

Cartographie thématique

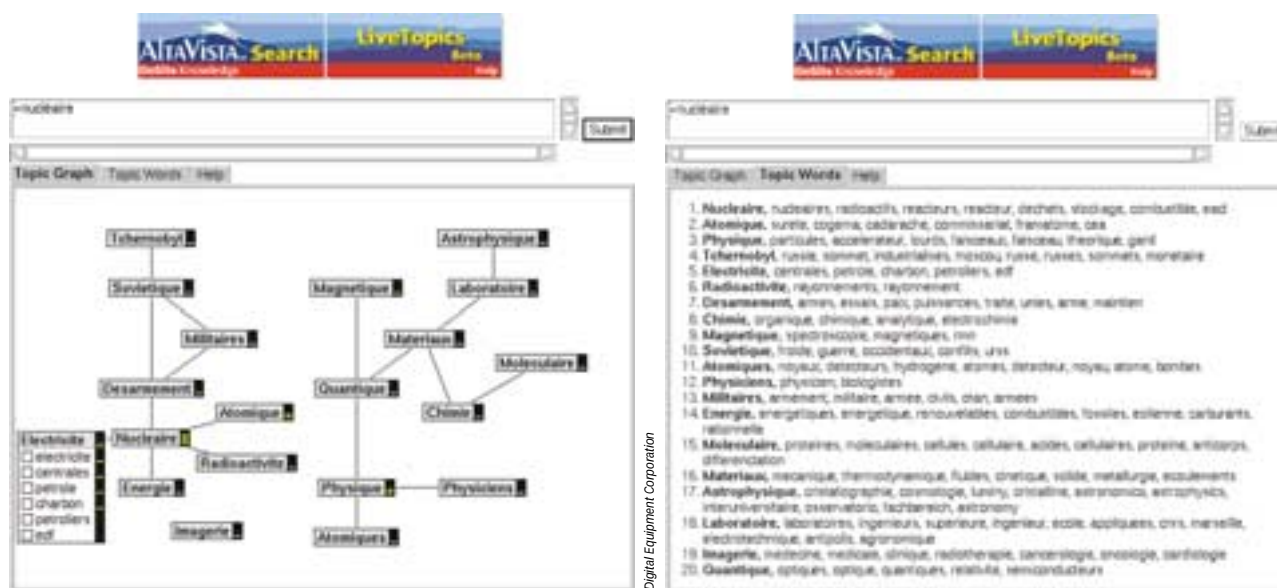
Dans une approche différente de celle des programmes qui aident à visualiser la classification *a priori* des thèmes d'une recherche, tels ceux décrits dans cet article, nous avons conçu un logiciel, nommé *LiveTopics*, qui propose à l'utilisateur, submergé par l'avalanche de documents renvoyés par sa requête, une cartographie thématique et dynamique des documents. À l'aide de cette carte, l'utilisateur appréhende rapidement le contenu des documents. En incluant ou en excluant certains des thèmes de la carte, il affine sa recherche.

Ce logiciel est aujourd'hui proposé sur le moteur de recherche *AltaVista* de la Société *Digital Equipment*. Lorsque l'utilisateur reçoit plus d'une vingtaine de réponses, il peut demander à visualiser une synthèse thématique de l'ensemble des réponses. Par exemple, la recherche du mot «nucléaire» renvoie des thèmes variés où l'on

retrouve ce qui concerne la physique, l'énergie, l'électricité et l'astrophysique (voir la figure). L'utilisateur passe ensuite en revue chacun des thèmes et les mots qui s'y rapportent. Il peut alors modifier sa requête d'un simple clic de la souris, en incluant ou en excluant certains mots, et lancer une nouvelle recherche qui pourra de nouveau donner lieu à une analyse thématique.

Fondé sur une analyse statistique du texte intégral des documents, *LiveTopics* n'utilise aucune analyse linguistique ni sémantique, ce qui rend possible l'analyse de documents indépendamment de leur langue, pourvu que cette langue s'écrive avec des caractères latins. Une évolution vers des caractères d'autres langues est à l'étude, ainsi qu'une version de la technique adaptée au monde des réseaux d'entreprises.

François BOURDONCLE, École des mines de Paris



Dans le moteur de recherche *AltaVista* (<http://www.altavista.digital.com/>), la demande «nucléaire» renvoie 11 877 documents. Lorsque on lance *LiveTopics*, une carte des thèmes apparaît. À gauche, on trouve le thème *Nucléaire* relié à *Énergie*, *Électricité*, *Radioactivité*, mais aussi le thème du désarmement, lié à *Soviétique*, *Tchernobyl* et

Militaires. À droite de la carte apparaissent plutôt les thèmes liés à la recherche : *Quantique*, *Magnétique*, *Physique*, *Chimie*, etc. Notez le thème *Imagerie* qui est isolé du reste. Sur l'écran de droite, on trouve la liste des mots associés aux thèmes de la demande «nucléaire». On peut les exclure ou les inclure pour une nouvelle recherche.

Le groupement de documents n'est pas une panacée, car il n'explique pas comment les documents sont trouvés. En revanche, quand le formulaire d'entrée encourage les utilisateurs à décomposer leurs demandes en plusieurs groupes de mots clés apparentés, une interface graphique peut indiquer les voies de recherche obtenues et les endroits où elles sont apparues dans les documents trouvés. Si tous les sujets se retrouvent au même endroit, le document a plus de chances d'être pertinent, et le programme lui accorde une priorité supérieure. Alice peut n'avoir aucune idée précise des sujets qui doivent apparaître dans le document ou du degré de proximité qui doit les

unir, mais elle a des chances de reconnaître ce qu'elle désire lorsqu'elle le voit et peut alors affiner sa demande. Cette technique, que j'ai nommée *TileBars*, peut guider les utilisateurs lorsqu'ils choisissent les documents qu'ils souhaitent examiner et les conduit plus rapidement aux endroits les plus pertinents.

L'étude de ces nouvelles interfaces utilisateur et techniques d'analyse de texte commence seulement. D'autres techniques combinent méthodes statistiques et empirisme pour résumer automatiquement des documents et les insérer dans un ensemble de catégories déjà existant. Elles proposent des synonymes pour des mots clés et répondent à des questions simples.

Marti HEARST est informaticien au Centre de recherche *Xerox*, à Palo Alto.

Ramana RAO, Jan O. PEDERSEN, Marti A. HEARST et Jock D. MACKINLAY et al., *Rich Interaction in the Digital Library*, in *Communications of the ACM*, vol. 38, n° 4, pp. 29-39, avril 1995.

Stuart K. CARD, George G. ROBERTSON et William YORK, *The WebBook and the Web Forager: An Information Workspace for the World-Wide Web*, in *Proceedings of the ACM/SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, avril 1996. Présent sur le réseau à l'adresse : <http://www.acm.org/sigs/sigchi/ch96/proceedings/papers/Card/skc1txt.html>

Les systèmes sécurisés

MARK STEFIK

La mise en place de dispositifs de protection numérique des droits d'auteur permettra aux auteurs de publier sereinement leurs œuvres sur l'Internet.

La banalisation de l'informatique a rapidement conduit les utilisateurs d'ordinateurs à penser que toutes les œuvres numériques – les programmes informatiques, les livres, les revues, la musique et la vidéo – pouvaient être copiées. Certaines figures de l'informatique ont même proclamé que la simplicité de la reproduction marquait la fin des droits d'auteur : « l'information veut être libre », disaient-ils. Autrement dit, il serait impossible de s'opposer à la diffusion de l'information, et tout ce qui peut s'exprimer sous forme de bits pourrait être impunément reproduit.

Cette idée est à la base de la création du réseau Internet : dans la nouvelle ère numérique où nous entrons, l'accès à l'information numérisée serait universel, et l'œuvre de n'importe quel auteur devrait être accessible à n'importe qui, n'importe où et n'importe quand.

Toutefois, par expérience, les utilisateurs du réseau savent que ce dernier ne contient pas de grandes œuvres ni d'informations capitales ; ce qu'on y trouve est souvent de faible valeur. La raison en est que les auteurs ou les éditeurs ne peuvent vivre en faisant cadeau de leur travail. Il suffit aujourd'hui de taper quelques touches d'un clavier pour copier un paragraphe, une revue, un livre, voire l'œuvre d'une vie. La duplication incontrôlée a tant modifié l'équilibre du contrat social entre les créateurs et les consommateurs d'œuvres numériques que la plupart des éditeurs et des auteurs refusent de laisser circuler le meilleur de leurs travaux sous forme numérique.

Toutefois, en coulisse, la technique se prépare à modifier cet équilibre. Au cours des toutes dernières années, plusieurs sociétés, telles *Folio*, *IBM*, *Intertrust*, *NetRights*, *Xerox* ou *Wave Systems*, ont mis au point des programmes et des



Jeff Brice

machines qui permettent aux éditeurs de spécifier les modalités d'utilisation des œuvres numériques et de contrôler leur utilisation. Certains législateurs pensent que ce changement est si radical que les éditeurs auront un pouvoir excessif, et que les consommateurs et les bibliothécaires seront lésés.

Pourtant les besoins des consommateurs seront satisfaits, même avec les nouveaux systèmes : la technique introduisant davantage de sécurité, des œuvres de meilleure qualité pourront être proposées sur le réseau. Bien protégées, des auteurs célèbres auront peut-être envie de publier directement sur le réseau. L'accès à l'information ne sera peut-être pas gratuit, mais il pourrait être bien moins onéreux qu'aujourd'hui, car les coûts d'impression et de distribution sont considérablement réduits par la mise des informations sur le réseau ; les éditeurs pourraient répercuter ces économies aux consommateurs.

Les systèmes « sécurisés » sont la clé de cette mutation technique : ce sont des machines et des programmes dont le fonctionnement se conforme à des règles, nommées droits d'usage, qui définissent le coût d'accès aux œuvres numérisées et les modalités de leur utilisation. Un ordinateur sécurisé, par exemple, refuserait de faire des copies non autorisées ou de transmettre des documents sonores ou graphiques à un utilisateur qui n'aurait pas payé pour les acquérir.

Les types de systèmes sécurisés sont variés : des lecteurs sécurisés peuvent lire des livres numérisés, des magnétophones et des magnétoscopes sécurisés peuvent lire des enregistrements audio et vidéo, des imprimantes sécurisées peuvent faire des copies contenant des signes (des « filigranes ») qui indiquent le droit d'utilisation et de reproduction accordé, et des serveurs sécurisés peuvent se charger de vendre des œuvres numériques sur l'Internet. Bien que complexes, les techniques d'élaboration des systèmes sécurisés ont une fonction simple : elles permettent aux éditeurs de distribuer leurs produits – sous forme cryptée –, de sorte que ceux-ci ne soient affichés ou imprimés que par des machines sécurisées. Les premiers systèmes sécurisés équiperont des imprimantes ou des lecteurs numériques de poche – avec un surcoût pour l'utilisateur, qui accéderait alors à des documents bien plus intéressants qu'aujourd'hui. Puis, comme toujours, l'avènement des nouvelles techniques en ferait chuter les prix. Naturellement les éditeurs pourront encore donner gratuitement l'accès à certaines œuvres : les serveurs sécurisés les transmettront alors sans imposer les protocoles spécifiques des œuvres protégées.

Comment les systèmes sécurisés savent-ils les règles qui doivent être respectées ? La Société *Xerox* et d'autres ont tenté de décrire les montants des droits et les modalités précises d'utilisation dans un langage formel qui sera interprété par les systèmes sécurisés. Le recours à un tel langage est essentiel pour l'édition électronique : les vendeurs et les acheteurs ne pourront négocier et parvenir à des accords que si l'ensemble des actions autorisées ou interdites est explicitement défini. Les droits numériques sont de plusieurs types. Les droits de transport concer-