Le langage XML

JON BOSAK • TIM BRAY

Le réseau Internet s'est considérablement développé grâce à la mise au point de systèmes «hypertexte» qui gèrent les images, les textes, les sons. Le langage XML accentuera ce développement.

Quelques indices nous suffisent pour reconstruire une information : en jetant un coup d'œil à cette page, et en repérant des mots imprimés en gros caractères, suivis de colonnes en petits caractères, nous déduisons que nous avons sous les yeux le début d'un article. De même, un simple coup d'œil sur une liste de produits d'épicerie montre que celle-ci est une liste de courses. En consultant quelques rangées de chiffres, on identifie un relevé de compte bancaire.

Les ordinateurs n'ont pas cette capacité : on doit leur indiquer précisément la nature des objets qu'on leur fait manipuler, les liens entre ces objets et ce qu'ils sont censés en faire. Le langage XML (*eXtensible Markup Language*, soit «langage de balisage extensible») est conçu justement pour rendre l'information autodescriptive. Ce changement de la communication entre les ordinateurs, simple en apparence, pourrait étendre l'utilisation de l'Internet à nombre d'activités humaines, au-delà de la simple transmission d'information. Depuis la création de XML, au début de 1998, par le Consortium W3C *(World Wide Web Consortium)*, il s'est répandu comme une traînée de poudre dans les milieux scientifiques et industriels.

Cet engouement s'explique par l'espoir que XML comblera certaines des principales lacunes du réseau Internet. L'une d'elles est la lenteur : sur le réseau Internet, l'information est censée circuler à la vitesse de la lumière, mais, en pratique, elle arrive aux utilisateurs à la vitesse de l'escargot. En outre, bien qu'une quantité phénoménale d'information soit disponible, la recherche de l'information adéquate est souvent fastidieuse.

Ces deux problèmes sont en grande partie dus à la nature du langage principal du réseau, le langage HTML (l'acronyme de *Hypertext Markup Language*, soit «langage de balisage hypertexte»). Bien que celui-ci soit le langage de publication électronique le plus populaire jamais inventé, il reste superficiel : il décrit seulement comment les programmes de navigation sur le réseau doivent organiser le texte, les images et les boutons sur les pages qu'ils affichent. Cette superficialité de HTML le rend facile à apprendre, mais cette simplicité a un prix.

Notamment, avec HTML, on crée difficilement des sites qui ne se limitent pas à envoyer des documents à ceux qui en font la demande, tels des télécopieurs améliorés. Les particuliers et les entreprises veulent des sites qui enregistrent les commandes des clients, qui transmettent des dossiers médicaux ou, même, qui gèrent des usines ou des instruments scientifiques à l'autre bout du monde. Le langage HTML n'est pas conçu pour de telles tâches.

Aujourd'hui, un médecin peut récupérer un dossier médical à l'aide d'un navigateur, mais il ne peut ensuite l'envoyer électroniquement à un collègue spécialiste afin qu'il l'intègre directement à la base de données de son hôpital. Son ordinateur ne sait pas quoi faire de l'information, qui n'est que du <H1>bla bla </H1> <BOLD>bla bla bla<p>. Le problème avec le «Ce Que Vous Voyez Est Ce Que Vous Obtenez» (la traduction française de *What You See Is What You Get* ou WYSIWYG), c'est que ce que vous voyez n'est que ce que vous obtenez.

Les étiquettes entre crochets, dans l'exemple précédent, sont des balises. Le langage HTML ne possède pas de balise qui signale une réaction allergique à des médicaments, par exemple ; sa raideur est une autre de ses limites. L'adjonction d'un nouveau type de balises impose des démarches administratives qui peuvent prendre des années, si bien que peu s'y aventurent. Pourtant, chaque application, et pas seulement l'échange de dossiers médicaux, nécessite ses propres balises.

Cette raideur explique également la lenteur actuelle des bibliothèques en ligne, des catalogues de vente par correspondance ou des divers sites interactifs. Quand on veut modifier les quantités commandées ou le mode d'acheminement des marchandises, le serveur doit renvoyer une nouvelle page, dont du texte et des graphiques déjà reçus. Pendant ce temps, votre puissant ordinateur attend bêtement, parce qu'il ne connaît que les <H1> et les <BOLD>, et non les prix et les options de transport.

De là résultent également les recherches frustrantes sur le réseau. Parce qu'il n'existe pas de moyen de signaler qu'une donnée est un prix, vous ne pouvez pas utiliser les informations tarifaires dans vos recherches.

Comment faire du neuf avec du vieux ?

En théorie, le problème serait résolu si des balises décrivaient le contenu de l'information, et non ce à quoi elle ressemble. Par exemple, au lieu de classer les différentes composantes d'une commande d'après les rubriques «caractères gras», «paragraphe», «ligne» et «colonne» (ce que fait le langage HTML), on les dénommerait «prix», «taille»,

«quantité» et «couleur». De la sorte, un programme identifierait un document contenant ces rubriques comme étant une commande, et il le gérerait de manière souple : il pourrait l'afficher, le rentrer dans un système de gestion ou vous faire livrer une nouvelle chemise demain.

Dès 1996, un groupe de travail comprenant une dizaine de membres du Consortium W3C a exploré une telle solution. L'idée, bien que puissante, n'était pas entièrement originale : depuis des générations, les éditeurs annotent les manuscrits pour guider les typographes. Ce système de balisage (*markup* en anglais) s'est progressivement formalisé jusqu'à ce qu'en 1986, après des années de travail, l'Organisation des normes internationales (ISO) approuve un système pour la création de nouveaux langages de balisage.

Nommé SGML (pour *Standard Generalized Markup Language*, soit «langage de balisage généralisé standard»), ce métalangage (langage de description de langages) a fait la preuve de son utilité dans l'édition. Le langage HTML lui-même a été défini à partir du SGML. Le seul inconvénient du SGML est qu'il est trop général, plein d'astuces destinées à minimiser les frappes, à une époque où chaque octet avait un coût non négligeable. Sa complexité est excessive pour les programme de navigation.

Notre groupe de travail a créé XML en épurant le SGML pour obtenir un métalangage plus simple. XML est un ensemble de règles que chacun peut suivre pour créer un langage de balisage. Les règles de XML garantissent qu'un seul programme simple et compact, souvent nommé analyseur, est capable de traiter tous ces nouveaux langages.

Reprenons l'exemple du médecin qui veut envoyer un dossier médical à un spécialiste. Si le corps médical utilise XML pour fabriquer un langage de balisage des dossiers médicaux (plusieurs groupes étudient ce cas précis), le courrier électronique d'un médecin signalera une réaction à des médicaments par un code tel que <patient> <nom> Moreau </nom> <allergie médicamenteuse> pénicilline </allergie médicamenteuse> </patient>. Un programme très simple permet à un ordinateur de comprendre cette notation médicale standard et d'ajouter cette statistique vitale à une base de données.

Comme le langage HTML, qui permet à tous les utilisateurs d'ordinateurs de lire les documents sur l'Internet, XML pourrait être le nouvel espéranto informatique. De surcroît, à la différence de la plupart des formats de données informatiques, les

Illustrations de Brucie Rosch

1. LE LANGAGE XML est un pont entre des systèmes informatiques hétérogènes. Il permet la recherche d'informations, l'échange de données scientifiques, de produits commerciaux et de documents multilingues avec une vitesse et une facilité accrues.

annotations XML ont un sens pour les humains, puisqu'elles sont composées de texte ordinaire.

Le potentiel unificateur de XML résulte de quelques règles bien choisies. L'une d'entre elles stipule que les balises vont presque toujours par paires. Comme les parenthèses, elles encadrent le texte auquel elles se rapportent. Comme des guillemets, les paires de balises peuvent être emboîtées.

La règle d'emboîtement simplifie les documents XML en leur conférant une structure d'arbre. À la manière d'un arbre généalogique, chaque graphique ou morceau de texte du document représente un parent, un enfant ou un frère ; les parentés sont sans ambiguïté. Les arbres ne peuvent pas représenter n'importe quelle information, mais ils représentent la plupart des types d'information nécessaires au fonctionnement des ordinateurs. De surcroît, les arbres sont informatiquement très pratiques. Si un relevé de compte bancaire se trouve sous forme d'un arbre, on écrit très simplement un petit programme qui ordonne les transactions ou qui affiche seulement les chèques tirés avant une certaine date.

Le potentiel unificateur de XML s'accroît de son association avec Unicode, un système de codage de caractères qui autorise le mélange de textes dans les principales écritures de la planète. En

Bases de données XML

Peut-on considérer un répertoire de quelques milliers ou millions de documents XML comme une base de données et y retrouver des informations avec la même efficacité qu'avec des systèmes classiques de gestion de bases de données? Imaginons que, demain, des milliers de musées dans le monde publient leurs catalogues de collections et d'expositions en XML et placent ces documents sur des serveurs Web. Si l'on demande à un moteur de recherche : «Existe-t-il un portrait de Blaise Pascal, et si oui qui en est l'auteur?», obtiendrons-nous la réponse «Philippe de Champaigne» sans qu'elle soit noyée sous un déluge d'informations non pertinentes ayant trait à tous les Pascal de l'Univers, ou, au moins, à toutes les œuvres, à tous les commentaires ou biographies écrits à propos de Blaise Pascal? Avec son système de balisage, XML peut grandement faciliter la recherche, même si, à elle seule, cette norme ne résout pas toutes les difficultés de la recherche de l'information en ligne.

Pour améliorer la recherche d'information sur le Web, plusieurs approches non exclusives sont envisageables. L'une d'elles, mentionnée par J. Bosak et T. Bray, est celle des «métadonnées», la transposition au monde numérique des pratiques des bibliothécaires et des documentalistes : à chaque document, on associe une «fiche signalétique». Ces fiches peuvent être rangées dans un catalogue (une base de données) d'où elles peuvent être rapidement retrouvées. Cette approche, appliquée au Web, ne convainc pas, loin s'en faut : nous ne sommes pas tous des documentalistes et, s'il fallait systématiquement créer et gérer des fiches descriptives précises pour chaque document numérique mis sur le réseau, le coût et la complexité seraient rédhibitoires. Or, si ces fiches sont trop simplifiées, elles ne permettent plus cette précision dans la recherche d'information qui est l'objectif visé.

Une variante consiste à indexer les documents par les termes figurant dans certains éléments bien choisis des documents eux-mêmes : les métadonnées descriptives doivent être trouvées dans les documents. Si toutes les informations pertinentes figurent explicitement dans les documents et y sont identifiées par des balises, il n'est nul besoin de créer une fiche descriptive avec un nom d'auteur, un titre, une date de publication, des mots clefs, etc.

Le problème devient alors de partager des systèmes de balisage communs pour ces éléments devant servir à l'indexation des documents, tout en sachant que ces balises doivent aussi être utilisées directement dans les documents les plus divers, aux structures les plus variées et écrits dans des langues différentes. Il reste aussi à inventer des méthodes de recherche qui exploitent ces index créés à partir de ces fragments de documents et qui restent efficaces lorsque le nombre de documents indexés devient très grand.

Une autre approche consiste à structurer différemment l'hypertexte que forme le Web. Aujourd'hui les liens relient entre eux des fichiers, et, pour un ordinateur, un lien du Web signifie seulement une adresse physique de fichier sur le réseau. Si vous voyez dans une page Web la phrase «Vous pouvez contacter l'auteur…», avec le mot auteur mis en valeur par une couleur et un soulignement, vous savez qu'il s'agit d'un lien qui matérialise la relation qui existe entre le document que vous lisez et son auteur, représenté par son adresse de courrier électronique ou sa page d'accueil. Pourtant vous ne pouvez pas demander à un serveur Web de vous retrouver l'adresse électronique des auteurs des pages hébergées : l'information existe, mais sous une forme inexploitable pour un programme. Or, XML permet de typer les liens, c'est-à-dire de spécifier leur sémantique, soit directement, soit en les décrivant comme des relations entre documents dans le système descriptif du langage.

Si les auteurs utilisent des liens qui correspondent à des relations prédéfinies (être auteur de, être une biographie de, être localisé à, etc.), le Web se transforme en une sorte de «graphe conceptuel», et il permet au moins la recherche des groupes de documents par l'exploitation de ces relations explicites définies entre eux.

Ces approches et plusieurs autres sont activement explorées par les spécialistes des bases de données et de la recherche d'information en ligne. Il est d'ailleurs probable qu'aucune méthode ne satisfera à elle seule tous les besoins et que l'on utilisera une panoplie de méthodes complémentaires. Une chose est sûre en tout cas : l'arrivée de XML a relancé et orienté la recherche dans ce domaine et sera la source de grands progrès en matière de recherche d'information en ligne.

Alain MICHARD, INRIA, Rocquencourt

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

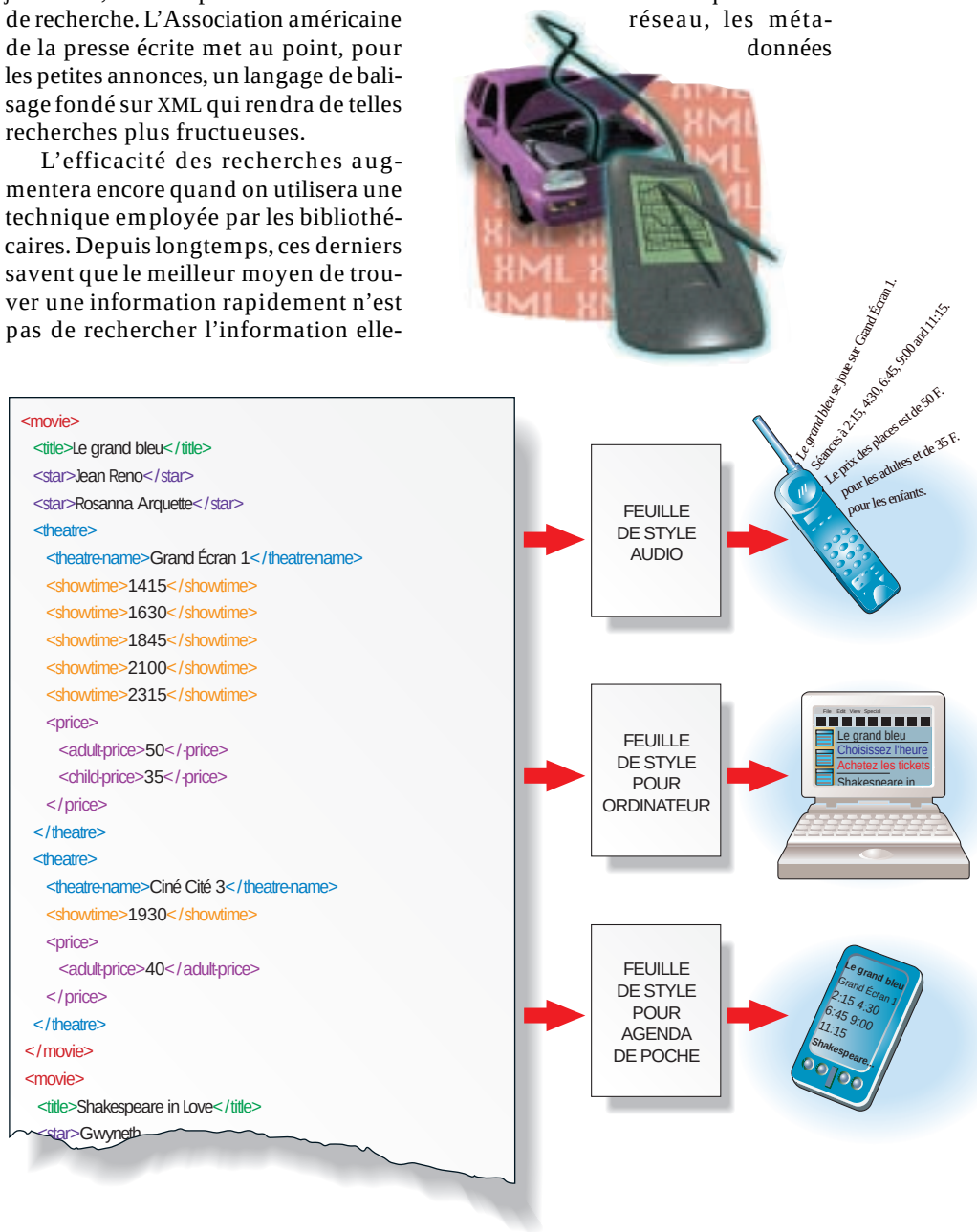
L'avènement de XML devrait accélérer les échanges sur le réseau Internet. Aujourd'hui, les ordinateurs connectés au réseau, quelle que soit leur puissance, récupèrent des formulaires, les remplissent et leur font faire la navette jusqu'à un serveur, jusqu'à l'aboutissement de la requête. L'information 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.

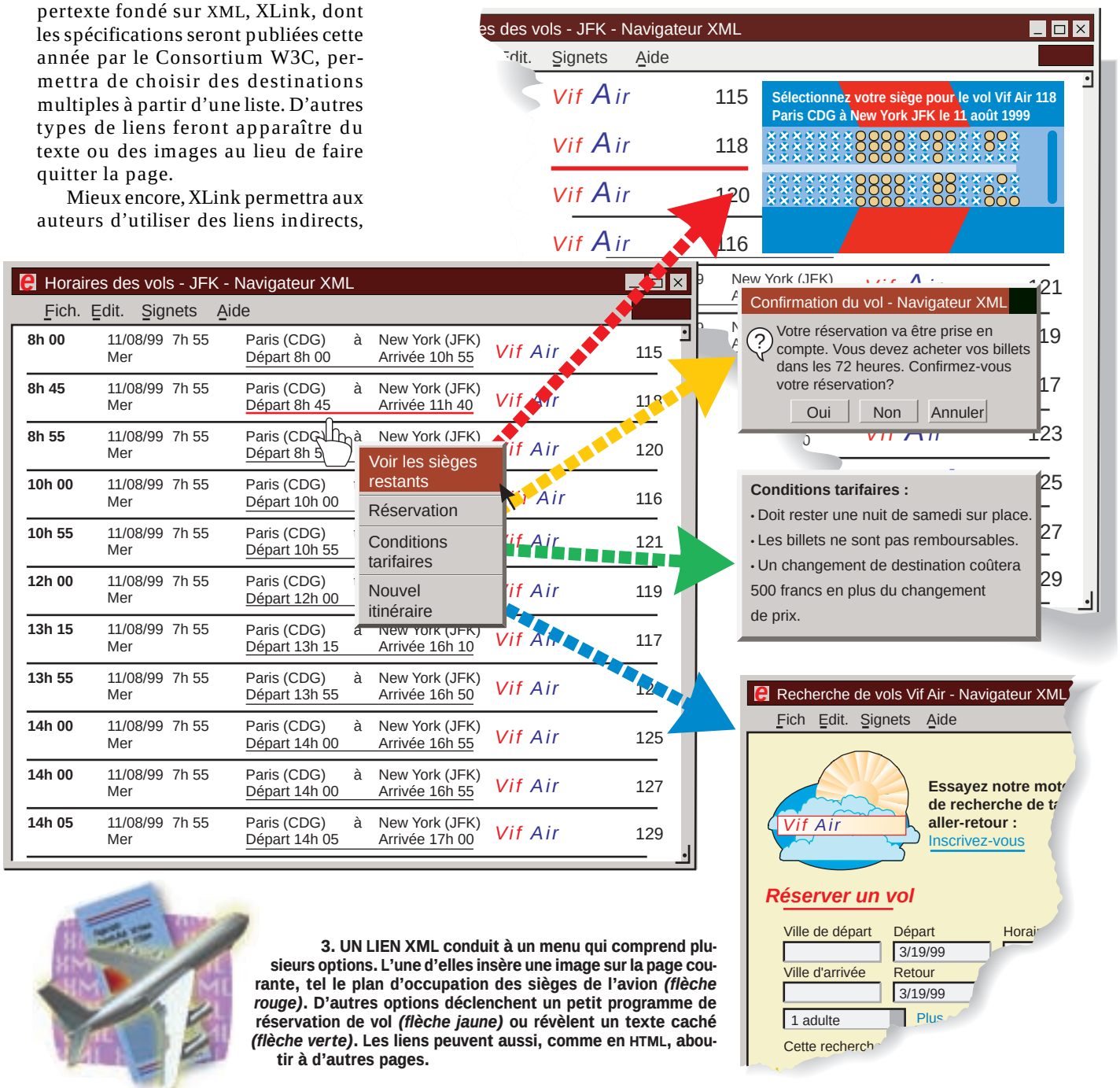
D'autres méthodes que les moteurs de recherche permettent de retrouver de l'information. Le réseau est avant tout un hypertexte dont les millions de pages sont reliées par des liens, ces mots soulignés sur lesquels on clique pour passer de page en page. XML rendra plus efficace l'utilisation des liens : 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.



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)*.

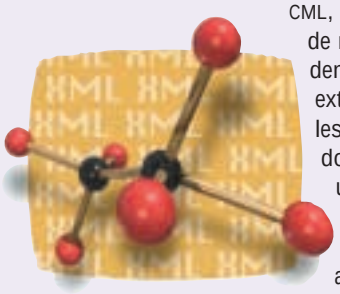
Les concepteurs d'un nouveau langage doivent s'accorder sur trois points : quelles balises seront acceptées, comment les éléments balisés pourront être emboîtés et comment ils seront traités. Les deux premières spécifications, le vocabulaire du langage et sa structure, sont codifiées dans une définition 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.

Les nouveaux langages des sciences

Grâce 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