
Norme internationale



8879

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION • МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ • ORGANISATION INTERNATIONALE DE NORMALISATION

Traitement de l'information — Systèmes bureautiques — Langage normalisé de balisage généralisé (SGML)

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

Première édition — 1986-10-15

(incorporant l'Amendement 1 : 1988)

ITeH STANDARD PREVIEW
(standards.iteh.ai)

[ISO 8879:1986](https://standards.iteh.ai/catalog/standards/sist/bbf25d89-cdc0-4eac-8ec4-00cfb29943ad/iso-8879-1986)

<https://standards.iteh.ai/catalog/standards/sist/bbf25d89-cdc0-4eac-8ec4-00cfb29943ad/iso-8879-1986>

CDU 681.3.06 : 655.53

Réf. n° : ISO 8879-1986 (F)

Descripteurs : traitement de l'information, traitement de textes, échange d'information, documentation, document, structure logique, programmation des calculateurs, langage artificiel, langage de programmation, SGML.

Prix basé sur 164 pages

Avant-propos

L'ISO (Organisation internationale de normalisation) est une fédération mondiale d'organismes nationaux de normalisation (comités membres de l'ISO). L'élaboration des Normes internationales est confiée aux comités techniques de l'ISO. Chaque comité membre intéressé par une étude a le droit de faire partie du comité technique créé à cet effet. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO, participent également aux travaux.

Les projets de Normes internationales adoptés par les comités techniques sont soumis aux comités membres pour approbation, avant leur acceptation comme Normes internationales par le Conseil de l'ISO. Les Normes internationales sont approuvées conformément aux procédures de l'ISO qui requièrent l'approbation de 75 % au moins des comités membres votants.

La Norme internationale ISO 8879 a été élaborée par le comité technique ISO/TC 97, *Systèmes de traitement de l'information*. Cette version française incorpore l'Amendement 1 qui a été publié séparément en anglais en 1988.

<https://standards.iteh.ai/catalog/standards/sist/bb25d89-cdc0-4eac-8ec4-00c1b29943ad/iso-8879-1986>
ISO 8879:1986

L'attention des utilisateurs est attirée sur le fait que toutes les Normes internationales sont de temps en temps soumises à révision et que toute référence faite à une autre Norme internationale dans le présent document implique qu'il s'agit, sauf indication contraire, de la dernière édition.

Page blanche

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 8879:1986

<https://standards.iteh.ai/catalog/standards/sist/bbf25d89-cdc0-4eac-8ec4-00cfb29943ad/iso-8879-1986>

Table des matières

0 Introduction	1
0.1 Le contexte	1
0.2 Objectifs	2
0.3 Organisation	4
1 Objet	4
2 Domaine d'application	5
3 Références	5
4 Définitions	5
5 Notation	23
5.1 Unités lexicales syntaxiques	23
5.2 Symboles d'ordre et de choix	23
6 Structuration en entités	23
6.1 Document SGML	23
6.2 Entités SGML	24
6.3 Séparateur S	24
6.4 Fin d'entité	24
6.5 Déclaration SGML déduite	24
6.6 Entités textuelles	24
7 Structuration en éléments	25
7.1 Prologue	25
7.2 Élément document	25
7.2.1 Limites	25
7.3 Élément	25
7.3.1 Minimisation par omission de balises	25
7.3.1.1 Omission d'une balise de début	25
7.3.1.2 Omission d'une balise de fin	26
7.3.2 Minimisation du balisage par le texte	26
7.3.3 Quantités	26
7.4 Balise de début	26
7.4.1 Minimisation	26
7.4.1.1 Balise de début vide	26
7.4.1.2 Balise de début non fermée	27
7.4.1.3 Balise de début autorisant une balise de fin nulle	27
7.4.2 Quantités	27
7.5 Balise de fin	27
7.5.1 Minimisation	27
7.5.1.1 Balise de fin vide	27
7.5.1.2 Balise de fin non fermée	27
7.5.1.3 Balise de fin nulle	27
7.6 Contenu	28
7.6.1 Limites d'enregistrement	28
7.7 Spécification de type de document	29
7.8 Spécification de l'identificateur générique (IG)	29
7.8.1 Fonctionnalité niveau	29
7.8.1.1 Identificateur générique complet	29

7.8.1.2	Radical de niveau	29
7.9	Liste de spécifications d'attributs	29
7.9.1	Minimisation	29
7.9.1.1	Omission de la spécification d'attribut	29
7.9.1.2	Omission du nom d'attribut	30
7.9.2	Quantités	30
7.9.3	Spécification de valeur d'attribut	30
7.9.3.1	Minimisation	30
7.9.4	Valeur d'attribut	30
7.9.4.1	Exigences syntaxiques	30
7.9.4.2	Attribut fixé	30
7.9.4.3	Nom d'entité générale	31
7.9.4.4	Notation	31
7.9.5	Quantités	31
8	Instructions de traitement	31
8.1	Quantités	31
9	Structures communes	31
9.1	Données textuelles remplaçables	31
9.2	Données textuelles	32
9.2.1	Caractère SGML	32
9.2.2	Caractère de fonction	32
9.3	Nom	32
9.3.1	Quantités	32
9.4	Appels d'entité	32
9.4.1	Quantités	33
9.4.2	Limites	33
9.4.3	Appels d'entités obscurs	33
9.4.4	Appel d'entité désignée	33
9.4.4.1	Déclaration d'entité applicable	33
9.4.5	Fin d'appel	34
9.4.6	Appels abrégés	34
9.4.6.1	Chaîne de référence équivalente	34
9.5	Appel de caractère	35
9.6	Reconnaissance des délimiteurs	35
9.6.1	Modes de reconnaissance	35
9.6.2	Contraintes liées au contexte	36
9.6.3	Ordre de reconnaissance	37
9.6.4	Délimiteurs commençant par le même caractère	37
9.6.5	Appels abrégés avec séquences de blancs	37
9.6.5.1	Quantités	37
9.6.6	Caractères constitutifs de noms	37
9.7	Suppression du balisage	37
9.8	Capacité	37
10	Déclarations de balisage : généralités	38
10.1	Parties de la déclaration	39
10.1.1	Séparateur de paramètre	39
10.1.2	Quantités	39
10.1.3	Groupe	39
10.1.3.1	Quantités	39
10.1.4	Séparateur de déclarations	39
10.1.5	Type d'élément associé	40
10.1.6	Identificateur externe	40
10.1.6.1	Quantités	40
10.1.6.2	Capacités	40
10.1.7	Contenu littéral minimum	40
10.1.7.1	Quantités	40

10.2	Identificateur public formel	40
10.2.1	Identificateur de dépositaire	40
10.2.1.1	Identificateur de dépositaire ISO	41
10.2.1.2	Identificateur de dépositaire enregistré	41
10.2.1.3	Identificateur de dépositaire non enregistré	41
10.2.2	Identificateur de texte	41
10.2.2.1	Classe de texte public	41
10.2.2.2	Description de texte public	41
10.2.2.3	Langue du texte public	41
10.2.2.4	Séquence de désignation de texte public	42
10.2.2.5	Version affichée de texte public	42
10.3	Déclaration de commentaire	42
10.4	Déclaration de section marquée	42
10.4.1	Quantités	42
10.4.2	Spécification de mot-clé de statut	42
10.5	Déclaration d'entité	43
10.5.1	Nom d'entité	43
10.5.1.1	Quantités	43
10.5.1.2	Capacités	43
10.5.2	Texte d'entité	43
10.5.3	Texte brut	44
10.5.4	Texte entre crochets	44
10.5.4.1	Quantités	44
10.5.5	Spécification d'entité externe	44
11	Déclarations de balisage : Définition de type de document	45
11.1	Déclaration de type de document	45
11.2	Déclaration d'élément	45
11.2.1	Type d'élément	45
11.2.1.1	Élément hiérarchisé	45
11.2.1.2	Quantités	46
11.2.2	Minimisation par omission de balises	46
11.2.3	Contenu déclaré	46
11.2.4	Modèle de contenu	46
11.2.4.1	Connecteur	47
11.2.4.2	Symbole d'occurrence	47
11.2.4.3	Modèle de contenu ambigu	47
11.2.4.4	Groupe de balises textuelles	47
11.2.4.5	Quantités	48
11.2.5	Exceptions	48
11.2.5.1	Inclusions	48
11.2.5.2	Exclusions	48
11.3	Déclaration de liste de définitions d'attributs	49
11.3.1	Quantités	49
11.3.2	Nom d'attribut	49
11.3.3	Valeur déclarée	49
11.3.4	Valeur par défaut	49
11.3.4.1	Quantités	50
11.3.4.2	Capacités	50
11.4	Déclaration de notation	50
11.4.1	Attributs du texte	50
11.4.1.1	Nom de notation associé	50
11.4.1.2	Spécification d'attribut du texte	50
11.5	Déclaration d'assignation d'appel abrégé	50
11.6	Déclaration d'utilisation d'appel abrégé	51
11.6.1	Utilisation dans une déclaration de type de document	51
11.6.2	Utilisation dans une instance de document	51
11.6.3	Table de correspondance active	51

12 Déclarations de balisage : Définition de processus de liens	51
12.1 Déclaration de type de lien	51
12.1.1 Spécification de lien simple	52
12.1.1.1 Limites	52
12.1.2 Spécification de lien implicite	52
12.1.3 Spécification de lien explicite	52
12.1.3.1 Limites	52
12.1.4 Sous-ensemble de déclaration de type de lien	52
12.1.4.1 Entités	52
12.1.4.2 Attribut de lien	52
12.1.4.3 Lien simple	52
12.2 Déclaration d'ensemble de liens	53
12.2.1 Spécification d'élément source	53
12.2.2 Règle de liaison explicite	53
12.2.3 Déclaration d'ensemble de liens ID	54
12.3 Déclaration d'utilisation d'ensembles de liens	54
12.4 Ensemble de liens en cours	54
13 Déclaration SGML	54
13.1 Jeu de caractères du document	55
13.1.1 Description de jeu de caractères	55
13.1.1.1 Jeu de caractères de base	55
13.1.1.2 Portion de jeu de caractères décrite	55
13.1.2 Identification de caractères non-SGML	56
13.2 Ensemble de capacités	56
13.3 Portée de la syntaxe concrète	56
13.4 Syntaxe concrète	57
13.4.1 Syntaxe concrète publique	57
13.4.2 Identification de numéro de caractère esquivé	58
13.4.3 Jeu de caractères de référence de la syntaxe	58
13.4.4 Identification de caractère de fonction	58
13.4.5 Règles de désignation	58
13.4.6 Jeu de délimiteurs	59
13.4.6.1 Délimiteurs généraux	59
13.4.6.2 Délimiteurs d'appel abrégé	59
13.4.7 Usage de noms réservés	60
13.4.8 Ensemble de quantités	60
13.5 Utilisation des fonctionnalités	61
13.5.1 Fonctionnalités de minimisation du balisage	61
13.5.2 Fonctionnalités type de lien	61
13.5.3 Autres fonctionnalités	61
13.6 Information spécifique à l'application	61
14 Syntaxe concrète de référence et syntaxe concrète minimale	61
15 Conformité	62
15.1 Document conforme SGML	62
15.1.1 Document SGML de base	62
15.1.2 Document SGML minimal	62
15.1.3 Variante de document conforme SGML	62
15.2 Application conforme SGML	62
15.2.1 Conventions d'application	63
15.2.2 Conformité des documents	63
15.2.3 Conformité de la documentation	63
15.3 Système conforme SGML	63
15.3.1 Conformité de la documentation	63
15.3.2 Conformité par rapport à la déclaration de système	63
15.3.3 Gestion de la syntaxe concrète de référence	63
15.3.4 Gestion de l'ensemble des capacités de référence	63

15.3.5	Cohérence du parsing	63
15.3.6	Conventions d'application	64
15.4	Parser SGML de validation	64
15.4.1	Reconnaissance d'erreur	64
15.4.2	Identification des messages SGML	65
15.4.3	Contenu des messages SGML	65
15.5	Exigences concernant la documentation	65
15.5.1	Identification de la norme	65
15.5.2	Identification de structures SGML	65
15.5.3	Terminologie	65
15.5.4	Variante de la syntaxe concrète	65
15.6	Déclaration de système	65
15.6.1	Syntaxes concrètes acceptées	66
15.6.1.1	Changements dans la syntaxe concrète	66
15.6.2	Services de validation	66
15.6.3	Gestion de SDIF	67

Annexes

A	Introduction au balisage généralisé	68
A.1	Le balisage	68
A.2	Le balisage descriptif	69
A.3	Balisage rigoureux	71
A.4	Conclusion	74
A.5	Remerciements	74
A.6	Bibliographie	75
B	Concepts de base	76
B.1	Documents, définitions de types de documents et procédures	76
B.1.1	Documents	76
B.1.2	Définitions de types de documents	76
B.1.3	Procédures	77
B.2	Balisage	77
B.3	Distinguer le balisage du texte	78
B.3.1	Balises descriptives	78
B.3.2	Autre balisage	79
B.3.3	Limites d'enregistrement	80
B.3.3.1	Limites d'enregistrement dans les données	80
B.3.3.2	Limites d'enregistrements dans le balisage	80
B.4	Structure du document	80
B.4.1	Définitions de types de document	81
B.4.2	Déclarations d'éléments	81
B.4.2.1	Modèles de contenu	81
B.4.2.2	Connecteurs et symboles d'occurrence	82
B.4.2.3	Appels d'entités dans les modèles	83
B.4.2.4	Groupes de noms	83
B.4.2.5	Caractères du texte	83
B.4.2.6	Contenu vide	84
B.4.2.7	Données non-SGML	84
B.4.2.8	Résumé des délimiteurs de modèle	84
B.5	Attributs	85
B.5.1	La spécification des attributs	85
B.5.1.1	Noms	85
B.5.1.2	Valeurs d'attributs	85
B.5.2	Déclaration des attributs	86
B.5.2.1	Syntaxe des définitions d'attributs	86

B.5.2.2	Valeurs d'attributs complexes	87
B.5.2.3	Groupe d'unités lexicales nominales	88
B.5.2.4	Changer les valeurs par défaut	88
B.6	Entités	88
B.6.1	Syntaxe pour les appels d'entités	88
B.6.2	Déclaration d'entités	89
B.6.2.1	Instructions de traitement	90
B.6.2.2	Entités contenant des appels d'entités	90
B.6.2.3	Entités externes	90
B.6.2.4	Entités publiques	90
B.7	Caractères	91
B.7.1	Classification de caractères	91
B.7.2	Appels de caractères	92
B.7.3	Caractères délimiteurs utilisés comme données	93
B.8	Sections marquées	94
B.8.1	Ignorer une section marquée	94
B.8.2	Versions d'un document unique	95
B.8.3	Sections protégées du parsing	95
B.8.4	Sections temporaires	96
B.8.5	Spécification de mot-clé	96
B.8.6	Définition d'une section marquée en tant qu'entité	96
B.9	Attributs « identificateurs uniques »	97
B.10	Attributs d'appel de contenu	97
B.11	Exceptions au modèle de contenu	98
B.11.1	Éléments inclus	98
B.11.2	Éléments exclus	98
B.12	Déclaration de type de document	99
B.13	Contenu textuel	99
B.13.1	Représentations du contenu textuel	100
B.13.1.1	Données textuelles (PCDATA, CDATA et RCDATA)	100
B.13.1.2	Données non-SGML (NDATA)	100
B.13.2	Notations de contenu textuel	101
B.13.2.1	Notations pour les données textuelles	101
B.13.2.2	Notation de données non-SGML	102
B.13.2.3	La spécification des notations de contenu textuel	102
B.14	Personnalisation	103
B.14.1	La déclaration SGML	103
B.14.1.1	Fonctionnalités optionnelles	103
B.14.1.2	Variante de la syntaxe concrète	103
B.14.2	Impact de la personnalisation	104
B.15	Conformité	104
C	Concepts complémentaires	105
C.1	Fonctionnalités de minimisation du balisage	105
C.1.1	SHORTTAG : Balises avec omission de balisage	106
C.1.1.1	Balises abrégées non fermées	106
C.1.1.2	Balises vides	106
C.1.1.3	Minimisation des attributs	106
C.1.2	OMITTAG : Les balises peuvent être omises	107
C.1.2.1	Principes de l'omission de balisage	107
C.1.2.2	Indication de la minimisation	108
C.1.2.3	Omission de balise de fin : intrusion d'une balise de début	109
C.1.2.4	Omission de balise de fin : Balise de fin d'un élément englobant	109

C.1.2.5	Omission de balise de début : élément obligatoire d'après le contexte	109
C.1.2.6	Combinaison avec la minimisation par abréviation de balises	110
C.1.2.7	Considérations de minimisation du balisage	110
C.1.3	SHORTREF : Les délimiteurs d'appels abrégés peuvent remplacer des appels d'entités complets	110
C.1.3.1	Frappe au clavier : WYSIWYG généralisé	111
C.1.3.2	Exemple de saisie clavier : définition d'une table de correspondance d'appels abrégés	111
C.1.3.3	Exemple de saisie clavier : activation d'une table de correspondance d'appel abrégé	112
C.1.3.4	Exemple de tableau	113
C.1.3.5	Exigences particulières	114
C.1.4	DATATAG : Les données peuvent également constituer une balise	114
C.1.5	RANK : les niveaux peuvent être omis des balises	117
C.2	Fonctionnalités LINK : SIMPLE, IMPLICIT et EXPLICIT	118
C.2.1	Définitions de processus de lien	118
C.3	Autres fonctionnalités	118
C.3.1	CONCUR : Des instances de document peuvent apparaître de façon concurrente	119
C.3.2	SUBDOC : Des entités sous-documents imbriqués peuvent apparaître	119
C.3.3	FORMAL : Les identificateurs publics sont formels	120
D	Texte public	121
D.1	Jeux d'éléments	121
D.1.1	Types d'éléments communs	121
D.1.2	Types d'éléments pro forma	121
D.2	Notations de contenu textuel	121
D.3	Variantes de la syntaxe concrète	122
D.3.1	Syntaxes concrètes multicodes	122
D.4	Jeux d'entités	122
D.4.1	Considérations générales	123
D.4.1.1	Format des déclarations	123
D.4.1.2	Jeux d'entités affichables correspondants	124
D.4.1.3	Noms d'entités	124
D.4.1.4	Organisation des jeux d'entités	125
D.4.2	Caractères alphabétiques	125
D.4.2.1	Alphabet latin	125
D.4.2.2	Caractères de l'alphabet grec	128
D.4.2.3	Caractères de l'alphabet cyrillique	130
D.4.3	Usage général	132
D.4.3.1	Caractères graphiques numériques et spécialisés	132
D.4.3.2	Caractères diacritiques	133
D.4.3.3	Caractères utilisés dans l'édition	133
D.4.3.4	Caractères utilisés pour le dessin d'encadrés et de lignes	135
D.4.4	Usage technique	136
D.4.4.1	Général	137
D.4.4.2	Symboles grecs	138
D.4.4.3	Autres symboles grecs	139
D.4.5	Symboles mathématiques supplémentaires	139
D.4.5.1	Symboles ordinaires	139
D.4.5.2	Opérateurs binaires et de grande taille	140
D.4.5.3	Relations	141

iTeh STANDARD PREVIEW
(standards.iteh.ai)

<https://standards.iteh.ai/catalog/standards/sist/8879-1986-iso-8879-1986>

D.4.5.4	Négations	143
D.4.5.5	Flèches	144
D.4.5.6	Délimiteurs ouvrant et fermant	145
E	Exemples d'application	146
E.1	Définition de type de document	146
E.2	Computer Graphics Metafile	149
E.3	Extension de codes indépendante de la configuration	149
E.3.1	Fonctionnalités d'extension de codes	150
E.3.1.1	Éviter des erreurs de reconnaissance de délimiteurs	151
E.3.1.2	Élimination des dépendances vis-à-vis de la configuration et des codes	153
F	Considérations relatives à l'implémentation	154
F.1	Un modèle de parsing SGML	154
F.1.1	Entrée physique	154
F.1.1.1	Entités	154
F.1.1.2	Limites d'enregistrement	154
F.1.2	Modes de reconnaissance	154
F.1.3	Minimisation du balisage	155
F.1.4	Traduction	156
F.1.5	Analogie avec un langage de programmation	156
F.2	Initialisation	156
F.2.1	Assignation de procédure initiale	156
F.2.2	Spécification de processus de lien	156
F.2.3	Instances de document concurrentes	157
F.3	Assignation de procédure dynamique	157
F.4	Traitement des erreurs	157
G	Classification de conformité et certification	158
G.1	Code de classification	158
G.1.1	Code de fonctionnalité	158
G.1.2	Code de validation	159
G.1.3	Code de syntaxe	160
G.2	Considérations de certification	160
H	Base théorique du modèle de contenu SGML	161
H.1	Notation de groupe modèle	161
H.2	Application de la théorie des automates	161
H.3	Différence par rapport à la théorie des automates	162
I	Variantes non conformes	163
I.1	Identificateurs génériques de longueur fixe	163
I.2	Délimiteur unique	163

Figures

1	Classes de caractères : Syntaxe abstraite	32
2	Classes de caractères : Syntaxe concrète	34
3	Ensemble des délimiteurs de référence : Général	36
4	Ensemble des délimiteurs de référence : Appels abrégés	38
5	Ensemble des capacités de référence	57
6	Ensemble des quantités de référence	60
7	Syntaxe concrète de référence	62
8	Exemple-type de déclaration SGML pour un document SGML de base	64
9	Balisage d'un élément	79
10	Balise de début avec 2 attributs	85
11	Syntaxe concrète de base multicode	123
12	Graphics Metafile Attributes (1 sur 2) : Encodage et vue	150
13	Graphics Metafile Attributes (2 sur 2) : Taille et rotation	152
14	Caractères de fonction pour syntaxes concrètes multicode indépendantes de la configuration	153
15	Classification de conformité FSV	159

Traitement de l'information — Systèmes bureautiques — Langage normalisé de balisage généralisé (SGML)

0 Introduction

La présente norme internationale décrit les spécifications d'un langage de description de document nommé SGML, *Standard Generalized Markup Language*, soit « langage normalisé de balisage généralisé ». SGML peut être utilisé dans l'édition, au sens le plus général de ce terme, qui s'étend de l'édition traditionnelle sur support unique jusqu'à l'édition multimédia à partir de bases de données. SGML peut également être utilisé pour traiter les documents en environnement bureautique, lorsque l'on recherche à la fois la possibilité de relecture par des humains et de transfert vers des systèmes de publication.

0.1 Le contexte

Un document peut être analysé de façon abstraite comme une structure constituée de plusieurs types d'éléments. L'auteur organise son livre en le découpant en chapitres qui contiennent des paragraphes et, le cas échéant, des illustrations qui contiennent à leur tour des légendes. Le secrétaire de rédaction d'un magazine le structure en articles, qui contiennent des paragraphes contenant eux-mêmes des mots, etc.

Les programmes prennent en compte ces éléments de différentes façons. On peut avoir recours à un programme de formatage pour mettre en valeur les titres avec un type de caractères approprié, aérer les paragraphes en laissant du blanc, révéler au lecteur par toute une série de procédés visuels la structure du document ou d'autres attributs. On peut envisager d'attribuer aux mots d'un titre, à l'aide d'un programme de recherche documentaire, une signification supplémentaire au moment de la création d'un lexique.

La relation entre les attributs d'un document et son traitement automatique semble aujourd'hui aller de soi, mais il n'en a pas toujours été ainsi ; les premières

méthodes de traitement automatique du texte avaient tendance à la masquer. Avant l'automatisation de la composition, un secrétaire d'édition « codait » un manuscrit en y inscrivant des instructions de traitement spécifiques qui permettaient d'obtenir la présentation voulue, après exécution par un compositeur. La relation entre les instructions et la structure du document n'existait que pour le secrétaire d'édition.

Cette approche s'est maintenue dans les premiers systèmes informatiques avec l'insertion d'un « balisage » déclenchant des actions particulières dans le fichier document lisible par la machine. Ce balisage est, lui encore, constitué d'instructions de traitement spécifiques, mais elles ont cette fois été transcrites dans un langage propre au programme de formatage, et non plus dans celui de la personne qui compose. Sans transformation préalable complète des codes qui y figurent, il est difficile de se servir du fichier en question pour d'autres applications ou de le transférer sur tout autre type d'ordinateur.

Le développement de méthodes pour résoudre ce problème est allé de pair avec la progression du degré d'expertise acquise par les utilisateurs et la montée en puissance des traitements de texte. Les « appels macros », ou « appels de formats » ont été utilisés pour identifier les endroits d'un document qui doivent être soumis à un traitement particulier dans un document. Les instructions de traitement proprement dites étaient stockées à l'extérieur du document, dans des « procédures » (ou « définitions de macro », ou « formats stockés »), où elles pouvaient être plus facilement modifiées.

Bien que les macros puissent être appelées n'importe où dans le document, les utilisateurs s'aperçurent que la plupart d'entre elles étaient placées au début ou à la fin de certains éléments du document. Il parut donc naturel de choisir des noms de macros qui identifiaient

les classes d'éléments de façon générique, plutôt que des noms invoquant un traitement particulier (par exemple, « titre » plutôt que « format-17 ») ; c'est ainsi que l'on commença à pratiquer le « codage générique » (ou « balisage généralisé »).

Le codage générique a constitué une étape décisive dans l'élaboration de systèmes de traitement automatisé du texte reflétant la relation naturelle qui unit les attributs d'un document et leur traitement. L'apparition, au début des années 1970, des « langages de balisage généralisé » a renforcé cette tendance en fournissant au codage générique le fondement d'un langage formel. Un langage de balisage généralisé respecte deux principes fondamentaux :

- a) Le balisage descriptif est prédominant, et il se distingue des instructions de traitement.

Le balisage descriptif comprend à la fois des identificateurs génériques et d'autres attributs d'éléments du document qui déclenchent les instructions de traitement. Les instructions de traitement, qui peuvent être écrites dans n'importe quel langage, sont généralement stockées à l'extérieur du document, dans des procédures.

En examinant le fichier source pour analyser le balisage et en reconnaître les différents éléments, le système de traitement exécute les procédures associées à chaque élément et à chaque attribut pour ce traitement. D'autres traitements sont possibles : des procédures différentes peuvent être associées aux mêmes éléments et attributs sans modifier le balisage du document.

Lorsqu'on doit entrer directement une instruction de traitement dans un document, il faut la délimiter d'une autre manière que le balisage descriptif de façon à en faciliter la localisation et la modification pour d'autres traitements.

- b) Le balisage est défini formellement pour chaque type de document.

Un langage de balisage généralisé formalise le balisage d'un document en incorporant des « définitions de type de document ». Les définitions de type (de document) spécifient (comme les grammaires formelles) les éléments et les attributs qui peuvent se trouver dans un document et l'ordre dans lequel ils peuvent apparaître. Cette information permet de déterminer si le balisage d'un document donné est correct (c'est-à-dire correspond à la définition de type de document) ainsi que de compléter le balisage manquant, celui-ci pouvant en effet être déduit sans ambiguïté des balises présentes.

NOTE — On trouvera une introduction plus détaillée aux concepts de codage générique et au langage normalisé de balisage généralisé dans l'annexe A.

0.2 Objectifs

Le Langage normalisé de balisage généralisé normalise l'application des concepts de codage générique et de balisage généralisé. Il fournit une syntaxe cohérente et non ambiguë pour décrire tout ce qu'un utilisateur choisit d'identifier à l'intérieur d'un document. Le langage comprend :

- une « syntaxe abstraite » pour le balisage descriptif des éléments d'un document.

- une « syntaxe concrète de référence » qui relie la syntaxe abstraite à des caractères particuliers servant de délimiteurs et à des quantités. Les utilisateurs peuvent définir d'autres syntaxes concrètes pour répondre à leurs propres besoins.

- des déclarations de balisage permettant à l'utilisateur de définir un vocabulaire spécifique d'identificateurs génériques et d'attributs pour différents types de documents.

- des moyens d'utiliser des contenus textuels arbitraires. Dans le balisage généralisé, on appelle « données » tout ce qui n'est pas défini par le langage de balisage. Il peut s'agir de « notations de contenu textuel » nécessitant une interprétation différente de celle du texte général : formules, illustrations, alphabets non latins, texte déjà formaté, ou graphiques.

- des appels d'entités : il s'agit d'une technique indépendante du système utilisé qui permet d'appeler des contenus situés à l'extérieur du corps du document, comme des chapitres séparés, des polices casseau, des photographies, etc.

- des délimiteurs particuliers encadrant les instructions de traitement, qui servent à les distinguer du balisage descriptif. Les instructions de traitement peuvent être introduites dans les cas qui ne peuvent être traités par les procédures, mais elles peuvent alors être facilement retrouvées voire modifiées ultérieurement au moment d'envoyer le document vers un système de traitement différent.

Mais pour qu'un langage de balisage généralisé puisse être accepté comme norme, il faut qu'il possède plus que les fonctionnalités strictement indispensables. Il doit être doté de propriétés métalinguistiques, afin de pouvoir être utilisé dans de multiples environnements. Voici résumées les exigences les plus importantes auxquelles SGML doit pouvoir répondre, et les moyens mis en œuvre pour y parvenir :

- a) Les documents « balisés » selon le langage doivent pouvoir être traités par un grand nombre de systèmes et de logiciels de traitement de texte.

La forme complète du langage, munie de toutes ces fonctionnalités optionnelles, offre un degré de généralité et de flexibilité qui lui permet d'être exploitée par des systèmes sophistiqués ; il n'est pas nécessaire que des systèmes moins puissants soient dotés de toutes ces fonctionnalités. Pour faciliter l'échange entre systèmes différents, une « déclaration SGML » décrit chaque fonctionnalité de balisage ou toutes les variantes par rapport à la syntaxe concrète utilisée dans un document.

- b) Le parc installé de plusieurs millions de systèmes de saisie de texte doit pouvoir continuer à être utilisé.

Les documents SGML, munis de leur syntaxe concrète de référence, peuvent facilement être saisis et compris par tous, sans l'assistance d'une machine. Par conséquent :

— Il n'est pas nécessaire d'attendre que naisse et que soit acceptée une nouvelle génération de matériels pour utiliser SGML — on peut se contenter des logiciels de traitement de documents sur les machines existantes.

— La migration vers cette nouvelle génération de matériel (quand viendra son heure) sera d'autant plus facile que les utilisateurs se seront déjà familiarisés avec SGML.

- c) Il ne doit pas y avoir la moindre dépendance par rapport à un jeu de caractères spécifique, étant donné que les documents doivent pouvoir être tapés sur des machines différentes.

Le langage ne dépend d'aucun jeu de caractères spécifique. Tout jeu de caractères qui utilise des combinaisons d'éléments binaires pour représenter les lettres, les chiffres, les espaces et les délimiteurs est admis.

- d) Il ne doit pas y avoir la moindre dépendance par rapport à un logiciel, à un système ou à un matériel donnés.

Le balisage généralisé est avant tout descriptif et donc par essence non soumis à de telles dépendances. S'il faut insérer de temps en temps une instruction de traitement spécifique, on veillera à la placer entre des délimiteurs de façon à ce que l'on puisse aisément la retrouver et la convertir lors d'un échange, ou pour le cas où un traitement différent rendrait cette instruc-

tion non pertinente.

Les appels à des parties externes à un document sont exprimés de façon indirecte. Les correspondances avec les dispositifs de stockage des systèmes existants s'expriment par des « déclarations d'entités externes » qui apparaissent au début du document, là où elles peuvent être facilement modifiées lors des échanges.

La syntaxe concrète peut être modifiée dans la déclaration SGML afin de pouvoir s'accorder avec l'ensemble des caractères réservés de n'importe quel système.

- e) Rien ne doit être lié à une langue particulière.

Les caractères disponibles pour les noms peuvent s'enrichir des caractères qui sont spécifiques à une langue particulière. Les identificateurs génériques, les noms d'attributs et les autres noms utilisés dans le balisage descriptif sont définis par l'utilisateur dans les déclarations d'éléments et d'entités.

Les noms de déclarations et les mots clés utilisés dans les déclarations de balisage peuvent également être modifiés.

Les répertoires de caractères multiples comme ceux utilisés dans les documents multilingues sont autorisés.

- f) Le langage doit s'accorder avec les conventions habituelles utilisées en dactylographie et en traitement de texte.

Les « appels abrégés » et les « balises textuelles » peuvent être tapés sur un clavier normal de machine à écrire ou d'ordinateur. Un texte courant contenant des paragraphes et des citations peut être interprété comme texte SGML bien qu'il puisse être saisi sans aucun balisage explicite.

- g) Le langage doit être indépendant du flux de données ou de l'organisation physique des fichiers.

Le langage de balisage possède un modèle de stockage virtuel dans lequel les documents sont formés d'une ou de plusieurs entités de stockage, chacune étant faite d'une séquence de caractères. Tout accès réel à un fichier est effectué par le système de traitement, qui peut décider si la séquence de caractères doit être considérée comme continue, ou si elle doit refléter les limites physiques des enregistrements.