

Économie et portabilité : une chaîne d'édition électronique destinée à la dissémination de l'information primaire

Pier Luigi ROSSI
ORSTOM
Centre de recherche d'Ile-de-France
32, avenue Henri Varagnat
93143 Bondy cedex
courrier électronique : rossi@orstom.fr

La fonction idéale d'une bibliothèque est donc un peu semblable à celle du bouquiniste chez qui on fait des trouvailles et seul le libre accès aux rayons le permet.

*Umberto ECO
De Bibliotheca*

● Résumé

Qu'il s'agisse de documents édités (livres, périodiques, ...) ou de littérature grise (thèses, rapports, ...) les documents se concentrent et se conservent dans les bibliothèques ou les centres de documentation (nationaux, de sociétés, d'institutions, d'unités de recherche, ...).

Les documents sont décrits et répertoriés par des catalogues et/ou des bases de données bibliographiques informatisées (**information secondaire**) : l'accès à l'information se fait **par leur intermédiaire**.

L'utilisateur qui a pris connaissance de l'existence du **document primaire** et veut le consulter doit se rendre sur les lieux de conservation ou doit obtenir une copie (un prêt) à partir de ces lieux ou, éventuellement, en s'adressant à l'auteur.

Le passage d'un support papier à un support électronique pour les documents permet d'introduire une dématérialisation des lieux de conservation (**bibliothèque virtuelle**) ainsi que des pratiques sensiblement différentes en ce qui concerne l'accès et la diffusion des documents.

Dans la mesure où le document est disponible en texte intégral, les éléments qui le décrivent et son contenu sont associés : la notion d'information secondaire tend à disparaître.

Via les autoroutes de l'information, grâce au texte intégral, les utilisateurs retrouvent les documents qui les intéressent et peuvent directement en disposer.

Le document se "granularise" : on passe de l'"**objet-document**" (livre, numéro de périodique, rapport, ...) à l'"**unité-document**" (chapitre, article, page(s)).

Dans le cadre des activités d'ingénierie de l'information de notre Institut, nous allons présenter notre expérience en ce qui concerne la réalisation de documents électroniques produits à partir de supports numériques et/ou de supports papier ainsi que la réalisation de collections de documents accessibles en texte intégral et diffusées soit sur cédérom soit via les autoroutes de l'information (bibliothèques virtuelles). Ces réalisations sont toutes basées sur l'utilisation du logiciel Adobe Acrobat.

Compte tenu de notre expérience en ce qui concerne la valorisation de l'information documentaire produite par nos partenaires de la zone tropicale, nous allons montrer les apports stratégiques de ces technologies vis à vis de l'accès, de la diffusion et de la dissémination de l'information primaire produite par les pays en développement.

● Introduction

L'informatisation des catalogues des bibliothèques et des centres de documentation a permis de donner un accès souple et performant, devenu de plus en plus un accès en ligne, à une masse d'information concernant les "objets" présents dans les fonds. Ces systèmes d'information secondaire fournissent une description fine des "objets" et permettent de conduire des recherches pertinentes par rapport aux besoins des utilisateurs.

La réalisation de documents électroniques produits à partir de supports numériques et/ou de supports papier ainsi que la réalisation de collections de documents accessibles en texte intégral et diffusées soit sur cédérom soit via les autoroutes de l'information conduisent producteurs et usagers aux bibliothèques virtuelles.

"La rapide évolution des réseaux télématiques et de l'édition électronique entraînent la production et l'accès à une masse considérable d'information en dehors des murs de la bibliothèque" [CRIM].

● Catalogues et information secondaire

Les années '80 constituent l'âge de la maturité [Lesk] en ce qui concerne l'informatisation des fonds documentaires et/ou la réalisation de bases de données bibliographiques.

Au cours de cette période, grâce à une diffusion massive des outils (machines et logiciels), la description sur ordinateur des fonds devient possible aussi bien dans les pays en développement que dans les pays développés. Elle se réalise dans des structures dont la taille, les moyens et la notoriété sont extrêmement variées.

La production de l'information secondaire se fait par une description des "objets" et de leur contenu. Ces opérations comportent des coûts de production souvent très élevés liés à la nécessité de disposer de personnels qualifiés ainsi que de réaliser une structuration et un contrôle de l'information tout en assurant sa qualité.

L'information produite est structurée en champs : les moteurs d'interrogation permettent de retrouver rapidement et de façon pertinente les notices correspondant aux documents disponibles. A partir des fichiers il est possible de produire également des bibliographies, organisées selon différents critères, pouvant être diffusées sous la forme d'ouvrages ou de numéros thématiques de périodiques. La consultation de ces ressources peut se faire en ligne, bien que cette possibilité soit souvent réduite à des bases de données bibliographiques commercialisées via des serveurs professionnels.

Une contrainte forte réside dans les moyens d'accès aux documents. Après la consultation de ces sources d'information l'utilisateur dispose de listes de documents pertinents. Pour accéder aux documents il doit se rendre (ou rester) sur le lieu possédant les documents, ou bien il doit faire appel à un prestataire de services pour obtenir une reproduction, généralement payante, des documents, ou bien il doit essayer d'obtenir, après une localisation, un prêt par un établissement possédant les documents. A ce stade, l'information primaire est dans les "murs de la bibliothèque" [CRIM].

Une solution intéressante concernant les procédures d'accès à l'information est celle fournie par les *Current Contents on diskette* (CCOD) produits et commercialisés par l'Institute for Scientific Information (ISI). Un nombre important de périodiques scientifiques sont répertoriés dans les bases des CCOD. Pour chaque article sont indiquées les coordonnées des auteurs pouvant fournir les tirés à part. L'utilisateur dispose, sous la forme d'une base de données bibliographiques, d'une "revue de presse" hebdomadaire des articles parus. Il peut imprimer des notices, alimenter un fichier bibliographique propre, éditer une demande de tiré à part qu'il doit envoyer à l'auteur. Par ce mode d'accès aux documents, dont le coût se limite aux frais postaux, l'utilisateur reçoit, dans la plupart

des cas, une réponse positive à sa requête.

● Texte intégral et information primaire

Parallèlement au développement de l'informatisation des fonds documentaires, les "éditeurs d'information" (logiciels de traitement de texte, tableurs, logiciels de publication assistée par ordinateur, ...) ont gagné une place essentielle et indispensable sur pratiquement tous les postes de travail informatiques. La plus grande partie des documents est donc conçue et réalisée sur ordinateur. De ce fait les "objets" sont devenus pour l'essentiel électroniques, bien que certaines parties comme les images, les éléments cartographiques, les illustrations puissent encore intégrer les documents selon des procédés typographiques.

Dans la mesure où les documents sont conçus sur ordinateur, plusieurs éléments permettant de les décrire ainsi que le contenu même de ces objets sont implicitement et directement disponibles : le texte intégral est devenu l'essence même des "objets". Les coûts liés à la nécessité d'une production de l'information secondaire tendent largement à disparaître.

Au cours des années '90 le protocole Internet s'impose comme le fédérateur des réseaux informatiques et la "toile" devient un des principaux vecteurs de l'information.

Les pages accessibles via le réseau créent une nouvelle forme d'"objets" fondés sur les techniques de l'hypertexte et de l'hypermédia. Les serveurs fournissent des nouveaux points d'accès aux systèmes d'information secondaire mais fournissent, également et surtout, directement les "objets" qui étaient répertoriés par les systèmes d'information secondaire. Les "**objets-document**" conçus sur ordinateur (livres, numéros de périodiques, rapports, ...) destinés à l'impression pour diffuser l'information peuvent désormais être hébergés sur les serveurs tout en gardant leur forme électronique originelle.

Compte tenu de leur structure, chapitre, article, page, paragraphe, "matérialisée" par l'outil informatique, leur mise en accès détermine leur "granularisation" lors de la consultation et de l'usage ("**unité document**").

Les moteurs de recherche fournissant un accès aux pages du réseau et/ou au contenu des fichiers disponibles sur un/des serveur(s) se multiplient. Des versions gratuites ou payantes sont disponibles.

L'utilisateur, après une navigation sur les pages d'un serveur ou après une recherche sur des index globaux ou spécifiques à un serveur, accède directement aux documents. Ces derniers paraissent sous une forme "simplifiée" (HTML/ASCII) ou sous une forme basée sur leur "apparence typographique" (postscript, pdf, ...) et sont généralement disponibles dans leur intégralité.

La mise en accès, sur les autoroutes de l'information, du support électronique des documents introduit une dématérialisation des lieux de conservation et d'accès des documents. Elle introduit et "matérialise" le concept de **bibliothèque virtuelle**.

● Documents électroniques, bibliothèques virtuelles : notre expérience

Dans le cadre des activités que nous menons pour la diffusion de l'information produite par notre Institut et de celle produite par les partenaires de la zone tropicale, nous nous intéressons aux technologies permettant la production et la diffusion des documents électroniques.

Pour la réalisation des projets de mise en accès des documents électroniques, plusieurs critères de qualité/efficacité doivent être pris en compte :

un format de diffusion universel (multiplateformes informatiques), compact et optimisé pour les

réseaux Internet/Intranet,

l'intégrité visuelle des documents : conservation de la mise en page, des graphiques, des tableaux, des images,....,

l'indexation et la recherche de l'information en texte intégral,

la possibilité d'ajouter des enrichissements hypertexte et hypermédia,

la possibilité d'ajouter des annotations dans le document en cours d'élaboration au niveau de chaque intervenant,

l'ergonomie d'utilisation du logiciel,

l'ergonomie de lecture à l'écran des documents électroniques,

des coûts d'acquisition limités,

une solution pouvant disposer d'une "base installée" très étendue.

Après une veille technologique d'environ une année (1996), réalisée sur Internet ainsi que sur des revues spécialisées, nous nous sommes orientés vers le logiciel Acrobat de la société Adobe. Cette solution répond aux différents critères indiqués ci-dessous et son prix de revient est très faible. La version Acrobat 3.0 assure également la numérisation des documents en mode multipages avec une solution de reconnaissance optique des caractères dont les résultats nous paraissent excellents. La possibilité d'associer image originelle et texte reconnu fournit une solution associant la "justesse de l'image" à la "présence du texte" (indexation en texte intégral, exportation avec "contrôle sur image"). Le prix d'une version disposant de ces fonctionnalités est actuellement d'environ 150 FF pour les organismes du "monde Education-Recherche".

Avec les solutions proposées par Acrobat nous réalisons et nous diffusons des documents électroniques selon trois directions complémentaires :

réalisation de documents multimédia pour la présentation et la valorisation de la production scientifique de l'Institut. Il s'agit de documents illustrant des programmes de recherche, de catalogues, de plaquettes présentant les produits/activités de l'Institut,

réalisation de documents électroniques à caractère scientifique : actes de colloques, ouvrages,

numérisation de collections de documents à partir du support papier avec accès à l'information en texte intégral et en champs structurés.

Les collections de documents produites sont rendues accessibles en format pdf via nos serveurs Internet/Intranet et sur supports cédérom. Les utilisateurs accèdent aux documents primaires soit via les liens hypertexte des pages de présentation soit via un moteur de recherche en texte intégral représentant le "point d'entrée" de la bibliothèque virtuelle.

Le moteur de recherche que nous utilisons est une version gratuite, récupérée sur Internet, issue des logiciels de recherche en texte intégral produits par la société Verity. Nous l'avons implanté sur des stations de travail SUN utilisant Solaris 2.5.1 comme système d'exploitation. Il permet de réaliser des recherches sur les systèmes d'index produits par Acrobat, liste les résultats en fonction du "poids" du document et repère les mots recherchés sur les pages pertinentes.

● Conclusion

Les pratiques de l'accès et de la diffusion de l'information peuvent être transposées pour les organismes et les utilisateurs des pays en développement. Alors que l'information secondaire, notamment celle produite par les pays développés, est souvent disponible sous la forme de bibliographies et de cédéroms, l'accès à l'information primaire est extrêmement difficile dans ces pays. L'utilisateur doit se rendre dans les bibliothèques, souvent dans les villes ou les capitales, qui disposent, dans la plupart des cas, de collections d'ouvrages et de périodiques de taille limitée et pour lesquelles les acquisitions sont des événements rares. D'autre part les organismes producteurs d'information rencontrent d'énormes difficultés pour la diffusion des documents. S'ils arrivent à assurer les frais d'impression/reproduction, ils arrivent rarement à surmonter les frais postaux pour l'acheminement des versions papier.

Compte tenu de notre expérience en ce qui concerne l'accès en ligne aux bases de données bibliographiques [Rossi] ainsi que en ce qui concerne la production et la diffusion des documents électroniques, il nous semble que les technologies liées à l'usage d'Acrobat permettent d'introduire des solutions intéressantes et particulièrement économiques pour l'accès et la diffusion de l'information primaire produite par les partenaires de la zone tropicale.

Ces solutions offrent une prise en main extrêmement facile et conviviale des nouvelles technologies liées aux concepts d'hypertexte, hypermédia, recherche en texte intégral, élaboration de l'information en groupes de travail, numérisation des documents, reconnaissance optique des caractères. Elles permettent également d'associer étroitement les spécialistes de l'information des pays en développement à la réflexion qui est menée dans le cadre de la structuration et de la diffusion de l'information.

En exploitant les vecteurs de diffusion de documents électroniques - les supports de masse (disquettes, cédérom, cédéworm, ...), le courrier électronique, les serveurs Internet-Intranet - les collections de documents actuellement réalisés sur ordinateur ainsi que les collections de documents anciens, numérisés à partir du support papier, permettront la réalisation d'innombrables bibliothèques virtuelles, d'innombrables point d'accès pour la dissémination de la richesse et de la mémoire de la production scientifique des pays en développement.

● Bibliographie

Anonyme. Portable document formats. Online and CD-ROM review. 1996, vol. 20, n° 1, p 33-35.

Adam Lishan. The hybrid CD-ROM : potential applications in Africa. Online and CD-ROM review. 1997, vol. 21, n° 4, p 211-216.

Centre de recherche informatique de Montréal (CRIM). La bibliothèque virtuelle... ou les nouveaux pouvoirs de l'information. http://www.crim.ca/doc/biblio_virt_doc.html

Eco Umberto. De Bibliotheca. Caen : L'Echoppe, 1986.- 31 pages, ISBN 2-905567-09-X.

Guédon Jean-Claude. L'édition savante et l'autoroute électronique.
<http://www.droit.umontreal.ca/crdp/fr/equipes/technologie/conferences/ae/guedon.html>

Gouvernement du Canada. Guide d'Internet. <http://canada.gc.ca/programs/guide/index.html>

Howkins John (ed.) Valantin Robert (ed.). Le développement à l'âge de l'information : quatre scénarios pour l'avenir des technologies de l'information et des communications.

<http://www.idrc.ca/books/836/>

Lesk Michael. The Seven Ages of Information Retrieval,
<http://community.bellcore.com/lesk/ages/ages.html>

Odlyzko Andrew. Sur la route de l'édition électronique.
<http://www-mathdoc.ujf-grenoble.fr/textes/Odlyzko/odlyzko-fr/odlyzko-fr.html>

Rossi Pier Luigi. Servers and online bibliographic databases in developing countries : the African reality - IN: Raitt, D.I. (ed.) - Online information 92. Oxford : Learned Information, 1992, p. 431-435.