CONTENTS

Foreword ix
Preface xi

Chapter 1 **Introduction** 1

Chapter 2 **Information Retrieval Background** 9

Computers in Information Retrieval 10
Mechanized Documentation since 1965 11
Aspects of Mechanization 12
Key Questions 15
Relevance of Linguistics 17
Experiments in Information Retrieval 18
Experimental Methodology 21
Performance Measures 22
Current State of Experiments 24
Information Retrieval Models 26
Recent Developments 27
Present State of Play 29

Chapter 3 **Linguistics** 31

Theoretical Linguistics 31
Views of Grammar 34

Components of Grammars 36
Computational Linguistics 39
ALPAC Findings 41
Contributions to Information Science 43

Chapter 4 **Language in Documentation** 45

Linguistic Components of Information Retrieval Systems 45
Natural and Indexing Languages 46
Input Texts and Index Descriptions 47
Linguistic Problems 50
Units of Input and Description 51
Information Retrieval Needs and Linguistic Solutions 53
Analysis, Description, and Searching 54
Strong and Weak Approaches 56
Linguistic Implications 60
Available Linguistic Tools 60
Characteristics of Indexing Languages: Coyaud 63
Some Examples 65
Specification and Classification in Indexing Languages: Sharp 70
Postcoordination 74
Relations in Indexing Languages: Šreider *et al.* 76

Chapter 5 **Syntax** 79

Syntax in Natural Language 79
Phrase Structure and Dependency 81
Automatic Phrase-Structure Analysis 85
Parsing Strategies 87
Transformational Grammar 91
Form of Deep Structures 95
Transformational Parsing 97
Approaches Emphasizing Rule Form 97
Approaches Emphasizing the Deep–Surface Distinction 99
Syntax in Information Retrieval 103
Input Text Syntax 105
Partial Parsing 106
Statistical Syntax 110
Exploiting Syntactic Information 111
Index Description Syntax 112
Postcoordinate Syntax 113
Syntax in Searching 114

The SYNTOL Project 115
Syntax in Information Retrieval: Conclusion 118

Chapter 6 **Semantics** 120

Semantics in Natural Language 120
Semantic Notation 121
Semantic Component of a Transformational Grammar 123
Semantic Aspects of Information Retrieval 125
Semantic Input Analysis 128
Derivative and Assignment Indexing 129
Extraction Procedures 131
Statistical Extraction 134
Statistical Extraction: Conclusion 137
Semantic Indexing Procedures 139
Keyword Normalization 141
Term Assignment 141
Indexing for Substitution and Expansion 143
Term Weighting 146
Index Languages 148
Dictionary, Thesaurus, Classification 149
Vocabulary Formation 151
Vocabulary Requirements 152
Automatic Dictionary Formation 155
Index Language Classification 156
Automatic Term Classifications 159
Classification Base 162
Classification Procedures 163
Term Associations 164
Term Structures 165
Experiments with Automatic Term Structures 166
Value of Automatic Term Structures 168
Semantics in Searching 170
Automatic Classification in Linguistics 171
Semantics in Information Retrieval: Conclusion 173

Chapter 7 **Fact Retrieval** 174

Fact Retrieval and Document Retrieval 174
Early Systems 177
Semantic Networks 182
Theorem Proving and Fact Retrieval 189

Chapter 8 **Conclusion** 195

References 201

Author Index 227
Subject Index 234