

Contents

Preface vii
Introduction xi

Chapter 1 Scripts and Languages 1

Writing Systems, Scripts, and Languages 2
Categories of Writing Systems 3
Characters and Glyphs 6
Chapter Summary 10

Chapter 2 Characters and Terminology in Computing 11

Bits and Bytes 12
A Brief History of Character Encoding Standards 13
 ASCII 14
 ISO 8859 16
 Other 8-Bit Encoding Schemes 21
 Multi-octet Encoding Schemes 28
Problems with Encodings 31
 Unicode 31
 Transcoding 34
Chapter Summary 37

Chapter 3 The SAS Approach to Encodings 41

Encoding in SAS 6 42
 Support for European Locales 42
 Host-to-Host Translation Tables 42
 Transport-Format Translation Tables 43
 Support for Asian Locales 45

Encoding in SAS 8	46
Encoding in SAS 9	47
Encoding of External Files	50
Encoding of SAS Data Sets	51
Encoding of RDBMS Tables	54
Encoding of Output	64
The SAS/GRAPH Approach to Encoding	64
SAS Fonts	66
Transcoding Revisited	67
How Do We Know Which Encodings Are Compatible with Each Other?	67
How to Check Character Encodings	79
Chapter Summary	82

Chapter 4 Troubleshooting Encoding Problems and FAQ 85

General Remarks	86
SAS Encoding and Locale-Related System Options	86
Operating System-Specific Options	88
Windows	88
UNIX	92
X Window System	94
z/OS	103
The Next Steps	104
How to Handle Specific Issues	104
Transcoding Problems	105
Multilingual Data Handling	115
Garbage Characters in Output or Display	121
Problems with RDBMS Access	127
Chapter Summary	132

Conclusion 135

Glossary 137

References 155

- References and Further Reading 156
 - Books 156
 - Articles and Papers 159
- Online Resources 160
 - Scripts and Locales 160
 - General Introductions and Tutorials 160
 - Early History of Character Set Standardization 160
 - Code Charts 161

Index 163

