filing and thesauri, upon which the following bodies were represented:

British Association for Information and Library Education and Research

CAB International

CILIP — Chartered Institute of Library and Information Professionals

INSPEC

Society of Indexers

Stationery Office

Co-opted members

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 1 November 2005

First published November 2005

Amendments issued since publication

Amd. No.	Date	Comments

The following BSI references relate to the work on this British Standard:
Committee reference IDT/2/2
Draft for comment 04/30086620

ISBN 0 580 46798 8

Contents

		Page
Committees responsible		Inside front cover
Foreword		ii
Introduction		1
1	Scope	1
2	Definitions	1
3	Symbols, abbreviations and other conventions	6
Bibliography		9
Table 1 — Symbols and abbreviations		7

Foreword

This part of BS 8723 has been prepared by Subcommittee IDT/2/2. Together with BS 8723-2 it supersedes BS 5723:1987, which is withdrawn. When published, BS 8723-4 will revise and supersede BS 6723:1985 which will then be withdrawn. The other parts of BS 8723 will cover new scope, not previously published in a British Standard.

The other parts of BS 8723 are as follows:

—*Part 2: Thesauri;*

—*Part 3: Vocabularies other than thesauri*[1)];

—*Part 4: Interoperation between vocabularies*[1)];

—*Part 5: Interoperation between vocabularies and other components of information storage and retrieval systems*[1)].

A single index is provided, published with BS 8723-2.

Part 2 covers thesauri, designed for situations in which human indexers analyse documents and express their subjects using thesaurus terms, before searchers retrieve the documents with the same vocabulary.

Part 3 covers other types of structured vocabulary, including those for applications such as automatic categorization schemes and free-text search aids.

Part 4 applies to situations in which more than one language or vocabulary is in use, but access to all resources is needed through the one vocabulary chosen by the user.

Part 5 sets out the protocols and formats needed for the exchange of vocabulary data.

Part 1 together with part 2 of BS 8723 correspond broadly to ISO 2788-1986, since this is equivalent to the withdrawn BS 5723. Part 1 together with part 4 will correspond to ISO 5964:1985 (equivalent to BS 6723:1985), while extending the scope of the latter considerably. ISO 5964:1985 will remain current as an international Standard.

BS 8723-1 covers the development and maintenance of structured vocabularies rather than how to use them in indexing, which is covered by BS ISO 999.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

Compliance with a British Standard does not of itself confer immunity from legal obligations.

[1)] In preparation.

Summary of pages

This document comprises a front cover, an inside front cover, pages i and ii, pages 1 to 9 and a back cover.

The BSI copyright notice displayed in this document indicates when the document was last issued.

Introduction

Today's information retrieval systems are able to provide many facilities. As well as thesaural guidance they may offer free text searching, filtering by date or by form, automatic clustering and categorization, automatic summarization, identification of names of people and organizations, etc. While aiming to mesh with that diversified environment, this standard focuses on retrieval by subject. Within that still broad scope, it is particularly concerned with the controlled vocabulary approach to indexing and retrieval.

The precursor standards BS 5723 and BS 6723 were concerned only with thesauri. Furthermore, they were developed in a largely paper-based era. BS 8723 acknowledges that the environment has changed and that new needs have emerged. Today's thesauri are mostly electronic tools; they are built and maintained with the support of software and need to integrate with other software, such as search engines and content management systems. Whereas in the past thesauri were designed for information professionals trained in indexing and searching, today there is a demand for vocabularies that untrained users will find to be intuitive. There is also a need for search aids in contexts where "full text" is not available, such as museum collections and image databases. As the Internet and other networks allow simultaneous searching across resource collections that have been indexed using different vocabularies, there is a need to have the means of "translating" search queries across boundaries.

BS 8723 still gives most attention to thesauri (part 2), including their use within computerized systems. It describes, in less detail (part 3), some of the other types of vocabulary with which thesauri may need to interface. Part 4 deals with interoperability between different vocabularies — either different language versions of one multilingual thesaurus, or multiple vocabularies in one natural language. Part 5 presents the formats and protocols needed for data exchange in support of thesauri and other structured vocabularies.

1 Scope

This part of BS 8723 sets out the definitions, symbols and abbreviations and other conventions for information retrieval applications of structured vocabularies.

The BS 8723 series gives recommendations for the development and maintenance of structured vocabularies of subject terms and concepts for information retrieval. It also gives guidance on the exchange of data between multiple vocabularies and between a vocabulary and the software with which it interacts.

BS 8723 applies to vocabularies used for retrieving information from text resources, such as knowledge bases, bibliographic databases or collections of full text documents, and also catalogues of images or artefacts, such as those in a museum collection. Structured vocabularies are used in post-coordinate retrieval systems, hierarchical directories, pre-coordinate indexes and classification systems.

BS 8723 does not apply to the preparation of back-of-the-book indexes, although many of its recommendations may be useful for that purpose. It does not cover the structure of authority files of proper names of people, organizations and places, while recognizing that these may sometimes interact with vocabularies of subject terms.

2 Definitions

For the purposes of this British Standard, the following definitions apply. Where a term used in a definition is also defined in this clause, the term is italicized.

2.1
array
group of *sibling terms*

EXAMPLE In the following, outerwear and underwear are *sibling terms* in the same array.

clothing
  outerwear
    overcoats
  underwear

2.2
characteristic of division
attribute by which a concept can be subdivided into an array of narrower concepts each having a distinct value of that attribute

EXAMPLE In the following, age group is the characteristic by which the concept of people is divided:

people
(people by age group)
children
youths
adults

2.3
classification
grouping together of similar or related things and the separation of dissimilar or unrelated things and the arrangement of the resulting groups in a logical and helpful sequence

2.4
classification scheme
schedule of *concepts*, arranged by *classification*

NOTE A classification scheme may also include an index.

2.5
complex concept
concept that combines two or more simpler *concepts*

EXAMPLE human resource management combines the idea of people with their usefulness as resources requiring management.

NOTE Complex concepts are sometimes expressed in a single word, but are more often conveyed by a *multi-word term*.

2.6
concept
unit of thought

NOTE Concepts can often be expressed in a variety of different ways. They exist in the mind as abstract entities independent of *terms* used to express them.

2.7
controlled vocabulary
prescribed list of *terms* or headings each one having an assigned meaning

NOTE Controlled vocabularies are designed for use in classifying or *indexing documents* and for searching them.

2.8
document
item that can be classified or indexed in order that it may be retrieved

NOTE This definition refers not only to written and printed materials in paper or microform versions (for example, books, journals, diagrams, maps), but also to non-printed media, machine-readable and digitized records, Internet and intranet resources, films, sound recordings, people and organizations as knowledge resources, buildings, sites, monuments, three-dimensional objects or realia; and to collections of such items or parts of such items.

2.9
enumerative classification scheme
classification scheme in which all the *concepts* available for use are listed in the *schedules*

NOTE Compare with *synthetic classification scheme*.

2.10
facet
grouping of *concepts* of the same inherent category

EXAMPLE 1 Animals, mice, daffodils and bacteria could all be members of a living organisms facet.

EXAMPLE 2 Digging, writing and cooking could all be members of an activities facet.

EXAMPLE 3 Paris, the United Kingdom and the Alps could all be members of a places facet.

NOTE Examples of categories that may be used for grouping *concepts* into facets are: activities, disciplines, people, materials, living organisms, objects, places and times.

2.11
facet analysis
analysis of subject areas into constituent *concepts* grouped into *facets*

2.12
facet indicator
notational device that indicates the start of a new *facet* within a synthesized compound classmark

NOTE Examples of facet indicators are the 0 in the Dewey Decimal Classification [1], and parentheses and quotation symbols in the Universal Decimal Classification [BS 1000M]. In the past the term *facet indicator* has been used as synonymous with *node label* but that usage is deprecated by this standard, to avoid confusion.

2.13
faceted classification scheme
classification scheme in which subjects are analysed into their constituent *facets*

NOTE *Schedules* are compiled for each *facet*, and *terms* or *notations* from these may be combined according to prescribed rules to express a *complex concept*.

2.14
homograph
one of two or more words that have the same spelling, but different meanings

EXAMPLE The word "bank" could refer to a financial institution or the side of a river.

2.15
indexing
intellectual analysis of the subject matter of a *document* to identify the *concepts* represented in it, and allocation of the corresponding *preferred terms* to allow the information to be retrieved

NOTE The term "subject indexing" is often used for this *concept*, but as BS 8273 does not deal with the indexing of other elements such as authors or dates, "indexing" is sufficient.

2.16
loan term
term borrowed from another language that has become accepted in the borrowing language

EXAMPLES glasnost, gourmets.

2.17
monohierarchical structure
hierarchical arrangement of *concepts*, in a thesaurus or classification scheme, in which each *concept* can have only one broader *concept*

EXAMPLE In a monohierarchical structure, the *concept* pianos cannot be listed under keyboard instruments as well as under stringed instruments; a choice has to be made of one of these *concepts* to determine its placing.

NOTE Compare with *polyhierarchical structure*. In a monohierarchical structure, each *concept* can occur at only one place in the hierarchy and other broader term relationships have to be shown as related term relationships.

2.18
multi-word term
term consisting of more than one word

EXAMPLE human resource management.

NOTE Multi-word terms typically label *complex concepts* and are admissible in a *thesaurus* as *preferred terms*.

2.19
node label
label inserted into a hierarchical or classified display to show how the terms have been arranged

NOTE A node label contains one of two different types of information: either a) the name of a *facet* to which following *terms* belong; or b) the attribute or *characteristic of division* by which an *array* of *sibling terms* has been sorted or grouped. See examples in BS 8723-2:2005, Clause **9**.

2.20
non-preferred term
term that is not assigned to documents but is provided as an entry point in a thesaurus or alphabetical index

EXAMPLE

hounds
USE dogs

NOTE A non-preferred term is followed by a reference to the appropriate *preferred term* or *preferred terms*. Non-preferred terms are sometimes known as lead-in terms, entry terms or non-descriptors.

2.21
notation
set of symbols representing *concepts* or relations

EXAMPLE showing notation in the left-hand column:

H000 photography
H001 . . photographic equipment
H002 . . . camera accessories
H003 . . . . flash guns
H004 . . . . tripods
H005 . . . cameras and camera components
H006 . . . . camera components
H007 . . . . . camera lenses
H008 . . . . . camera viewfinders
H009 . . . . cameras

NOTE A notation may be used to sort and/or locate *concepts* in a pre-determined systematic order, and optionally to display how *concepts* have been structured and grouped. A *notation* can provide the link between alphabetical and systematic lists in a *thesaurus*.

2.22
paradigmatic relationship
relationship between *concepts* which is inherent in the *concepts* themselves

NOTE Such relationships are shown in a *structured vocabulary*, independently of any indexed *document*. For a more complete discussion of paradigmatic and *syntagmatic relationships*, see BS 8723-2:2005, **8.1.1**. A paradigmatic relationship is sometimes known as an a priori relationship.

2.23
polyhierarchical structure
hierarchical arrangement of *concepts*, in a *thesaurus* or *classification scheme*, in which each *concept* can have more than one broader *concept*

EXAMPLE In a polyhierarchical structure, organs (musical instruments) may be listed under keyboard instruments as well as under wind instruments.

NOTE Compare with *monohierarchical structure*. In a polyhierarchical structure, a single *concept* can occur at more than one place in the hierarchy. Its attributes and relationships, and specifically its narrower and related *terms*, are the same wherever it occurs.

2.24
post-coordinate indexing
system of *indexing* in which the subject of a *document* is analysed into its constituent *concepts* by an indexer but the *preferred terms* so allocated are not combined until they are selected by a user at the search stage

2.25
pre-coordinate indexing
system of *indexing* in which the *preferred terms* allocated to a particular *document* are syntactically combined in one or more sequences representing the only combinations available for retrieval purposes

2.26
preferred term
term used to represent a *concept*

EXAMPLES schools; school uniform; costs of schooling; teaching.

NOTE A preferred term should preferably be a noun or noun phrase. Preferred terms are sometimes known as descriptors.

2.27
quasi-synonym
one of two or more *terms* whose meanings are generally regarded as different in ordinary usage but which may be treated as labels for the same *concept*, for the purposes of a given *controlled vocabulary*

EXAMPLES diseases, disorders; earthquakes, earth tremors.

2.28
schedule
terms, *notations*, cross-references and *scope notes* set out to exhibit the content and structure of a *structured vocabulary*

2.29
scope note
note which defines or clarifies the meaning of a *concept* as it is used in the *structured vocabulary*

NOTE A *descriptor* used to label a *concept* may have several meanings in normal usage. A scope note may restrict the *concept* to only one of these meanings, and may refer to other *concepts* that are incuded or excluded from the scope of the *concept* being defined.

2.30
search thesaurus
vocabulary intended to assist searching even though it has not been used to index the *documents* being searched

NOTE Search thesauri are designed to facilitate choice of *terms* and/or expansion of search expressions to include *terms* for broader, narrower or related *concepts*, as well as *synonyms*. Optionally, a normal *thesaurus* may be used as a search thesaurus.

2.31
sibling term
one of two or more *terms* with the same immediate broader *term*

EXAMPLE In the following, outerwear and underwear are sibling terms in the same *array*:

clothing
  outerwear
    overcoats
  underwear

2.32
specificity
capability of a *structured vocabulary* to express a subject in depth and in detail

NOTE Specificity has an important influence on retrieval performance, as it determines the accuracy with which *concepts* may be pinpointed, and consequently the facility to exclude unwanted *documents*.

2.33
structured vocabulary
set of *terms*, headings or *concept* codes and their inter-relationships which may be used to support information retrieval

NOTE A structured vocabulary may also be used for other purposes. In the context of information retrieval, the vocabulary should be accompanied by rules for how to apply the *terms*.

2.34
subject heading list
list of *terms* used to represent the subjects of *documents*

NOTE The principles for constructing subject heading lists differ from the principles of *thesaurus* construction. Subject heading lists may have provision for the construction of *pre-coordinated indexing* strings including headings and one or more levels of subheading.

2.35
synonym
one of two or more *terms* whose meanings are considered to be the same in a wide range of contexts

EXAMPLES HIV, human immunodeficiency virus; guarantees, warranties.

NOTE Abbreviations and their full forms may be treated as synonyms.

2.36
syntagmatic relationship
relationship between concepts that exists only because they occur together in a *document* being indexed

NOTE Such relationships are not generally valid in contexts other than the *document* being indexed, and therefore they do not form part of the structure of a *thesaurus*. For a more complete discussion of syntagmatic and *paradigmatic relationships*, see BS 8723-2:2005, **8.1.1**. A syntagmatic relationship is sometimes known as an a posteriori relationship.

2.37
synthetic classification scheme
classification scheme in which users can synthesize *terms* or *notation* for *complex concepts* from lists of simpler *concepts*

NOTE Compare with *enumerative classification scheme*.

2.38
term
word or phrase used to label a *concept*

NOTE Terms in a **thesaurus** can be either *descriptors* or *non-descriptors*.

2.39
thesaurus
controlled vocabulary in which *concepts* are represented by *preferred terms*, formally organized so that *paradigmatic relationships* between the *concepts* are made explicit, and the *preferred terms* are accompanied by lead-in entries for *synonyms* or *quasi-synonyms*

NOTE The purpose of a thesaurus is to guide both the indexer and the searcher to select the same *preferred term* or combination of *preferred terms* to represent a given subject.

2.40
vocabulary control
restriction of choice of indexing *terms* to those in a specified list

NOTE This restriction increases the likelihood of indexers and searchers choosing the same *term* to label a *concept*.

3 Symbols, abbreviations and other conventions

3.1 For the purposes of BS 8723, the symbols, abbreviations and conventions in Table 1 apply. The table shows the abbreviations which are used in examples of thesaurus records as tags to prefix terms and notes. Each tag indicates the relationship or function of the term or note which follows.

Table 1 — Symbols and abbreviations

Symbol	Tag	Meaning
Descriptive elements		
	SN	Scope note
	DEF	Definition
	HN	History note
Codes		
	CC	Classification code or notation
Relationships		
→	**USE**	Use; the term that follows the tag is the preferred term that should be used in place of the non-preferred term that precedes the tag
=	**UF**	Use for; the term that follows the tag is a non-preferred term for which the preferred term preceding the tag should be used instead
	USE+	The two or more preferred terms following the tags should be used together to represent the indicated concept
	UF+	The non-preferred term that follows should be represented by a combination of preferred terms including the preferred term that precedes the tag
	TT	Top term; the preferred term that follows the tag represents the broadest concept in a hierarchy to which the specific concept belongs
<	**BT**	Broader term; the term that follows the tag represents a concept having a wider meaning
	BTG	Broader term (generic)
	BTI	Broader term (instantial)
–<	**BTP**	Broader term (partitive)
>	**NT**	Narrower term; the term that follows the tag refers to a concept with a more specific meaning
	NTG	Narrower term (generic)
	NTI	Narrower term (instantial)
>–	**NTP**	Narrower term (partitive)
—	**RT**	Related term; the term that follows the tag is associated, but is not a synonym, a quasi-synonym, a broader term or a narrower term

3.2 The abbreviations listed as tags in the second column of Table 1 have acquired status as generally recognized conventions, and they occur in many published thesauri. They have mnemonic value, but apply only in the English language. A corresponding set of language-specific abbreviations is well established in many other languages, including French and German. If a more neutral system of abbreviation is required, the language-independent symbols in the left-hand column are recommended.

3.3 The following conventions are also used in examples throughout BS 8273.

a) Preferred terms are printed in the case they would normally have in full text. Typically that means lower case throughout, but upper case for abbreviations and acronyms, and initial capital letters for proper names. A different typeface is used to distinguish the examples from the text of the standard.

EXAMPLE

cars
animals

b) Non-preferred terms are printed in italics.

EXAMPLE

cars
UF *automobiles*
animals
UF *fauna*
World Health Organization
UF *WHO*

Bibliography

Standards publications

BS 1000M (all parts), *Universal decimal classification — International medium edition — English text*.

BS ISO 999, *Information and documentation — Guidelines for the content, organization and presentation of indexes*.

Other documents

[1] OCLC (Online Computer Library Center). *Dewey Decimal Classification — 22nd edition*. Dublin, Ohio: OCLC, 2003. ISBN 0-910608-70-9 (4-volume set).

BSI — British Standards Institution

BSI is the independent national body responsible for preparing British Standards. It presents the UK view on standards in Europe and at the international level. It is incorporated by Royal Charter.

Revisions

British Standards are updated by amendment or revision. Users of British Standards should make sure that they possess the latest amendments or editions.

It is the constant aim of BSI to improve the quality of our products and services. We would be grateful if anyone finding an inaccuracy or ambiguity while using this British Standard would inform the Secretary of the technical committee responsible, the identity of which can be found on the inside front cover. Tel: +44 (0)20 8996 9000. Fax: +44 (0)20 8996 7400.

BSI offers members an individual updating service called PLUS which ensures that subscribers automatically receive the latest editions of standards.

Buying standards

Orders for all BSI, international and foreign standards publications should be addressed to Customer Services. Tel: +44 (0)20 8996 9001. Fax: +44 (0)20 8996 7001. Email: orders@bsi-global.com. Standards are also available from the BSI website at http://www.bsi-global.com.

In response to orders for international standards, it is BSI policy to supply the BSI implementation of those that have been published as British Standards, unless otherwise requested.

Information on standards

BSI provides a wide range of information on national, European and international standards through its Library and its Technical Help to Exporters Service. Various BSI electronic information services are also available which give details on all its products and services. Contact the Information Centre. Tel: +44 (0)20 8996 7111. Fax: +44 (0)20 8996 7048. Email: info@bsi-global.com.

Subscribing members of BSI are kept up to date with standards developments and receive substantial discounts on the purchase price of standards. For details of these and other benefits contact Membership Administration. Tel: +44 (0)20 8996 7002. Fax: +44 (0)20 8996 7001. Email: membership@bsi-global.com.

Information regarding online access to British Standards via British Standards Online can be found at http://www.bsi-global.com/bsonline.

Further information about BSI is available on the BSI website at http://www.bsi-global.com.

Copyright

BSI
389 Chiswick High Road
London
W4 4AL