

# D

## Infoset

---

Vous trouverez ici une description détaillée du modèle Infoset du W3C. Elle vient en complément de l'introduction faite à ce modèle dans le chapitre 2.

### **Unité d'information de type document**

L'unité d'information document est celle par laquelle commence le document XML. Souvent, elle est appelée prologue.

Cette unité d'information a les propriétés suivantes :

- Des enfants pouvant être: une unité d'information de type élément qui correspond à la racine du document XML, une instruction de traitement par instruction de traitement se trouvant en dehors du contenu même du document XML et de la DTD éventuellement présente en début de document XML, une unité d'information de type déclaration de type de document quand une telle déclaration est présente dans le prologue du document XML.
- Le cas échéant une propriété notations : constituée exactement d'une unité d'information de type notation par notation déclarée dans la DTD.
- Des entités non analysées : exactement une par entité non analysée déclarée dans la DTD.
- Un URI de base : celui de l'entité document.

- Un type de codage des caractères : celui déclaré dans l'entité document (concrètement, le type de codage des caractères est déclaré *via* la déclaration XML sous la forme `<?xml version="1.0" encoding="utf-8"?>` par exemple).
- Un indicateur de complétude dont la valeur peut être *yes* ou *no*. Il indique si le document XML a déjà été analysé et canonisé ou si des valeurs par défaut restent à découvrir au travers de la DTD ou du schéma.
- Un attribut de version représentant la version de XML utilisée dans le document.
- Un indicateur de traitement des déclarations d'entités qui ne fait pas partie vraiment du document mais que le parseur renseigne pour indiquer s'il a lu ou non la DTD.

## Unité d'information de type élément

Les unités d'information de type élément correspondent aux balises contenues dans un document XML. En particulier, l'élément racine est une unité d'information de type élément qui fait l'objet d'une propriété de l'unité d'information de type document.

Cette unité d'information a les propriétés suivantes :

- Un espace de noms qui est renseigné quand l'élément est rattaché à un espace de noms spécifique.
- Un nom local qui est le nom de l'élément diminué de l'éventuel préfixe représentant un espace de noms et du signe « deux-points » qui va avec.
- Le cas échéant un préfixe qui est celui associé à l'espace de noms.
- Un ensemble d'enfants qui représentent chacun l'une des unités d'information constitutives du corps du document XML. Ensemble ordonné selon leur ordre d'apparition dans le document.
- Un ensemble d'unités d'information de type attribut, exactement une par attribut porté explicitement ou implicitement par l'élément.
- Un ensemble d'unités d'information de type attributs représentant des déclarations d'espaces de noms : exactement une par déclaration d'espace de noms, y compris si un espace de noms par défaut est déclaré.
- Une liste d'espaces de noms valides pour l'élément courant.
- L'URI de base valable pour l'élément.
- Le parent de l'élément courant.

## Unité d'information de type attribut

Les unités d'information de type attribut correspondent aux attributs portés par les éléments. Elles sont utilisées pour tous les attributs, y compris ceux définis avec une valeur par défaut dans la DTD ainsi que l'attribut spécial de déclaration d'espace de noms.

Cette unité d'information a les propriétés suivantes :

- Le cas échéant un nom d'espace de noms.
- Un nom local.
- Un préfixe.
- Une valeur normalisée : il s'agit de la valeur de l'attribut suite à sa mise en forme lexicale conformément aux règles de normalisation édictées par XML ou les catalogues de types de XML Schema.
- Un indicateur permettant de savoir si l'attribut était présent dès le départ dans le document XML où s'il a été ajouté par le parseur (ce qui arrive quand l'attribut a une valeur par défaut attribuée par la DTD ou le schéma XML).
- Un indicateur du type de l'attribut.
- Un indicateur de référence pour les attributs qui sont de type identificateur ou référence à une entité ou une notation.
- Un indicateur permettant de connaître l'élément propriétaire de l'attribut.

## Unité d'information de type instruction de traitement

Ce type d'unité d'information correspond aux instructions de traitement présentes dans le document XML.

Il a les propriétés suivantes :

- Une cible.
- Un contenu.
- Un URI de base.
- Le cas échéant une notation correspondant au nom de la cible.
- Un parent qui est l'unité d'information élément de type document ou définition de type de document.

## Unité d'information de type référence d'entité non résolue

L'unité d'information de type référence d'entité non résolue est utilisée quand le processeur XML n'a pas expansé une entité externe.

Cette unité d'information a les propriétés suivantes :

- Le nom de l'entité référencée.
- Le cas échéant l'identifiant système de l'entité.
- Le cas échéant l'identifiant public de l'entité.
- Le cas échéant l'URI de base de l'identifiant public.
- L'unité d'information parent qui la contient.

## Unité d'information de type caractère

Il existe une unité d'information de type caractère par caractère de donnée du document, même ceux qui apparaissent sous la forme de référence de caractère.

Cette unité d'information a les propriétés suivantes :

- Le code ISO 10646 du caractère.
- Un booléen qui indique si un espace blanc valide fait partie du contenu du document ou non.
- Un parent qui est une unité d'information de type élément.

## Unité d'information de type commentaire

Il existe une unité d'information de type commentaire par commentaire présent dans le document, à l'exception de ceux se trouvant dans la DTD du document.

Cette unité d'information a les propriétés suivantes :

- Un contenu constitué de la chaîne de caractères formant le commentaire à proprement parler.
- Un parent qui est une unité d'information de type élément ou document.

## Unité d'information de type déclaration de document

L'unité d'information de type déclaration de document existe quand le document XML a une déclaration de type de document. Les éventuelles déclarations d'entités et notations définies dans la déclaration de type de document sont des propriétés de l'unité d'information de type document et non de celle que nous sommes en train d'écrire.

Cette unité d'information a les propriétés suivantes :

- Le cas échéant un identificateur système permettant de référencer l'éventuel corps de la DTD qui peut être stocké dans un fichier à part.
- Le cas échéant un identificateur public.
- Des enfants composés des unités d'information de type instruction de traitement présentes dans la DTD.
- L'unité d'information parent de type document.

## Unité d'information de type entité non analysée

Une unité d'information de type entité non analysée existe pour chaque entité générale déclarée dans la DTD.

Cette unité d'information a les propriétés suivantes :

- Le nom de l'entité.
- Le cas échéant un identificateur système.
- Le cas échéant un identificateur public.
- Le cas échéant l'URI de base de l'identificateur système.
- Le cas échéant un nom de notation.

## Unité d'information de type notation

Une unité d'information de type notation existe pour chaque notation définie dans la DTD.

Cette unité d'information a les propriétés suivantes :

- Un nom.
- Le cas échéant un identificateur système.
- Le cas échéant un identificateur public.
- Le cas échéant l'URI de base de l'identificateur système.

## Unité d'information de type espace de noms

Chaque élément du document a comme propriété une unité d'information de type espace de noms pour chaque espace de noms en vigueur au niveau de l'élément.

Cette unité d'information a les propriétés suivantes :

- Le préfixe associé à l'espace de noms.
- Le nom de l'espace de noms que ce préfixe représente.