

HTML, comme dans la plupart des programmes de traitement de texte, un document est en général dans une seule langue : français, japonais, arabe... Si le programme ne sait pas lire les caractères de la langue de l'utilisateur, ce dernier ne peut pas utiliser le programme. Pis encore, les logiciels destinés à être utilisés à Taiwan sont souvent incapables de lire les textes issus de la Chine continentale, à cause d'incompatibilités du codage des caractères. Au contraire, un programme qui comprend correctement le XML peut faire face à n'importe quelle combinaison de ces jeux de caractères. XML permet l'échange d'information non seulement entre des systèmes informatiques différents, mais aussi au travers des barrières nationales et culturelles.

La fin de l'attente

L'avènement de XML devrait accélérer les échanges sur le réseau Internet. Aujourd'hui, les ordinateurs connectés au réseau, quelle que soit leur puissance, récupèrent des formulaires, les remplissent et leur font faire la navette jusqu'à un serveur, jusqu'à l'aboutissement de la requête. L'information structurelle et sémantique qu'ajoute XML permet aux ordinateurs de faire une bonne partie du traitement localement, sans intervention des serveurs distants, qui se trouvent soulagés d'autant. En outre, le trafic sur le réseau sera également réduit.

Pourquoi ce gain de rapidité? Imaginez que vous vous connectiez au site Internet d'une agence de voyages afin de connaître tous les vols Paris-New York du 11 août. Vous recevez une liste plus longue que ce que votre écran peut afficher. Comment préciser votre choix? En spécifiant l'heure approximative de départ, la gamme de prix ou la compagnie aérienne que vous désirez emprunter. Toutefois, pour obtenir cette liste écourtée, vous devrez envoyer une nouvelle requête au serveur de l'agence de voyages et attendre la réponse. Si la longue liste de vols avait été envoyée en XML, l'agence de voyages aurait pu vous fournir, en même temps que les caractéristiques des vols, un petit programme (en langage Java) que vous auriez pu utiliser sur votre machine pour faire un tri à votre convenance en quelques microsecondes, sans solliciter à nouveau le serveur (voir la

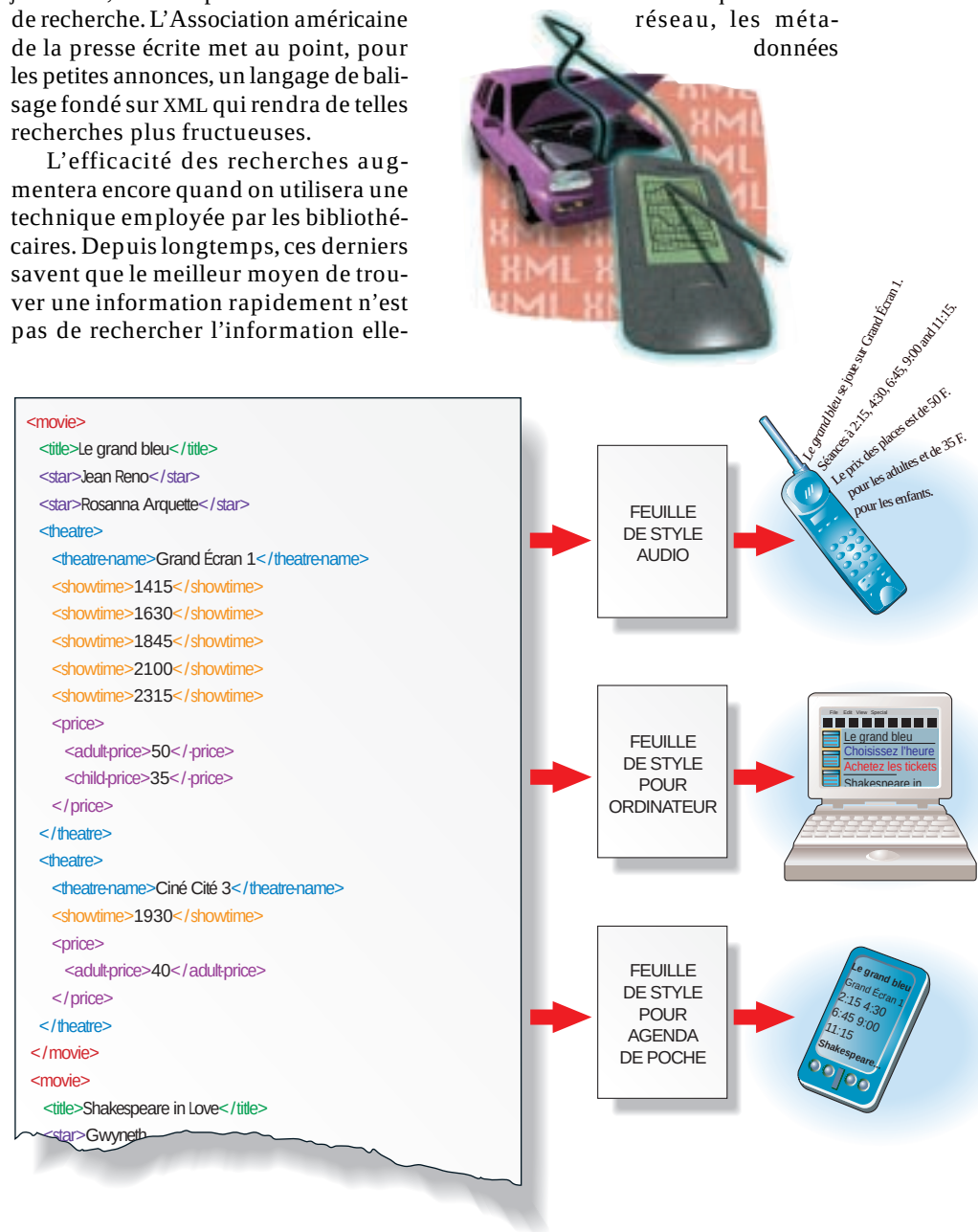
figure 3). Ce gain de temps, multiplié par plusieurs millions d'utilisateurs, sera la cause d'une amélioration notable du réseau Internet.

Plus l'information circulant sur le réseau sera étiquetée avec des balises XML spécifiques à chaque corps de métier, plus on trouvera facilement des informations précises. Aujourd'hui une recherche d'«offres d'emploi d'ingénieur système» récupère beaucoup de publicités, mais peu d'offres, parce que la plupart sont dans les services de petites annonces des sites des divers journaux, hors de portée des moteurs de recherche. L'Association américaine de la presse écrite met au point, pour les petites annonces, un langage de balisage fondé sur XML qui rendra de telles recherches plus fructueuses.

L'efficacité des recherches augmentera encore quand on utilisera une technique employée par les bibliothécaires. Depuis longtemps, ces derniers savent que le meilleur moyen de trouver une information rapidement n'est pas de rechercher l'information elle-

même, mais plutôt des ensembles de données plus spécifiques, qui orientent vers les sources utiles. Ces informations sur l'information, dont les catalogues de bibliothèque sont un bon exemple, sont des métadonnées.

Depuis le début du projet XML, on envisage la création d'un standard pour les métadonnées, fondé sur le langage XML lui-même. Le «système descriptif de ressources» RDF (*Resource Description Framework*), publié en février 1999, est l'équivalent pour les données Internet des catalogues pour les livres de bibliothèque. Sur le réseau, les métadonnées



2. RÉDIGÉ DANS LE LANGAGE XML, un document qui contient les programmes de cinéma d'une ville peut être visualisé sur des machines différentes. Des «feuilles de style» filtrent, réarrangent et affichent l'information : elles engendrent des documents comprenant des graphiques pour les ordinateurs de bureau, des fichiers composés exclusivement de texte pour les agendas de poche ou des messages vocaux pour les téléphones.

rendront la recherche de l'information bien plus rapide et précise qu'elle ne l'est aujourd'hui. Comme le réseau n'a pas de bibliothécaires et que tous les auteurs de site souhaitent que leur site soit connu, nous prévoyons que le RDF accroîtra le trafic Internet lorsque son potentiel sera reconnu.

D'autres méthodes que les moteurs de recherche permettent de retrouver de l'information. Le réseau est avant tout un hypertexte dont les millions de pages sont reliées par des liens, ces mots soulignés sur lesquels on clique pour passer de page en page. XML rendra plus efficace l'utilisation des liens : un standard d'hypertexte fondé sur XML, XLink, dont les spécifications seront publiées cette année par le Consortium W3C, permettra de choisir des destinations multiples à partir d'une liste. D'autres types de liens feront apparaître du texte ou des images au lieu de faire quitter la page.

Mieux encore, XLink permettra aux auteurs d'utiliser des liens indirects,

qui dirigeront vers des rubriques dans une base de données centrale plutôt que vers des pages cibles elles-mêmes. Lorsque l'adresse d'une page changera, l'auteur pourra mettre à jour tous les liens qui s'y rapportent en modifiant un seul enregistrement dans la base de données. Cette procédure devrait contribuer à faire disparaître les erreurs fréquentes de type «404 File Not Found» qui se produisent quand les liens hypertexte sont cassés.

Grâce à un traitement plus efficace, à des recherches plus précises et au système de liens perfectionné, la structure

du réseau sera transformée, et ses utilisateurs bénéficieront de moyens nouveaux d'accéder à l'information.

Une nécessaire refonte

Ce perfectionnement nécessite du soin et du travail. Si XML permet la mise au point de langages sur mesure, la conception de bons langages reste difficile. La conception n'est en outre qu'un premier stade : pour que la signification d'une balise soit claire, on doit rédiger des explications et créer de petits programmes pour que les ordinateurs puissent les intégrer.

Horaires des vols - JFK - Navigateur XML

Heure	Date	Heure	Destination	Compagnie	Prix
8h 00	11/08/99	7h 55	Paris (CDG) à New York (JFK)	Vif Air	115
8h 45	11/08/99	7h 55	Paris (CDG) à New York (JFK)	Vif Air	118
8h 55	11/08/99	7h 55	Paris (CDG) à New York (JFK)	Vif Air	120
10h 00	11/08/99	7h 55	Paris (CDG) à New York (JFK)	Vif Air	116
10h 55	11/08/99	7h 55	Paris (CDG) à New York (JFK)	Vif Air	121
12h 00	11/08/99	7h 55	Paris (CDG) à New York (JFK)	Vif Air	119
13h 15	11/08/99	7h 55	Paris (CDG) à New York (JFK)	Vif Air	117
13h 55	11/08/99	7h 55	Paris (CDG) à New York (JFK)	Vif Air	122
14h 00	11/08/99	7h 55	Paris (CDG) à New York (JFK)	Vif Air	125
14h 00	11/08/99	7h 55	Paris (CDG) à New York (JFK)	Vif Air	127
14h 05	11/08/99	7h 55	Paris (CDG) à New York (JFK)	Vif Air	129

Confirmation du vol - Navigateur XML

Voire réservation va être prise en compte. Vous devez acheter vos billets dans les 72 heures. Confirmez-vous votre réservation?

Conditions tarifaires :

- Doit rester une nuit de samedi sur place.
- Les billets ne sont pas remboursables.
- Un changement de destination coûtera 500 francs en plus du changement de prix.

Recherche de vols Vif Air - Navigateur XML

Essayez notre moteur de recherche de voyage aller-retour : [Inscrivez-vous](#)

Réserver un vol

Ville de départ : Départ : Horaire :

Ville d'arrivée : Retour :

Cette recherche



3. UN LIEN XML conduit à un menu qui comprend plusieurs options. L'une d'elles insère une image sur la page courante, tel le plan d'occupation des sièges de l'avion (flèche rouge). D'autres options déclenchent un petit programme de réservation de vol (flèche jaune) ou révèlent un texte caché (flèche verte). Les liens peuvent aussi, comme en HTML, aboutir à d'autres pages.

S'il suffisait d'étiqueter un bon de commande avec une balise «commande» pour que l'ordinateur puisse le traiter, nous n'aurions pas besoin de XML. Nous n'aurions même pas besoin de programmeurs, les ordinateurs se débrouilleraient tout seuls.

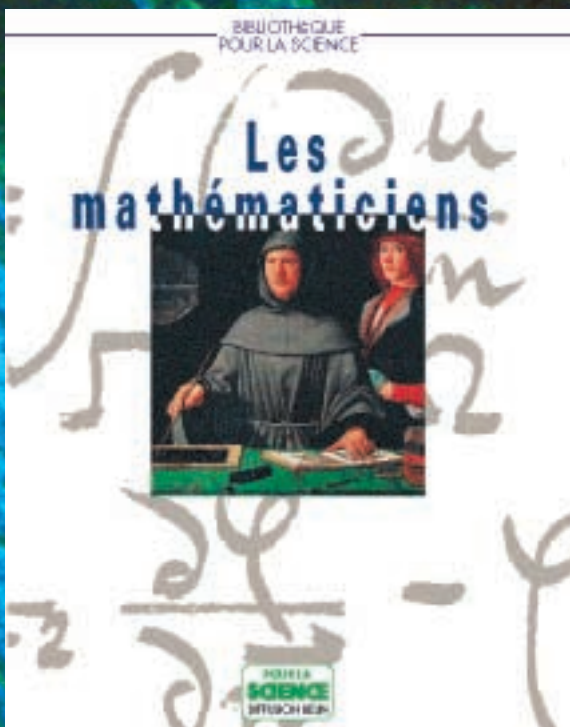
Le langage XML n'est pas magique, mais il est efficace ; il définit des règles de base qui éliminent une couche de programmation correspondant à des détails, de sorte que des personnes partageant les mêmes intérêts peuvent se concentrer sur la véritable difficulté : s'entendre sur la façon dont ils souhaitent représenter l'information qu'ils échangent couramment. Ce problème n'est pas simple, mais il n'est pas nouveau.

Ces harmonisations sont incontournables, car la prolifération de systèmes informatiques incompatibles a engendré des retards, des coûts et de la confusion dans la plupart des activités humaines. Les utilisateurs de l'informatique veulent échanger des idées ou faire des affaires sans utiliser tous les mêmes ordinateurs ; des langages d'échange spécifiques à chaque activité y contribuent grandement. De fait, l'avalanche d'acronymes se terminant par ML témoigne de la créativité qui s'exprime grâce à XML en sciences ou dans le monde des affaires (voir l'encadré de la page 52).

Les concepteurs d'un nouveau langage doivent s'accorder sur trois points : quelles balises seront acceptées, comment les éléments balisés pourront être emboîtés et comment ils seront traités. Les deux premières spécifications, le vocabulaire du langage et sa structure, sont codifiées dans une définition de type de document (ou DTD). Le standard XML n'impose pas de telles définitions, mais la plupart des nouveaux langages en comporteront probablement, parce qu'elles simplifient le travail des programmeurs. Ces derniers doivent en effet écrire des programmes qui comprennent le balisage et s'en servent intelligemment.

Les programmeurs auront également besoin d'un ensemble de directives qui décrivent en langage intelligible ce que signifient toutes les balises. Le langage HTML, par exemple, est défini par une DTD et par quelques centaines de pages de descriptions que consultent les programmeurs quand ils écrivent des programmes de navigation ou d'autres programmes pour le réseau Internet.

BIBLIOTHÈQUE POUR LA SCIENCE



LES MATHÉMATICIENS

Pour comprendre le processus de la création mathématique, il faut saisir ce qui fait l'originalité d'un mathématicien. Cet ouvrage réunit les portraits de quinze grands mathématiciens. Leurs rapports avec leur époque, le regard qu'ils portent sur leur activité sont plus qu'anecdotiques : ils révèlent la genèse des idées nouvelles et mettent en lumière le caractère audacieux, dynamique et imaginaire de l'invention mathématique.

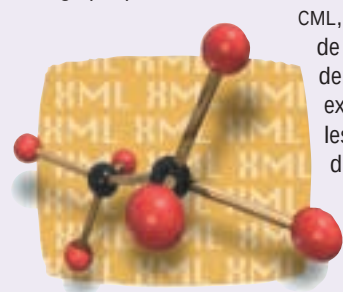
Voir bon de commande p. 98

POUR LA
SCIENCE
DIFFUSION BELIN

Les nouveaux langages des sciences

Gâce à XML, les scientifiques échangent plus facilement des théories, des calculs et des résultats expérimentaux. Depuis longtemps, les mathématiciens étaient frustrés par l'inaptitude des navigateurs à afficher les expressions mathématiques autrement que sous forme d'images. Le langage MathML, opérationnel depuis le printemps 1999, leur permet de couper des équations dans les pages Internet et de les coller directement dans leurs logiciels de calcul formel, en vue de les utiliser pour des calculs ou pour des graphiques.

Les chimistes vont plus loin encore : ils ont mis au point de nouveaux navigateurs pour leur langage CML (*Chemical Markup Language*), fondé sur XML, qui donne des représentations graphiques de la structure moléculaire des composés décrits dans les pages Internet.



CML, de même que le langage de l'astronomie, en cours de mise au point, aide les chercheurs à explorer rapidement une multitude de citations de revues pour en extraire les articles correspondant à leur sujet d'étude : les astronomes, par exemple, peuvent rentrer les coordonnées célestes d'une galaxie, afin de récupérer une liste d'images, d'articles et de données instrumentales sur l'objet en question.

XML sera utile pour la réalisation d'expériences aussi bien que pour l'analyse de leurs résultats. En 1998, les ingénieurs de la NASA ont entrepris de mettre au point l'AIML (*Astronomical Instrument ML*, soit « langage ML pour les instruments astronomiques »), qui commande le télescope infrarouge SOFIA lors de ses vols stratosphériques à bord d'un Boeing 747. À terme, AIML permettra aux astronomes du monde entier de commander des télescopes et, peut-être même, des satellites, par l'intermédiaire d'un simple logiciel de navigation sur le réseau Internet.

Pendant ce temps, sur Terre, les généticiens utiliseront bientôt BSML (*Biosequence ML*), afin d'échanger et de manipuler les quantités énormes d'information qui résultent des innombrables projets de séquençage du génome. Un navigateur BSML, mis au point et distribué gratuitement par la Société *Visual Genomics*, permet aux chercheurs de fouiller de gigantesques bases de données de code génétique et d'afficher les fragments trouvés sous forme de cartes et de tableaux, bien plus explicites que les suites de lettres habituelles.

Question de style

Les utilisateurs, eux, sont préoccupés de ce que font les programmes, et non de leur description. Ils veulent que les programmes affichent de manière intelligible l'information codée en XML, mais les balises XML ne donnent pas d'indications sur la présentation de l'information sur l'écran ou sur le papier.

Cette caractéristique est d'ailleurs un avantage pour les éditeurs, qui veulent souvent réutiliser le contenu d'un document, c'est-à-dire décliner une publication sous une multitude de formes, imprimées ou électroniques. En balisant le contenu d'un texte pour en décrire son sens, XML autorise cette publication multiple, indépendamment du type de support. Les éditeurs peuvent ensuite appliquer des règles organisées en feuilles de style pour mettre en forme automatiquement le passage sous diverses formes. Le standard pour les feuilles de style XML, en cours de mise au point, est XSL (pour *eXtensible Stylesheet Language*).

Les versions les plus récentes de certains navigateurs Internet peuvent lire un document XML, récupérer la feuille de style appropriée, et l'utiliser pour trier et formater l'information sur l'écran. L'utilisateur ne verra pas la différence entre des informations codées par HTML ou par XML, mais l'accès aux sites XML sera bien plus rapide.

Les malvoyants gagneront beaucoup de cette approche de l'édition, car les feuilles de style leur permettront de présenter un document XML en braille ou sous forme parlée. De même, les pages Internet pourront être lues à voix haute par les ordinateurs de bord des voitures (voir la figure 2).

Un traducteur universel

Bien que le réseau Internet ait été créé par le milieu universitaire et scientifique, c'est l'industrie et le commerce qui, attirés par les potentialités du système, ont suscité son expansion fulgurante. Le commerce électronique s'est imposé, mais les transactions entre entreprises se développent au

moins aussi rapidement. Les industries aimeraient utiliser le réseau pour régler leur production sur les commandes qu'elles enregistrent, mais cet ajustement, qui limiterait les stocks, impose une automatisation des chaînes de production et des communications avec les fournisseurs de matières premières ou de pièces détachées. Or, les interactions de systèmes informatiques sont gênées par l'hétérogénéité des traitements.

Depuis des siècles, les affaires se font par l'échange de documents standardisés : bons de commande, factures, reçus. Ces documents assurent les interactions des parties concernées. Chaque document contient exactement ce que son destinataire a besoin de savoir, et rien de plus. De même, l'échange de documents reste le fondement des transactions commerciales en ligne, mais le langage HTML n'a pas été conçu pour ces échanges. En revanche, XML a été conçu pour l'échange de documents, et nous ne doutons pas que le commerce électronique mondial reposera largement sur la circulation de transactions matérialisées par des millions de documents XML échangés par l'Internet.

Ainsi, le réseau renforcé par XML sera plus rapide, plus convivial et mieux adapté aux affaires. En contrepartie, les concepteurs des sites auront plus de travail. Des bataillons de programmeurs seront nécessaires pour tirer un plein parti des nouveaux langages XML. Bien que les jours des autodidactes du réseau ne soient pas encore comptés, l'espèce est en danger. Les concepteurs de demain devront être versés non seulement dans la production de textes et de graphiques, mais encore dans la construction de systèmes combinant à plusieurs niveaux des arbres de données, des structures de liens hypertexte, des métadonnées et des feuilles de style, qui, ensemble, constituent l'infrastructure robuste du réseau de deuxième génération.

Jon BOSAK, qui travaille pour la Société *Sun Microsystems*, est président du groupe de coordination XML du Consortium W3C. Tim BRAY, fondateur de la Société *Textuality*, est coéditeur de la norme XML 1.0.

Alain MICHARD, *XML : langage et applications*, Eyrolles, 1998.

Des liens pour cet article sont disponibles sur notre site : <http://www.pourlascience.com>

Cônes et toxines

FRÉDÉRIC LE GALL • PHILIPPE FAVREAU • GEORGES RICHARD

Pour paralyser leurs proies, ces coquillages utilisent des venins qui contiennent d'innombrables toxines dont l'action est très spécifique. Certaines ont été isolées et seraient utilisées comme analgésiques et comme anticonvulsivants chez l'homme.

En 1796, lors d'une vente aux enchères, une œuvre du peintre flamand Vermeer, la *Femme en bleu lisant une lettre* est vendue 43 florins. À la même enchère, un spécimen de *Conus cedonulli*, un mollusque marin de cinq centimètres de longueur, est vendu plus de six fois plus cher !

La famille des Conidae suscitait déjà l'enthousiasme chez les collectionneurs et les artistes. Selon le naturaliste Lamarck, «le genre cône est le plus beau, le plus étendu et l'un des plus considérables de la classe des univalves. C'est celui qui renferme les coquilles les plus précieuses et en même temps les plus remarquables, soit par la régularité de leur forme, soit par l'éclat et l'admirable variété de plusieurs d'entre elles. Aussi sont-elles très recherchées des amateurs».

Aujourd'hui, si les collectionneurs sont toujours aussi nombreux, c'est le venin produit par les Conidae, ou cônes, qui intéresse les biologistes et surtout les neurobiologistes.

Les toxines contenues dans ces venins, nommées conotoxines, perturbent le fonctionnement du système nerveux ou des muscles. Les cônes utilisent une panoplie de conotoxines qui, agissant en synergie, paralysent les proies.

Ces toxines ont des cibles spécifiques et sont utilisées par les biologistes pour étudier certaines fonctions du système nerveux et pour caractériser de nouveaux types de récepteurs et de canaux ioniques. Leur spécificité d'action physiologique fait de ces toxines des agents utilisables dans le traitement de la douleur et des convulsions, notamment dans les épilepsies.



1



2



3



4



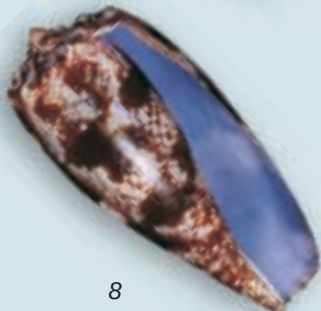
5



6



7



8



9



10



11



12



13



14



15