

# Contents

Acknowledgments .....	v
About This Book .....	xi
How to Use This Book .....	xiii
Notations and Conventions .....	xv
<i>Codes</i> .....	xv
<i>Images</i> .....	xv
<i>Names</i> .....	xvi
<i>Names List Annotations</i> .....	xvii
<b>1.0 Introduction</b> .....	<b>1</b>
1.1 Background .....	1
1.2 Conformance .....	2
1.3 Coverage .....	2
<i>Statistics of Version 1.0</i> .....	4
<i>Future Plans</i> .....	4
1.4 The Unicode Consortium .....	5
The Unicode Technical Committee .....	5
<b>2.0 General Principles of the Unicode Standard</b> .....	<b>7</b>
2.1 Architectural Context of a Character Encoding .....	7
<i>Text Elements, Code Elements, and Text Processes</i> .....	7
<i>Plain and Fancy Text</i> .....	9
2.2 The Basic Design of the Unicode Character Encoding .....	10
<i>Unicode Principles</i> .....	10
<i>Glyphs</i> .....	12
2.3 Codespace Allocation .....	13
2.4 Character Properties, Controls, and Sequences .....	14
<i>Properties</i> .....	15
<i>Control Characters</i> .....	15
<i>Paragraph/Line Separators</i> .....	16
<i>Ordering of Character Sequences</i> .....	16
<i>Alternate Spellings</i> .....	17
2.5 Non-spacing Marks .....	18
<i>Sequence of Base Letters and Non-spacing Marks</i> .....	19
<i>Spacing Clones of European Diacritical Marks</i> .....	19
<i>Multiple Non-spacing Marks</i> .....	19

2.6 Formal Requirements for Unicode Conformance .....	21
Examples of Conformant and Non-conformant Processing .....	24
3.0 Character Blocks .....	27
3.1 General Scripts .....	28
ASCII U+0000 → U+007F .....	29
Latin1 Characters U+0080 → U+00FF .....	32
European Latin U+0100 → U+017F .....	33
Extended Latin U+0180 → U+01FF .....	34
Standard Phonetic U+0250 → U+02AF .....	36
Modifier Letters U+02B0 → U+02FF .....	38
Generic Diacritical Marks U+0300 → U+036F .....	40
Greek U+0370 → U+03FF .....	42
Cyrillic U+0400 → U+04FF .....	44
Armenian U+0530 → U+058F .....	46
Hebrew U+0590 → U+05FF .....	47
Arabic U+0600 → U+06FF .....	49
Devanagari U+0900 → U+097F .....	53
Bengali U+0980 → U+09FF .....	60
Gurmukhi U+0A00 → U+0A7F .....	61
Gujarati U+0A80 → U+0AFF .....	62
Oriya U+0B00 → U+0B7F .....	63
Tamil U+0B80 → U+0BFF .....	64
Telugu U+0C00 → U+0C7F .....	65
Kannada U+0C80 → U+0CFF .....	66
Malayalam U+0D00 → U+0D7F .....	67
Thai U+0E00 → U+0E7F .....	68
Lao U+0E80 → U+0EFF .....	70
Tibetan U+1000 → U+105F .....	71
Georgian U+10A0 → U+10FF .....	73
3.2 Symbols .....	74
General Punctuation U+2000 → U+206F .....	75
Superscripts and Subscripts U+2070 → U+209F .....	79
Currency Symbols U+20A0 → U+20CF .....	80
Diacritical Marks for Symbols U+20D0 → U+20FF .....	81
Letterlike Symbols U+2100 → U+214F .....	82
Number Forms U+2150 → U+218F .....	83
Arrows U+2190 → U+21FF .....	84
Mathematical Operators U+2200 → U+22FF .....	85
Miscellaneous Technical U+2300 → U+23FF .....	87

<i>Pictures for Control Codes U+2400 → U+243F</i> .....	88
<i>Optical Character Recognition U+2440 → U+245F</i> .....	89
<i>Enclosed Alphanumerics U+2460 → U+24FF</i> .....	90
<i>Form and Chart Components U+2500 → U+257F</i> .....	91
<i>Blocks U+2580 → U+259F</i> .....	92
<i>Geometric Shapes U+25A0 → U+25FF</i> .....	93
<i>Miscellaneous Dingbats U+2600 → U+26FF</i> .....	94
<i>Zapf Dingbats U+2700 → U+27BF</i> .....	96
3.3 CJK Phonetics and Symbols .....	98
<i>CJK Symbols and Punctuation U+3000 → U+303F</i> .....	99
<i>Hiragana U+3040 → U+309F</i> .....	100
<i>Katakana U+30A0 → U+30FF</i> .....	101
<i>Bopomofo U+3100 → U+312F</i> .....	102
<i>Hangul Elements U+3130 → U+318F</i> .....	103
<i>CJK Miscellaneous U+3190 → U+31FF</i> .....	104
<i>Enclosed CJK Letters and Ideographs U+3200 → U+32FF</i> .....	105
<i>CJK Squared Words U+3300 → U+337F</i> .....	106
<i>CJK Squared Abbreviations U+3380 → U+33FF</i> .....	107
<i>Korean Hangul Syllables U+3400 → U+3D2F</i> .....	108
3.4 CJK Ideographs .....	110
<i>Chinese/Japanese/Korean Ideographs U+4000 → U+8BFF</i> .....	111
3.5 Private Use Area .....	118
<i>Private Use Area U+E800 → U+FDFF</i> .....	119
3.6 Compatibility Zone .....	120
<i>Compatibility Zone U+FE00 → U+FFEF</i> .....	121
<i>Special U+FFF0 → U+FFFD</i> .....	123
3.7 Code Charts .....	125
3.8 Block-by-Block Charts .....	171
<b>4.0 Character Properties</b> .....	<b>389</b>
4.1 Numeric .....	390
<i>Decimal Digits</i> .....	391
<i>Digits</i> .....	395
<i>Numbers</i> .....	397
4.2 Space Characters .....	400
4.3 Dashes .....	401
4.4 Line Breaking .....	402
4.5 Non-spacing Marks .....	403
4.6 Directional Character Types .....	407

<b>5.0 Internal Mapping Tables</b> .....	411
5.1 Composite Character Mappings .....	412
<i>Unicode Decomposition Mapping—Precomposed Letters</i> .....	413
<i>Unicode Decomposition Mapping—Ligatures &amp; Digraphs</i> .....	419
5.2 Compatibility Mappings .....	420
<i>Unicode Compatibility Zone Mapping</i> .....	421
5.3 Vietnamese Mappings .....	432
<i>Unicode Vietnamese Decomposition</i> .....	433
5.4 Polytonic Greek to Unicode Characters .....	436
<i>Unicode Polytonic Greek Decomposition</i> .....	437
5.5 Case Tables .....	440
<i>Unicode Upper-case to Lower-case Character Mapping</i> .....	441
<i>Unicode Lower-case to Upper-case Character Mapping</i> .....	453
<b>6.0 External Mapping Tables</b> .....	465
6.1 ISO Mapping Tables .....	466
6.2 Vendor Mapping Tables .....	494
Appendix A: Directionality .....	611
Appendix B: Implementation Guidelines .....	623
Appendix C: How to Deal with Unknown or Missing Characters .....	629
Appendix D: How to Submit New Characters for Consideration .....	631
Appendix E: Proposed Scripts for Future Versions of the Unicode Standard .....	635
Appendix F: Glossary .....	637
References .....	645
Index to Names of Characters .....	651
General Index .....	679