Appendix E: Table of Contents to Volume 1

Notations and Conventions xv
Codes xv
Images xv
Names xvi
Names List Annotations xvii

1.0 Introduction 1
1.1 Background 1
1.2 Conformance 2
1.3 Coverage 2
Statistics of Version 1.0 4
Future Plans 4
1.4 The Unicode Consortium 5
The Unicode Technical Committee 5

2.0 General Principles of the Unicode Standard 7
2.1 Architectural Context of a Character Encoding 7
Text Elements, Code Elements, and Text Processes 7
Plain and Fancy Text 9
2.2 The Basic Design of the Unicode Character Encoding 10
Unicode Principles 10
Glyphs 12
2.3 Codespace Allocation 13
2.4 Character Properties, Controls, and Sequences 14
Properties 15
Control Characters 15
Paragraph/Line Separators 16
Ordering of Character Sequences 16
Alternate Spellings 17
2.5 Non-spacing Marks 18
Sequence of Base Letters and Non-spacing Marks 19

Spacing Clones of European Diacritical Marks 19
Multiple Non-spacing Marks 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
Pictures for Control Codes U+2400 → U+243F 88
Optical Character Recognition U+2440 → U+245F 89
Enclosed Alphanumerics U+2460 → U+24FF 90
Form and Chart Components U+2500 → U+257F 91
Blocks U+2580 → U+259F 92
Geometric Shapes U+25A0 → U+25FF 93
Miscellaneous Dingbats U+2600 → U+26FF 94
Zapf Dingbats U+2700 → U+27BF 96
3.3 CJK Phonetics and Symbols 98
CJK Symbols and Punctuation U+3000 → U+303F 99
Hiragana U+3040 → U+309F 100
Katakana U+30A0 → U+30FF 101
Bopomofo U+3100 → U+312F 102
Hangul Elements U+3130 → U+318F 103
CJK Miscellaneous U+3190 → U+31FF 104
Enclosed CJK Letters and Ideographs U+3200 → U+32FF 105
CJK Squared Words U+3300 → U+337F 106
CJK Squared Abbreviations U+3380 → U+33FF 107
Korean Hangul Syllables U+3400 → U+3D2F 108
3.4 CJK Ideographs 110
Chinese/Japanese/Korean Ideographs U+4000 → U+8BFF 111
3.5 Private Use Area 118
Private Use Area U+E800 → U+FDFF 119
3.6 Compatibility Zone 120
Compatibility Zone U+FE00 → U+FFEF 121
Special U+FFF0 → U+FFFD 123
3.7 Code Charts 125
3.8 Block-by-Block Charts 171

4.0 Character Properties 389
4.1 Numeric 390
Decimal Digits 391
Digits 395
Numbers 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

5.0 Internal Mapping Tables 411
5.1 Composite Character Mappings 412
Unicode Decomposition Mapping—Precomposed Letters 413
Unicode Decomposition Mapping—Ligatures & Digraphs 419
5.2 Compatibility Mappings 420
Unicode Compatibility Zone Mapping 421
5.3 Vietnamese Mappings 432
Unicode Vietnamese Decomposition 433
5.4 Polytonic Greek to Unicode Characters 436
Unicode Polytonic Greek Decomposition 437
5.5 Case Tables 440
Unicode Upper-case to Lower-case Character Mapping 441
Unicode Lower-case to Upper-case Character Mapping 453

6.0 External Mapping Tables 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