

Contents

1	Introduction	1
2	Overview	4
	What is information and what can be done with it?	4
	Find a document	5
	Organize a collection	6
	Information structures supporting the collection	7
	A preliminary case study	8
	Key concepts	10
3	Text retrieval	11
	Query mechanics	11
	The problem is not the query	12
	Why is it hard to get high precision and recall?	15
	Document nearness query languages	16
	Index construction	19
	How does a boolean ISAR compare with a nearness-based ISAR? ..	19
	A user's real need generally requires multiple queries	23
	How do we know when to stop?	23
	Key concepts	24
	Further reading	25
	Major formative exercise	25
	Tutorial exercises	26
	Open discussion question	28
4	Describing information objects	29
	What is stored?	29
	How is an information object identified?	29
	Physical description of document	31
	Content description	31
	Choice of descriptors	33
	Properties of descriptors	34
	Engineering issues in surrogate design	35
	Bates's principles: Variety and uncertainty	37
	Push technologies	39
	Content analysis	40
	Key concepts	41
	Further reading	42
	Formative exercise	42

	Tutorial exercises	42
	Open discussion question:.....	43
5	Browsing: Hypertext	44
	Information spaces.....	44
	Hypertext implementation of information space	45
	Building a hypertext.....	46
	The possibility of automatic creation of a hypertext.....	48
	Concept map approach for link design	50
	Navigation	53
	Relationship between browsing and search	56
	Why can this work?.....	57
	Key concepts	58
	Further reading.....	58
	Major formative exercise.....	58
	Tutorial exercises	59
	Open discussion question:.....	59
6	World Wide Web	60
	How is the Web different from a library catalogue or a CD-ROM? ..	60
	Resource discovery.....	62
	The emerging order (sort of) in Web information space	65
	Portals.....	65
	Vertical portals.....	66
	How might this order develop?	67
	Pull and push – advertising.....	69
	Agents and the need for standards.....	70
	Key concepts	71
	Further reading.....	71
	Formative exercise	71
	Tutorial exercises	72
	Open discussion question	72
7	Structured documents: XML	73
	Complex objects	73
	Markup languages	74
	XML	77
	Applications of XML	81
	Automatic abstracting	84
	Information space with XML objects.....	85
	Key concepts	85
	Further reading.....	85
	Formative exercise	86
	Tutorial exercises	86
	Open discussion question	86
8	Controlled vocabulary.....	87
	Why controlled vocabulary?	87
	What do classification systems look like?	89
	How are classification systems organized?	92
	Classification systems are often hierarchical	95

Basis of hierarchy	96
Designing a classification scheme	98
Key concepts	100
Further reading	101
Formative exercise	101
Tutorial exercises	101
Open discussion question	101
9 Semantic dimensions	102
Place	102
Relationship among places	104
Multiple semantic dimensions	105
Absolute and relative dimensions	107
Use of semantic dimensions for understanding a whole information space	109
Navigation in a hierarchy	112
Visualization of relative dimensions	113
Geometries based on many dimensions	115
Key concepts	116
Further Reading	117
Formative exercise	117
Tutorial exercises	117
Open discussion question	118
10 Classification in context	119
Individuals and classes	119
Variability within classes	121
Classification is not the only way to organize information spaces	126
What is in the information space?	132
Systematic classification is an aspect of modern culture	134
Key concepts	136
Further reading	136
Formative exercises	136
Tutorial exercises	137
Open discussion question	137
11 Large classification systems	138
Introduction	138
What an indexing system involves	138
Requirements for schedules	139
Library of Congress system	139
Exhaustivity	139
Coding efficiency	140
Semantic coherence	140
Maintainability	141
Workability	141
The published schedule	142
Dewey Decimal System	143
Types of classification schemes	143
Developing faceted classification systems	147

Notation	153
Locating objects in physical space	153
Locating objects in information space	153
Key concepts	154
Further reading	154
Formative exercise	154
Tutorial exercises	155
Open discussion question	155
12 Descriptors	156
Introduction	156
Design requirements	157
Subject lists	158
Thesauri	160
Building a thesaurus	162
Automated construction of thesauri	164
Descriptors as primary organizers of a collection	167
Ontologies	167
Key concepts	171
Further reading	171
Formative exercise	171
Tutorial exercises	171
Open discussion question	172
13 Visualization	173
Information as a space	173
Navigation	174
Keeping close to real space	177
Example of close-to-real-space visualization	180
Discussion of close-to-real-space visualization design guidelines	183
Visualizing relative dimensions – enumerative systems	184
Key concepts	187
Further reading	187
Formative exercise	187
Tutorial exercises	188
Open discussion question	188
14 Archiving	189
What are archives?	189
How do we interpret an item of information?	190
Design of an archive	191
Obsolescence of media	193
Key concepts	194
Further reading	194
Formative exercise	194
Tutorial exercises	195
Open discussion question	195
15 Quality	196
Quality of content	196

Legal issues – overview	198
Crime	199
Censorship.....	199
Libel	200
Privacy	200
Intellectual property.....	202
Napster.....	203
Key concepts	204
Further reading.....	204
Formative exercise	204
Tutorial exercises	205
Open discussion question	205
Major assignment	206
Appendix A: US Department of Labor Standard Industrial Classification division structure, 1987 edition.....	209
Appendix B: Excerpts from the Yahoo! classification system	214
Appendix C: Excerpt from the <i>Encyclopedia Britannica Propaedia</i>, 15th edition, 1992.....	216
Appendix D: Extracts from the Library of Congress system	218
Appendix E: Library of Congress schedules as of November 1999.....	224
Appendix F: Top levels of the Dewey Decimal System	226
Appendix G: Extract of subject terms from Sears' list, 1986 edition.....	228
Appendix H: Extract of subject terms from Library of Congress list	231
Appendix I: Selection from ACM keyword list	233
Appendix J: Extract of thesaurus terms from Sears' list, 1997 edition	235
Appendix K: Extract of thesaurus terms from INSPEC, 1999 edition.....	237
Appendix L: Images of spikes.....	239
Bibliography	245
Index	247