

The logo for Tirème SARL, featuring the company name in white, uppercase letters on a dark blue rectangular background.

SGML, Langage normalisé de balisage généralisé

Rédaction : Pierre Attar

Comme [XML](#), [SGML](#) est un langage de codage de données dont l'objectif est de permettre, dans un échange entre systèmes informatiques, de transférer, en même temps, des données textuelles et leurs structures. [SGML](#) était très utilisé avant qu'[XML](#) n'existe : c'est d'ailleurs à partir de ce langage (et en tenant compte de ses différentes utilisations) qu'a été élaboré le standard [XML](#).

On a dit beaucoup de choses sur [SGML](#) : que le langage était lourd, que les implémentations étaient difficiles, etc. De fait, ce n'était pas la technologie qui était lourde, mais les applications elles-mêmes. En effet, s'il est nécessaire de manipuler un manuel de maintenance aéronautique de quelques centaines de méga-octets, avec des arbres d'une profondeur allant jusqu'à des dizaines de niveaux, la technologie n'y fait rien (qu'elle soit [XML](#) ou [SGML](#)) : les manipulations sont de toutes les façons complexes !

C'est ce dont est en train de se rendre compte la communauté [XML](#), qui comprend que ce standard sera utilisé pour différentes classes d'applications : depuis l'échange de messages entre applications, avec des informations de l'ordre du kilo-octet jusqu'à des documents beaucoup plus lourds et beaucoup plus sémantiques, comme le sont un dictionnaire ou un ouvrage juridique.

En termes de technologie, la lourdeur de [SGML](#) est certainement liée à ses [DTD](#) qui mélangeaient deux type d'information : la notion de modèle de donnée et le format de codage des documents conformes à ce modèle. Par ailleurs, le langage lui-même proposait beaucoup de facilités de saisie simplifiées des noms d'élément : toute chose difficiles à prendre en compte dans les outils de traitement.

D'un point de vue technique donc, la grande avancée conceptuelle de [XML](#) est donc de différencier définition de modèle et format de codage des documents. C'est en cela que [XML](#) supprime des lourdeurs de [SGML](#).

L'avenir de [SGML](#) ? D'un point de vue technique, ce format peut sans problème être remplacé par [XML](#) : tout ce qui est nécessaire à [SGML](#) existe à peu près dans [XML](#). D'un point de vue pratique, les utilisateurs de [SGML](#) y gagneront en diversité d'outils utilisables et apprécieront à son juste avantage la prise en compte par [XML](#) d'[Unicode XML](#).

D'un point de vue plus stratégique, beaucoup se posent la question de passer de [SGML](#) à [XML](#), en raison des organismes de normalisation qui s'en occupent. Pour [SGML](#), c'est l'[ISO](#) qui assure la normalisation, avec ses processus de maturation d'une norme et ses représentations nationales assurant la qualité des normes ; en ce qui concerne la normalisation de [XML](#), c'est le travail du [W3C](#), un Consortium qui n'est représentatif que de

lui-même et qui, comme son nom l'indique, ne s'intéresse qu'au point de vue du Web. Cet argument devrait rapidement être balayé par la volonté commune des deux organismes de promouvoir XML au sein des normes documentaires de l'ISO.

Objectifs

L'objectif de SGML était de définir un format neutre de rétention d'information chez un photocomposeur. ce format neutre devait permettre aux éditeurs de mieux faire jouer la concurrence et d'autoriser un éditeur à renégocier ses contrats avec les photocomposeurs pour, par exemple, pouvoir transférer ses documents chez celui qui proposerait le meilleur prix. Par la suite, et comme pour XML, l'objectif de SGML fut de définir un format neutre de codage de document permettant de mettre en place des notions de réutilisation de la même information pour différentes publications, sur différents médias.

Aujourd'hui, l'objectif de SGML est le même que celui de XML et cette norme est appelée à être remplacée par XML.

D'un point de vue pratique, tout document SGML peut être transformé de façon totalement automatisée en document XML. Les DTD de SGML devront subir quelques modifications qui les rendront certainement parfois plus laxiste. En effet, XML ne permet pas de prendre en compte les mécanismes d'inclusion et d'exclusion" propres à SGML. Il permet aussi un typage des attributs plus restreint. Cette dernière limite est maintenant bien levée avec la standardisation des Schema.

Principes

À la différence d'XML, un document SGML doit toujours être valide au regard d'une DTD. Les concepteurs, en raison de leur orientation documentaire, s'intéressaient beaucoup à la notion de validation électronique et automatique. Cette validation au regard d'une DTD permet de s'assurer que l'information est saisie conformément au modèle et que l'on peut donc lui appliquer les traitements automatisés sans avoir à valider la structure du document au préalable. A noter que cela reste vrai pour XML mais que cette dernière recommandation supprime l'obligation de validation pour lecture car des processus d'assurance qualité peuvent remplacer cette validation à toutes les étapes d'un processus informatique.

Dans les DTD, une des erreurs de jeunesse de SGML était de définir, en même temps que le modèle, les variations possibles du format de codage des documents eux-mêmes : options de minimisation des balises et, si nécessaire, options de redéfinition des balises elles-mêmes. Du fait de la standardisation d'Unicode XML et des nouvelles capacités des ordinateurs actuels, toutes ces options ont été supprimées, pour arriver à l'obligation de "document bien formé", notion qui était, de fait, la sortie standard de tous les éditeurs SGML.

Enfin, d'autres mécanismes peu utilisés existaient, comme ceux d'inclusion et d'exclusion. Il n'ont pas été jugés utiles dans XML, car apportant de la complexité aux traitements sans pour autant faciliter la vie du rédacteur.

Recommandations(s)

■ ■ *Traitement de l'information - Systèmes bureautiques - Langage normalisé de balisage généralisé*

Standard ISO, version 8879:1986, du 17-08-1996

Document sur [http://www.iso.ch/iso/fr/CatalogueDetailPage.CatalogueDetail?
CSNUMBER=16387](http://www.iso.ch/iso/fr/CatalogueDetailPage.CatalogueDetail?CSNUMBER=16387)

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

Standard ISO, version 8879:1986, du 17-08-1996

Document sur [http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?
CSNUMBER=16387](http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=16387)