
Éditorial :

XML ou la démocratisation du web

Michel GOOSSENS

CERN, Division IT, CH-1211 Genève 23, Suisse, <Michel.Goossens@cern.ch>

À la fin de l'année 1996, les développeurs de pages pour la toile mondiale (le *Web*) étaient de plus en plus désespérés en voyant apparaître tous les jours de nouvelles balises non-standard dans tel ou tel butineur (*browser*). Les auteurs de pages devaient s'assurer continuellement que l'information présentée pouvait être visualisée de façon identique par tous les surfers du monde, chaque butineur ayant sa propre idée de la présentation d'une balise donnée.

Il n'était donc pas surprenant que quelques visionnaires se soient mis ensemble, pour proposer un langage « normalisé » mais extensible pour baliser l'information sur le Web : XML (*Extensible Markup language*, ou langage de balisage extensible). Le langage XML a vu officiellement le jour le 10 février 1998 avec la publication de la recommandation W3C¹. Dans son article culte : « XML, Java et le futur du Web » (<http://www.xml.com/pub/w3j/s3.bosak.html>) Jon Bosak a parfaitement décrit les questions à résoudre dans les trois domaines où HTML posait problème :

- *l'extensibilité* : HTML a un vocabulaire figé et l'utilisateur ne peut pas élargir le langage en inventant ses propres éléments ou attributs pour mieux qualifier sémantiquement ses données ;
- *la structure* : HTML n'autorise pas la construction de structures assez profondes pour permettre son utilisation avec les bases de données ou les hiérarchies orientés objets ;
- *la validation* : HTML ne permet pas aux applications clients de vérifier la validité structurelle des données reçues.

De son côté SGML, norme publiée par l'ISO en 1986, offre trop de possibilités optionnelles qui ne sont pas nécessaires pour les applications Web et qui seraient beaucoup trop coûteuses en ressources pour une utilisation dans les butineurs.

Jon Bosak et les co-concepteurs de XML se sont inspirés de la norme SGML tout en éliminant les parties considérées trop complexes ou inutiles pour le quotidien du

1. Le Consortium de la toile mondiale (*World Wide Web Consortium*) est une organisation réunissant quelques 350 organisations et entreprises actifs dans le domaine du Web. Il fut créé début 1995 et est basé au MIT (États unis), à l'INRIA (France) et au Keio (Japon). Les recommandations W3C sont l'équivalent d'une norme ISO, AFNOR, ANSI, etc. ; elles sont acceptées pas consensus entre *tous* les membres.

Web. La spécification XML (<http://www.w3.org/TR/REC-xml>, voir page 191) tient en une trentaine de pages A4, alors que la norme SGML en nécessite cinq fois davantage.

Ce nouveau langage fut adopté avec enthousiasme pratiquement instantanément par la communauté de l'internet. En fait, c'était le commencement d'une vraie démocratisation du Web, où tout le monde peut actuellement définir son propre langage, puis le traduire en HTML, \LaTeX , etc. d'une façon *standard* et en utilisant presque uniquement des utilitaires gratuits. En effet, en moins de deux ans, le nombre de programmes d'analyse pour XML disponibles gratuitement sur l'internet se compte en dizaines, alors que pour SGML nous avons dû attendre dix ans pour en avoir un seul à accès non-payant ! Aussi le nombre de livres parus depuis un an avec le mot « XML » dans le titre est-il en gros plus élevé que tous ceux traitant de SGML publiés depuis 1986.

Ce *Cahier* offre un premier contact avec le monde de XML, point de départ avec beaucoup de pointeurs vers des sources d'information complémentaires.

Nous commençons par une introduction à XML et son langage de feuille de style associé XSL (page 3). Ce chapitre doit être lu de pair avec la spécification XML dont on trouvera le texte complet en version bilingue français/anglais page 191 et suivantes. L'information pourra aussi être complétée par la foire aux questions (*Frequently Asked Questions*) de Peter FLYNN dont la traduction française est page 281 et suivantes.

La deuxième partie de ce *Cahier* contient des articles correspondant aux présentations faites à la journée XML du congrès GUTenberg GUT'99 le 19 mai 1999 à Lyon. On y trouvera une comparaison entre SGML et XML (page 127) par Sarra BEN LAGHA et ses collaborateurs, une introduction aux modèles objets pour les documents (page 155) par François ROLE et Philippe VERDRET, puis deux articles qui parlent de MathML, un vocabulaire XML pour les mathématiques. Le premier décrit une extension d'Oméga (page 173) par Yannis HARALAMBOUS et John PLAICE et le deuxième un compilateur d'expressions mathématiques générant du MathML (page 183) par Benjamin JENNES et Raphaël MARÉE.

J'espère que ce numéro dédié à XML permettra au maximum d'entre vous de « se lancer » dans l'aventure des langages structurés et extensibles de l'internet. Ainsi vous pourrez profiter pleinement de la flexibilité, de l'adaptabilité et de la richesse de cette nouvelle approche qui permet une optimisation de la communication électronique à tous les niveaux, dans toutes les langues et sur tous les équipements, du plus perfectionné au plus élémentaire.