

(PICS-1.1 "http://www.w3.org/PICS/vocab.html"

labels

by "paul@GoodMouseClicking.com"

for "http://www.w3.org/PICS"

generic

exp "1997.04.04T08:15-0500"

ratings (q 2 v 3)

Auteur de l'étiquette

URL de la page
étiquetéeCe terme signifie
que l'étiquette s'applique
à l'ensemble des fichiers
accessibles à l'adresse
<http://www.w3.org/PICS>Le document à cette
adresse, où URL, con-
tient la définition du
vocabulaire d'étique-
tage : par exemple, "q"
signifiera qualité litté-
raire et "v" violence.Cette étiquette
expire le 4 avril 1997.L'auteur de l'étiquette
a attribué une qualité
littéraire de 2 et un degré
de violence de 3 au site.**2. LA LECTURE d'une éti-
quette PICS est effectuée auto-
matiquement par les logiciels
de traitement d'étiquettes.
L'étiquette présentée ici note
la qualité littéraire et le degré
de violence du site Web
<http://www.w3.org/PICS>.**

qu'il n'approuve pas les critères qu'elles retiennent. Ainsi, des systèmes tels que *GroupLens* et *Firefly* recommandent des choix de livres, d'articles, de vidéos ou d'œuvres musicales fondés sur les appréciations de personnes partageant des goûts semblables : après que les utilisateurs ont évalué des documents traitant de sujets qui leur sont familiers, le logiciel compare leurs appréciations à celles d'autres utilisateurs et il privilégie ultérieurement les documents approuvés par les personnes qui partagent les appréciations que les utilisateurs portent sur d'autres documents. Les utilisateurs qui ont les mêmes goûts ou opinions n'ont pas à se connaître ; chacun peut participer anonymement à l'étiquetage, préservant ainsi la confidentialité de ses modes d'évaluation et de leur résultat.

L'emploi des étiquettes pose plusieurs problèmes politiques et sociaux. Tout d'abord, qui décide des procédures d'étiquetage et des étiquettes que l'on doit retenir ? En théorie, n'importe qui pourrait étiqueter un site et définir ses propres critères de filtrage. En pratique, l'attribution des étiquettes ou le choix des critères pourraient être imposés par les autorités. Cela s'est déjà produit pour la télévision : sous la pression du Conseil supérieur de l'audiovisuel, les chaînes ont entrepris d'évaluer l'adéquation de leurs émissions à l'âge des téléspectateurs.

L'auto-étiquetage obligatoire ne débouche pas sur la censure, tant que les individus peuvent choisir eux-mêmes quelles étiquettes ignorer. Malheureusement, les gens n'ont pas toujours ce pouvoir. Si un contrôle individuel amélioré est un argument contre

un contrôle central, il n'empêche pas sa mise en place. À Singapour et en Chine, par exemple, on expérimente des «garde-fous», matérialisés par des programmes et par des cartes électroniques qui interdisent l'accès à certains forums de discussions et sites du réseau.

Toutefois, même en l'absence d'une censure centrale, la mise en place d'un barème consensuel, quel qu'il soit, encouragera les utilisateurs à faire des choix laxistes qui ne refléteront pas les valeurs auxquelles ils croient. Aujourd'hui, nombre de parents qui n'approuvent pas obligatoirement les critères de classement des films, interdisent néanmoins à leurs enfants de voir des films «interdits aux moins de 12 ans» ou «interdits au moins de 16 ans» : ils ne prennent pas le temps d'évaluer eux-mêmes ces films.

Les organismes d'étiquetage doivent choisir leur vocabulaire avec discernement, de manière à correspondre aux critères de la plupart. Même ainsi, un vocabulaire unique ne peut pas répondre aux besoins de tous. Les étiquettes mentionnant uniquement le degré de pornographie des sites n'aident pas les utilisateurs préoccupés par le degré de violence verbale. De plus, aucun barème ne peut remplacer une évaluation précise : les comptes rendus de films publiés par les journaux sont bien plus éclairants que les codes parfois utilisés pour classer les œuvres cinématographiques.

De surcroît, les étiquetages, quels que soient les soins apportés à leur conception et à leur application, risquent d'étouffer les informations non commerciales. Un étiquetage bien fait nécessite du temps et de l'énergie : nombre

de sites d'intérêt limité ne seront probablement jamais étiquetés. Par souci de sécurité, certaines personnes bloqueront l'accès aux documents dépourvus d'étiquette ou dont les étiquettes ne leur semblent pas fiables. Pour elles, Internet fonctionnera davantage comme la télévision, offrant uniquement accès aux documents suffisamment porteurs d'audience pour qu'il soit financièrement rentable de les étiqueter.

Ce problème inévitable n'est pas dû à l'étiquetage : quel que soit le média utilisé, les gens tendent à éviter l'inconnu lorsqu'il y a des risques, et il est bien plus facile d'obtenir des informations sur des documents destinés à une grande audience que sur des sujets confidentiels.

Si le réseau Internet supprime les barrières techniques de communication, il n'élimine pas pour autant les coûts sociaux de cette communication. Les étiquettes peuvent réduire ces coûts, en aidant les utilisateurs à décider de la confiance qu'ils accordent aux logiciels ou aux sites du réseau. Aux utilisateurs de faire que les systèmes d'étiquetage guident leur exploration sans les empêcher de voyager.

Paul RESNICK, président du Consortium World Wide Web, est informaticien au laboratoire de la Société AT&T.

Jonathan WEINBERG, *Rating the Net*, in *Hastings Communications and Entertainment Law Journal*, vol. 19, mars 1997. Disponible sur le Web, à l'adresse : <http://www.msen.com/~weinberg/rating.htm>

Les spécifications du PICS sont consultables sur <http://www.w3.org/PICS>

Le multilinguisme

BRUNO OUDET

La multitude de langues parlées à travers le monde résistera-t-elle à l'impérialisme de l'anglais sur l'Internet?

Ces dernières années, la culture américaine a accru son influence planétaire grâce au commerce international et aux productions cinématographiques. Alors que l'Internet atteint des endroits de plus en plus reculés du Globe, une question se pose : renforcera-t-il cette influence au point que l'anglais deviendra la langue dominante dans les échanges ? Ou l'univers de la communication s'enrichira-t-il d'une multitude de langues ? Certains observateurs pensent que les langues locales ne survivront pas sur l'Internet et que l'anglais dominera.

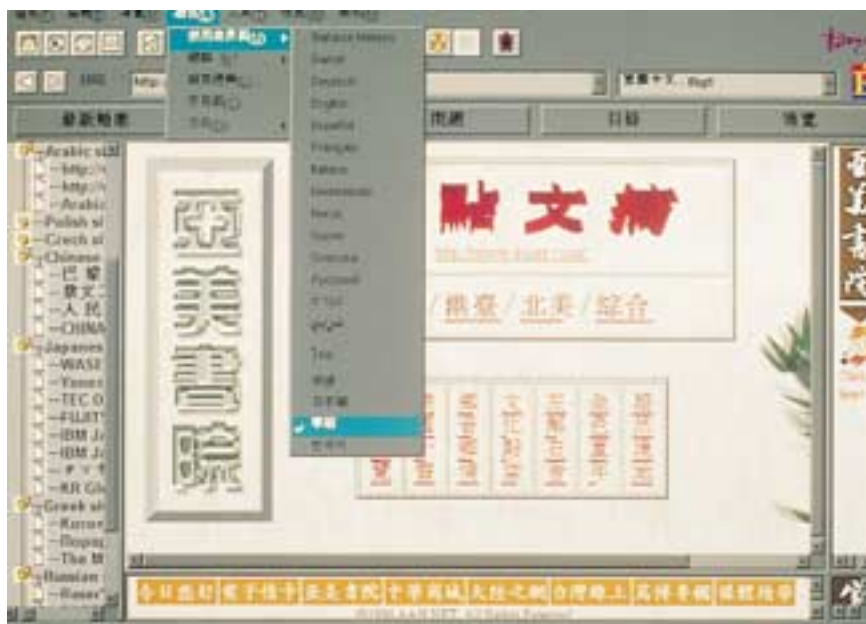
Une telle domination aurait des inconvénients. Comme la plupart des utilisateurs de l'Internet ne sont pas de langue anglaise, leur maîtrise de l'anglais est souvent limitée. Pour les discussions complexes, le recours à la langue maternelle s'impose ; mais si l'Internet ne permet pas des conversations multilingues, il ne sera pas le catalyseur attendu des communications internationales. Les erreurs et les malentendus deviendront fréquents, et nombre d'utilisateurs seront privés des fabuleuses potentialités des communications internationales.

Plusieurs forces détermineront la diversité linguistique du réseau. Aujourd'hui, 60 pour cent des ordinateurs reliés à l'Internet sont situés aux États-Unis. Hors d'Amérique du Nord et de quelques pays européens, la plupart des branchements sur l'Internet sont récents et encore limités. Cependant, le nombre de connexions augmente partout. Avec la chute constante du coût d'installation des réseaux de communication, la répartition des utilisateurs approchera bientôt celle des ordinateurs dans le monde.

Grâce à son faible coût et à sa facilité d'emploi, l'Internet permet à certains écrivains – particulièrement ceux qui utilisent l'alphabet latin – de publier leurs œuvres ou d'échanger des messages dans leur propre langue. Certains défenseurs des langues vernaculaires utilisent déjà ce réseau à leur avantage. Par exemple, bien que les Canadiens francophones ne représentent que cinq pour cent de la population francophone, 30 pour cent de toutes les pages publiées en français sur le réseau proviennent du Québec. L'extension mondiale de l'Internet favorise également une langue qui peut être, du moins superficiellement, comprise par le plus grand nombre. Ainsi, je crois que l'Internet véhiculera de nombreuses langues pour les communications locales, tandis que

l'anglais servira aux conversations internationales simples.

Les difficultés techniques de la communication dans la plupart des langues de la planète ne sont pas résolues. Les premières machines et les premiers logiciels furent conçus pour traiter des textes en anglais. Les difficultés subsistent même avec les caractères latins standards. Au début d'Arpanet – le réseau qui précéda Internet –, on ne pouvait transmettre que des messages électroniques codés en ASCII (l'acronyme de cette norme américaine de codage des caractères) à sept bits : dans ce code, chaque caractère est défini par une chaîne de sept chiffres binaires, ce qui permet le codage de 2^7 , soit 128 caractères. Aujourd'hui le nouveau protocole de transport de courrier (ESMTP) permet de traiter les huit bits nécessaires au codage en ISO-latin, l'une des normes proposées par l'Organisation internationale de la standardisation (ISO) : ce standard autorisant 256 caractères, on peut transmettre et afficher les signes diacritiques (par exemple, les accents sur les lettres) de toutes les langues d'Europe occidentale. Toutefois, comme de nombreux ordinateurs reliés par l'Internet ont des logiciels obsolètes, le huitième bit disparaît parfois, ce qui rend le message quasi incompréhensible. En 1995, sur les 12 000 utilisateurs qui recevaient les



1. SUR CE SITE, qui utilise Unicode, une norme de codage des caractères de toutes les langues du monde, l'utilisateur peut choisir la langue.



2. LA SOCIÉTÉ ACCENTSOFT propose des outils multilingues, tel le navigateur Web Mosaic Multingue.

nouvelles quotidiennes en français par l'Internet, 8 500 avaient réclamé une version codée en ASCII sept bits, plutôt que la version estropiée en ISO-latin.

Bien que certains logiciels récents puissent afficher leurs textes dans de nombreux alphabets différents, la plupart sont bilingues et traitent seulement une langue locale, par exemple le japonais, plus l'anglais. Les obstacles techniques à l'affichage d'autres alphabets sont-ils temporaires ? Unicode (ISO 10646), un codage de la plupart des caractères du monde, permet de coder presque n'importe quelle langue (mais sans résoudre la question de son affichage). Cependant un long chemin reste à parcourir avant que l'Internet ne devienne réellement multilingue et que l'on puisse, par exemple, insérer une citation en grec à l'intérieur d'un texte russe qui sera correctement affiché sur l'écran d'un utilisateur sud-américain. Des logiciels dotés de ces fonctionnalités apparaissent, mais, dans leur concurrence, les grands constructeurs de logiciels produisent sans cesse de nouvelles versions, ce qui ne laisse guère de chances de survie aux petites entreprises qui conçoivent des produits multilingues.

Puisque les interprètes aident à surmonter les barrières du langage, ne pourrait-on pas les utiliser pour traduire les documents échangés sur l'Internet ? Le volume et la diversité des échanges limite leur rôle. Seuls les programmes de traduction assistée par ordinateur nous rapprocheront d'un monde – peut-être utopique – où chaque participant d'une conférence virtuelle des Nations unies utiliserait sa langue maternelle, qui serait

alors traduite simultanément dans toutes les autres langues.

Malgré 50 ans de recherches, la traduction assistée par ordinateur reste insatisfaisante. Aujourd'hui, les systèmes de traduction ne sont utilisés qu'au Japon, au Canada et en Europe. Généralement, les interprètes électroniques sont uniquement bilingues, et ils doivent être extrêmement spécialisés pour livrer des traductions suffisamment bonnes pour être affinées par un être humain.

Le premier système public fut *Sys-tran*, qui traduisait 14 paires de langues dès 1983, sur le réseau Minitel. Utilisé par la Commission européenne, *Sys-tran* traduit aujourd'hui des centaines de milliers de pages par an. Le système *Meteo*, une autre grande réussite, traduit les bulletins météorologiques canadiens de l'anglais en français et inversement. Il traite quotidiennement 80 000 mots (environ 400 bulletins), avec seulement trois à cinq interventions humaines pour 100 mots.

Plusieurs groupes de recherche élaborent aujourd'hui une procédure à deux étapes dont bénéficiera la traduction multilingue. Dans cette procédure, on analyse d'abord entièrement le texte : on le décompose en ses éléments constitutifs (titre, paragraphe, phrase), et les ambiguïtés potentielles sont levées par un dialogue avec l'auteur. Le texte est alors traduit dans une forme intermédiaire abstraite, qui sert de base pour les traductions dans différentes langues. La dépense occasionnée est justifiée lorsque le texte doit être traduit en plus de dix langues. L'Université des Nations Unies, à Tokyo, a récemment proposé un projet qui devrait, en l'espace de dix ans, concrétiser ce système.

Toutefois, le réseau Internet ne deviendra véritablement multilingue que s'il existe une concertation internationale. Lui donnerons-nous une priorité suffisante ? La réponse à cette question est encore floue. Il nous est si facile de nous laisser dériver vers l'anglais en tant qu'unique langue commune.

Bruno OUDET est professeur à l'Université Joseph-Fourier de Grenoble et chercheur au Laboratoire Leibniz de l'IMAG. Il est actuellement président de la Société française de l'Internet.

Les interfaces de recherche

MARTI HEARST

Le rapide développement du réseau mondial gêne sa rationalisation et son organisation. De nouvelles interfaces faciliteraient son emploi.

Comment éviter de se perdre dans l'immensité du réseau mondial? Nombre de sites permettent de retrouver facilement certaines informations courantes et sommaires, tels des numéros de téléphone ou les cours de la Bourse, mais le réseau Internet est remarquable parce qu'il ne nous cantonne pas à un quotidien prosaïque et qu'il fait venir sur nos bureaux des informations extraordinairement variées. Pourtant, par manque d'organisation cohérente, le cyberspace devient de plus en plus confus. Les outils qui servent aujourd'hui à localiser un document dans l'Oregon, un catalogue en Grande-Bretagne ou une image au Japon sont souvent d'une lenteur désespérante.

Quelques algorithmes classent les résultats des recherches d'après leur pertinence, mais le remède à la lenteur du réseau résidera plus probablement dans l'emploi de nouveaux programmes nommés interfaces utilisateur. Aujourd'hui des logiciels d'analyse de textes et de manipulation de données hiérarchisées facilitent la recherche sur le réseau Internet ou dans de grandes bases de données, car la plupart des sites informatiques sont organisés en «pages», une structure simple et familière, mais inutilement restrictive. Dans l'avenir, cette présentation sera remplacée par des structures qui montreront les informations de plusieurs points de vue simultanés. Imaginons qu'Alice, dans la Creuse, se connecte au réseau afin de savoir quels types de bulbes comestibles, tels l'ail ou l'oignon, elle peut planter dans son jardin. Les réponses à sa question se trouvent certainement sur le réseau, mais où?

Alice dispose aujourd'hui de plusieurs techniques pour s'y retrouver, mais aucune n'est parfaitement satisfaisante. Elle peut d'abord demander à des amis s'ils connaissent des sites inté-

ressants. Elle peut aussi feuilleter des index du réseau, qui sont aujourd'hui de deux types : soit des listes construites manuellement, qui classent les sites par type de contenu, soit des moteurs de recherche qui repèrent rapidement des mots clés dans les index du réseau.

Grâce à leurs employés qui répertorient quotidiennement des centaines de documents, des sites, tels *Yahoo* et *Nomade*, proposent aux utilisateurs une table des matières. Pour y rechercher l'information, on choisit dans un menu (voir la figure 1) la catégorie qui semble la plus proche de celle que l'on cherche et l'on reçoit un sous-menu plus spécialisé ou une liste de sites dont le personnel de la société pense qu'ils appartiennent à cette catégorie. Tou-

tefois l'emploi de ces tables des matières n'est pas très simple, car les catégories ne sont pas toujours mutuellement exclusives : Alice doit-elle choisir la rubrique «Loisirs», ou bien la rubrique «Écologie», par exemple? Quel que soit son choix, le menu disparaît de l'écran, de sorte qu'en cas d'échec elle doit refaire toute la recherche, et répondre à toutes les questions posées par l'interface de recherche. Si l'un de ses choix n'aboutit pas (si la rubrique «Écologie» ne contient pas l'information cherchée, par exemple), elle doit repartir à zéro et en essayer une autre. Si l'information désirée est enfouie dans les profondeurs de la hiérarchie ou n'existe pas, la procédure peut être fastidieuse, voire exaspérante.



1. LE SYSTÈME de mise en page utilisé sur la plupart des sites du réseau impose des contraintes superflues aux services, tels *Yahoo* ou *Nomade*, qui tentent de structurer le contenu du réseau, afin de faciliter la recherche d'informations. Un utilisateur qui désire connaître les type de bulbes comestibles (ail, oignon, etc.) qu'il peut planter en automne, par exemple, peut choisir une catégorie. Plus il précise sa demande, plus il lui est difficile de se rappeler quelles catégories potentiellement utiles il n'a pas encore explorées.