

Les TICE au cœur de la documenta

Dominique Cavet

RESPONSABLE DU SECTEUR DOCUMENTATION
INSTITUT DE RECHERCHE POUR LE DÉVELOPPEMENT

Le nouveau contexte d'Internet ne change pas le rôle fondamental de la documentation : assurer une médiation entre des ressources documentaires et des usagers, en fonction de leurs besoins. L'appropriation des mécanismes de recherche du Web et la constitution de bibliothèques numériques représentent les deux moyens les plus adaptés.

Les outils de gestion documentaire des CDI ont été parmi les premiers logiciels introduits dans les établissements scolaires ; ils ont donné les premières occasions d'utiliser de manière pertinente, pour la pédagogie, des applications informatiques complexes mettant en œuvre des bases de données partagées et des fonctionnalités de réseau. Cependant, depuis qu'Internet a pris toute sa place dans la société, c'est sur un terrain beaucoup plus vaste que doivent se situer les évolutions de la documentation et des technologies qu'elle utilise.

Prise en compte d'un nouveau paysage documentaire

On ne va pas revenir ici sur la révolution qu'a représenté l'arrivée du numérique et d'Internet. Dans le domaine documentaire, les technologies de l'information et de la communication ont apporté des bouleversements multiples et la création d'un nouveau paysage documentaire jusqu'alors inconnu, avec plusieurs caractéristiques :

- **Un éclatement sans précédent des frontières.**

En passant du périmètre relativement clos d'un centre de documentation à celui du grand Internet mondial, on se situe de manière irréversible dans un univers très vaste, ouvert, aux contours non maîtrisés.

- **Une démultiplication des sources d'information.** Le Web a permis à de très nombreux acteurs, bien au-delà du monde des éditeurs tradition-

nels, de publier et de mettre en ligne informations et documents.

- **Une transformation de la notion même de document.** Les aspects positifs comme la mise en œuvre des liens hypertextuels de manière généralisée, que seul le Web pouvait réussir, ne compensent pas les effets négatifs : disparition du document au profit de la page HTML, structuration absente ou inadaptée, pérennité incertaine, manque de professionnalisme dans la mise en ligne.

- **Un impact très fort sur notre public.** Jamais les enseignants et les élèves n'ont été placés face à de tels outils d'accès à l'information et à une telle multitude de ressources utiles.

- **Des pratiques nouvelles induites par le Web.** Les moyens d'accès à cette information – navigation, zapping, recherche – sont directement induits par ces nouvelles technologies.

C'est tout l'environnement documentaire qui s'est transformé. Le rôle de la documentation est plus que jamais de se placer en intermédiaire (en médiateur) entre ce paysage documentaire qui a explosé et les utilisateurs potentiels de ces ressources, qui ont encore plus de besoins qu'auparavant.

La recherche sur Internet

Du fait de leur nouveauté, on s'est d'abord focalisé sur les technologies de recherche du Web. Elles sont tout à fait différentes de celles des outils documentaires traditionnels. Avec les contenus du Web, il ne s'agit pas de traiter un fonds documentaire aux contours connus ou même imaginés, mais de prendre en compte un univers complètement ouvert et sans limites précises. La logique de liens hypertextuels sur laquelle repose le Web permet théoriquement de mettre en relation tous les contenus en ligne. En s'appuyant sur ces liens, les grands moteurs de recherche tentent de proposer des services couvrant l'intégralité du Web, même si on sait qu'ils n'y arrivent que très partiellement. Le Web, dans son état actuel, n'aurait pas pu se développer sans ces moteurs, qui, pour beaucoup



d'internautes, en constituent l'unique point d'entrée...

Qu'est-ce qui différencie de manière significative les outils de recherche du Web et ceux qu'on utilise dans les bases de données documentaires ou les logiciels de gestion de fonds documentaires ?

Comme les contenus du Web ont par nature une forme numérique, les moteurs de recherche s'appuient sur le contenu textuel du document – le document *primaire* –, ou du moins des pages qui le constituent. Les logiciels documentaires, quant à eux, s'appuient principalement sur des données descriptives établies par les documentalistes à partir du document primaire, qu'on appelle aujourd'hui métadonnées et qu'on pourrait qualifier d'information *secondaire*. Selon les cas, la recherche porte sur les mots présents dans le document, ou sur des informations structurées décrivant le document (titre, auteurs, mots-clés, date, résumé...). La recherche plein texte correspondant au premier cas est devenue une référence incontournable, même si on en connaît les défauts et les limites. Les méthodes s'appuyant sur des notices descriptives, quant à elles, ne sont pas remises en question pour leur pertinence, mais ne peuvent pas se développer dans des proportions comparables pour le Web, puisqu'elles font appel à un travail de production d'une information supplémentaire sur chaque document, nécessitant le plus souvent une intervention humaine.

Apport des métadonnées

La recherche plein texte apporte des services inestimables, mais elle est pauvre du point de vue documentaire. On aimerait qu'elle s'enrichisse des éléments sémantiques que constituent les métadonnées, produites automatiquement par des outils adaptés ou manuellement par différents acteurs. Les outils de recherche du Web peuvent s'améliorer en s'appuyant à la fois sur le plein texte et sur les métadonnées. C'est l'enjeu du Web sémantique dont certains rêvent depuis quelques années. Pour cela, il faut que l'infor-

**Enrichir le Web
par des métadonnées,
première étape vers le Web
sémantique dont rêvent
les professionnels
de l'information.**

mation secondaire (les métadonnées) soit elle-même disponible sur le Web, et facilement identifiable par les moteurs de recherche. Plusieurs catégories de métadonnées peuvent y contribuer.

- Chaque document comporte en lui-même de l'information secondaire. Le titre, la date, le statut, le nom de l'auteur et de nombreuses autres informations ont une forte signification et sont utilisés dans la recherche s'ils sont compris par les automates comme autre chose que de simples mots du texte. Mais, pour se servir de ces métadonnées intrinsèques à chaque document, il faut tout d'abord que ces éléments figurent bien dans le texte ; il est tout à fait désolant que les documents publiés aujourd'hui sur le Web soient mis en ligne sans indication de date, de statut ou de droit d'usage, ou bien découpés en pages HTML mal reliées entre elles... Il faut également que ces éléments soient correctement identifiables. Les technologies de structuration des documents existent depuis longtemps, grâce à XML en particulier. Leur généralisation ne se fait cependant pas. C'est un élément de professionnalisation de l'édition sur Internet que les acteurs de la documentation doivent promouvoir.
- D'autres informations peuvent être produites par l'auteur ou l'éditeur du document au moment où il le met en ligne sur le Web : ce sont les métadonnées attachées au document. Elles peuvent reprendre certaines données déjà évoquées, mais également aller beaucoup plus loin dans la description et l'information sur le document. À qui

est-il destiné? À quoi peut-il servir? Comment peut-il être utilisé? Pour des documents pédagogiques, cette information est de toute première importance et ne peut pas, le plus souvent, être retrouvée automatiquement par un moteur dans le contenu textuel du document lui-même.

- D'autres acteurs interviennent éventuellement. À chaque fois que l'on constitue une notice documentaire relative à une ressource du Web, on dispose potentiellement de métadonnées, à condition de savoir les rendre disponibles aux outils de recherche. Les bases de données documentaires comme Éducasource constituent par exemple des réservoirs de métadonnées portant sur des ressources éducatives. Mais ces réservoirs ne sont pas encore, aujourd'hui, accessibles par d'autres outils.

Vers le Web sémantique?

L'enjeu des années à venir est d'arriver à mettre en réseau toutes ces sources de métadonnées pour les rendre interopérables et utilisables par des moteurs de recherche aux fonctionnalités améliorées, qui s'appuieraient ainsi non seulement sur le contenu propre des documents, mais également sur des propriétés plus fines des ressources, celles qui peuvent figurer dans des métadonnées. Quand les moteurs seront capables d'utiliser un réseau assez large de métadonnées, en scrutant les sources d'informations secondaires, et de les mettre en relation avec les documents primaires, on disposera d'outils beaucoup plus performants et susceptibles de rendre des services de qualité.

Bibliothèques numériques – Centres de documentation virtuels

Les outils de recherche ont pour vocation d'amener les utilisateurs vers les documents. Dans une démarche documentaire, on ne se contentera pas de cela; il faut également amener les documents vers les usagers, en fonction du public et du besoin. Constituer des fonds documentaires, sélectionner des documents et assurer les conditions de leur mise à disposition font partie des activités documentaires fondamentales, qu'il convient de transposer dans le contexte des technologies de l'information et de la communication.

Rassembler des contenus

Avec le numérique, constituer un fonds documentaire ne signifie pas rassembler des documents numériques dans un même site, mais organiser l'accès à un ensemble de ressources bien délimité, et garantir les conditions d'utilisation de ces contenus. Concernant les ressources du Web, des portails comme Éduduc jouent déjà ce

rôle, mais en donnant accès à des sites ou à des portions de sites et pas aux documents eux-mêmes, trop nombreux et souvent difficiles à identifier et à qualifier individuellement.

Aujourd'hui, les sources numériques de documents pédagogiques ou intéressants pour l'éducation ne croisent que partiellement le Web tel que les moteurs de recherche nous le montrent.

- Il y a tout d'abord les contenus numériques produits par des éditeurs, accessibles dans des sites à accès réservé, nécessitant un abonnement ou un autre mode de paiement.

- Les institutions et les communautés éducatives produisent également de nombreux documents et des ressources éducatives. Mais ces contenus sont le plus souvent mélangés à d'autres informations au sein de sites web dont les structures sont diverses.

- Il y a aussi des documents dans des sites personnels, des sites d'établissements scolaires, des intranets d'école ou de groupes d'enseignants...

Comment rassembler les documents qui intéressent nos publics dans des services documentaires adaptés? Les éditeurs tentent actuellement de le faire par des portails d'accès regroupant leurs productions, proposés sous forme d'abonnements globaux aux établissements. Dans une approche documentaire, on souhaiterait que ces ressources, lorsqu'elles sont acquises par des établissements ou des collectivités, soient associées à d'autres documents pour former des sortes de centres de documentation virtuels, adaptés aux différents types de publics du système éducatif.

L'exemple des publications scientifiques

L'exemple de l'enseignement supérieur et de la recherche, avec les publications scientifiques, est intéressant et instructif.

- Les éditeurs de revues scientifiques ont constitué, à partir de leurs publications, de grandes bibliothèques numériques accessibles par abonnement. Ils ne se sont pas limités à un portail d'accès unique mais ont veillé à ce que chaque revue, chaque numéro et chaque article soient accessibles individuellement, les restrictions d'accès ne portant pas sur les notices descriptives mais seulement sur le contenu édité. Il se développe par ailleurs des standards et des outils pour un système d'identification unique des documents sur le Web (DOI, sorte d'ISBN du Web pour les articles), ainsi que des mécanismes qui permettront de retrouver le lien vers le document numérique à partir d'une simple référence (citation, par exemple).

- Les universités et les communautés de chercheurs constituent de leur côté des bibliothèques numériques avec leurs propres productions:



mémoires, thèses, publications, cours... Pour assurer l'interopérabilité entre ces réservoirs de documents, elles s'appuient sur des standards ouverts (OAI-PMH¹: *Open Access Initiative – Protocol for Metadata Harvesting*) permettant de mettre à disposition et d'échanger entre elles les métadonnées descriptives de leurs ressources. Les archives Tematic² développées récemment avec un outil du CNRS en sont un exemple dans le domaine des TICE.

À partir de ces sources, qui représentent déjà en elles-mêmes de vraies bibliothèques numériques, des développements sont possibles : d'autres acteurs vont pouvoir constituer des sites thématiques ou disciplinaires, des moteurs de recherches spécialisés et des centres de documentation virtuels. Ces sites ne contiendront pas les documents eux-mêmes mais assureront un service d'accès adapté à un public et à un objectif donnés, tout en respectant l'origine de chaque document et même les conditions d'accès pour les ressources payantes.

C'est par des procédures analogues qu'on envisagera la constitution de services d'accès organisés à des ressources numériques pour l'éducation, s'appuyant sur des sources multiples incluant les éditeurs, les institutions et les autres communautés. Comme pour la production scientifique, cela suppose que ces sources s'appuient sur des standards de description communs, et sur des règles d'interopérabilité communes.

Chaque source peut garder ses caractéristiques propres et rester bien identifiée, mais elle doit faire partie d'un réseau de sources d'informations qui peuvent être combinées.

Si on qualifie ces services de « centres de documentation virtuels », c'est par analogie avec l'existant. C'est aussi pour donner à l'environnement documentaire qu'on souhaite fournir aux enseignants et aux élèves un cadre et une dimension plus appropriés que le grand Internet mondial. Mais on attend des technologies qu'elles aident à construire des outils novateurs prenant en compte le contexte de l'éducation. Ils doivent être paramétrables, en fonction des politiques documentaires des établissements, des niveaux et des publics ciblés. Les espaces numériques de travail pourront être un instrument pour les personnaliser et les adapter aux besoins et aux habitudes de chacun.

Organiser l'accès aux ressources numériques représente un des enjeux les plus importants dans le domaine de la documentation aujourd'hui. On ne peut pas se cantonner uniquement dans un rôle d'aide à la recherche documentaire, même si c'est très utile. Participer à la définition des politiques documentaires en matière de ressources numériques, mettre en œuvre à différents niveaux ces politiques en s'appuyant sur les technologies adaptées sont autant de fonctions qui relèvent des fondamentaux de la documentation. ●