
Modélisation et manipulation de documents XML

François Role

Hermès

Lavoisier

-005-502-1

2-005-502-1

Modélisation et manipulation de documents XML

François Role

**Hermès
Science**
publications

Table des matières

Préface	13
Joseph MARIANI	
Chapitre 1. Introduction	15
1.1. Modélisation et manipulation de documents XML.	16
1.2. Public concerné	18
1.3. Organisation de l'ouvrage	18
1.4. Objectifs de l'ouvrage	20
1.5. Nature des exemples utilisés	21
1.6. Conventions utilisées	21
Chapitre 2. Structure et balisage	23
2.1. Introduction.	23
2.2. Structures	23
2.3. Typologie des structures	25
2.3.1. Structure logique et structure physique	25
2.3.2. Structure spécifique et structure générique	27
2.4. Balisage et représentation textuelle des structures	29
2.4.1. Représentation des éléments	30
2.4.2. Représentation des relations hiérarchiques entre éléments	30
2.5. Analyse automatique du balisage	32
2.6. Balisage et interopérabilité : la nécessité d'une approche métalinguistique	36
2.6.1. Normalisation du contrôle de la bonne formation des documents.	36
2.6.2. Normalisation du contrôle de la validité des documents	37
2.7. Conclusion	38

Chapitre 3. Le document XML comme texte balisé	39
3.1. Introduction.	39
3.2. Le document XML comme texte balisé bien formé	41
3.3. Fonctions du balisage dans un document XML.	43
3.4. Balisage dédié à la représentation des structures logiques	44
3.5. Balisage lié à la représentation des structures de stockage	45
3.5.1. Appel d'entité	46
3.5.2. Appel de caractères (<i>character reference</i>)	46
3.6. Autres formes de balisage	47
3.6.1. Notation des commentaires	47
3.6.2. Notation des sections CDATA	47
3.6.3. Prologue XML.	48
3.6.4. Déclaration XML	48
3.6.5. Instructions de traitement	48
3.6.6. Déclaration de type de document.	49
3.7. Représentation des données textuelles (<i>character data</i>).	49
3.8. Conformité XML 1.0 et XML 1.1	51
3.9. Les espaces de noms et l'écriture de documents XML hétérogènes.	51
3.9.1. Déclaration des espaces de noms.	53
3.9.2. Utilisation de noms qualifiés	55
3.9.3. Importance des espaces de noms	56
3.10. Création de documents XML.	56
3.11. XML et les données semi-structurées.	57
3.12. Conclusion	58
Chapitre 4. De la représentation balisée aux modèles de données	59
4.1. Introduction.	59
4.2. XML Information Set (Infoset)	61
4.2.1. Unités d'information (<i>information items</i>) de l'Infoset	62
4.2.2. Unité de type document	62
4.2.3. Unité de type élément	63
4.2.4. Unité de type attribut	64
4.2.5. Unité de type caractère	64
4.2.6. Unité de type espace de noms.	64
4.2.7. Autres unités	65
4.2.8. Informations non mémorisées dans l'Infoset	65
4.2.9. Conclusion sur l'Infoset	67
4.3. Le modèle de données XPath/XQuery	67
4.3.1. Valeurs atomiques	68
4.3.2. Nœuds.	69
4.3.3. Séquences	72
4.3.4. Correspondance avec l'Infoset	73
4.4. Conclusion	73

Chapitre 5. Modélisation et validation de documents avec une DTD	75
5.1. Introduction : des documents bien formés aux documents valides	75
5.2. La notion de DTD (<i>Document Type Definition</i>)	78
5.3. Structure générale d'une DTD	80
5.4. Association d'une DTD à un document XML	81
5.5. Déclaration de type d'élément	83
5.5.1. Contenu élémentaire pur (<i>element content</i>)	84
5.5.2. Contenu dit mixte (<i>mixed content</i>)	85
5.5.3. Contenu de type quelconque : ANY	86
5.5.4. Contenu vide : EMPTY	86
5.5.5. Modèles de contenu déterministes	86
5.6. Définition de liste d'attributs	87
5.6.1. Normalisation des attributs	91
5.7. Déclaration d'entités	92
5.7.1. Déclaration d'entités générales	92
5.7.2. Déclaration d'entités paramètres	93
5.8. Conformité	94
5.9. Limites des DTD	94
Chapitre 6. Modélisation et validation de documents avec un schéma XML	97
6.1. Introduction	97
6.2. Modèle abstrait d'un schéma et validation	98
6.3. Représentation XML d'un schéma	101
6.4. Typage des éléments et des attributs	102
6.4.1. Hiérarchie de types	103
6.4.2. Types simples et types complexes	104
6.5. Types simples	105
6.5.1. Espace de valeurs et espace lexical d'un type de données	105
6.5.2. Catégories de types simples	105
6.5.3. Types prédéfinis	106
6.5.3.1. Types de chaînes	106
6.5.3.2. Types numériques	106
6.5.3.3. Types temporels	107
6.5.3.4. Autres types	107
6.5.4. Dérivation de types par les utilisateurs	107
6.5.4.1. Dérivation d'un type par restriction	108
6.5.4.2. Définition de types listes	111
6.5.4.3. Définition de types union	111
6.6. Types complexes	112
6.6.1. Dérivation de types complexes	112

6.6.2. Modèle de contenu d'un type complexe	116
6.6.3. Utilisation des types complexes	119
6.6.3.1. Types appropriés à des éléments ayant un contenu élémentaire (<i>element content</i>)	120
6.6.3.2. Types appropriés à des éléments ayant un contenu mixte (<i>mixed content</i>)	120
6.6.3.3. Types appropriés à des éléments ayant un contenu vide	121
6.6.3.4. Types appropriés à des éléments pouvant avoir un contenu quelconque	121
6.6.3.5. Types appropriés à des éléments ayant un contenu textuel pur	123
6.6.4. Syntaxe abrégée	123
6.7. Déclarations d'éléments	124
6.7.1. Choix d'un mode d'association entre type et élément	124
6.8. Déclarations d'attributs	126
6.9. Portée des déclarations d'éléments et d'attributs : déclarations locales et globales	126
6.10. Mécanismes de substitution	127
6.10.1. Utilisation de l'attribut <i>xsi:type</i>	128
6.10.2. Utilisation d'un groupe de substitution	129
6.10.3. Contrôler les possibilités de substitution	129
6.10.4. Substitution et éléments abstraits	130
6.11. Représentation des valeurs inconnues	130
6.12. Schémas XML et espaces de noms	131
6.13. Schémas XML et validation de documents hétérogènes	134
6.14. Expression de contraintes d'intégrité	140
6.14.1. Contraintes d'unicité	140
6.14.2. Intégrité référentielle	141
6.15. Conformité	143
6.16. Autres formalismes de validation	144
6.17. Bonnes pratiques d'écriture d'un modèle de document	148
6.17.1. Utilisation de commentaires	148
6.17.2. Écriture modulaire	151
6.17.3. Utilisation de patrons de conception (<i>design patterns</i>)	156
6.18. Conclusion	159
Chapitre 7. Interfaces de programmation	161
7.1. Introduction	161
7.2. API événementielles	163
7.2.1. SAX: un exemple d'API en mode <i>push</i>	163
7.2.2. StAX : un exemple d'API en mode <i>pull</i>	167
7.3. API basées sur des arbres (<i>tree-based API</i>)	168

7.3.1. Evolutions du DOM : du DOM niveau 1 au DOM niveau 3	174
7.3.2. Organisation modulaire	176
7.3.3. Conformité	177
7.4. Autres approches.	179
7.5. Critères de choix entre les différentes API	179
7.6. Conclusion	180

Chapitre 8. Désignation de fragments et navigation

dans la structure d'un document : de XPath à XPointer	183
8.1. Introduction.	183
8.2. Types des expressions XPath	185
8.3. Expressions de chemins	187
8.3.1. Axes XPath.	188
8.3.2. Chemins et pas.	192
8.3.3. Structure d'un pas de déplacement.	192
8.3.3.1. Notation des axes (<i>axis specifier</i>)	193
8.3.3.2. Test du type de nœuds à sélectionner (<i>node test</i>)	193
8.3.3.3. Notation des prédicats (<i>predicate</i>).	193
8.4. Evaluation des expressions XPath	194
8.5. Notation abrégée des expressions de chemin	199
8.6. Autres types d'expressions XPath	199
8.6.1. Expressions littérales	200
8.6.2. Références à des variables.	200
8.6.3. Appels de fonctions.	200
8.6.3.1. Fonctions de manipulation de chaînes	201
8.6.3.2. Fonctions permettant de retrouver des informations du modèle de données	201
8.6.3.3. Fonctions numériques	202
8.6.3.4. Fonctions de conversion de types	202
8.6.3.5. Fonctions booléennes	202
8.6.4. Expressions numériques	202
8.6.5. Expressions logiques	203
8.6.6. Expressions relationnelles	203
8.7. Règles de conversion de types.	203
8.8. Construction de liens avec XLink et XPointer	204
8.8.1. Utilisation de pointeurs étendus (XPointer).	204
8.8.2. XLink (XML Linking Language)	205
8.9. Conclusion	207

Chapitre 9. Transformation de documents avec XSLT	209
9.1. Introduction	209
9.2. Principe général d'une transformation XSLT	211
9.3. Un exemple simple de transformation avec XSLT	215
9.4. Motifs	220
9.5. Contenu des modèles et construction de l'arbre résultat	220
9.5.1. Éléments littéraux	221
9.5.2. Instructions XSLT	222
9.5.2.1. Détermination de la liste de nœuds courante	222
9.5.2.2. Création d'éléments avec <code>xsl:element</code>	223
9.5.2.3. Création d'attributs avec <code>xsl:attribute</code>	225
9.5.2.4. Copie de nœuds <code>xsl:copy</code>	225
9.5.2.5. Création de texte : <code>xsl:text</code> <code>xsl:value-of</code>	225
9.5.2.6. Création de commentaires et d'instructions de traitement	226
9.5.2.7. Traitements conditionnels : <code>xsl:if</code> et <code>xsl:choose</code>	226
9.5.2.8. Itérations : <code>xsl:for-each</code>	227
9.5.3. Regroupements <i>via</i> des clés XSLT	227
9.5.4. Réalisation de tris	229
9.5.5. Utilisation de variables	230
9.5.6. Traitement simultané de plusieurs documents source	231
9.6. Sérialisation	233
9.7. Gestion des espaces de noms	235
9.8. Règles prédéfinies	235
9.9. Traitement des espaces	236
9.10. Utilisation des modes	237
9.11. Choix des règles à appliquer	238
9.12. XSLT et XSL-FO	239
9.13. Utilisation de fonctions	240
9.14. Styles de programmation	241
9.15. Écriture modulaire des feuilles de style	242
9.15.1. Importation et inclusion de feuille de style	243
9.15.2. Utilisation de groupes d'attributs nommés	243
9.15.3. Utilisation de règles nommées	244
9.16. Autres bonnes pratiques d'écriture	245
9.17. Appliquer une feuille de style à un document	246
9.18. Feuille de style et validation	246
9.19. Conformité	246
9.20. Conclusion	247

Chapitre 10. Gestion et exploitation de documents XML dans des bases de données	249
10.1. Introduction	249
10.2. Utilisation de modèles de données non XML	251
10.2.1. Utilisation du modèle relationnel	251
10.2.2. Utilisation du modèle objet-relationnel avec le type XML	255
10.2.3. Utilisation d'un modèle-objet	256
10.3. Utilisation d'un modèle de données XML	257
10.4. XQuery : un langage de requêtes dédié à XML	259
10.4.1. Quelques exemples de requêtes	260
10.4.2. Contexte	263
10.4.2.1. Contexte statique	263
10.4.2.2. Contexte d'évaluation	263
10.4.3. Structure d'une requête XQuery	263
10.4.4. Types	264
10.4.5. Expressions du langage XQuery	265
10.4.5.1. Expressions littérales	265
10.4.5.2. Variables	266
10.4.5.3. Expressions parenthésées	266
10.4.5.4. Expressions arithmétiques	266
10.4.5.5. Expressions de chemins	266
10.4.5.6. Constructeurs	267
10.4.5.7. Opérations sur les séquences	268
10.4.5.8. Expressions FLWR	268
10.4.5.9. Expressions conditionnelles	271
10.4.5.10. Expressions de comparaison	271
10.4.5.11. Expressions logiques	272
10.4.5.12. Appels de fonctions	272
10.4.6. Sémantique formelle	274
10.4.7. Conformité	274
10.4.8. Problèmes posés par l'utilisation d'un modèle de données XML avec XQuery	274
10.5. Conclusion	275
Chapitre 11. XML et l'architecture du web	277
11.1. Introduction	277
11.2. XML et les architectures à base de services <i>web</i>	278
11.2.1. Description et création d'un service <i>web</i>	278
11.2.2. Représentation et transport des messages XML	282
11.2.3. XML et l'orchestration des services <i>web</i>	284
11.2.4. XML et les architectures orientées services (SOA)	287

12 Modélisation et manipulation XML

11.3. XML et les métadonnées du <i>web</i> sémantique	288
11.3.1. Les métadonnées du <i>web</i>	288
11.3.2. Le cadre RDF (<i>Resource Description Framework</i>)	290
11.3.2.1. Modèle et syntaxe concrète	290
11.3.2.2. Schémas RDF	292
11.3.3. OWL et la représentation d'ontologies.	293
11.4. Du <i>web</i> sémantique au services <i>web</i> : un exemple d'intégration des techniques.	296
11.5. Conclusion	301
Chapitre 12. Conclusion	303
Annexe. Les recommandations du consortium W3	307
Bibliographie	309
Index	321