and errors: a case study of novice
online search – Part 1 *Online*, **15**
indows based CD-ROMs: an early
-88.
n Interface Project – final report.
ign on CD-ROMs. *Online and CD-*
. Aldershot: Gower.
n: LITC.
study of information seeking and
nce, **39** (3) 161–96.
whither? *Journal of the American*
n retrieval. *Annual Review of In-*
Effective Human–Computer Inter-
, CO: Libraries Unlimited.
achieve more effective online help
published).
McGraw Hill.
keley, CA and London: Osborne
gn. *Journal of Documentation*, **49**
ibraries. In R. Biddiscombe (ed.),
ole of the Information Professional,
: user interfaces to online public
Science, **42** (2), 78–98.
ieval: a longitudinal study. *Journal*

5 Indexing and searching languages

INTRODUCTION

This chapter introduces a range of principles associated with the concept and use of indexing and searching languages. The primary focus of this chapter is on indexing languages that are used to represent subjects. Accordingly the chapter starts by exploring the concept of a subject. Next the types of indexing languages are introduced and, finally, a range of principles that apply to the application and use of indexing languages. At the end of this chapter you will:

- recognize the complexities associated with naming subjects and the need to identify relationships between subjects
- be familiar with different approaches to indexing and with associated concepts such as specificity and exhausitivity
- understand the difference between natural and controlled languages
- be aware of the principles of vocabulary control
- understand the principles of thesaurus construction
- be familiar with the search facilities that are available to support post-coordinate searching.

APPROACHES TO SUBJECT RETRIEVAL

Users often approach information sources not with names in mind, but with a question that requires an answer or a topic for study. Users seek documents or information concerned with a particular subject. This is a common approach to information sources and, in order to provide for it, the document or document

are retrieved.

What is a specific subject? A rabbit is a rabbit; but is it? Europeans will have in mind the European rabbit, Americans the cottontail; they belong to different genera. A rabbit is a concrete entity – that is, we can see it and pick it up (preferably not by its ears) and define it by its physical characteristics (long ears, furry, weighs around a kilogram) and behaviour (hopping movement, digs burrows, breeds freely). Abstract concepts can be more difficult to pin down. Some are fairly straightforward, like Music (encyclopaedia definition: 'the organized movement of sounds through a continuum of time'); some, like Geography ('the science that deals with the distribution and arrangement of all the elements of the earth's surface') look straightforward until we think of the vast scope of the subject; while Games defies definition – the philosopher Wittgenstein concluded after a long study that the subject could only be defined through its examples. Not only may subjects be in themselves difficult to define, we must remember that they do not exist in isolation in the way that named entities do. If we are looking for information on William Shakespeare, Mount Everest or Microsoft, we can be sure when we have found it that we have come to the right place as these are all 'classes of one'. Common subjects, on the other hand, form networks of conceptual relationships with other subjects. If we are trying to identify a rabbit, we may not be entirely sure that what we saw was not a hare; the reader in a library looking for the geography books is likely to be directed to at least three widely separated sets of shelves; and library readers everywhere are notorious for asking for the games section when what they are looking for is information on chess. Any system of subject retrieval must then have a mechanism for directing users to other, closely related, subjects.

WHY INDEX?

Now that cheap online storage and retrieval of full text are commonplace, the value of indexing has been questioned. If the individual words in a text are immediately accessible in any combination, why go to the trouble of constructing indexes at all? Why not simply search for combinations of words from the text? The assumption behind this attitude is that a text is 'about' what it mentions. Fairthorne put his finger on the weakness of this assumption:

> *Moby Dick* is about a whale, *Othello* is about a handkerchief, and about other things. The difficulties are to identify which of the things mentioned refer to relevant topics, and how to deal with topics of the document that are not mentioned explicitly... Parts of the document are not always what the entire document is about, nor is a document usually about the sum of the things it mentions.
>
> (Fairthorne, 1969, p. 79)

but nobody would suggest that ensure that a document's overal are adequately represented.

WHAT IS A DOCUMENT ABOUT?

Information retrieval is in gene than what it *means*. What is *mea* is inferential: a scientific paper tobacco smoking and lung canc other. Another argument state and not attempt to impose upc means. There is also the poin meaning is interactive (and to th between the text and the individ against indexers attempting to r it would simply take too long to is about by scanning it rapidly. a far closer study, as well as re

APPROACHES TO INDEXING

An indexing language can be d as access points in an index. A that are used by a searcher whe codes are assigned by an indexe language is used in indexing. access points to records during distinct from the searching lang two must be closely related. In controlled-indexing languages languages or derived-term syste

Controlled-indexing languages (a

With these languages a person Controlled-indexing languages much emphasis is placed upo *Normally* an authority list *ident* involves a person assigning ter

:hat items on specific subjects

)ut is it? Europeans will have
ıtail; they belong to different
'e can see it and pick it up
)hysical characteristics (long
)ur (hopping movement, digs
: more difficult to pin down.
)paedia definition: 'the organ-
time'); some, like Geography
'angement of all the elements
we think of the vast scope of
hilosopher Wittgenstein con-
only be defined through its
ı difficult to define, we must
way that named entities do.
kespeare, Mount Everest or
.t that we have come to the
ı subjects, on the other hand,
ıer subjects. If we are trying
that what we saw was not a
ɡraphy books is likely to be
shelves; and library readers
section when what they are
f subject retrieval must then
ɜely related, subjects.

l text are commonplace, the
lividual words in a text are
:o *the trouble of constructing*
ions of words from the text?
is 'about' what it mentions.
ıssumption:

erchief, and about other things.
ntioned refer to relevant topics,
ıre not mentioned explicitly . . .
re document is about, nor is a
.tions.

(Fairthorne, 1969, p. 79)

In other words, this paragraph has just mentioned a whale and a handkerchief, but nobody would suggest that it is *about* those things. It is the indexer's job to ensure that a document's overall topic and, perhaps, its major constituent themes are adequately represented.

WHAT IS A DOCUMENT ABOUT?

Information retrieval is in general concerned with what a text is *about* rather than what it *means*. What is meaning? One point of view holds that that meaning is inferential: a scientific paper may be *about* a statistical correlation between tobacco smoking and lung cancer; what it *means* is that the one may cause the other. Another argument states that indexers should adopt a neutral position and not attempt to impose upon the reader their views on what a document means. There is also the point of view – grounded in literary theory – that meaning is interactive (and to that extent subjective), the result of the interaction between the text and the individual reader. Perhaps the most powerful argument against indexers attempting to represent the meaning of documents is economic: it would simply take too long to do. A trained indexer can grasp what a document is about by scanning it rapidly. To attempt to extract its meaning would involve a far closer study, as well as requiring expert subject knowledge.

APPROACHES TO INDEXING

An indexing language can be defined as the terms or codes that might be used as access points in an index. A searching language can be defined as the terms that are used by a searcher when specifying a search requirement. If the terms or codes are assigned by an indexer when a database is created, then the indexing language is used in indexing. The same terms or codes may also be used as access points to records during searching. While the indexing language may be distinct from the searching language, clearly, if retrieval is to be successful, the two must be closely related. Indexing languages may be of two different types: controlled-indexing languages or assigned-term systems, and natural-indexing languages or derived-term systems. Each of these is briefly discussed below.

Controlled-indexing languages (assigned-term systems)

With these languages a person controls the terms that are used as index terms. Controlled-indexing languages may be used for names and other labels but much emphasis is placed upon languages with terms that describe subjects. Normally an authority list identifies the terms that may be assigned. Indexing involves a person assigning terms from this list to specific documents on the

basis of subjective interpretations of the concepts in the document; in this process the indexer exercises some intellectual discrimination in choosing appropriate terms.

There are two types of subject-based controlled-indexing languages: alphabetical-indexing languages and classification schemes. In alphabetical-indexing languages, such as are recorded in thesauri and subject headings lists, subject terms are the alphabetical names of subjects. Control is exercised over which terms are used, and relationships between terms are indicated, but the terms themselves are ordinary words. In classification schemes each subject is represented by a code or notation. Classification schemes are particularly concerned to place subjects in a framework that crystallizes their relationships one to another. More generally though, classification is implicit in all indexing. A document in which content is wholly or partially specified in the index term RABBITS is thereby classed with other documents to which the same specification has been applied. Controlled-indexing languages take the process of classification one stage further, by displaying semantic links, between rabbits and hares for example. Formal bibliographic classification schemes, such as the DDC and LCC classifications, display these relationships in a systematic manner. They are able, in addition, through their notation to exclude particular connotations of meaning: thus DDC's 599.322 denotes rabbits as zoological entities, but not as pets (which would be 636.9322).

Thesauri have always been a feature of the document management systems that have been designed to manage larger collections. They are increasingly featuring in OPACs and other information retrieval environments, and their applicability for Internet applications is of interest. Thesauri typically show the controlled indexing term, with related, narrower and broader terms, as shown in Figure 5.6. They may be displayed in a window during search strategy formulation, to aid a user in the selection of terms. Often terms can be selected from the thesaurus listing simply by clicking on them. Hypertext links in thesauri listings can be used to move between different occurrences of the same term in the list. Another application of thesauri is as a basis for automatic indexing. All terms in the documents that appear in the thesaurus will generate an entry in the inverted index. Related applications of thesauri are in the creation of semantic nets and semantic knowledge bases.

Natural-indexing languages (uncontrolled or derived-term systems)

These languages are not really a distinct or stable language in their own right, but rather are the 'natural' or ordinary language of the document being indexed. Strictly, natural language systems are only one type of derived-term system. A derived-term system is one where all descriptors are taken from the document

being indexed. Thus, auth
as natural language subjec
appear in the document r
traditionally been on the te
text of the document is use
using the full text of the d
some mechanism for dec
indexing of a given docum
based upon statistical anal
Natural language indexing
by the computer. The com
from a limited stop-list of
that have been listed in a c

Natural language indexi
sively in many information
CD-ROM, via the online se
in online public access cat
to be more consistent and
searcher, but research has
systems designers is that t
in the context of the huge
prohibitively expensive. On
as valuable in a supportive
do not need to navigate all t
effort is being directed t
manage this variability, eith
databases include terms fi
both alphabetical indexin
support searching on the t

FEATURES OF RETRIEV

EXHAUSTIVITY AND CONTEN

It was suggested above th
single words or phrases.
specified by anything less
in the way of footnotes and
of classic writers). Indexi
ignoring the non-essentials

being indexed. Thus, author indexes, title indexes and citation indexes, as well as natural language subject indexes, are derived-term systems. Any terms that appear in the document may be candidates for index terms. Emphasis has traditionally been on the terms in titles and abstracts, but increasingly the full text of the document is used as the basis for indexing. Natural language indexing using the full text of the document may be very detailed, and in some systems some mechanism for deciding which terms are the most important in the indexing of a given document may be appropriate. Such mechanisms are often based upon statistical analysis of the relative frequency of occurrence terms. Natural language indexing can be executed by a human indexer, or automatically by the computer. The computer might index every term in the document, apart from a limited stop-list of very common terms, or may only index those terms that have been listed in a computer-held thesaurus.

Natural language indexing and controlled language indexing are used extensively in many information retrieval applications. Both are used in retrieval on CD-ROM, via the online search services, in document management systems and in online public access catalogues. Controlled-indexing languages are claimed to be more consistent and therefore more efficient and straightforward for the searcher, but research has failed to prove this convincingly. The dilemma facing systems designers is that to offer anything other than natural language indexing in the context of the huge databanks available through the Internet would be prohibitively expensive. On the other hand, controlled language indexing is seen as valuable in a supportive environment for inexperienced users because they do not need to navigate all the variations inherent in natural language. Significant effort is being directed towards the development of system interfaces that manage this variability, either implicitly or explicitly, on behalf of the user. Many databases include terms from controlled indexing languages (often including both alphabetical indexing languages and classification schemes) and also support searching on the text of the record, thus covering all options.

FEATURES OF RETRIEVAL SYSTEMS

EXHAUSTIVITY AND CONTENT SPECIFICATION

It was suggested above that indexes attempt to specify content by means of single words or phrases. Clearly, the *whole* of the subject content cannot be specified by anything less than the complete text (and may well require more, in the way of footnotes and other glosses, as with the heavily annotated editions of classic writers). Indexing has to try to sum up the salient points, while ignoring the non-essentials. This can be done at a number of levels, which, even

Advantages of Uncontrolled Indexing Languages

- Low input cost
- Full database contents searchable
- No human indexing errors
- No delay in incorporating new terms
- High specificity gives precision. Excels in retrieving individual terms – names of persons, organizations, etc.
- Exhaustivity gives potential for high recall. Does not apply to title-only databases

Disadvantages of Uncontrolled Indexing Languages

- Greater burden on searcher, particularly with terms that have many synonyms and several species
- Information implicitly but not overtly included in text may be missed
- Absence of specific to generic linkage
- Vocabulary of discipline must be known
- Syntax problems. Danger of false drops through incorrect term association
- Exhaustivity may lead to loss of precision.

Advantages of Controlled Vocabulary Indexing Languages

- Eases searching through:
  - control of synonyms and near synonyms
  - qualification of homographs
  - provision of scope notes
  - display of broader, narrower and related terms
  - expresses concepts elusive in free text.
- Overcomes syntax problems with compound terms and other devices
- Normally avoids precision loss through over-exhaustivity
- Maps areas of knowledge

Disadvantages of Controlled Vocabulary Indexing Languages

- High input cost
- Possible inadequacies of coverage
- Human error in interpretation and application of index terms can occur
- Possible out-of-date vocabulary
- Difficulty of systematically incorporating all relevant relationships between terms
- Lack of specificity
- Lack of exhaustivity
- The searcher needs to become acquainted with the language

Figure 5.1 **Comparing uncontrolled and controlled indexing languages**

though they are presented here as distinct strata, form a continuum. *Exhaustivity* of indexing is the name given to the depth of indexing which it is the policy of a given indexing system to employ. Exhaustivity is therefore a management decision. The level of exhaustivity at which a system operates can either be built into the system (for example, by restricting the number of fields available for index terms), or it can be controlled operationally, by giving instructions to indexers.

Summarization refers to the process of conveying the overall subject content

example, RABBITS, or BREEDS C
zation is commonly applied
which convey information pe
normally have their own det
indexes and contents lists. L
nearly always indexes at the
indexes to periodical literatu
range of indexes – *Humaniti*
W. Wilson Company.

A second level of exhausti
collections or to periodical li
the text – often around six
words in the title and abstrac
at this level.

Even more exhaustive ar
documents, which should lis
The ultimate level of exha
retrieval systems any word o
systems have a *stop-list* of ve
cannot therefore be retrieve
an index.

Figure 5.3 gives example
levels of exhaustivity.

Compare the Abstract, D
PsycInfo define 'Identifiers' c
has 96. For Descriptors and

Specificity

Specificity is an aspect of con
of the system, and denotes
content when indexing. De
rabbits as domestic animals
specify individual breeds o
Angora rabbits or any other
rabbits kept as pets and ra
manual on keeping pet Angc
rabbits as domestic animals
searcher has to sift through

of a document in a single word or a short phrase or structured heading: for example, RABBITS, or BREEDS OF RABBITS – BREEDS. Indexing at the level of summarization is commonly applied to graphic material – photographs and the like – which convey information perceptually; and also – particularly – to books, which normally have their own detailed indexing systems in the form of back-of-book indexes and contents lists. Library catalogues and published bibliographies are nearly always indexes at the level of summarization. So, too, are some published indexes to periodical literature: for example, *British Humanities Index* and the range of indexes – *Humanities Index*, *Education Index*, etc. – published by the H. W. Wilson Company.

A second level of exhaustivity is found in many databases that are indexes to collections or to periodical literature, and select the most significant subjects in the text – often around six to twelve controlled descriptors. In addition, the words in the title and abstract are available for searching. Contents lists operate at this level.

Even more exhaustive are back-of-book type indexes: indexes to individual documents, which should list every subject discussed in the text (Figure 5.2). The ultimate level of exhaustivity is provided by the text itself. In full-text retrieval systems any word or phrase is potentially available for searching. (Most systems have a *stop-list* of very common words that have not been indexed and cannot therefore be retrieved.) At this point we have a concordance rather than an index.

Figure 5.3 gives examples of indexing the same journal article at different levels of exhaustivity.

Compare the Abstract, Descriptors, and Identifiers in each. (NB: ERIC and PsycInfo define 'Identifiers' differently.) ERIC's abstract has 44 words; PsycInfo's has 96. For Descriptors and Identifiers the word counts are 20 and 46.

Specificity

Specificity is an aspect of controlled language systems. It refers to the vocabulary of the system, and denotes the extent to which we are able to specify subject content when indexing. Dewey Decimal Classification, for example, specifies rabbits as domestic animals at class 636.9322. This class is, however, unable to specify individual breeds of rabbit: there are no subclasses for lop-eared or Angora rabbits or any other breed. Neither can this class distinguish between rabbits kept as pets and rabbits grown for meat or for their fur. A specialist manual on keeping pet Angora rabbits has to be classed with all other works on rabbits as domestic animals. This clearly makes searching less precise, as the searcher has to sift through a number of marginally relevant items all classed

1. Summarization

Subject heading:	Indexing (*supplied by the Library of Congress*)
Title:	Indexing books
Series title:	Chicago guides to writing, editing and publishing

2. Most significant subjects

Chapter headings:

1. Introduction to book indexing
2. The author and the index
3. Getting started
4. Structure of entries
5. Arrangement of entries
6. Special concerns in indexing
7. Names, names, names
8. Format and layout of the index
9. Editing the index
10. Tools for indexing

3. Detailed subject specification

Index (*part*):

Abbreviations
  alphabetizing, 130
  of company names, 177, 180–81
  cross-references to and from, 102, 128–29
  double-posting, 130
  explaining, 12, 70
  spelling out, 128-29
  for states in U.S., 175–76
access points
  converting subentries to main headings, 219
  main heading as primary, 77, 217
  multiple, with double posting, 75, 76, 221
accuracy of entries, assessing, 230
acronyms
  alphabetizing, 130

4. The full text

Text: Alphabetizing of Abbreviations and Acronyms

Abbreviations and acronyms should be alphabetized in the same way as the other entries in the index, whether letter-by-letter or word-by-word. They are not usually alphabetized as if they were spelled out. An exception that many publishers allow is that the abbreviation U.S. may be alphabetized as though spelled out. This allows a term like *U.S. Bureau of Reclamation* to interfile with other entries such as *United States Coast Guard.*

Figure 5.2 **Levels of exhaustivity within a single work**
***Source*: Mulvaney, 1994.**

Example from ERIC and PsycInfo, sh
by the two services.

ERIC:
EJ521883 PS524404
The Role of Emotion in Children's L
Crockenberg, Susan; Forgays, Debor
Merrill-Palmer Quarterly; v42 n1 pp.
Theme issue topic: 'Conflicts in Fami
ISSN: 0272-930X
Available from: UMI
Language: English
Document Type: RESEARCH REPOR
Journal Announcement: CIJAUG96
Abstract: Tested a process model f
behavior adjustment with a sample
conflict behavior and children's negati
behavioral adjustment. (MDM)
Descriptors: *Adjustment (to Environ
*Family Problems; *Models; Predict
Identifiers: *Marital Discord

PsycInfo:
01416789 1996-01718-002
The role of emotion in children's ur
Author: Crockenberg, Susan; Forgays
Author Affiliation: U Vermont, Dept
Journal: Merrill-Palmer Quarterly , V
Special Issue: Special Issue: Confl
research.
ISSN: 0272-930X
Document Type: Journal Article; Emp
Special Features: References
Record Type: Abstract
Language: English
Population Group: Human; Male; Fer
(6–12 yrs)); 300 (Adulthood (18 yrs
Abstract: (Presents a process mod
essential mediator between the conf
and fathers (aged 21–44 yrs) assess
list and questionnaires. Children view
Children's perceptions indicate abilit
conflict. Children report negative emc
Maternal conflict behavior and childr
children's behavioral adjustment. ((c)
Descriptors: *Child Attitudes; *Em
*Parents ; Adjustment; Childhood; P
Identifiers: perceptions of & emotion
behavioral adjustment, male vs fema
Subject Codes & Headings: 2820 (C
Release Date: 19970101

Figure 5.3

Example from ERIC and PsycInfo, showing how one research paper has been indexed and abstracted by the two services.

ERIC:

EJ521883 PS524404
The Role of Emotion in Children's Understanding and Emotional Reactions to Marital Conflict.
Crockenberg, Susan; Forgays, Deborah Kirby
Merrill-Palmer Quarterly; v42 n1 pp. 22–47 Jan 1996
Theme issue topic: 'Conflicts in Families and Between Children: Advances in Theory and Research.'
ISSN: 0272-930X
Available from: UMI
Language: English
Document Type: RESEARCH REPORT (143); JOURNAL ARTICLE (080)
Journal Announcement: CIJAUG96
Abstract: Tested a process model for the impact of children's exposure to marital conflict on their behavior adjustment with a sample of 28 couples and their 6-year-old children. Found that maternal conflict behavior and children's negative emotional reactions to fathers independently predicted children's behavioral adjustment. (MDM)
Descriptors: *Adjustment (to Environment); *Child Behavior; Conflict Resolution; *Emotional Response; *Family Problems; *Models; Predictor Variables; Sex Differences; *Young Children
Identifiers: *Marital Discord

PsycInfo:

01416789 1996-01718-002
The role of emotion in children's understanding and emotional reactions to marital conflict.
Author: Crockenberg, Susan; Forgays, Deborah Kirby
Author Affiliation: U Vermont, Dept of Psychology, Burlington, VT, USA
Journal: Merrill-Palmer Quarterly , Vol 42(1) , 22–47 , Jan , 1996
Special Issue: Special Issue: Conflicts in families and between children: Advances in theory and research.
ISSN: 0272-930X
Document Type: Journal Article; Empirical Study
Special Features: References
Record Type: Abstract
Language: English
Population Group: Human; Male; Female **Age Group:** 100 (Childhood (birth–12 yrs)); 180 (School Age (6–12 yrs)); 300 (Adulthood (18 yrs & older)) **Population Location:** USA
Abstract: (Presents a process model that identifies children's processing of marital conflict as an essential mediator between the conflict and behavioral maladjustment. 28 mothers (aged 27–42 yrs) and fathers (aged 21–44 yrs) assessed their 6-yr-old child's adjustment with the Child Behavior Checklist and questionnaires. Children viewed videotapes of their parents working toward conflict resolution. Children's perceptions indicate ability to distinguish behavior of fathers from mothers during marital conflict. Children report negative emotional reactions to mothers when fathers exhibit negative behavior. Maternal conflict behavior and children's negative emotional reactions to fathers independently predict children's behavioral adjustment. ((c) 1997 APA PsycINFO, all rights reserved)
Descriptors: *Child Attitudes; *Emotional Responses; *Marital Conflict; *Human Sex Differences; *Parents ; Adjustment; Childhood; Psychosocial Development; School Age Children; Adulthood
Identifiers: perceptions of & emotional reactions to mother's vs father's behavior during marital conflict, behavioral adjustment, male vs female 6 yr olds & their 21–44 yr old parents
Subject Codes & Headings: 2820 (Cognitive & Perceptual Development)
Release Date: 19970101

Figure 5.3 **Indexing at different levels of exhaustivity**

at the same place. Specificity thus improves the precision of a search: that is, its ability to sift out unwanted material.

Special systems (i.e. systems confined to one subject area or other field) often use differential levels of specificity. Topics that are central to the subject field are indexed at a higher level of specificity than peripheral subjects. For example, if Domestic Animals is a system's principal subject field, it would be quite likely to make specific provision for the various breeds of rabbit. If the subject field is something remote, however, there might be no specific provision even for rabbits: we might have to include them under a more general term, like Pets.

Specificity and exhaustivity are related to the extent that in practice greater exhaustivity needs to be matched by greater specificity in the indexing terms. Most book indexes, for example, are both specific and exhaustive. The combination of specificity and exhaustivity is often referred to as *depth of indexing*.

Complex topics

A final set of definitions concerns the way complex topics are handled. A document may not be simply about rabbits or apples or chess; it may very well deal with some more precise aspect, like breeds of rabbits, or the effect of pre-storage heat treatment on the shelf life of apples. The traditional method of dealing with complex topics has been to encapsulate as much of the topic as possible into a single heading, RABBITS - BREEDS for example, or APPLES - SHELF LIFE - *EFFECT OF* - HEAT TREATMENT. Indexes of this kind are known as *pre-coordinate* indexes, because the topics that comprise a heading are strung together or coordinated by the indexer in advance of any searches that may be carried out on any of the topics represented within the heading. These indexes require elaborate rules for the consistent construction of headings, and will be considered in Chapter 6. Because of this lack of flexibility many systems employ a quite different method of handling complex topics. Here the subject of a document is represented by a number of one-concept terms – these are the descriptors in Figure 5.3 – and the searcher is able to combine as many or as few of them as are required, using Boolean logic: for example, CHILD ATTITUDES AND MARITAL CONFLICT. Systems employing this method of indexing and searching are known as *post-coordinate* systems. The earliest of these systems were card-based, using specialized stationery and other equipment, but nearly all systems in use today are computerized.

User-friendliness

In 1960 Calvin Mooers expounded his famous law:

An information system
and troublesome for a
have it.

The corollary of this is tha
increased if steps are tak
factors influencing user-frie

- *Accessibility*: the service
- *Ease of use*: the service
users. While great strid
Chapter 4), there is a
capabilities. The more
complex the instructio
error on the part of the
- *System error*: as oppose
and output errors cau
automatic indexing and
- *Form of output*: output r
surrogates; if the latter
convenient output is n
with manually searched
- *Delay*: whether in acces

SEARCH FACILITIES IN

SEARCH LOGIC

Search logic is the means
matched for successful ret
most systems. It may be u
indexing languages, or bot
the concepts present in th
more search terms may b
the search statement. Sear
of all synonyms and related
search-term combinations.
with natural language term
variations and near-synony
evolved one at a time, an
specifies a search stateme

> An information system will tend not to be used whenever it is more painful and troublesome for a customer to have information than for him not to have it.

The corollary of this is that the use of information systems and services will be increased if steps are taken to improve their user-friendliness. Some of the factors influencing user-friendliness are:

- *Accessibility*: the service should be physically accessible to users.
- *Ease of use*: the service should be within the intellectual capabilities of its users. While great strides have been made in improving user interfaces (see Chapter 4), there is always a trade-off between ease of use and system capabilities. The more powerful the functionality of a system, the more complex the instructions and protocols for using it, and its proneness to error on the part of the user.
- *System error*: as opposed to user error. This includes system malfunctions, and output errors caused by inadequacies in the system, as with some automatic indexing and retrieval systems.
- *Form of output*: output may be in the form of actual documents, or document surrogates; if the latter, output may or may not be downloadable. The least convenient output is non-downloadable document surrogates, for example, with manually searched catalogues.
- *Delay*: whether in accessing the service, or in obtaining the search results.

SEARCH FACILITIES IN POST-COORDINATE SEARCHING

SEARCH LOGIC

Search logic is the means of specifying combinations of terms that must be matched for successful retrieval. Boolean search logic is employed in searching most systems. It may be used to link terms from either controlled- or natural-indexing languages, or both. The logic is used to link the terms that describe the concepts present in the statement of the search. As many as 20 to 30 or more search terms may be linked together by search logic in order to frame the search statement. Search logic permits the inclusion in the search statement of all synonyms and related terms, and also specifies acceptable and unacceptable search-term combinations. Search strategies often need to be more complex with natural language terms, in order to accommodate all the potential spelling variations and near-synonyms. In an online search the search statements are evolved one at a time, and feedback is available at each stage. The searcher specifies a search statement and the computer responds with the number of