

3. Des fonds éditoriaux structurés pour améliorer la lisibilité du Web éducatif

Nicole Rodriguez

SCÉRÉN-CNDRP

SERVICE DES PRODUCTIONS IMPRIMÉES ET NUMÉRIQUES

Les pratiques éditoriales et les pratiques documentaires restent encore trop séparées. Leur rencontre dans le document numérique contribuerait pourtant à organiser le Web éducatif en fonds éditoriaux structurés, pour faciliter l'accès des utilisateurs. Une évolution vers la construction de « pôles de ressources » se profile ainsi : il s'agit de structurer le Web en « services d'accès sur Internet à des fonds éditoriaux dont l'organisation est adaptée aux besoins d'un public circonscrit ».

L'avantage du document numérique sur Internet est double : il est immédiatement accessible, et il peut contenir en lui-même sa description documentaire et sa structure éditoriale, voire sémantique, avantage sous-exploité par les éditeurs.

Parallèlement, la force des moteurs de recherche est double : ils actualisent en permanence leurs connaissances sur des millions de pages dans des délais relativement courts, et ils accèdent au contenu de ces pages par une indexation en plein texte, indexation qui progresse en qualité sur le plan de l'analyse lexicale et sur le plan sémantique, si l'on croit les fournisseurs de ces technologies.

Mais la faiblesse de l'ensemble réside dans la difficulté que rencontre l'utilisateur pour trouver le document dont il a besoin.

L'utilisateur au centre du dispositif

Lors d'une recherche sur Internet, en effet, l'utilisateur exprime une requête à l'aide d'un formulaire de recherche ; il s'adresse à un automate,

le moteur de recherche, qui explore un fonds chaotique formé de milliards de pages et fournit d'immenses listes de résultats, qu'il faut explorer un par un...

Lors d'une recherche de documents dans des fonds éditoriaux locaux, l'enseignant, après avoir exploré sa bibliothèque personnelle, se dirige vers le CDI de son établissement, puis vers les fonds de son environnement proche (IUFM, CDDP, CRDP, bibliothèques spécialisées...). Il formule une demande spontanément ou en s'appuyant sur ses outils professionnels – le programme dans une discipline et une classe données, par exemple. Il s'adresse alors à un professionnel de la documentation, qui interprète cette demande et la convertit en stratégie de recherche documentaire à l'aide d'outils qu'il maîtrise. Le fonds éditorial exploré est construit et organisé en fonction de besoins pédagogiques. Au final, une sélection de documents, limités en nombre, est proposée.

La différence entre ces deux situations de recherche est bien connue... Elle est rappelée pour introduire cette question : comment la recherche sur Internet pourrait-elle approcher l'efficacité d'une recherche effectuée par un professionnel de la documentation ?

Une séparation fatale entre le document et son indexation

La constitution et l'organisation de fonds documentaires étaient jusqu'à présent effectuées par des professionnels, bibliothécaires ou documentalistes, qui repéraient, sélectionnaient, indexaient les documents publiés, gérant ainsi leurs fonds en bases de données bibliographiques. Appliquée sur Internet, cette démarche a produit une multitude de *bookmarks*, de sites portails thématiques ou disciplinaires, structurés ou non en bases de données. Effectuée dans le but de faciliter l'accès de l'utilisateur aux documents, elle s'est vite révélée extrêmement limitée et a, par sa prolifération, souvent égaré l'utilisateur.

« Ces descriptions sont destinées à injecter de l'intelligence humaine consommable par des moteurs de recherche... »

L'expérience des sites Éducasource et Éduclic, conduite par le CNDP et le réseau des CRDP à la demande du Ministère, illustre bien les limites rencontrées. Outre l'ampleur de la tâche – vider la mer avec ses mains –, deux obstacles majeurs ont été rencontrés. D'autre part, l'hétérogénéité du corpus, son étendue, échappant aux règles habituelles de l'édition, ont rendu fluctuante la définition de l'unité documentaire à traiter : visait-on la construction d'un annuaire, et dans ce cas attaquaient-on au niveau du site, du sous-site, du corpus homogène de documents... ou privilégiait-on l'accès au document lui-même, en prenant le risque d'avoir à prospecter tous les océans de la planète ?

L'indexation automatique : à la volée, ou à l'aveuglette ?

De leur côté, les moteurs de recherche sur Internet se sont développés. En attaquant directement le contenu, ils offrent un accès régulièrement actualisé à l'utilisateur. Mais ils le font avec une « intelligence » limitée.

En effet, ils indexent indistinctement des pages, quelle que soit leur forme ou leur nature, pages d'accueil de sites, de rubriques, listes de liens, sommaires, pages de contenu... sans être aptes à reconstituer l'entité document...

L'indexation automatisée est encore imparfaite malgré les progrès effectués en indexation lexicale et sémantique. De plus, elle ne peut porter que sur des éléments présents dans le texte. Ainsi, dans des contenus traitant du programme de SVT, pourront être reconnues des thématiques comme « préparation d'une sortie de classe » ou « êtres vivants », mais ni la classe, ni la discipline ne seront mentionnées, dans le cas fréquent où ces informations figurent dans une page antérieure sous forme de lien vers le document.

L'indexation automatisée ou l'intelligence documentaire

Les importants travaux conduits par les organismes de normalisation et de standardisation, y

compris l'engagement du W3C vers le Web sémantique, révèlent une mutation considérable dans la façon d'aborder ces questions.

La piste poursuivie consiste à enrichir les flots de pages indexées automatiquement de leur structure éditoriale et documentaire. Ainsi, le moteur est à même de reconstituer « le document » placé dans la toile d'araignée hypertextuelle. Il agit donc sur un contenu délimité par le contour du document complet – et non plus seulement sur celui de la page. L'indexation automatique, appliquée jusqu'ici à un texte où chaque chaîne de caractères était équivalente à une autre, est guidée par une hiérarchisation de l'information et l'utilisation de descriptions documentaires (catalogages, thésaurus, classifications documentaires) lui donnant sens. La pertinence des résultats de recherche s'en trouve doublement améliorée.

Cela signifie que chaque document ou « objet pédagogique » admis dans le corpus publié est accompagné d'un ensemble de descriptions documentaires, appelées « métadonnées ». Comme ces descriptions sont destinées à injecter de l'intelligence humaine consommable par des moteurs de recherche, il est nécessaire d'en normaliser la forme afin qu'elle soit décriptable par tous ces automates. C'est pourquoi sont actuellement développés de nombreux « schémas de métadonnées¹ » sur lesquels les organismes de normalisation et les prescripteurs éducatifs travaillent activement. Conjointement, les technologies d'indexation et de recherche automatique évoluent dans le sens de la prise en compte des principaux schémas de métadonnées.

Est-ce vider la mer avec ses mains que d'organiser le fonds documentaire éducatif ?

1. Dublin Core Educative, Learning Objects Metadata (LOM) sont les noms de schémas qui reviennent le plus fréquemment dans les milieux éducatifs.



Deux tâches nouvelles pour l'éditeur

La fonction d'indexation, assumée traditionnellement a posteriori par le documentaliste, s'intègre désormais à l'acte d'édition. Le document devra être indexé et structuré² d'une part, d'autre part incorporé dans un fonds éditorial partagé.

Plusieurs contraintes en découlent, qui nécessiteront une prescription acceptée ou élaborée par une communauté d'éditeurs ainsi qu'une coordination assurée par des professionnels de la documentation pédagogique, à savoir :

- le choix d'un schéma de métadonnées ;
- un habillage du schéma, c'est-à-dire la four-niture aux éditeurs des langages et des vocabulaires documentaires communs devant être utilisés pour indexer le fonds ;
- une définition des critères d'appartenance au fonds, des types d'objets éditoriaux qui le constituent et de l'unité documentaire sur lesquels appliquer des métadonnées ;
- des consignes d'application ;
- l'assistance et la formation des éditeurs à l'application ;
- l'apposition de métadonnées sur chaque document dans le flux de production ;
- une politique de mise à niveau des fonds actuels ;
- des procédures de contrôle et de validation.

Des fonds à vocation éducative existent déjà. Il suffit de parcourir les sites du Ministère, des académies, du CNDP, des CRDP, de l'INRP... les pages éducatives des grands sites institutionnels (Culture, BNF, musées, Sénat, Assemblée nationale...), les sites associatifs et bien d'autres encore, pour s'en convaincre. Mais ils se sont constitués dans la spontanéité, l'ignorance des autres et sans règles de mutualisation. La question est alors d'en améliorer la qualité éditoriale et la lisibilité, d'en étudier les manques, de les étendre pour couvrir l'ensemble des besoins éducatifs et de mieux prendre en compte la dimension éditoriale de ce nouveau support.

En effet, la publication sur Internet, encore fortement marquée par l'édition imprimée, est appelée à évoluer progressivement vers des « objets pédagogiques » davantage fragmentés, incorporés dans des cursus d'apprentissage interactifs, et déposés dans des fonds partagés auxquels ils s'agrègent. Cette évolution est vraiment perceptible sur les sites associatifs, dans les rubriques des groupes académiques disciplinaires ou certaines pages personnelles, où la publication est effectuée directement par des auteurs de ter-

rain. La relation de l'éditeur aux auteurs s'en trouve modifiée. L'établissement de partenariats avec ces groupes constitués d'auteurs, quasi-éditeurs, ressentis parfois comme concurrentiels, semble être un facteur important de l'adéquation entre l'offre et les besoins.

La prise en compte du point de vue de l'utilisateur accédant simultanément à des fonds éditoriaux de sources hétérogènes devra conduire l'éditeur à inscrire ses publications dans des chartes éditoriales définissant des typologies d'objets édités, des modes de présentation et de circulation dans les documents, en trouvant un compromis entre son image éditoriale et une certaine standardisation.

**« Un premier pas
serait de dégager des
programmes d'enseignement
un langage d'interrogation
présenté à l'utilisateur pour
sa recherche... »**

Vers des pôles publics de ressources pédagogiques

Il serait illusoire d'imaginer construire un univers parfait pour l'ensemble de « la » communauté éducative. Il va donc falloir segmenter celle-ci en établissant des communautés d'utilisateurs et des priorités : enseignant dans sa classe, personnels d'encadrement dans un établissement scolaire, etc. Sinon, en voulant satisfaire tout le monde, on court le risque de ne satisfaire personne.

En effet, pour chacune des communautés retenues, il faut évaluer les besoins en matière de ressources pédagogiques et définir des scénarios d'accès adaptés. Ainsi, à partir des programmes et des documents d'accompagnement, le profil « enseignant dans sa classe par disciplines et niveaux », naturellement prioritaire, peut être couvert facilement. L'élaboration de langages d'interrogation ne nécessitant pas de connaissances documentaires est une piste à développer. Par exemple, un premier pas serait de dégager des programmes d'enseignement un langage d'interrogation présenté à l'utilisateur pour sa recherche, d'autant que des outils informatiques maintenant disponibles offrent une assistance lexicale et sémantique à l'analyse de contenus.

Il serait illusoire également d'imaginer que l'ensemble des éditeurs éducatifs va s'enflammer spontanément pour la mutualisation de leurs fonds dans un pôle public... Cependant, à l'initiative du Ministère, l'idée fait son chemin, comme on le constate avec les exemples des ENS, CNS et KNE³. Le rôle de la prescription est donc essentiel : chartes éditoriales, typologies d'objets, couverture des besoins... n'existeront pas par génération spontanée. Il faudra les définir, les faire

2. Il ne s'agit pas, bien sûr, de la structure éditoriale offerte au lecteur, mais de son explicitation sous une forme compréhensible par un automate grâce à des schémas XML.

3. Cf. *Les Dossiers de l'ingénierie éducative*, n° 46, « Les nouveaux espaces numériques d'éducation »

connaître, aider à les appliquer. Des actions comme celle du Schéma de l'édition numérique pour l'enseignement (SCHENE) mis en place par la direction de la Technologie au Ministère sont, de ce point de vue, intéressantes. La réalisation effective de tels fonds suppose la présence d'un maître d'ouvrage et l'adhésion d'une communauté de producteurs/éditeurs de documents pédagogiques bien décidés à mutualiser l'ensemble de leur production et à en accepter les contraintes.

Il va falloir raisonner en sphères de syndication d'éditeurs. La première, naturelle, concerne la sphère institutionnelle. S'il est créé à l'initiative du Scérén, le fonds sera évidemment constitué de ressources placées sur les sites du CNDP et des CRDP, dont les éditeurs, organisés en réseau, sont animés d'une volonté commune d'avancer. On peut donc espérer une réalisation relativement rapide grâce à une mutualisation de l'élaboration des chartes éditoriales et documentaires et à une application immédiate, qui pourrait servir de préfiguration et de base de tests. À l'initiative du Ministère – à travers le projet MURENE⁴ –, le fonds en sera plus riche et plus complet, les corpus des sites du Ministère, des académies... y trouvant automatiquement leur place. Une deuxième sphère devra être prévue dès le départ sous la forme d'une ouverture progressive à des partenariats avec les organismes et éditeurs, institutionnels, privés ou associatifs, publiant sur le Web. Restent les éditeurs qui publient hors Web ou ceux qui ne souhaiteront ou ne pourront pas s'intégrer dans ce type de partenariat mais seraient favorables à l'inscription de leurs ressources dans le pôle. Les méthodes centralisées de sélection et de description telles qu'elles sont pratiquées dans Éducasource garderaient toute leur utilité pour leurs fonds.

Dans tous les cas, le CNDP et le réseau des CRDP présentent la particularité d'être simultanément des éditeurs sous tutelle du Ministère et des professionnels de la documentation. Ils sont également chargés d'une mission de formation, notamment sur les techniques d'indexation et d'interrogation documentaires. De ce fait, ils sont des interlocuteurs privilégiés du Ministère, notamment de la direction de la Technologie, pour la prescription éditoriale et documentaire et l'accompagnement de son application auprès des éditeurs éducatifs.

La clé de la réussite: la simplicité

Afin de ne pas alourdir le travail de chacun, les prescriptions devront être simples, claires, rapides à appliquer. Il faudra, notamment, s'employer à optimiser l'utilisation des fonctions automatisées des moteurs afin d'alléger au mieux les



« Des fonds à vocation éducative existent déjà. Il suffit de parcourir les sites du Ministère, [...] les pages éducatives des grands sites institutionnels [...] et bien d'autres encore, pour s'en convaincre. »

travaux d'indexation des producteurs de contenus, sauf à faire fuir les éditeurs devant des schémas de métadonnées trop complexes. Seront également fournis des outils qui les aideront à appliquer les métadonnées ou à les extraire de bases de données existantes.

La plus grande attention sera également portée à l'élaboration d'interfaces de recherche accessibles à tout utilisateur. Les travaux et recherches n'en sont qu'à leur début; l'histoire technologique des prochaines années ne fait que commencer, avec le Web sémantique, les ontologies... Mais, d'ores et déjà, les améliorations peuvent être significatives.

Enfin, on le sait, le développement des usages marque toujours des retards sur celui des TIC, retards résultant de la priorité donnée aux développements technologiques, sans réelle prise en compte des périodes incompressibles de formation, d'appropriation et de maturation. Une attention particulière sera donc apportée à l'accompagnement et la formation des usagers, éditeurs ou consommateurs de documents. Avec Dominique Borne, doyen de l'IGEN, gageons que l'« *on ne pourra pas longtemps demander aux collectivités de poser les tuyaux sans rien mettre dedans*⁵ » ! ●

4. MURENE: projet de mutualisation des ressources pour l'éducation conduit par la direction de la Technologie du Ministère.

5. Dépêche de l'AEF, Nantes, 5 octobre 2004.