

# **Introduction to Modern Information Retrieval**

G. G. Chowdhury



Library Association Publishing  
London

# Contents

<b>Preface</b>	<b>xv</b>
A blend of the traditional and the new	xv
A detailed coverage map of this book	xvi
Acknowledgments	xix
<b>1 Basic concepts of information retrieval systems</b>	<b>1</b>
Introduction	1
Purpose	2
Functions	3
Components	3
Kinds of information retrieval systems	4
Design issues	5
Design phases	6
References	10
<b>2 Database technology</b>	<b>12</b>
Introduction	12
Data	12
The database	12
Records and fields	13
Properties of databases	14
Kinds of databases	14
Database technology	16
The development of databases in an information retrieval environment	17
Discussion	23
References	23
<b>3 Bibliographic formats</b>	<b>25</b>
Introduction	25
Bibliographic records	26
Integrated database approach	28
ISO 2709: format for bibliographic information interchange	29
MARC format	32
UNIMARC format	37
CCF	40
MIBIS format	42

	<b>ABNCD format</b>	<b>44</b>
	<b>Discussion</b>	<b>47</b>
	<b>References</b>	<b>54</b>
<b>4</b>	<b>Subject analysis and representation</b>	<b>56</b>
	Introduction	56
	Cataloguing and classification	57
	Subject analysis	68
	Subject indexing	69
	Exhaustivity and specificity	69
	Manual indexing	71
	Pre-coordinate indexing systems	73
	Post-coordinate indexing systems	81
	Problems of manual indexing	83
	Theory of indexing	83
	Discussion	84
	References	84
<b>5</b>	<b>Automatic indexing and file organization</b>	<b>87</b>
	Introduction	87
	The process of indexing	87
	Automatic classification	90
	Index file organization	91
	Inverted file	92
	Sequential access	97
	Alternative text retrieval structures	106
	Discussion	115
	References	115
<b>6</b>	<b>Vocabulary control</b>	<b>118</b>
	Introduction	118
	Controlled vs. natural indexing	119
	Vocabulary control tools	121
	Guidelines for developing a thesaurus	138
	Criteria for evaluating a thesaurus	140
	Software for development and maintenance and use of online thesauri	140
	Discussion	141
	References	141
<b>7</b>	<b>Abstracts and abstracting</b>	<b>144</b>
	Abstracts	144
	Types of abstract	144
	Qualities of abstracts	146
	Uses of abstracts	147

The art of abstracting	148
Automatic abstracting	152
Discussion	156
References	156
<b>8 Searching and retrieval</b>	<b>158</b>
Introduction	158
The search strategy and its prerequisites	158
The pre-search interview	159
The searching process	159
Retrieval models	160
Search facilities offered by most text retrieval software systems	169
References	177
<b>9 Users of information retrieval</b>	<b>179</b>
Introduction	179
Users and their nature	179
Types of information needs	180
Information needs in different areas of activity	182
Information seeking behaviour of users	187
What we need to know about users	188
User studies	192
Possible sources of information about users	197
Discussion	198
References	198
<b>10 Evaluation of information retrieval systems</b>	<b>200</b>
Introduction	200
The purpose of evaluation	200
Evaluation criteria	200
Design of an evaluation programme	209
The steps of evaluation	210
Simulation studies for retrieval effectiveness	211
References	213
<b>11 Evaluation experiments</b>	<b>215</b>
Introduction	215
The Cranfield tests	216
MEDLARS	221
The SMART retrieval experiment	222
The STAIRS project	225
Discussion	227
TREC: the Text Retrieval Conferences	229
References	233

<b>12</b>	<b>Online information retrieval</b>	<b>235</b>
	Introduction	235
	Features of online searching	235
	Online search services	235
	Growth and development of online searching	236
	Suppliers of online search services	237
	Elements of an online search	237
	Human-computer interactions	241
	Sample dialogues in online searches	247
	Discussion	248
	References	249
<b>13</b>	<b>CD-ROM information retrieval</b>	<b>250</b>
	Introduction	250
	Background	250
	CD-ROM technology	252
	Physical characteristics of CD-ROM	255
	Logical layout of the discs	255
	Stages of CD-ROM production	256
	Advantages of CD-ROM	260
	Disadvantages of CD-ROM	261
	CD-ROM: hardware issues	263
	CD-ROM networking	265
	CD-ROM retrieval software	267
	CD-ROM user interface	269
	CD-ROM applications	271
	Criteria for selection of CD-ROM databases	272
	Training CD-ROM users	274
	Discussion	276
	References	276
<b>14</b>	<b>Trends in CD-ROM and online information retrieval</b>	<b>278</b>
	Introduction	278
	CD-ROM vs. online information retrieval	278
	Developments in CD-ROM technology	280
	Hardware for CD-ROM: trends	281
	Current alternatives	281
	Recent developments: Electronic Resource Library (ERL) and SilverLinker	283
	References	284
<b>15</b>	<b>Multimedia information retrieval</b>	<b>286</b>
	Introduction	286

Multimedia systems	286
The technology	287
Multimedia information retrieval	288
Interactive multimedia	289
Multimedia applications	289
Application development tools	293
Discussion	293
References	294
<b>16 Hypertext and hypermedia systems</b>	<b>296</b>
Introduction	296
The history of hypertext	297
Some important terms	298
Hypertext definition and meaning	299
Hypertext and conventional text retrieval systems	301
Advantages of hypertext	302
Design features	303
Components of hypertext	304
Hypertext reference model	305
Hypertext software	305
Open hypertext systems	306
Hypermedia systems	308
Access to hypermedia systems	308
Types of hypermedia system	309
Hypermedia tools and techniques	309
Collaborative hypermedia systems	310
Hypertext and hypermedia applications	310
Evaluation of hypertext and hypermedia information retrieval systems	312
Hypertext markup language	313
Discussion	313
References	313
<b>17 Intelligent information retrieval</b>	<b>316</b>
Introduction	316
Intelligent retrieval systems	317
Artificial intelligence	317
Expert systems	318
Kinds of expert systems	319
Components of expert systems	319
Historical development of expert systems	321
Development methodology and approaches	321
Knowledge elicitation and representation methods	323
Inference strategies	324
End-user modelling and interfaces	324

Development tools	326
Expert systems for library and information services	326
Discussion	329
References	329
<b>18 Natural language processing and information retrieval</b>	<b>333</b>
Introduction	333
Natural language understanding	333
Syntactic analysis	334
Semantic analysis	342
Pragmatic knowledge	351
References	355
<b>19 Natural language interfaces</b>	<b>357</b>
Introduction	357
Natural language interface	357
CANSEARCH	358
PLEXUS	363
IOTA	368
IRUS	368
I <sup>3</sup> R	369
Fuzzy set systems	369
Discussion	369
References	370
<b>20 Natural language text processing and retrieval systems</b>	<b>372</b>
Introduction	372
Text processing and retrieval	373
Text processing systems	374
Sublanguage analysis	377
Sublanguage analyses in chemical information retrieval	377
Sublanguage analyses in patent information retrieval	380
Discussion	382
References	382
<b>21 Problems and prospects of natural language processing systems</b>	<b>387</b>
Introduction	387
Problem areas	387
Lexical analysis stage	387
Syntactic analysis stage	388
The constraints of natural language	389

Nature of knowledge bases	389
The sublanguage approach	390
Conclusions	390
References	392
<b>22 The Internet and information retrieval</b>	<b>394</b>
Introduction	394
The Internet	395
Internet services and tools	396
The World Wide Web (WWW)	397
The impact of the Internet	398
The Internet and government	400
Information retrieval issues	401
Retrieval features of some web search engines	403
The future of the online and CD-ROM industry in the Internet age	414
Intelligent search agents	415
Discussion	416
References	417
<b>23 Trends in information retrieval</b>	<b>420</b>
Introduction	420
Evaluation of information retrieval systems	422
Developments related to the input subsystem	423
Searching and retrieval	425
Full-text retrieval	428
Software development	429
User interfaces	430
Information retrieval standards and protocols	430
The digital global library	431
Intelligent information retrieval	435
Intelligent hypertext and hypermedia systems	435
Conclusions	436
References	438
<b>Index</b>	<b>443</b>