

Australian Standard[®]

**Information processing—
SGML support facilities—
Techniques for using SGML**

Part 1: SGML tutorial

This Australian Standard was prepared by Committee IT/1, Information Systems—Interconnection. It was approved on behalf of the Council of Standards Australia on 27 May 1991 and published on 9 August 1991.

The following interests are represented on Committee IT/1:

AUSSAT

Australian Association of Permanent Building Societies

Australian Bankers Association

Australian Bureau of Statics

Australian Committee of Directors and Principals

Australian Computer Society

Australian Computer Users Association

Australian Information Industry Association

Australian Telecommunications Users Group

Australian Vice Chancellors Committee

Confederation of Australian Industry

CSIRO—Institute of Information and Communication Technologies

Department of Defence

Department of Industry, Technology and Commerce

Information Exchange Steering Committee

Life Insurance Federation of Australia

OTC

Telecom Australia

Review of Australian Standards. To keep abreast of progress in industry, Australian Standards are subject to periodic review and are kept up to date by the issue of amendments or new editions as necessary. It is important therefore that Standards users ensure that they are in possession of the latest edition, and any amendments thereto.

Full details of all Australian Standards and related publications will be found in the Standards Australia Catalogue of Publications; this information is supplemented each month by the magazine 'The Australian Standard', which subscribing members receive, and which gives details of new publications, new editions and amendments, and of withdrawn Standards.

Suggestions for improvements to Australian Standards, addressed to the head office of Standards Australia, are welcomed. Notification of any inaccuracy or ambiguity found in an Australian Standard should be made without delay in order that the matter may be investigated and appropriate action taken.

Australian Standard[®]

**Information processing—
SGML support facilities—
Techniques for using SGML**

Part 1: SGML tutorial

First published as AS 3967—1991.

Appending:
Amdt 1—1993

PUBLISHED BY STANDARDS AUSTRALIA
(STANDARDS ASSOCIATION OF AUSTRALIA)
1 THE CRESCENT, HOMEBUSH, NSW 2140

ISBN 0 7262 7030 X

PREFACE

This Standard was prepared by the Standards Australia Committee on Information Technology—Interconnection. It is identical with and has been reproduced from ISO/IEC/TR 9573:1988, *Information processing—SGML support facilities—Techniques for using SGML*.

Under arrangements made between Standards Australia and the international Standards bodies, ISO and IEC, as well as certain other Standards organizations, users of this Australian Standard are advised of the following:

- (a) Copyright is vested in Standards Australia.
- (b) The number of this Standard is not reproduced on each page; its identity is shown only on the cover and title pages.

For the purpose of this Australian Standard, the ISO/IEC text should be modified as follows:

- (i) *Terminology* The words 'Australian Standard' should replace the words 'International Standard' wherever they appear.
- (ii) *References* The references to International Standards should be replaced by references to Australian Standards as follows:

<i>Reference to International Standard or other Publication</i>		<i>Australian Standard</i>	
ISO		AS	
646	Information processing—ISO 7-bit coded character set for information interchange	1776	Information processing—7-bit coded character set for information interchange
2022	Information processing—ISO 7-bit and 8-bit coded character sets—Code extension techniques	1953	Information processing—ISO 7-bit and 8-bit coded character sets—Code extension techniques
4873	Information processing—ISO 8-bit code for information interchange—Structure and rules for implementation	3716	Character-sets and information coding—ISO 8-bit code for information interchange—Structure and rules for implementation
6937	Information processing—Coded character sets for text communication	2793	Information processing—Coded character sets for text communication
8632	Information processing systems—Computer graphics—Metafile for the storage and transfer of picture description information	3603	Computer graphics—Metafiles for the storage and transfer of picture description information
8632-2	Part 2: Character encoding	3603.2	Part 2: Character encoding
8632-4	Part 4: Clear text encoding	3603.4	Part 4: Clear text encoding
8859	Information processing—8-bit single-byte coded graphic character sets	—	
8859-6	Part 6: Latin/Arabic alphabet	—	
8879	Information processing—Text and office systems—Standard Generalized markup Language (SGML)	3514	Information processing—Text and office systems—Standard generalized markup language (SGML)
9069	Information processing—SGML support facilities—SGML Document Interchange Format (SDIF)	—	
9070	Information processing—SGML support facilities—Registration procedures of public text owner identifiers	—	
JIS			
X 0201	(formerly C 6220) Code for Information Interchange	—	
X 0208	(formerly C 6226) Code of the Japanese Graphic Character Set for Information Interchange	—	

CONTENTS

	<i>Page</i>
0 Introduction	7
1 Scope and Field of Application	7
2 References	8
3 Definitions	8
4 Document Analysis for the Creation of a Document Type Definition	8
4.1 General notes	8
4.1.1 Steps of document analysis	9
4.1.2 Design issues	9
4.2 Example of the analysis of body text	11
4.2.1 Headed sections	13
4.2.2 Paragraph and note elements	13
4.2.3 List elements	14
4.2.3.1 Separately identified elements	14
4.2.3.2 Common list element	15
4.2.3.3 Discussion of the two approaches to lists	15
4.2.4 Footnote element	16
4.2.5 Figure element	16
4.2.6 Floating elements	16
4.2.7 Reference elements	17
4.2.8 Summary of the definition of the body element	17
4.2.9 Additional elements	18
4.2.10 Markup minimization	18
4.3 Example of the analysis of title matter	18
4.3.1 Title element	19
4.3.2 Instlist element	20
4.3.3 Authlist element	21
4.3.4 Date element	22
4.3.5 Abstract element	23
4.3.6 Article element	23
4.3.7 Complete definition of the title matter	24
4.4 Example of the analysis of tabular material	25
4.5 Example of the analysis of a simple mathematical formula	27
4.6 Example of the analysis of an auction catalogue	28
4.7 The DTD with relation to the size of a document	29
5 General Document	29
5.1 General notes	29
5.2 Document structure	29
5.3 Description of the tags	37
5.3.1 Major document elements	37
5.3.2 Title page elements	38
5.3.3 Front matter elements	39
5.3.4 Back matter elements	40
5.3.5 Headed sections	40
5.3.6 Topics	40
5.3.7 Tabular material	41
5.3.8 Figure	41
5.3.9 Phrases	42
5.3.10 Lists	43
5.3.11 References	45
5.3.12 Basic elements	46

5.4	Alphabetical list of the generic identifiers and their attributes	48
5.5	The document type definition	49
5.6	Commentary on the document type definition	53
5.6.1	Parameter entity declarations	53
5.6.2	Element and attribute declarations	53
5.6.3	Short references	53
5.7	National versions of the elements of the general document	54
5.7.1	French	54
5.7.2	German	57
5.7.3	Swedish	60
5.7.4	Danish	63
5.7.5	Dutch	66
5.7.6	Generic identifiers	69
5.8	Extensions to the general document	70
5.8.1	Extending the bibliography to include a bibliography list	70
5.8.2	Extending the DTD to include tables	71
6	Letter and Memorandum	72
6.1	Constituent parts of the documents	72
6.1.1	Common document elements	72
6.1.2	Elements specific to a business letter	74
6.1.3	Elements specific to an inter-office memorandum	75
6.1.4	Alphabetical list of the tags	75
6.2	Entity set	75
6.3	SGML features	76
6.4	Typical document type declarations	76
6.4.1	The document type declaration for letters	76
6.4.2	The document type declaration for memoranda	76
6.5	The common document type definition	76
6.6	A sample letter	77
6.7	A sample inter-office memorandum	78
7	Spreadsheet	78
7.1	Constituent parts of a spreadsheet	79
7.1.1	Elements of a spreadsheet	79
7.1.2	Elements of a formula	80
7.1.3	Alphabetical list of the tags	82
7.1.4	Entity sets	82
7.1.4.1	Parameter entities	83
7.1.4.2	Entities used with short references	73
7.1.4.3	Short references, notations, maps, etc	83
7.1.5	SGML features	83
7.2	The document type definition	83
7.3	A sample spreadsheet	85
8	Mathematics	89
8.1	Concepts	89
8.2	User-interface	89
8.3	Symbols	90
8.4	Description of the elements	90
8.4.1	Formula and formula reference	90
8.4.2	Formula content	91
8.5	Short references	98
8.6	Complex examples	98
8.7	Element and entity definition	99

	<i>Page</i>
9 Tables	104
9.1 Table elements	104
9.2 Element definition	105
9.3 Example of a marked-up table	105
10 Integration of Computer Graphics Metafiles	106
11 Japanese	111
11.1 Special considerations	111
11.1.1 Horizontal and vertical typesetting	111
11.1.2 Sentences including words in the Latin alphabet	111
11.1.3 Ruby	113
11.2 The SGML declaration	113
11.3 The document type definition and source document	115
12 Linguistic Application — Scandinavian Runes	117
12.1 Use of tags and entities	117
12.2 Character extension techniques	119
12.3 Character extension techniques mapped onto tags using short references	120
12.4 Definition of a special character set in the SGML declaration	121
13 European Multi-lingual Document	123
14 The Use of Short References	124
14.1 Single level quotation in English	124
14.2 Limitations of the use of short references to be observed	125
15 Mixing Text in languages Written from Left to Right and from Right to Left	126
15.1 Right to left base text mixed with text written from left to right	127
15.2 Left to right base text mixed with text written from right to left	128
15.2.1 Use of tags	128
15.2.2 Use of an 8-bit Latin/Arabic character set	129

Figures

1 Sample pages from the body of an International Standard	12
2 Sample title page	19
3 Overall structure of the basic document	30
4 Structure of the h2 element	31
5 Structure of the phrase (%m.ph), paragraph (%m.p), and paragraph sequence (%m.pseq) models	32
6 Structure of the section subelements	33
7 Structure of all phrases	34

	<i>Page</i>
8 Structure of all paragraph/section subelements	35
9 Structure of all floating elements	37
10 Overall structure of a spreadsheet	79
11 Structure of a formula in a spreadsheet	81
12 Japanese: horizontal typesetting (text written from left to right)	112
13 Japanese: vertical typesetting (text written from top to bottom)	113
14 Japanese: source document	116

Information processing—SGML support facilities—Techniques for using SGML

0 Introduction

ISO 8879, *Information processing — Text and office systems — Standard Generalized Markup Language (SGML)*, states the rules for the description and markup of documents for their publication and interchange. A basic document type, primarily for computer-assisted publishing, is provided as an example in clause E.1 of ISO 8879, but is not explained there. It is explained in this Technical Report to assist comprehension. Also given is a variety of examples on the use of SGML. Thus this Technical Report is complementary to ISO 8879, its principal purpose being to assist in the adoption of the language.

1 Scope and Field of Application

This Technical Report complements ISO 8879 by providing additional tutorial information. It is not intended, and should not be regarded, as an extension, modification, or interpretation of ISO 8879. The SGML language contains a number of components, some of which are optional features. The tutorial information covers the main components of the language only; notably a discussion of LINK, CONCUR, and DATATAG is outside the scope of this Technical Report.

The intended audience is mainly document type designers already familiar with the basic concepts of SGML, but requiring more tutorial information on techniques for using SGML for various applications. Subclauses 5.3 and 8.4 are written in the style of a "User Guide" and can be used as a basis for end-user documentation. For an introductory tutorial on SGML the annexes of ISO 8879 can be used.

This Technical Report includes notes on the analysis of a document prior to the writing of a formal document type definition, and a series of examples.

The principal example is for a general document type, formally defined as an example in clause E.1 of ISO 8879. Others of a general nature are for letter and memorandum, spreadsheet, mathematics, and the mixing of text and graphics. Those for language applications include Scandinavian runes, Japanese, a European multilingual document, and mixing text in languages written from left to right and from right to left.

NOTE — Throughout this Technical Report terms like "is keyed in", and "with keyboarding" are used. This does not necessarily imply that the markup is to be added explicitly by a user; for text entry one would expect structured, context sensitive, editors to be used, or the markup added by application programs, e.g. in the case of interchange of spreadsheets.