ISO 8879: 1986-10(E)

Information processing; Text and office systems; Standard Generalized Markup Language (SGML)

Contents

0 Introduction **1**
- **0.1** Background 1
- **0.2** Objectives 2
- **0.3** Organization 3

1 Scope **4**

2 Field of Application **4**

3 References **5**

4 Definitions **5**

5 Notation **20**
- **5.1** Syntactic Tokens 21
- **5.2** Ordering and Selection Symbols 21

6 Entity Structure **21**
- **6.1** SGML Document 21
- **6.2** SGML Entities 21
  - **6.2.1** S Separator 22
  - **6.2.2** Entity End 22
  - **6.2.3** Implied SGML Declaration 22
- **6.3** Non-SGML Data Entity 22

7 Element Structure **22**
- **7.1** Prolog 22
- **7.2** Document Element 22
  - **7.2.1** Limits 22
- **7.3** Element 22
  - **7.3.1** Omitted Tag Minimization 22
    - **7.3.1.1** Start-tag Omission 22
    - **7.3.1.2** End-tag Omission 23
  - **7.3.2** Data Tag Minimization 23
  - **7.3.3** Quantities 23
- **7.4** Start-tag 23
  - **7.4.1** Minimization 23
    - **7.4.1.1** Empty Start-tag 23
    - **7.4.1.2** Unclosed Start-tag 24
    - **7.4.1.3** NET-enabling Start-tag 24
  - **7.4.2** Quantities 24
- **7.5** End-tag 24
  - **7.5.1** Minimization 24
    - **7.5.1.1** Empty End-tag 24
    - **7.5.1.2** Unclosed End-tag 24
    - **7.5.1.3** Null End-tag 24
- **7.6** Content 24
  - **7.6.1** Record Boundaries 25
- **7.7** Document Type Specification 25
  - **7.7.1** General Entity References 25
  - **7.7.2** Parameter Entity References 25
- **7.8** Generic Identifier (GI) Specification 26
  - **7.8.1** Rank Feature 26
    - **7.8.1.1** Full Generic Identifier 26
    - **7.8.1.2** Rank Stem 26
- **7.9** Attribute Specification List 26

7.9.1 Minimization ... 26
7.9.1.1 Omitted Attribute Specification ... 26
7.9.1.2 Omitted Attribute Name ... 26
7.9.2 Quantities ... 26
7.9.3 Attribute Value Specification ... 26
7.9.3.1 Minimization ... 27
7.9.4 Attribute Value ... 27
7.9.4.1 Syntactic Requirements ... 27
7.9.4.2 Fixed Attribute ... 27
7.9.4.3 General Entity Name ... 27
7.9.4.4 Notation ... 27
7.9.4.5 Quantities ... 27

8 Processing Instruction ... 27
8.1 Quantities ... 28

9 Common Constructs ... 28
9.1 Replaceable Character Data ... 28
9.2 Character Data ... 28
9.2.1 SGML Character ... 28
9.2.2 Function Character ... 28
9.3 Name ... 28
9.3.1 Quantities ... 28
9.4 Entity References ... 28
9.4.1 Quantities ... 28
9.4.2 Limits ... 28
9.4.3 Obfuscatory Entity References ... 29
9.4.4 Named Entity Reference ... 29
9.4.5 Reference End ... 29
9.4.6 Short Reference ... 29
9.4.6.1 Equivalent Reference String ... 29
9.5 Character Reference ... 30
9.6 Delimiter Recognition ... 30
9.6.1 Recognition Modes ... 30
9.6.2 Contextual Constraints ... 31
9.6.3 Order of Recognition ... 32
9.6.4 Delimiters Starting with the Same Character ... 32
9.6.5 Short References with Blank Sequences ... 32
9.6.5.1 Quantities ... 32
9.6.6 Name Characters ... 32
9.7 Markup Suppression ... 32
9.8 Capacity ... 32

10 Markup Declarations: General ... 34
10.1 Parts of Declarations ... 34
10.1.1 Parameter Separator ... 34
10.1.2 Parameter Literal ... 34
10.1.2.1 Quantities ... 34
10.1.3 Group ... 34
10.1.3.1 Quantities ... 34
10.1.4 Declaration Separator ... 34
10.1.5 Associated Element Type ... 35
10.1.6 External Identifier ... 35
10.1.6.1 Quantities ... 35
10.1.6.2 Capacities ... 35
10.1.7 Minimum Literal ... 35
10.1.7.1 Quantities ... 35
10.2 Formal Public Identifier ... 35
10.2.1 Owner Identifier ... 35
10.2.1.1 ISO Owner Identifier ... 35
10.2.1.2 Registered Owner Identifier ... 35

12.1.4.3 Simple Link ... 45
12.2 Link Set Declaration ... 45
12.2.1 Source Element Specification ... 46
12.2.2 Result Element Specification ... 46
12.3 Link Set Use Declaration ... 46
12.3.1 Use in Link Type Declaration ... 46
12.3.2 Use in Document Instance ... 46
12.3.3 Current Link Set ... 46

13 SGML Declaration ... 47
13.1 Document Character Set ... 47
13.1.1 Character Set Description ... 47
13.1.1.1 Base Character Set ... 47
13.1.1.2 Described Character Set Portion ... 47
13.1.2 Non-SGML Character Identification ... 48
13.2 Capacity Set ... 48
13.3 Concrete Syntax Scope ... 48
13.4 Concrete Syntax ... 49
13.4.1 Public Concrete Syntax ... 49
13.4.2 Shunned Character Number Identification ... 49
13.4.3 Syntax-reference Character Set ... 50
13.4.4 Function Character Identification ... 50
13.4.5 Naming Rules ... 50
13.4.6 Delimiter Set ... 51
13.4.6.1 General Delimiters ... 51
13.4.6.2 Short Reference Delimiters ... 51
13.4.7 Reserved Name Use ... 51
13.4.8 Quantity Set ... 51
13.5 Feature Use ... 52
13.5.1 Markup Minimization Features ... 52
13.5.2 Link Type Features ... 52
13.5.3 Other Features ... 52
13.6 Application-specific Information ... 53

14 Reference and Core Concrete Syntaxes ... 53

15 Conformance ... 53
15.1 Conforming SGML Document ... 53
15.1.1 Basic SGML Document ... 53
15.1.2 Minimal SGML Document ... 53
15.1.3 Variant Conforming SGML Document ... 53
15.2 Conforming SGML Application ... 53
15.2.1 Application Conventions ... 53
15.2.2 Conformance of Documents ... 53
15.2.3 Conformance of Documentation ... 53
15.3 Conforming SGML System ... 53
15.3.1 Conformance of Documentation ... 54
15.3.2 Conformance to System Declaration ... 54
15.3.3 Support for Reference Concrete Syntax ... 54
15.3.4 Support for Reference Capacity Set ... 55
15.3.5 Consistency of Parsing ... 55
15.3.6 Application Conventions ... 55
15.4 Validating SGML Parser ... 56
15.4.1 Error Recognition ... 56
15.4.2 Identification of SGML Messages ... 56
15.4.3 Content of SGML Messages ... 56
15.5 Documentation Requirements ... 56
15.5.1 Standard Identification ... 56
15.5.2 Identification of SGML Constructs ... 56
15.5.3 Terminology ... 57
15.5.4 Variant Concrete Syntax ... 57

12.1.4.3 Simple Link 45
12.2 Link Set Declaration 45

15.6 System Declaration 57
15.6.1 Concrete Syntaxes Supported 57
15.6.1.1 Concrete Syntax Changes 57
15.6.1.2 Character Set Translation 57
15.6.2 Validation Services 58

Annexes

A Introduction to Generalized Markup 59

A.1 The Markup Process 59
A.2 Descriptive Markup 60
A.3 Rigorous Markup 62
A.4 Conclusion 64
A.5 Acknowledgments 65
A.6 Bibliography 65

B Basic Concepts 66

B.1 Documents, Document Type Definitions, and Procedures 66
B.1.1 Documents 66
B.1.2 Document Type Definitions 66
B.1.3 Procedures 67
B.2 Markup 67
B.3 Distinguishing Markup from Text 68
B.3.1 Descriptive Markup Tags 68
B.3.2 Other Markup 69
B.3.3 Record Boundaries 69
B.3.3.1 Record Boundaries in Data 70
B.3.3.2 Record Boundaries in Markup 70
B.4 Document Structure 70
B.4.1 Document Type Definitions 70
B.4.2 Element Declarations 71
B.4.2.1 Content Models 71
B.4.2.2 Connectors and Occurrence Indicators 71
B.4.2.3 Entity References in Models 72
B.4.2.4 Name Groups 72
B.4.2.5 Data Characters 73
B.4.2.6 Empty Content 73
B.4.2.7 Non-SGML Data 73
B.4.2.8 Summary of Model Delimiters 74
B.5 Attributes 74
B.5.1 Specifying Attributes 74
B.5.1.1 Names 74
B.5.1.2 Attribute Values 75
B.5.2 Declaring Attributes 75
B.5.2.1 Attribute Definition Syntax 75
B.5.2.2 Complex Attribute Values 76
B.5.2.3 Name Token Groups 77
B.5.2.4 Changing Default Values 77
B.6 Entities 78
B.6.1 Entity Reference Syntax 78
B.6.2 Declaring Entities 78
B.6.2.1 Processing Instructions 79
B.6.2.2 Entities with Entity References 79
B.6.2.3 External Entities 79
B.6.2.4 Public Entities 80
B.7 Characters 80

15.6 System Declaration ... 57
15.6.1 Concrete Syntaxes Supported ... 57
15.6.1.1 Concrete Syntax Changes ... 57
15.6.1.2 Character Set Translation ... 57
15.6.2 Validation Services ... 58

Annexes

A Introduction to Generalized Markup ... **59**

A.1 The Markup Process ... 59
A.2 Descriptive Markup ... 60
A.3 Rigorous Markup ... 62
A.4 Conclusion ... 64
A.5 Acknowledgments ... 65
A.6 Bibliography ... 65

B Basic Concepts ... **66**

B.1 Documents, Document Type Definitions, and Procedures ... 66
B.1.1 Documents ... 66
B.1.2 Document Type Definitions ... 66
B.1.3 Procedures ... 67
B.2 Markup ... 67
B.3 Distinguishing Markup from Text ... 68
B.3.1 Descriptive Markup Tags ... 68
B.3.2 Other Markup ... 69
B.3.3 Record Boundaries ... 69
B.3.3.1 Record Boundaries in Data ... 70
B.3.3.2 Record Boundaries in Markup ... 70
B.4 Document Structure ... 70
B.4.1 Document Type Definitions ... 70
B.4.2 Element Declarations ... 71
B.4.2.1 Content Models ... 71
B.4.2.2 Connectors and Occurrence Indicators ... 71
B.4.2.3 Entity References in Models ... 72
B.4.2.4 Name Groups ... 72
B.4.2.5 Data Characters ... 73
B.4.2.6 Empty Content ... 73
B.4.2.7 Non-SGML Data ... 73
B.4.2.8 Summary of Model Delimiters ... 74
B.5 Attributes ... 74
B.5.1 Specifying Attributes ... 74
B.5.1.1 Names ... 74
B.5.1.2 Attribute Values ... 75
B.5.2 Declaring Attributes ... 75
B.5.2.1 Attribute Definition Syntax ... 75
B.5.2.2 Complex Attribute Values ... 76
B.5.2.3 Name Token Groups ... 77
B.5.2.4 Changing Default Values ... 77
B.6 Entities ... 78
B.6.1 Entity Reference Syntax ... 78
B.6.2 Declaring Entities ... 78
B.6.2.1 Processing Instructions ... 79
B.6.2.2 Entities with Entity References ... 79
B.6.2.3 External Entities ... 79
B.6.2.4 Public Entities ... 80
B.7 Characters ... 80

B.7.1 Character Classification ... 80
B.7.2 Character References ... 81
B.7.3 Using Delimiter Characters as Data ... 82
B.8 Marked Sections ... 83
B.8.1 Ignoring a Marked Section ... 83
B.8.2 Versions of a Single Document ... 84
B.8.3 Unparsable Sections ... 84
B.8.4 Temporary Sections ... 85
B.8.5 Keyword Specification ... 85
B.8.6 Defining a Marked Section as an Entity ... 85
B.9 Unique Identifier Attributes ... 86
B.10 Content Reference Attributes ... 86
B.11 Content Model Exceptions ... 87
B.11.1 Included Elements ... 87
B.11.2 Excluded Elements ... 87
B.12 Document Type Declaration ... 88
B.13 Data Content ... 88
B.13.1 Data Content Representations ... 89
B.13.1.1 Character Data (PCDATA, CDATA, and RCDATA) ... 89
B.13.1.2 Non-SGML Data (NDATA) ... 89
B.13.2 Data Content Notations ... 90
B.13.2.1 Notations for Character Data ... 90
B.13.2.2 Notations for Non-SGML Data ... 91
B.13.2.3 Specifying Data Content Notations ... 91
B.14 Customizing ... 92
B.14.1 The SGML Declaration ... 92
B.14.1.1 Optional Features ... 92
B.14.1.2 Variant Concrete Syntax ... 92
B.14.2 Impact of Customization ... 92
B.15 Conformance ... 93

C Additional Concepts ... **94**

C.1 Markup Minimization Features ... 94
C.1.1 SHORTTAG: Tags With Omitted Markup ... 94
C.1.1.1 Unclosed Short Tags ... 95
C.1.1.2 Empty Tags ... 95
C.1.1.3 Attribute Minimization ... 95
C.1.2 OMITTAG: Tags May be Omitted ... 96
C.1.2.1 Tag Omission Concepts ... 97
C.1.2.2 Specifying Minimization ... 97
C.1.2.3 End-tag Omission: Intruding Start-tag ... 98
C.1.2.4 End-tag Omission: End-tag of Containing Element ... 98
C.1.2.5 Start-tag Omission: Contextually Required Element ... 99
C.1.2.6 Combination with Short Tag Minimization ... 99
C.1.2.7 Markup Minimization Considerations ... 99
C.1.3 SHORTREF: Short Reference Delimiters May Replace Complete Entity References ... 100
C.1.3.1 Typewriter Keyboarding: Generalized WYSIWYG ... 100
C.1.3.2 Typewriter Keyboarding Example: Defining a Short Reference Map ... 100
C.1.3.3 Typewriter Keyboarding Example: Activating a Short Reference Map ... 101
C.1.3.4 Tabular Matter Example ... 102
C.1.3.5 Special Requirements ... 103
C.1.4 DATATAG: Data May Also be a Tag ... 103
C.1.5 RANK: Ranks May be Omitted from Tags ... 106
C.2 LINK Features: SIMPLE, IMPLICIT, and EXPLICIT ... 107
C.2.1 Link Process Definitions ... 108
C.3 Other Features ... 108

C.3.1 CONCUR: Document Instances May Occur Concurrently 108
C.3.2 SUBDOC: Nested Subdocument Entities May Occur 109
C.3.3 FORMAL: Public Identifiers are Formal 109

D Public Text **110**

D.1 Element Sets 110
D.1.1 Common Element Types 110
D.1.2 Pro Forma Element Types 110
D.2 Data Content Notations 110
D.3 Variant Concrete Syntaxes 111
D.3.1 Multicode Concrete Syntaxes 111
D.4 Entity Sets 111
D.4.1 General Considerations 112
D.4.1.1 Format of Declarations 112
D.4.1.2 Corresponding Display Entity Sets 113
D.4.1.3 Entity Names 113
D.4.1.4 Organization of Entity Sets 114
D.4.2 Alphabetic Characters 114
D.4.2.1 Latin 114
D.4.2.2 Greek Alphabetic Characters 117
D.4.2.3 Cyrillic Alphabetic Characters 119
D.4.3 General Use 121
D.4.3.1 Numeric and Special Graphic Characters 121
D.4.3.2 Diacritical Mark Characters 123
D.4.3.3 Publishing Characters 123
D.4.3.4 Box and Line Drawing Characters 125
D.4.4 Technical Use 126
D.4.4.1 General 126
D.4.4.2 Greek Symbols 128
D.4.4.3 Alternative Greek Symbols 129
D.4.5 Additional Mathematical Symbols 130
D.4.5.1 Ordinary Symbols 130
D.4.5.2 Binary and Large Operators 130
D.4.5.3 Relations 131
D.4.5.4 Negated Relations 133
D.4.5.5 Arrow Relations 134
D.4.5.6 Opening and Closing Delimiters 135

E Application Examples **136**

E.1 Document Type Definition 136
E.2 Computer Graphics Metafile 140
E.3 Device-Independent Code Extension 140
E.3.1 Code Extension Facilities 140
E.3.1.1 Avoiding False Delimiter Recognition 141
E.3.1.2 Eliminating Device and Code Dependencies 143

F Implementation Considerations **145**

F.1 A Model of SGML Parsing 145
F.1.1 Physical Input 145
F.1.1.1 Entities 145
F.1.1.2 Record Boundaries 145
F.1.2 Recognition Modes 145
F.1.3 Markup Minimization 146
F.1.4 Translation 147
F.1.5 Command Language Analogy 147

F.2 Initialization 147
F.2.1 Initial Procedure Mapping 147
F.2.2 Link Process Specification 147
F.2.3 Concurrent Document Instances 147
F.3 Dynamic Procedure Mapping 148
F.4 Error Handling 148

G Conformance Classification and Certification 149

G.1 Classification Code 149
G.1.1 Feature Code 149
G.1.2 Validation Code 150
G.1.3 Syntax Code 151
G.2 Certification Considerations 151

H Theoretical Basis for the SGML Content Model 152

H.1 Model Group Notation 152
H.2 Application of Automata Theory 152
H.3 Divergence from Automata Theory 153

I Nonconforming Variations 154

I.1 Fixed-length Generic Identifiers 154
I.2 Single Delimiter 154

Figures

1 Character Classes: Abstract Syntax 29
2 Character Classes: Concrete Syntax 30
3 Reference Delimiter Set: General 31
4 Reference Delimiter Set: Short References 33
5 Reference Capacity Set 49
6 Reference Quantity Set 52
7 Reference Concrete Syntax 54
8 Typical SGML Declaration for Basic SGML Document 55
9 Element Markup 69
10 Start-tag with 2 Attributes 75
11 Multicode Basic Concrete Syntax 112
12 Graphics Metafile Attributes (1 of 2): Encoding and View 141
13 Graphics Metafile Attributes (2 of 2): Size and Rotation 142
14 Function Characters for Device-Independent Multicode Concrete Syntaxes 143
15 FSV Conformance Classification 150