Contents

Preface ix

Chapter 1 **Introduction to Text Processing**

1-1 An Overview of Text Processing 2
1-2 Text Processing Primitives 4
1-3 Programming Languages and Packaged Programs for Text Processing 6
1-4 The Algorithmic Approach to Text Processing 7
1-5 Computer Elements for Text Processing: I/O, Storage, and CPU 11
1-6 Special Considerations for Text Processing 16
Questions and Exercises 19

Chapter 2 **Introduction to PL/I for Text Processing**

2-1 Numbers, Strings, and Declarations 21
2-2 Elementary Input and Output 27
2-3 Arithmetic and Assignment 34
2-4 Complete Programs and Control Statements 38
2-5 Character String Manipulation: Concatenation, Substrings, and Assignment 50
2-6 Built-In Functions for String Manipulation 53
2-7 Extending the Built-In Functions 57
2-8 Advanced Input and Output with Applications in Bibliographic Search and Dictionary Building 61
2-9 Hints for Diagnosing Errors in PL/I Programs: SUBRG, STRG, and Program Trace 72
Questions and Exercises 75

Chapter 3 **Introduction to SNOBOL for Text Processing**

3-1 Numbers, Strings, Variables, and Arrays 79
3-2 Elementary Input and Output 82
3-3 Expressions and Assignment 83
3-4 Complete Programs and Control Statements 86
3-5 Character String Manipulation: Pattern Matching and Conditional Assignment 89
3-6 Built-In and Programmer-Defined Functions for String Manipulation 97
3-7 Hints for Diagnosing Errors in SNOBOL Programs: Program Trace 100
Questions and Exercises 101

Chapter 4 **Overview of Text Processing Packages and Applications**

4-1 KWIC for Concordance Generation 106
4-2 FAMULUS for Bibliographic Information Retrieval 113
4-3 CMS Editor for Interactive Text Editing 125
4-4 SCRIPT for Text Formatting 135
Questions and Exercises 141

Chapter 5 **Literature Review**

5-1 Programming Languages for Text Processing 144
5-2 Text Editing, Formatting, and Bibliographic Information Retrieval Packages 146
5-3 Concordance Generation Packages and Applications 147
5-4 Content Analysis, Machine Translation, and Other Research Areas 148
5-5 Additional Bibliographies and Other Sources 149
5-6 Conclusion 150

Answers to Selected Exercises 153

Appendix A **EBCDIC and ASCII Character Sets** 157

Appendix B **Selected Conventions for Tape and Disk I/O** 160

Appendix C **Glossary of Text Processing Primitives** 162

References 165

Index 169