

Structuration du savoir et organisation des collections dans les répertoires du Web

On a comparé le World Wide Web à une bibliothèque dévastée par des vandales qui auraient détruit le catalogue et jeté des centaines de milliers de documents pêle-mêle sur le sol, après en avoir arraché les pages liminaires¹. Cette image désolante, reflet de la réalité d'il y a quelques années, n'a plus le même impact aujourd'hui, alors que sont faits d'importants efforts d'organisation des ressources numériques.

Michèle Hudon

École de bibliothéconomie
et des sciences de l'information
Université de Montréal
michele.hudon@umontreal.ca

L'explosion de l'utilisation du Web par le grand public nous permet d'assister à un phénomène intéressant, celui de la prolifération des offres d'accès thématiques aux ressources d'Internet. On pensera aux répertoires généraux de type Yahoo, ainsi qu'au nombre croissant de portails et de bibliothèques virtuelles spécialisés qui facilitent le repérage de ressources ayant fait l'objet d'une sélection et d'une évaluation.

Une organisation physique et conceptuelle minimale des ressources d'Internet est désormais considérée comme essentielle. Pourquoi ? Pour les mêmes raisons que l'on ne pourrait laisser des collections documentaires traditionnelles sans organisation dès que le nombre de documents qu'elles contiennent dépasse un seuil au-delà duquel il devient difficile et onéreux de les exploiter efficacement. Ceux qui cherchent des ressources portant sur un sujet particulier ont peu de temps à consacrer à une errance sur les réseaux qui devient vite coûteuse. Lorsqu'il s'est agi d'organiser les res-

sources disponibles sur le Web, les méthodes et les instruments utilisés en milieu traditionnel depuis plus d'un siècle n'ont pas été retenus. Aux schémas de classification documentaire on a préféré des structures faites de « catégories » représentant des disciplines (Sciences naturelles, Sciences sociales), des grands thèmes (Loisirs, Éducation) et des formes de documents (Dictionnaires, Ouvrages de référence). Aux listes de vedettes-matières et aux thésaurus, on a préféré la recherche libre par mots-clés dans le plein texte des ressources ou dans les métadonnées. Pour justifier ces choix, on invoque encore la complexité et l'inflexibilité des schémas traditionnels, en soutenant qu'ils reflètent une structure du savoir statique et dépassée par l'évolution rapide des disciplines. Qu'en est-il vraiment ? Les schémas traditionnels de représentation des connaissances et d'organisation documentaire ne sont-ils vraiment pas appropriés au nouvel environnement informationnel ? Les nouveaux outils de structuration des savoirs et de naviga-

1. Michael Gorman, « The corruption of cataloging », *Library Journal*, September 15, 1995, p. 33.

Michèle Hudon est titulaire d'un Ph. D. en sciences de l'information (université de Toronto). Elle a assumé la responsabilité des services techniques de Bibliothèque de l'Université du Québec à Hull, avant de dispenser des cours en techniques de la documentation dans différents collèges du Canada. Depuis 1997, elle est professeure adjointe à l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal. Elle a publié plusieurs articles sur les problèmes d'indexation, ainsi qu'un ouvrage *Le thésaurus : conception, élaboration, gestion, paru en 1995*.

tion dans les collections virtuelles sont-ils aussi conviviaux et efficaces qu'on le dit ? Peut-on espérer un rapprochement, pour le grand bénéfice des utilisateurs, entre les tenants de la systématisation et ceux de la flexibilité ?

Pour alimenter la réflexion sur ces questions auxquelles aucune réponse définitive ne peut encore être apportée, nous rappellerons d'abord les caractéristiques et les fonctions essentielles des schémas de classification documentaire traditionnels. Nous décrirons ensuite un nouveau mode d'organisation systématique des collections virtuelles générales et spécialisées, avant d'évoquer quelques tentatives d'application de structures traditionnelles en bibliothèque virtuelle.

Structures et organisation traditionnelles

En milieu documentaire traditionnel, l'utilisation d'un schéma de classification permet d'offrir simultanément à l'utilisateur deux types d'accès aux collections : un accès systématique et un accès physique. On croit à tort dans certains milieux que la fonction essentielle de la classification documentaire est liée à l'organisation physique des documents sur les rayons d'une bibliothèque. La fonction de rangement, qui assure le contact direct entre utilisateurs et documents, est pourtant subordonnée à la fonction d'organisation systématique, laquelle permet au

Figure 1. Exemple de structure hiérarchique logique (CDD)

600	Sciences appliquées - Technologie
610	Sciences médicales - Médecine
616	Maladies
616.7	Maladies de l'appareil locomoteur
616.72	Maladies des articulations
616.722	Arthrite
616.722 [03]	[Dictionnaires]

chercheur de repérer les documents pertinents qu'il ne connaît pas encore.

Les schémas de classification de type philosophique (la Classification décimale de Dewey [CDD] et la Classification décimale universelle [CDU] par exemple) tentent de reproduire les structures d'organisation du savoir que proposent les théoriciens et autres spécialistes des domaines couverts. La classification fonctionne d'ailleurs jusqu'à un certain point comme l'ouvrage théorique, en identifiant et en nommant les concepts et en les mettant en relation les uns avec les autres dans une structure logique². Cette structure est fondée sur un ensemble complexe de relations hiérarchiques pures (*i.e.* des relations de genre à espèces) ou contextuelles (des relations de cause à effet, par exemple) dont le rôle est d'amener progressivement le chercheur vers les documents qui sont susceptibles de contenir la réponse à sa question. Ainsi, l'individu en quête d'un lexique de termes relatifs à l'arthrite est-il d'abord dirigé dans la section de la collection qui contient les ressources

traitant de sciences appliquées, laquelle contient les ressources en sciences de la santé. De là, il devient relativement aisé pour notre chercheur de repérer l'ensemble des documents traitant de maladies, de maladies de l'appareil locomoteur, de

maladies des articulations et spécifiquement d'arthrite. La CDD et la CDU permettent même la spécification du genre de document, isolant ainsi les dictionnaires des autres ouvrages sur un sujet.

Dans l'exemple de la figure 1, la structure conceptuelle est exposée dans la colonne de droite, chaque retrait marquant le passage à un niveau hiérarchique inférieur. Dans la colonne de gauche,

les indices numériques ne sont que la représentation symbolique de cette organisation conceptuelle. L'indice de classification sert d'adresse ; il a permis d'aller ranger le document sur les rayons et il permettra aussi d'aller l'y rechercher.

La structure hiérarchique est efficace pour représenter les connaissances. Elle offre de façon économique une information relativement complète sur chacun des concepts et des phénomènes étudiés, permettant ainsi de mieux les cerner et de les visualiser. Le schéma de classification hiérarchique n'est toutefois pas exempt de faiblesses ; il est le produit

La fonction
de rangement
est subordonnée à la
fonction d'organisation
systématique,
laquelle permet
au chercheur de repérer
les documents pertinents
qu'il ne connaît
pas encore

2. Barbara Kwasnik, « The role of classification in knowledge representation and discovery », *Library Trends*, 1999, t. 48, n° 1, p. 25.

STRUCTURATION DU SAVOIR DANS LES RÉPERTOIRES DU WEB

de son époque et il reflète nécessairement une vision du savoir qui ne saurait être exempte d'influences et de préjugés. On lui reproche notamment sa rigidité et son incapacité à intégrer des phénomènes qui peuvent être vus sous plusieurs angles. Le degré de profondeur requis dans les schémas de classification encyclopédiques, et, par conséquent, le nombre de paliers à traverser avant d'atteindre le niveau de spécificité voulu, rendent la tâche de l'utilisateur de plus en plus complexe. Et si la structuration des connaissances et l'organisation de la documentation peuvent se faire efficacement dans un domaine du savoir établi et plutôt stable, il en va tout autrement lorsqu'il s'agit de représenter un domaine en émergence ; on voit alors se multiplier les classes aux limites mal définies, peu utiles au repérage.

Structures et organisation de la bibliothèque virtuelle

La collection de la bibliothèque virtuelle est faite de liens actifs vers des ressources numériques portant sur une ou plusieurs thématiques définies. Les bibliothèques virtuelle et traditionnelle ont beaucoup en commun. On y fait de la sélection (ou filtrage) de ressources, de l'organisation (sur la base du traitement descriptif et analytique des éléments constitutifs de la collection) et du développement d'outils de repérage. Au niveau du service offert, cependant, la différence est grande entre elles. La simple existence de la bibliothèque virtuelle est elle-même un service à l'explorateur du Web, certes, mais le degré de personnalisation de ses services à l'utilisateur n'est pas comparable à celui qu'offre encore la bibliothèque traditionnelle. La bibliothèque virtuelle offre peu d'information descriptive sur ses collections ; ceux qui la gèrent considèrent que la description du contenu et des caractéristiques d'une ressource n'est pas essentielle, à partir du moment où l'on peut y accéder

d'un clic de souris. Ce mode de fonctionnement force évidemment l'utilisateur à un va-et-vient incessant et peu efficace entre l'outil d'accès à la collection et la ressource. Et puisque la bibliothèque virtuelle ne fournit pas d'accès physique direct à sa collection, il faut nécessairement passer par une structure d'organisation conceptuelle avant d'accéder aux documents.

Organisation de type hiérarchique

Les créateurs et diffuseurs d'information sur le Web reconnaissent l'intérêt d'une organisation de type hiérarchique. Ils ont adopté avec enthousiasme ce mode d'organisation, comme en témoigne la prolifération des sites-répertoires comme Yahoo ou La Toile du Québec et les nombreuses tentatives de structuration des instruments d'accès aux collections de bibliothèques virtuelles. Les structures d'organisation utilisées

comparaison des catégories de départ offertes par Yahoo France³ et par La Toile du Québec⁴, deux répertoires francophones populaires et fréquemment consultés.

On notera déjà que ces répertoires répartissent différemment les ressources auxquelles ils donnent accès. L'exploration de ressources dans le domaine de l'éducation révèle des différences encore plus marquées au deuxième niveau de division ; Yahoo y propose une vingtaine de sous-catégories, La Toile dix seulement (figure 3 page suivante).

Règne de l'ambiguïté

Les dix sous-catégories de La Toile méritent qu'on s'y attarde un peu. Où trouvera-t-on, par exemple, les associations et les organismes s'intéressant à l'enseignement collégial ? Une vérification rapide montre en effet que les ressources qui pourraient être assignées simultanément aux deux sous-catégories n'apparaissent que

Figure 2. Catégories de départ dans Yahoo France et dans La Toile du Québec

Yahoo France	La Toile du Québec
Actualités et médias	Actualités et médias
Art et culture	Arts et Culture
Commerce et économie	Commerces (<i>sic</i>) et économie
Divertissement	Éducation
Enseignement et formation	Gouvernement
Exploration géographique	Informatique et Internet
Informatique et multimédia	Sciences et santé
Institutions et politique	Société
Références et annuaires	Sports et loisirs
Santé	Tourisme, villes et régions
Sciences et technologies	
Sciences humaines	
Société	
Sports et loisirs	

dans la majorité des bibliothèques virtuelles ressemblent peu aux schémas de classification documentaire traditionnels. Dans la bibliothèque virtuelle générale, les catégories de premier niveau ne sont pas établies sur une base strictement disciplinaire : on y retrouve un mélange de général et de spécifique, de disciplines, de domaines d'intérêt, de thématiques précises, de types de documents. La figure 2 permet une

comparaison des catégories de départ offertes par Yahoo France³ et par La Toile du Québec⁴, deux répertoires francophones populaires et fréquemment consultés.

3. <http://www.yahoo.fr>
[consulté le 30 septembre 2000].

4. <http://www.toile.qc.ca>
[consulté le 30 septembre 2000].

Figure 3. Le domaine Éducation dans Yahoo France et dans La Toile du Québec

<p>Yahoo > Enseignement et formation (extraits)</p> <ul style="list-style-type: none"> Actualités et médias Conseil et orientation Emploi Enseignement primaire Enseignement professionnel Enseignement religieux Enseignement secondaire Guides et annuaires Web Organismes Technologies d'enseignement 	<p>La Toile du Québec > Éducation (intégral)</p> <ul style="list-style-type: none"> Associations Enseignement collégial Enseignement primaire et secondaire Enseignement universitaire Formation professionnelle Magazines Organismes Presse étudiante Ressources Technologie éducative
--	---

d'inclure les subdivisions communes Associations, Organismes, Magazines, etc., sous chacun des ordres d'enseignement ? Notons aussi que dans l'ensemble d'une douzaine de ressources énumérées directement sous Éducation, donc des ressources qui devraient être générales, on trouve des liens vers des sites se rapportant à des concours, à la lecture rapide, à l'enseignement à la maison, au transport scolaire : ces liens n'ont donc pas trouvé leur place dans la structure existante. Dans ces nouvelles structures, catégories et sous-catégories sont rarement mutuellement exclusives ; l'ambiguïté y règne et l'utilisateur ne peut être certain d'avoir repéré tout ce qui est disponible ou ce qu'il y a de mieux.

La recherche de ressources spécifiques sur le théâtre pour enfants, nouvel exemple, révèle des problèmes d'un autre ordre : le manque de cohérence et l'absence de prévisibilité. Les treize ressources pertinentes sélectionnées par La Toile du Québec sont dispersées au sein de cinq grandes catégories et de neuf sous-catégories, dont : « Arts et Culture > Édition et littérature > Genres > Littérature jeunesse » ; « Arts et cultures > Théâtre > Compagnies et troupes » ; « Commerces (*sic*) et économie > Produits et services > Enfants » ; « Société > Enfants et ados ». Aucune des ressources n'est classifiée à plus d'un endroit et il va sans dire que ces catégories très diverses ne sont liées par aucun renvoi. Yahoo France semble au départ

plus logique : toutes les ressources sur le théâtre pour enfants sont regroupées sous « Société > Groupes et communautés > Enfants > Yahoo pour les enfants > Art et baz'art > Spectacles ». Les ressources concernant le cinéma pour enfant sont cependant ailleurs, sous « Art et culture > Arts et spectacles », là où on trouvera également des informations sur les spectacles de marionnettes. La mise en parallèle des liens classifiés dans « > Art et baz'art > Spectacles » et dans « > Arts et spectacles > Marionnettes » révèle que quelques liens seulement paraissent sur les deux listes.

À juste titre, on pourrait penser que les clients de bibliothèques virtuelles spécialisées bénéficieraient d'une organisation plus cohérente, plus normalisée et plus prévisible. Cela n'est pourtant pas le cas. Les modes d'organisation des bibliothèques virtuelles spécialisées diffèrent quant à leur extension et à leur profondeur, allant de la simple liste alphabétique de thèmes (figure 4) à la

structure complexe basée sur une organisation des connaissances validée par les spécialistes du domaine (figure 5), en passant par la structure développée sur le tas dans un domaine multidisciplinaire (figure 6).

Convivialité et ordre de façade

Les schémas d'organisation de la bibliothèque virtuelle sont conviviaux, flexibles, et plus à même que les schémas traditionnels de faire une place, ne serait-ce que temporaire, à de nouveaux centres d'intérêt. Mais leurs faiblesses ne sont pas négligeables. Ces nouvelles structures se développent rarement sur plus de

Figure 4. Exemple d'une liste alphabétique de thèmes

<p>Domaine Gestion – École des Hautes Études Commerciales, Montréal* (extraits)</p> <ul style="list-style-type: none"> Marketing Monnaie, voir Taux de change Nouvelles, voir Actualités Organismes internationaux Pays Plan d'affaires Production Qualité <p><small>* http://www.hec.ca/biblio/sites/interet.html [consulté le 30 septembre 2000].</small></p>
--

trois niveaux de profondeur ; elles n'offrent donc pas la spécificité requise dans les grandes collections pour éviter que le nombre de ressources groupées au plus bas niveau de la hiérarchie ne soit trop élevé. On ne perçoit pas dans ces structures de souci de normalisation terminologique ou structurelle ; les gestion-

Figure 5. Exemple d'une structure complexe validée par les experts du domaine

<p>Domaine Ingénierie Engineering Electronic Library* (extraits)</p> <ul style="list-style-type: none"> 600 Mechanical engineering 700 Electrical engineering 720 Computers and Data processing 723 Computer software, Data handling and Applications 723.4 Artificial intelligence 723.4.1 Expert systems 800 Chemical engineering <p><small>* http://eels.lub.lu.se/ [consulté le 30 septembre 2000].</small></p>
--

STRUCTURATION DU SAVOIR DANS LES RÉPERTOIRES DU WEB

Figure 6. Exemple d'une structure adaptée à un domaine multidisciplinaire

<p>Domaine Développement international Centre de recherche sur le développement international, Canada* (extraits)</p> <hr/> <p>Droits de la personne - Démocratie Généralités Démocratie Droits de la personne Enfants Droit international, Traités, Conventions Réseaux Économie et finances Environnement</p> <hr/> <p>* http://www.acdi-cida.gc.ca/virtual.nsf/pages/index_f.htm [consulté le 30 septembre 2000].</p>

naires de bibliothèques virtuelles se préoccupent peu d'uniformisation, forçant leurs clients à réapprendre à chercher lorsqu'ils vont de l'une à l'autre. Les nouveaux schémas d'organisation ne sont pas cohérents dans l'arrangement des catégories et des sous-catégories et l'absence de relations logiques entre les différentes composantes de la structure rendent les hiérarchies très peu informatives. Quant à l'absence de renvois d'orientation, elle a des conséquences certaines sur le rappel de ressources pertinentes.

Alors que les schémas de classification traditionnels veulent ressembler à des classifications du savoir, les schémas d'organisation de la bibliothèque virtuelle n'ont aucune prétention de la sorte. Ces nouvelles structures sont résolument utilitaires et ne prétendent pas à la stabilité. Cela justifie-t-il cette apparente désinvolture face aux principes théoriques de la classification, notamment en ce qui concerne les bases de division et l'ordre d'énumération ? Le manque d'uniformité entre structures couvrant un même domaine montre bien qu'elles sont tout aussi marquées par leur environnement, et même parfois par une vision individuelle du monde, que ne le sont les schémas traditionnels. Il pourrait bientôt être difficile de convaincre les chercheurs avertis que les nouvelles structures d'organisation ne sont pas qu'un produit improvisé par des amateurs et qu'elles leur sont vraiment utiles.

Structures traditionnelles en bibliothèques virtuelles

Les gestionnaires de bibliothèques virtuelles affirment volontiers que les accès qu'ils offrent sont plus intuitifs que ceux qu'offre la bibliothèque traditionnelle. Il ne semble pourtant pas assuré que l'utilisateur hésitera moins devant la structure d'organisation des ressources proposée par les concepteurs de la Toile du Québec que devant celle que proposent les rédacteurs de la CDD ou de la CDU. Pour l'utilisateur moyen, l'effort à fournir dans la bibliothèque virtuelle pour obtenir de l'information pertinente et valable n'est pas beaucoup moindre que celui qui doit être fourni dans un milieu traditionnel ; il pourrait même être plus grand, puisqu'en bibliothèque virtuelle on peut rarement compter sur l'assistance d'un intermédiaire spécialiste du repérage d'information.

L'ignorance aurait-elle été un facteur déterminant dans la décision des premiers gestionnaires de bibliothèques virtuelles de ne pas utiliser les schémas de classification traditionnels ? Les concepteurs du modèle Yahoo vantent leur approche déductive (« *bottom up* ») dans le développement des hiérarchies, l'opposant à l'approche inductive traditionnelle⁵

5. Anne Callery, *Yahoo! Cataloging the Web*, University of California, Santa Barbara, 1996, <http://www.library.ucsb.edu/untangle/callery.html> [consulté le 30 septembre 2000].

(« *top down* »). Mais en réalité, aucun des grands schémas d'organisation documentaire n'est plus développé selon une approche inductive stricte (la classification de la bibliothèque du Congrès ne l'a d'ailleurs jamais été). Les schémas d'organisation les plus performants sont justement développés par une combinaison d'approches qui tient compte à la fois de l'évolution bibliographique (ce qui est offert) et de l'évolution scientifique (ce qui est demandé).

Les schémas de classification documentaire traditionnels ne sont certes pas tous appropriés à l'organisation de la bibliothèque virtuelle, mais il vaut la peine d'examiner la pertinence d'utiliser les plus performants d'entre eux. La CDD, par

Les gestionnaires de bibliothèques virtuelles affirment volontiers que les accès qu'ils offrent sont plus intuitifs que ceux qu'offre la bibliothèque traditionnelle

exemple, est déjà utilisée dans 135 pays sur les cinq continents, n'est pas liée à une langue particulière, est mise à jour régulièrement et fait l'objet d'un effort de révision important qui vise à la rendre en même temps moins américaine et plus acceptable par d'autres sociétés. La CDD est également utilisée dans des projets de recherche élaborés ayant pour objet la classification automatique de ressources numériques⁶.

6. Information sur la CDD disponible à <http://www.oclc.org/dewey/index.htm> [consulté le 30 septembre 2000].

Les bibliothèques traditionnelles qui ont mis sur pied des bibliothèques virtuelles ont souvent choisi d'en organiser les ressources à l'aide de schémas de classification déjà bien établis. Un nombre croissant de bibliothèques publiques virtuelles choisissent d'utiliser la CDD de façon explicite dès la page d'accueil. C'est le cas par exemple des Bibliothèques virtuelles de l'Association des bibliothèques départementales de prêt⁷ (ADBDP) et de Québec-Chaudière Appalaches⁸. Les indices ne comptent pas plus de cinq chiffres, mais ils sont tous significatifs.

Ce mode d'organisation offre le grand avantage de faciliter l'intégration de ressources documentaires traditionnelles et de ressources numériques dans un même domaine du savoir. Dans ce nouveau milieu comme en milieu traditionnel, l'usager n'a pas besoin de comprendre les principes de la classification et de connaître la structure complète du schéma de classification utilisé pour en retirer les bénéfices. Cependant, l'utilisation d'une structure classificatoire traditionnelle n'est évidemment pas intuitive pour l'usager. Dans des sites développés avec soin, on fournira donc aussi un accès alphabétique à la structure, par la voie d'un index de termes de la langue naturelle. En cliquant sur le terme « Écoles à la maison », par exemple, l'usager

7. <http://www.savoie-biblio.com/bv/index.htm> [consulté le 30 septembre 2000].

8. <http://www.bibliotheques.qc.ca/frames36.htm> [consulté le 30 septembre 2000].

Figure 7. Exemple d'organisation systématique doublée d'un index alphabétique

Information sur le Canada par matière*	
Organisation systématique (extrait)	Index alphabétique (extrait)
1 Philosophie	Écoles
2 Religion	Écoles à la maison
3 Sciences sociales	Écoles maternelles
37 Éducation	Écoles secondaires
370 Éducation	Enfants
370.113 Enseignement professionnel	Enseignants
370.7 Enseignants – Formation	Enseignants – Formation
371 Écoles	Enseignement professionnel
371.042 Écoles à la maison	Enseignement secondaire
371.1 Enseignants	Enseignement supérieur,

* <http://www.nlc-bnc.ca/caninfo/fcaninfo.htm> [consulté le 30 septembre 2000].

sera amené à l'indice 371.042 (figure 7), sous lequel il trouvera une liste de ressources pertinentes sur ce sujet.

Ce type d'organisation rappelle celle du catalogue classifié, considéré par plusieurs comme l'instrument le plus logique et le plus efficace d'accès thématique aux collections documentaires.

Des enseignements réciproques

Les nouveaux schémas d'organisation des ressources d'information n'ont pas encore démontré clairement qu'ils sont plus faciles à exploiter et plus puissants que les schémas traditionnels. Mais les défenseurs des schémas traditionnels doivent aussi prouver que leurs structures peuvent s'adapter à l'environnement virtuel, qu'elles restent plus efficaces que leurs concurrentes et surtout que

leur usage se justifie autrement que pour la simple raison qu'on les utilise déjà depuis plus de cent ans dans des types de bibliothèques déjà très diversifiés. L'expansion de l'univers virtuel suscite de nouvelles opportunités pour des structures d'organisation dynamiques et capables de s'ajuster à cette évolution rapide du monde de l'information. Il existe déjà des milliers de portails et de bibliothèques virtuelles sur le Web, et nous avons enfin la chance de voir comment les spécialistes d'un domaine qui s'improvisent classificateurs structurent le savoir et organisent les ressources documentaires de leur spécialité. Certains des nouveaux schémas d'organisation valent la peine qu'on les regarde de près : ils pourraient nous aider à rajeunir la terminologie et à redonner de la flexibilité structurelle à nos schémas traditionnels.

Octobre 2000