TABLE OF CONTENTS

Chapter 1: The Nature of Information Retrieval

Introduction 1
The Distinction Between Data Retrieval and Information Retrieval 2
Consequences of the Data/Document Distinction 5
The Trial-and-Error Nature of Information Retrieval 6
The Goals of Document Retrieval: Prediction Criteria and Futility Points 9
The Significance of the Futility Point Criterion 10
Satisfying the Prediction Criterion 11
Large Systems and Inefficient Search Term Use 13
Further Aspects of Anchoring 16
The Primacy of Term Selection Order 18
Associative Searching and the Prediction Criterion 19
Prediction Criteria and Futility Points, an Objection 21
The Central Issue of Information Retrieval 22
Notes 24

Chapter 2: The Principal Formal Models of Information Retrieval 27

Model 1 28
Model 2 31
Model 3 32
Model 4 37
Model 5 38
Model 6 41
Model 7 43
Model 8 45
Model 9 46
Model 10 49
Model 11 55
Model 12 62
Hybrid Models 64
Notes 65
Chapter 3: The Evaluation of Information Retrieval Systems 69

Physical Access vs. Logical (Intellectual) Access 69
What Should be Evaluated 71
How Should Retrieval Effectiveness be Measured? 73
Measuring Precision and Recall 80
Previous Recall/Precision Studies 84
Consequences of the Lack of Reliable Recall/Precision Studies 85
A Test of Retrieval Effectiveness 87
The Test Environment 88
The Experimental Protocol 88
Conduct of the Test 89
Test Results 94
Other Findings 96
The Retrieval Effectiveness of Lawyers vs. Paralegals 96
Computing Precision and Recall at Different Levels 98
Why was Recall so Low? 100
Deterioration of Recall as a Function of File Size 104
Discussion 106
Major Findings of the STAIRS Study 109
Notes 113

Chapter 4: Language and Representation: The Central Problem in Information Retrieval 121

The Central Issue of Information/Document Retrieval 121
Semiotics 123
The Sign 123
Signification 124
Expressions 125
Contents 126
Signification and Communication 128
Signified as Cultural Unit 129
Problems with Mentalistic Theories of Meaning 132
Consequences of Mentalistic Semiotics 134
The Alternative to Mentalistically Based Semiotics 136
Useful Aspects of Semiotic Theory 136
The Implementational Interpretation of Significance 139
The Tool Analogy 139
George Kingsley Zipf 141
Ludwig Wittgenstein 143
The Structure of Language 150
The Consequences of Wittgenstein's Theory of Language for Information Retrieval 155
The Nature of Document Representation 155
The Structure of Subject Searching 164
Subject Description and Ordinary Language 171
Improving Searching on Information Retrieval Systems 175
Information Retrieval: the Adaptive/Communication Model 178
A. The "Learnable" Retrieval System 178
B. The "Contextualization" of Subject Description 183
C. Information Retrieval as a Communication Process 188
An Indicator of Retrieval System Effectiveness 193
The Study of Language and Activities: More Recent Directions 194
Information Retrieval: Implications and Future Directions 200
Do We Have a Complete Typology of Illocutionary Acts? 204
Implementing an Information Retrieval system Based on Illocutionary Acts 207
Another Look at Simple Full-Text Retrieval 209
A Final Note 211
Notes 213

Chapter 5: Communication and Adaptation 249

Relevance Feedback in Information Retrieval 249
Problems with Simple Relevance Feedback Models 250
The Operational Definition of "About" Revisited 252
The Genetic Algorithm: Relevance Feedback Redux 254
The Genetic Algorithm: Detail 256
A Test of the Genetic Algorithm in Information Retrieval 261
Discussion 266
Conclusion 270
Notes 271

Chapter 6: Information Retrieval and the Nature of Scientific Theory 277
Scientific Theory vs. Experimental Law 277
The Constituents of Scientific Theory 279
Scientific Models 281
The Full Universe and the Sea of Air 283
Science vs. Pseudoscience 285
Science as Puzzle Solving 288
The Scientific Discipline 292
Information Retrieval and Scientific Theory 296
The Educational Antecedents of Information Retrieval Research 301
The Information Retrieval Paradigm: Some Tentative Conclusions 305
Notes 309

Chapter 7: Conclusion 319
The Problem of Evaluation 323
Artificial Intelligence and Information Retrieval: Waiting for Godot 324
The Magic of Language 326
Notes 330

Index 331