

|                                |   |
|--------------------------------|---|
| <b>Statut du document</b>      | <i>W3C Final Draft</i>  |
| <b>Titre</b>                   | <i>XML Schema Part 1: Structures</i> (Schéma XML : 1- Structures)   |
| <b>Mot clé</b>                 | Structuration logique de contenu  |
| <b>Source</b>                  | W3C   |
| <b>Date de publication</b>     | 7 avril 2000  |
| <b>Nombre de pages</b>         | 118   |
| <b>Langue</b>                  | Anglais   |
| <b>Lien avec autres normes</b> | Basée sur la syntaxe de XML 1.0 dont une version française est disponible à <a href="http://babel.alis.com/web_ml/xml/REC-xml.fr.html">http://babel.alis.com/web_ml/xml/REC-xml.fr.html</a> .<br>Est complétée par une partie 2 portant sur les types de donnée.  |
| <b>Situation actuelle</b>      | Version finale de Draft en consultation jusqu'à mai 2000  |
| <b>Description</b>             | <p>Les schémas XML offrent une façon améliorée, plus riche et plus souple, de définir la structure d'un document par rapport aux conventionnelles DTD (définitions de type de document). On veut notamment pouvoir vérifier des contraintes et valider aussi le contenu, pas seulement la structure. Un certain nombre de construits (concepts dérivés d'autres concepts mieux connus et définis) peuvent servir à contraindre et documenter le sens, l'usage et les associations entre les parties constituantes d'un document : les types de donnée (<i>datatypes</i>), les éléments et leur contenu, les attributs et leurs valeurs, les entités et leur contenu et leur notation, les valeurs par défaut. Le langage de schéma XML répond au besoin d'unifier le vocabulaire de documentation des schémas afin de mieux les définir, les décrire et les cataloguer en vue de la réutilisation.</p> <p>Le langage de schéma crée du nouveau en tirant profit de la double nature du schéma qui ne peut éviter d'être exprimé sous forme d'une définition de schéma, se constituant par ce fait même en instance de document. Cette astuce rend difficile son explication, mais permet d'ajouter des construits ou façons de concevoir devant conduire à une exploitation plus rationnelle des définitions de structure. Les organisations schématisent ces définitions en recourant à des construits comme l'héritage multiple des propriétés qui a été popularisée par l'approche orientée objet. La notion de schéma vise donc à moderniser certains aspects de développement des « DTD » avec des techniques plus avancées et déjà bien maîtrisés. En fait le concept n'est pas nouveau pour les DTD qui, dans le monde SGML, disposent de <i>formes architecturales</i> définies dans la norme HyTime pour faire ce que les schémas XML visent maintenant en terme d'héritage, de réutilisation et de validation du contenu. En d'autres mots, le langage de schéma XML vient améliorer certaines caractéristiques des définitions de structures afin de les constituer comme un actif gérable efficacement.</p> <p><b>1- Définitions</b></p> <p>Le traitement par un parseur de document XML peut se faire selon deux degrés de contrainte :</p> <ul style="list-style-type: none"> <li>- le document <i>bien-formé</i> n'est évalué qu'en fonction de la syntaxe générale de XML ;</li> <li>- le document <i>valide</i> est évalué en fonction d'un modèle particulier en plus des contraintes de syntaxe générale.</li> </ul> <p>Ce sont les documents valides et l'ajout de nouvelles sortes de contraintes et</p> |

d'un grand détail qui ont motivé la conception du langage des schémas.

Les auteurs du langage des schémas proposent une série de construits ou termes dérivés avec une définition ou une explication :

- **Schéma (Le)** : Une enveloppe contenant toutes les définitions et déclarations constituant un schéma de document.
- **Définition de type de donnée** : un type (i.e. contrainte sur contenu) qui s'applique aux données textuelles dans une instance de document, qu'il s'agisse d'une valeur d'attribut ou du contenu d'un élément.
- **Définition d'archétype** : un ensemble complet de contraintes pour les éléments dans des instances de document, applicable tant aux valeurs des attributs qu'au contenu des éléments.
- **Déclaration de type d'élément \*** : une association entre un nom donné à un élément et un archétype (analogue à une déclaration de DTD).
- **Déclaration d'attribut \*** : une association entre un nom donné à un attribut et un type de donnée, ainsi que les contraintes sur l'occurrence (obligatoire, par défaut, etc). Cette association est interne au contexte d'un archétype.
- **Type de contenu** : Soit un type de donnée, soit un modèle de contenu, ne s'applique qu'au contenu des éléments dans les instances, à l'exclusion de leurs attributs.
- **Modèle de contenu d'élément** : un type (i.e. contrainte sur contenu) qui s'applique au contenu des éléments dans une instance de document, à l'exclusion des déclarations d'attributs.
- **Contenu limité aux éléments** : composants en vue de constituer les modèles de contenu limités à des éléments (i.e. pas des données textuelles). Inclut des capacités pour regrouper, mettre en ordre séquentiel, déclarer des types d'éléments et réutiliser ceux qui existent par référence.
- **Définition de groupe d'attributs** : une association entre un nom et une collection réutilisable de déclarations d'attributs.
- **Groupe de modèles nommé** : un nom est donné à chaque groupe de modèles utilisé lors des développements afin d'en faciliter la réutilisation.
- **Raffinement d'archétype** : Un archétype peut être défini comme un raffinement d'un ou de plusieurs archétypes existants, de façon à en hériter le type de contenu, les attributs ou les deux.
- **Schéma source** : permet d'étendre un schéma en développement avec les définitions ou les déclarations (ou les deux) provenant d'un schéma externe, et ce en maintenant la dépendance avec le schéma source (et les modifications qui pourraient y être apportées).
- **Schéma inclus** : permet d'intégrer les définitions ou les déclarations ou les deux d'un schéma externe dans le schéma en cours de développement, comme si elles avaient été créées localement, sans maintien d'une dépendance avec le schéma source et en assumant toute responsabilité de les modifier.

Seuls les deux construits suivis d'un astérisque s'appliquent aux instances, ce qui retire l'utilité de les nommer. Tous les autres construits sont nommés de sorte qu'on puisse y référer : ainsi une définition d'attribut peut référer à un type de donnée, un modèle de contenu peut référer à un type d'élément, etc.

Tous ces noms ont besoin de quelque chose pour éviter les collisions de noms. C'est le **Contexte de symboles**, une zone spéciale du schéma dans laquelle sont rassemblés toutes les abstractions nommées (au-dessus d'*attribut* et de *type d'élément*). C'est ainsi que chaque *archétype* définit son contexte local de symboles pour les éléments et les attributs.

Le langage de schéma XML présente une *syntaxe abstraite* qui offre une spécification formelle de l'information fournie pour chaque déclaration et chaque définition dans le langage de schéma. La *syntaxe concrète*, i.e. les noms exacts d'élément et d'attribut dans un schéma, est la forme dans laquelle le langage de schéma est utilisé par les auteurs de schéma.

## 2- Des construits pour les schémas

La façon de réutiliser un construit est d'y référer par son nom façonné conformément au protocole *XML Namespaces*. Sont réutilisables les types, éléments et attributs, en particulier :

- Combinaisons de contraintes et d'ensembles d'information de schéma (appelées SC).
- Types de donnée simples dont l'expression consiste entièrement de données caractères, et qui peuvent s'appliquer tant aux valeurs d'attributs qu'au contenu d'élément sous forme de données textuelles.
- Définition d'archétype : regroupement de tous les éléments d'instance de document en incluant tous les SC pertinents.
- Déclaration d'attributs qui associe un nom apparaissant parmi les instances avec un ou des SC spécifiant les occurrences d'attribut et leur domaine de valeurs.
- Définition de groupe d'attributs pour nommer un groupe d'attributs réutilisable.
- Modèle de contenu d'élément : une séquence d'éléments et de données caractères qui peut contenir de tout, être vide ou mixte, ou ne contenir que des éléments.
- Contenu mixte : mélange de données caractères et d'éléments, sans ordre ni indication d'occurrence, dans une instance de document.
- Contenu réservé aux éléments : i.e. pas de données caractères.
- Groupe de modèles nommé.
- Déclaration de type d'élément : indication d'un schéma source ou inclus par son nom pour qu'il fasse partie d'un archétype.

Tous ces construits doivent être vus comme des façons de spécifier un modèle, ce qui se fera normalement par **raffinement d'archétype**. Le mot raffinement souligne que la définition d'un archétype peut emprunter aux schémas qui sont nommés, ou à d'autres archétypes en spécifiant quelles parties sont réutilisées. Un archétype est interprété par la machine comme l'imposition de contraintes sur des instances d'élément dans un univers clos (XML 1.0), raffiné (ce qui est exigé, rien d'exclus), ou ouvert (ce qui est exigé de base, plus les contraintes résultant du raffinement). Les contraintes peuvent être de deux types : *explicites* (dans la définition de l'archétype) ou *héritées*. L'ensemble des deux constituent les contraintes *effectives*.

La composition de schéma prend appui sur le contexte de dénomination de XML (XML Namespaces) pour désigner tous les construits utiles à une

|  |  |
|--|--|
|  | composition modulaire des schémas avec des règles d'emprunt souples, y compris des entités non-XML et leurs notations. |
|--|--|

## Remarque

### Lexique anglais-français

|                                |  |
|--------------------------------|--|
| Archetype refinement           | Raffinement d'archétype                    |
| Attribute group definition     | Définition de groupe d'attributs           |
| Bounds (above, below)          | Délimitations (supérieure, inférieure)     |
| Built-in                       | Fondamentales                              |
| Character data                 | Données caractères                         |
| Construct                      | Construit (façon de concevoir)             |
| Element content model          | Modèle de contenu d'élément                |
| Element-only content           | Contenu réservé aux éléments               |
| Markup constructs              | Balises                                    |
| Name space                     | Espace nominatif, contexte de dénomination |
| Named model group              | Groupe de modèles nommé                    |
| Namespace compatible (NC) name | Nom conforme à un libellé XML?             |
| Picture (COBOL)                | Masque                                     |
| Primitive                      | Primaire                                   |
| Schema document                | Document de schéma                         |
| Schema import                  | Schéma source                              |
| Schema inclusion               | Schéma inclus                              |
| Schema language                | Langage de schéma                          |
| Symbol space                   | Contexte de symboles, espace de symboles   |
| Wrapper                        | Enveloppe                                  |

|  |  |
|--|--|
| <b>Rédacteur :</b>                     | Richard Parent                             |
| <b>Organisation source :</b>           | Secrétariat du Conseil du trésor           |
| <b>Date de publication :</b>           | 21 août 2000                               |
| <b>Raison d'être :</b>                 | Connaissance technologique                 |
| <b>Programme gouvernemental :</b>      | Inforoutes et ressources informationnelles |
| <b>Nom du modificateur :</b>           | Richard Parent                             |
| <b>Date de dernière modification :</b> | 20 octobre 2000                            |
| <b>Note numéro :</b>                   | <b>9</b>                                   |