

**ORGANISATION DES CONNAISSANCES
ET DES RESSOURCES DOCUMENTAIRES**

*De l'organisation hiérarchique centralisée
à l'organisation sociale distribuée*

MICHÈLE HUDON

WIDAD MUSTAFA EL HADI

*We invest so much time in making sure our world isn't miscellaneous in part because disorder is inefficient, but also because it feels bad*¹ (Weinberger, 2007, 12)

Ce numéro thématique des *Cahiers du numérique* consacré à l'organisation des connaissances présente quelques modèles et systèmes d'organisation du savoir et de la documentation parmi les plus courants, qu'ils soient traditionnels ou novateurs. Il propose des éléments de réponse à trois questions fondamentales :

1. D'où venons-nous ?
2. Où en sommes-nous ?
3. Où allons-nous ?

Sous l'influence des réseaux sociaux et participatifs en pleine expansion, devons-nous bientôt constater la désuétude de tout processus d'organisation systématique des connaissances et de la documentation tel que mené depuis des générations ? Saurons-nous au contraire tirer profit de la tradition et de la systématisation pour créer des modèles et des outils d'organisation capables de s'ajuster aux besoins en constante évolution et à un appareillage technologique de plus en plus imposant ?

En guise d'introduction, le présent article décrit brièvement le processus d'organisation des connaissances avant de lier ce dernier plus directement à l'organisation des documents et de l'information dans les milieux traditionnels (bibliothèques surtout) et dans l'environnement numérique contemporain.

Instinct, organisation et classification

De tout temps, l'homme a cherché à organiser son environnement, condition essentielle à ses apprentissages, à son évolution et à sa survie. C'est généralement sans même s'en rendre compte qu'il manipule simultanément plusieurs principes et plusieurs modes d'organisation en les adaptant à son environnement, à un événement ou à un problème. Confronté quotidiennement à une multitude d'objets, d'individus et de situations jamais encore rencontrés, l'humain peut les reconnaître et réagir adéquatement en analysant leurs caractéristiques et en les comparant à ce qu'il connaît déjà.

1. « Nous investissons beaucoup de temps à nous assurer que notre monde n'est pas désordonné en partie parce que le désordre est inefficace, mais aussi parce que le désordre réduit notre sensation de bien-être » (trad. M. Hudon).

Bien qu'il puisse s'actualiser sous plusieurs formes, le processus d'organisation suppose presque toujours une opération de classification. On ne saurait malgré tout considérer comme équivalents et interchangeable l'organisation et la classification, la classification n'étant qu'une méthode d'organisation parmi d'autres. La classification est certainement l'une des méthodes les plus raffinées de segmentation de la réalité, de mise en ordre et de production de cadres de référence utiles ; tout processus d'analyse, de déduction, d'apprentissage et de mémorisation en fait usage. La classification est également inhérente au processus de définition, lequel consiste à déterminer à la fois ce qu'est une entité et ce qu'elle n'est pas, une distinction nécessaire à la compréhension du monde dans lequel nous vivons et à la communication de ce que nous savons.

La classification est l'opération qui consiste à organiser des entités en classes, de sorte que les entités semblables ou parentes soient regroupées et séparées des entités non semblables ou étrangères. La classe est donc un ensemble d'entités (objets physiques ou virtuels, individus, attributs, etc.) présentant une ou plusieurs caractéristiques communes. Les classes les plus générales (par exemple, « Êtres vivants », « Édifices », « Aliments », etc.) regroupent un très grand nombre d'entités qui ne partagent qu'un nombre restreint de caractéristiques essentielles ; les classes les plus spécifiques (par exemple « Mères françaises âgées de 43 ans habitant la banlieue parisienne », « Hôtels de ville de la Montérégie », « Pamplemousses à chair rose cultivés sur la côte ouest de la Floride », etc.) regroupent un petit nombre d'entités ayant en commun un grand nombre de caractéristiques essentielles.

La classification engage deux principes fondamentaux : le regroupement (*grouping*) et le classement (*ordering*). L'acte de regroupement consiste à rapprocher les entités qui possèdent au moins une caractéristique commune qu'on appelle parfois la caractéristique de division. La deuxième étape du processus est moins instinctive ; elle exige en effet la mise en séquence des groupes ainsi constitués. Les difficultés sont évidentes : puisque chaque entité possède nécessairement plusieurs caractéristiques essentielles, elle appartient potentiellement et simultanément à plusieurs groupes distincts : tout être humain appartient, par exemple, à plusieurs groupes selon que le principe de regroupement privilégié soit le sexe, l'âge, la nationalité, la profession, etc. De plus, puisqu'existe une variété de relations possibles entre divers groupes, l'idée de reconstituer un classement idéal, absolu, relève au mieux de l'utopie. Il existe bien évidemment plusieurs façons correctes de classer, ce qui ne signifie pas, loin s'en faut, que toutes les classifications soient appropriées à tous les environnements ou capables de répondre à tous les besoins de toutes les clientèles. La classification est un acte cognitif à finalité pragmatique qui doit

permettre d'identifier, de reconnaître, d'interpréter, de comparer, de déduire et de prédire, souvent dans un environnement particulier.

L'acte de classer, aussi invisible et banal soit-il, n'est pourtant pas sans conséquence. Bowker et Star (1999) considèrent que la classification est au cœur de toute vie sociale et insistent à juste titre sur les dimensions éthiques de l'acte. Le choix d'une caractéristique de division, la création de regroupements et la séquence de présentation de ceux-ci valorisent nécessairement un point de vue particulier et condamnent à l'invisibilité d'autres critères valables d'analyse et de regroupement (Bowker et Star, 1999, 5-6). Il peut en résulter des déductions ou interprétations erronées, et dans les cas extrêmes, une discrimination liée au placement d'une entité ou d'un individu dans un groupe particulier. En outre, le problème des entités, individus et situations jugés inclassables est critique, car si la classification permet de regrouper les entités, elle doit en même temps assurer à chaque entité un positionnement approprié dans la structure d'organisation ainsi constituée.

La classification est donc une façon de voir et de décrire (Kwasnik, 1999). Il n'existe pas de structure d'organisation « naturelle » et toute structure classificatoire ne peut être qu'inventée et construite, jamais découverte. De plus, l'établissement d'un cadre ou d'une structure classificatoire témoigne toujours de la rencontre et du choc d'impératifs philosophiques, technologiques, sociaux, économiques et politiques (Svenonius, 1991). Les structures classificatoires ainsi créées peuvent être de facture simple ou complexe, généreuses ou économes en information de nature descriptive, flexibles ou rigides, ouvertes ou fermées à l'intégration de nouvelles entités. Mais elles ne sont pas en elles-mêmes bonnes ou mauvaises ; c'est en considérant leur pertinence et leur efficacité dans la poursuite d'un objectif prédéfini qu'elles peuvent être évaluées.

Organisation des connaissances et classification documentaire

D'abord limité à la caractérisation de phénomènes observables et au rangement d'objets concrets, le processus d'organisation, réalisé sous forme d'une mise en ordre peu complexe ou au contraire d'une activité classificatoire sophistiquée, s'est rapidement révélé essentiel dans le monde des idées et des connaissances, donnant naissance à divers systèmes dits « philosophiques » d'organisation du savoir. Aristote, Hegel, Bacon, tous ceux qui se sont donné pour mission de réfléchir sur l'être humain et sur le monde dans lequel il vit ont toujours pratiqué volontiers, et de manière subjective, la catégorisation et la classification. La tâche ne pouvait être simple : tout comme il n'existe pas de

structure d'organisation naturelle, il ne peut exister de structure idéale et absolue, d'organisation des connaissances.

L'univers des connaissances est une énigme ; personne ne peut jamais le percevoir dans sa totalité² et Miksa en parle comme d'un concept et non comme d'une réalité (1992, 101). Cet univers abstrait est, tout comme le monde naturel, en constantes transformation et réorganisation. Pour percevoir l'univers des connaissances, pour le « dompter » aussi, il nous faut le rendre concret, lui dessiner des frontières, lui donner des attributs, des composantes distinctes mais reliées les unes aux autres. Constituée au fil des temps et des découvertes, la masse des savoirs doit être organisée pour que nous puissions y retrouver ce qui est nécessaire pour comprendre, pour poser des hypothèses et pour ajouter des connaissances nouvelles à celles qui existent déjà et qui, bien qu'elles puissent tomber en désuétude, continuent d'exister et s'avèrent essentielles à l'interprétation du monde au sein duquel nous vivons.

Connaissances (une entité) et classification (un processus) sont parfois présentées comme étant si intimement liées qu'elles paraissent indissociables. Kwasnik (1999) décrit trois formes possibles d'interaction. L'harmonisation connaissance/classification peut être complète, comme c'est le cas dans le tableau périodique des éléments, par exemple, et générer une structure qui reste pertinente et utile pendant une très longue période. Dans le tableau périodique, le positionnement de chaque élément permet d'en décrire les propriétés et de comprendre les liens qu'il entretient avec les autres éléments. Le tableau périodique décrit ce qui est déjà connu et prédit ce qui y sera éventuellement ajouté, remplissant parfaitement bien sa mission d'information et de communication scientifique (Kwasnik, 1999, 69). Plus souvent, les transformations dans l'univers des connaissances précèdent et rendent nécessaires des transformations identiques dans les structures classificatoires qui en sont le reflet ; dans les cas extrêmes, les structures classificatoires doivent être reconstruites, comme on dut le faire au siècle dernier pour représenter l'importance de l'impact sur le monde scientifique et dans la société toute entière des rapides développements technologiques. Finalement, les chercheurs savent bien que l'acte même de classifier est lui-même générateur de connaissances, puisqu'il permet de visualiser des liens réels ou potentiels entre entités et entre phénomènes qui n'auraient pu autrement être perçus.

Les modèles philosophiques d'organisation des connaissances ont à leur tour inspiré les gestionnaires d'information et de ressources documentaires en

2. Nous paraphrasons ici Jean-Paul Riopelle (peintre Québécois, 1923-2002) dont nous retenons la phrase clé suivante : « La nature reste une énigme : on ne la perçoit jamais dans sa totalité » (1993).

bibliothèques, dans les services de documentation, dans les centres d'archives et dans les institutions muséales, lorsqu'il s'est agi d'organiser des collections de documents pour en faciliter l'exploitation. Les structures d'organisation des connaissances les plus courantes ont d'abord été reproduites avec plus ou moins de succès dans des schémas de classification encyclopédiques, outils longtemps jugés essentiels aux opérations de classification, de repérage et de localisation documentaires.

Le processus de classification documentaire

Dans la classification documentaire, ce sont les documents qui constituent les entités à regrouper. La classification documentaire est l'opération qui consiste à réunir en classes les documents liés par une caractéristique intrinsèque (par exemple, le sujet dont elles traitent) ou extrinsèque (par exemple, leur format), en les séparant des documents avec lesquels ils n'ont rien en commun et n'entretiennent aucun lien.

Dès l'Antiquité, la nécessité de disposer d'un système qui permettrait de repérer et de localiser un document particulier se fit sentir à partir du moment où un grand nombre de documents devenaient disponibles dans un même lieu, dans une même collection. L'organisation systématique des collections fut longtemps basée sur des critères physiques, tels la forme et le matériau, bien qu'on puisse observer à toutes les époques des tentatives de segmentation en grandes catégories représentant des sujets ou des genres (histoire, poésie, mythologie, etc.). L'organisation des documents en fonction de ce dont ils traitent plutôt que de ce à quoi ils ressemblent a donc toujours été vue plus ou moins clairement comme un objectif à atteindre, mais il fallut attendre la Renaissance pour que se répandent les systèmes d'organisation des collections basés sur d'autres critères que la forme, la couleur, le nom de l'auteur ou la date d'acquisition.

À partir du XVI^e siècle, le classement des documents en fonction des sujets dont ils traitent s'implante progressivement, mais demeure une entreprise individuelle, rarement soumise à validation externe ou à normalisation. L'invention de l'imprimerie et la multiplication des ouvrages à organiser en collections et en sous-collections rendent impératif le développement de systèmes d'organisation plus performants, aptes à servir des utilisateurs autres que leurs seuls créateurs. Trois initiatives influenceront plus particulièrement l'évolution de la classification documentaire. Konrad Gessner (1516-1565), pionnier de la recherche bibliographique, organise le savoir en 21 classes distinctes, une division que l'on trouve encore aujourd'hui dans quelques grands outils de classification. Sir Francis Bacon (1561-1626) propose dans sa classification des sciences (1605) une segmentation des collections en œuvres de

mémoire, d'imagination et de raison ; cette segmentation sera reprise deux siècles plus tard par Thomas Jefferson d'abord, par Melvil Dewey ensuite. Et chez Gabriel Naudé (1600-1653) se dessinent déjà les divisions disciplinaires qui structurent aujourd'hui les systèmes encyclopédiques de classification documentaire : histoire, sciences humaines, droit, médecine, mathématiques, philosophie et théologie.

Au début du XIX^e siècle, une multitude de systèmes de classification développés localement coexistent encore. Ces systèmes ont en commun d'être peu accessibles à l'utilisateur des collections documentaires et c'est en réaction à ce fait que Melvil Dewey (1851-1931) explore la possibilité de simplifier et de systématiser le rangement des documents. Dewey implante la classification « relative », plaçant tout document d'une collection en relation non seulement avec celui qui le précède et avec celui qui le suit immédiatement, mais également avec tous les autres ouvrages faisant partie de la collection. Dewey développe ainsi l'idée d'un contemporain, William T. Harris, de classifier non pas le livre mais plutôt son contenu, donc les éléments d'information et de connaissance qu'il véhicule. Reprenant en les inversant les trois classes proposées par Bacon en 1605, il reproduit l'organisation des connaissances en vigueur dans la deuxième moitié du XIX^e siècle. Les disciplines liées à la raison y occupent la plus grande place, se répartissant en cinq grandes classes : philosophie, théologie, sciences sociales, linguistique, sciences pures. L'imagination se voit consacrer trois classes : sciences appliquées et technologie, beaux-arts et musique, littérature. La géographie et l'histoire occupent l'unique classe consacrée à la mémoire. L'indice de classification est conçu de telle sorte qu'il renvoie non seulement à la localisation exacte du document physique, mais également à la place et à l'importance relative du sujet dont il traite dans l'univers des connaissances.

Jusque-là, la classification des documents sur le modèle d'une structure conventionnelle d'organisation du savoir, poursuivait un double objectif : d'une part, l'éducation des masses par l'utilisation d'une carte de navigation des connaissances accumulées au fil des temps et, d'autre part, le repérage et la localisation de documents susceptibles de fournir la réponse à une question précise. Le deuxième objectif se révéla beaucoup plus difficile à atteindre, le nombre et la variété des questions potentielles dépassant de loin le nombre et la variété des constituantes de l'univers du savoir que les plus érudits purent longtemps se targuer de connaître. L'exercice devint de plus en plus complexe et l'on put observer dans la seconde partie du XX^e siècle une transformation rapide du processus de classification documentaire et un réalignement de ses objectifs, le second prenant nettement le pas sur le premier (Miksa, 1991, 117) ; ce deuxième objectif oriente aujourd'hui l'ensemble des opérations de description et de

recherche effectuées sur l'information et la documentation et est considéré comme une constituante fondamentale des sciences de l'information.

Une autre tendance, plus marquée en Amérique du nord qu'en Europe il est vrai, allait éloigner encore davantage la classification de son objectif premier. Pour pouvoir mieux rendre compte de l'évolution scientifique et sociale rapide de la fin du XIX^e et du début du XX^e siècles, et pour assurer aux structures classificatoires la profondeur requise afin de satisfaire des besoins très précis, les classificateurs se tournèrent vers les documents eux-mêmes plutôt que vers les modèles d'organisation des connaissances existants pour identifier les nouveaux sujets, leur trouvant ensuite une place, plus ou moins appropriée, dans la structure classificatoire en usage. Le sujet en vint à être considéré comme un attribut du document, au même titre que sa forme ou le titre de la collection à laquelle il appartient, plutôt que comme un élément se rattachant logiquement à l'univers des connaissances (Miksa, 1991, 111) ; c'est ainsi qu'au XXI^e siècle, on propose désormais des interfaces d'interrogation présentant le sujet (et les termes qui le décrivent) comme une « facette » du document.

Les schémas de classification documentaires

La classification des documents se fait à l'aide de structures classificatoires, que nous appelons schémas de classification documentaire, dont les contenus, la portée et l'organisation interne peuvent varier considérablement. Certaines structures reflètent l'organisation de la totalité de l'univers des connaissances : on les qualifie d'encyclopédiques. D'autres se limitent à la mise en ordre logique des sujets les plus courants dans un domaine d'activités particulier : ce sont des structures spécialisées. D'autres encore, les schémas de classification utilitaires, ne servent qu'à l'organisation de ressources constituant une collection particulière.

Le schéma de classification documentaire est un inventaire structuré de sujets organisés logiquement en domaines, sous-domaines, disciplines, sous-disciplines, et ainsi de suite. Cette organisation logique se reflète dans une notation plus ou moins signifiante, laquelle, en plus de représenter un sujet traité de façon générale ou spécifique, indique également dans quel contexte et dans quelle perspective il a été traité. Le système de classification est un outil essentiel à l'atteinte de trois objectifs des milieux documentaires : dans sa fonction cognitive, il permet l'organisation des concepts, des idées et des sujets ; dans sa fonction bibliographique, il permet la systématisation des représentations documentaires ; dans sa fonction bibliothéconomique, il permet l'organisation des documents-objets eux-mêmes (Satija, 1997). Notons que le rangement physique des documents, lorsqu'il permet une exploration logique et

même pédagogique de la collection, contribue à la réalisation de la mission d'éducation des bibliothèques, ces universités du peuple.

La majorité des schémas de classification documentaire se caractérisent par leur structure hiérarchique basée sur l'inclusion de sujets spécifiques à des sujets plus généraux. Cette organisation hiérarchique repose sur trois types de relations entre entités : la relation d'hypéronymie (de genre à espèce), la relation méronymie (de tout à partie) et la relation d'instance (de classe à entité unique) sont perçues, comprises et interprétables par l'esprit humain. Le schéma de classification documentaire ne peut jouer son rôle que s'il est complet (c'est-à-dire qu'il intègre tous les sujets qui ont été abordés ou pourraient éventuellement être abordés dans un ou plusieurs documents), flexible (c'est-à-dire qu'il s'ajuste aux besoins et aux habitudes de recherche des utilisateurs, au volume des collections à classer, à la rapidité d'évolution du domaine d'activités représenté, aux fonctionnalités informatiques disponibles, etc.) et hospitalier (c'est-à-dire qu'il autorise l'intégration à la structure existante, sans qu'il soit nécessaire de la modifier ou de la repenser, des sujets nouveaux qu'engendrent les développements disciplinaires et l'évolution des savoirs).

L'utilisation de schémas de classification modelés sur une structure conventionnelle d'organisation des connaissances présente plusieurs avantages. Lorsqu'ils sont encyclopédiques, ces schémas reflètent l'organisation des disciplines enseignées aux niveaux supérieurs. Ce parti pris est fréquemment critiqué mais convenons avec Maniez que, puisque « toute partition ou division rationnelle s'opère à partir d'un point de vue [...] mieux vaut partir du point de vue le plus universel » (2002, 262) ; la structure disciplinaire est un cadre familier à la majorité des chercheurs d'information.

Les schémas de classification documentaire les plus courants offrent des niveaux de spécificité incomparables pour la représentation des sujets. La terminologie, la définition des concepts, la syntaxe de combinaison des concepts pour former des sujets et la structure relationnelle liant les sujets entre eux y sont contrôlées, cohérentes et prévisibles. C'est le cas notamment dans le système de classification des documents établi en 1873 par Melvil Dewey, développé et perfectionné durant le 20^e siècle par une succession de rédacteurs issus de divers milieux, et désormais utilisé sur cinq continents et dans plus de 130 pays ; sans en avoir le statut officiel, la classification décimale de Dewey (CDD) en est venue à jouer un rôle de norme internationale pour l'organisation des collections documentaires physiques et virtuelles, permettant de ce fait l'interopérabilité des systèmes d'exploitation de ces collections ; lorsque plusieurs institutions utilisent un même schéma de classification, l'accès à leurs collections respectives en est nécessairement facilité, puisque toutes les ressources traitant d'un même sujet sont repérées à l'aide d'une seule stratégie

de recherche établie à partir des termes ou de la notation proposés pour le représenter (mai, 2003). Deux autres schémas de classification, la classification décimale universelle (CDU) et la classification de la bibliothèque du congrès (LCC) sont également très largement utilisés, le premier en Europe et en Afrique surtout, le second presque exclusivement en Amérique du nord dans les bibliothèques universitaires et de recherche.

La classification décimale universelle, conçue par Paul Otlet et Henri de Lafontaine à la fin du XIX^e siècle, avait comme fonction principale de structurer le contenu de la bibliographie universelle à la constitution de laquelle ils venaient de s'attaquer, et non pas de servir au rangement d'une collection de documents. Dérivée de la 5^e édition de la CDD dont elle reprit en les enrichissant et en les normalisant la structure de base, la notation et les tables de subdivisions communes, la CDU est une structure classificatoire des plus intéressantes tant au plan de sa philosophie qu'à celui de ses applications potentielles. Les classes énumérées sont des éléments de construction avec lesquels on peut former des classes composées et plus complexes, ce qui permet l'expression de sujets par des combinaisons non prévues par les concepteurs du système. La souplesse de la CDU est l'une de ses caractéristiques les plus intéressantes : on y appliquait déjà les principes de l'analyse par facettes bien avant leur formalisation par Ranganathan. La CDU permet les modifications de l'ordre de citation des parties de l'indice selon les exigences locales, prévoit des accès multiples et propose des mécanismes qui permettent son adaptation à des circonstances particulières. Les concepts étant représentés par une notation symbolique, laquelle demeure constante dans ses diverses versions linguistiques, la CDU est apte à jouer, comme la CDD d'ailleurs, un rôle de langage de commutation. Pour des raisons qui ont peu à voir avec sa logique et son efficacité cependant, la présence globale de la CDU dans les milieux documentaires et sur les réseaux d'information se fait malheureusement de plus en plus discrète.

Ces schémas de classification n'ont pas que des qualités et ne présentent pas que des avantages ; la littérature des sciences de l'information s'est d'ailleurs beaucoup intéressée récemment à leurs faiblesses, réelles ou imaginées. Au fil des décennies, le nombre de structures classificatoires encyclopédiques distinctes en usage dans les institutions documentaires s'est considérablement réduit. La transformation en norme de quelques systèmes de classification, la CDD et la CDU notamment, ont eu pour effet certain d'engendrer conservatisme et inertie (De Grolier, 1988). On reproche aux classifications documentaires encyclopédiques de refléter trop fidèlement une vision du monde datant du XIX^e siècle, occidentale, de l'hémisphère nord et souvent même proprement états-unienne. Les grandes divisions disciplinaires n'y sont

plus appropriées et certains domaines, l'informatique et les études environnementales notamment, y sont de fait peu ou très mal représentés ; c'est le cas également pour tous les domaines inter- et multidisciplinaires qui ne peuvent logiquement s'intégrer à la macrostructure existante. Leurs structures hiérarchiques unidimensionnelles peu flexibles ne permettent pas de réorganiser le savoir selon une perspective différente de celle qui a été imposée par le créateur ou le gestionnaire du système. On considère les schémas de classification comme des outils de travail lourds et complexes, dont la gestion et l'utilisation coûtent cher, et qu'il est difficile d'exploiter de manière optimale.

Sa grande visibilité expose évidemment la classification décimale de Dewey aux critiques les plus virulentes. On s'attaque à sa macrostructure (la définition et l'ordre des grandes classes), à l'organisation de domaines du savoir particuliers, la littérature ou la médecine, par exemple, et même à la localisation de sujets précis (l'anatomie humaine, par exemple, que l'on doit chercher en médecine, dans la classe technologie, plutôt qu'en biologie, avec les autres sciences naturelles). La structure de la CDD se prête mal à l'intégration de sujets inter- ou multidisciplinaires et le contexte au sein duquel le système a été conçu se reflète, encore aujourd'hui, dans le traitement accordé aux sujets qui concernent les femmes et les groupes qualifiés de « minoritaires » de par leur orientation sexuelle, leur origine ethnique ou leur statut politico-socio-économique. Devant l'explosion de l'utilisation de la CDD, en particulier au sein de réseaux d'information multinationaux, un effort particulier est fait par les rédacteurs pour rendre le système plus convivial et plus représentatif de l'organisation du savoir dans les cultures autres qu'américaines, et même autres qu'occidentales, mais on doit s'attendre à ce que le processus d'universalisation de la CDD soit très long. La tâche n'est pas simple. Les domaines du droit, de l'éducation et de l'histoire, notamment, sont particulièrement difficiles à internationaliser ; le système d'options proposées pour accommoder les différences culturelles et donner plus d'importance à un sujet d'intérêt local n'est ni entièrement satisfaisant, ni efficace.

Dans tous les domaines, l'adoption d'un produit, d'une méthode, d'un système, etc. comme norme ne signifie pas nécessairement que le produit, la méthode, le système qui ont été choisis étaient les meilleurs disponibles. C'est sans doute le cas au domaine de la classification documentaire. Il n'est pas rare de voir des chercheurs d'information et des professionnels de la documentation se rebeller contre une structure classificatoire qui ne répond pas à leurs besoins. Ils se tournent alors vers des systèmes spécialisés (ne couvrant que quelques sujets par exemple), utilitaires, simplifiés et parfois même personnalisés, assumant le risque de se détacher du peloton, de s'éloigner de la structure conventionnelle d'organisation des connaissances pour mieux satisfaire des

besoins immédiats et locaux, et réduisant ainsi les occasions d'échanges et de communication d'information à plus grande échelle. Ces systèmes d'utilisation restreinte et très marqués par la temporalité sont condamnés à brève échéance (Bowker et Star, 1999, 15) : qui connaît encore, autrement que par certains détails historiques, les nombreux systèmes de classification qui coexistaient au début du XIX^e siècle ?

Ailleurs, et particulièrement dans les nouveaux environnements virtuels, on abandonne la classification et ses outils au profit de modes d'organisation plus souples et plus immédiatement familiers aux chercheurs d'information. L'ordre alphabétique remplace souvent la mise en ordre logique d'un ensemble de ressources documentaires. Or, l'ordre alphabétique, certes facile à maintenir et mécaniquement réalisable, favorise davantage les intérêts du propriétaire de la collection ou du gestionnaire d'une base de données que ceux de ses clients. L'efficacité de la méthode alphabétique est trop dépendante des aléas de la terminologie et des problèmes causés par la synonymie et la polysémie, même au sein d'une seule langue naturelle ; elle est inapplicable là où plusieurs langues naturelles coexistent. L'ordre alphabétique ne peut assurer à lui seul l'efficacité de la navigation dans une collection et l'efficacité des résultats d'une recherche d'information sur un sujet particulier, sauf dans les cas où il est combiné à une structuration logique et révélatrice des liens entre concepts et entre sujets, comme c'est le cas dans les répertoires de vedettes-matière ou les thésaurus de descripteurs.

Le débat opposant les partisans de l'organisation logique aux partisans de l'accès alphabétique est maintenant centenaire. Les premiers affirment qu'une mise en ordre logique, surtout si elle est normalisée, reste accessible puisqu'elle est instinctive et a de grandes chances de correspondre à l'ordre logique établi par l'utilisateur pour traiter et mobiliser les connaissances qu'il a lui-même emmagasinées. Les seconds croient au contraire que l'utilisateur non averti ne peut deviner quelles caractéristiques de division ont été choisies, quel point de vue a été adopté, quels types de relations structurelles ont été privilégiées (Svenonius, 1992, 8).

Puissance et limites de la hiérarchie

Maniez parle de la hiérarchie comme du « socle naturel de l'organisation des connaissances » (2002, 278). Si le recours à une structure hiérarchique n'est pas essentiel au succès d'une opération de classification, on constate toutefois qu'une majorité de systèmes d'organisation des connaissances philosophiques, médicales, sociales et bibliographiques favorisent la hiérarchie plutôt que toute autre forme de mise en ordre logique. La classification hiérarchique est la forme

la plus complexe de mise en ordre des composantes de l'univers des connaissances, et par extension des ressources documentaires qui les décrivent. La structure hiérarchique, utilisée déjà par Aristote sous sa forme la plus rigide, est particulièrement efficace et utile dans la poursuite d'objectifs d'éducation et de compréhension.

Lorsque des entités d'abord, des classes ensuite, sont organisées de façon hiérarchique « pure », selon le modèle aristotélicien de la division logique, chaque entité ou chaque classe représente nécessairement un type particulier ou une instance particulière de celle qui la précède immédiatement (comme dans la série Fruits > Agrumes > Pamplemousses (*Citrus x paradisi*) > Pamplemousses roses > Pamplemousses roses de la Floride). La structure fournit alors de façon économique, claire et efficace un ensemble de renseignements sur une entité ou une classe, en fonction de la localisation de celle-ci dans la structure et de sa relation à ce qui l'entoure. Puisque la relation générique (hyperonymique) est la seule qui soit utilisée, il devient possible d'affirmer que tout ce qui est vrai de l'entité de niveau supérieur l'est également des entités de niveaux inférieurs qui lui sont directement rattachées ; ce principe d'héritage implique que ce qui est vrai de l'agrume l'est aussi du citron, de l'orange et du pamplemousse, les caractéristiques essentielles de l'entité inclusive étant nécessairement présentes chez les entités incluses. La structure hiérarchique pure est donc un formidable outil d'organisation, à condition que les critères différenciateurs et l'ordre de succession de ces critères soient choisis avec pertinence (Maniez, 2002, 235-239). Selon Kwasnik, ce modèle d'organisation est le plus à même de favoriser la découverte, la compréhension et même l'extension d'un champ de la connaissance puisqu'elle contribue à la consolidation de l'existant et révèle les vides à combler (1999, 70).

Or, s'il est relativement aisé de créer des sous-ensembles dans un ensemble d'objets concrets (par exemple, des ustensiles de cuisine, des blocs LEGO, ou les instruments de travail du charpentier), il s'avère beaucoup plus difficile de structurer des concepts abstraits, souvent mal définis et dont on connaît rarement les caractéristiques essentielles.

Aussi puissant et efficace soit-il, le modèle hiérarchique générique ne présente pas que des avantages. Il est rigide et n'admet qu'une seule façon de décrire et de classer une entité, niant la possibilité pour un objet ou un concept d'appartenir simultanément à plusieurs classes. En outre, il convient davantage à la structuration de domaines de la connaissance établis qu'à l'organisation de domaines en émergence dont les bases conceptuelles et terminologiques sont encore instables.

La classe fut longtemps considérée comme une entité à frontières non poreuses et au contenu bien défini, comme si une structure d'organisation en classes préexistait dans la nature, ce qui n'est pas le cas. La définition de ce qu'est une classe s'est considérablement assouplie au XX^e siècle, sous l'influence de travaux portant sur la relativité sémantique réalisés notamment en anthropologie et en linguistique. Il est maintenant courant de définir la classe comme un regroupement à contours perméables et au contenu variable en fonction d'un contexte ou d'un point de vue. Par ailleurs, la complexité croissante de l'univers des connaissances ne permet plus d'affirmer avec certitude, comme le faisaient Aristote et ses disciples, qu'une entité appartient toujours ou jamais, et sans conteste, à une classe particulière.

Cette souplesse introduite dans l'organisation en classes peut sembler compliquer la tâche d'organiser l'univers des connaissances, mais elle ouvre des perspectives intéressantes. Il devient possible d'imaginer d'autres formes d'organisation hiérarchique, l'arborescence (*tree structure*) par exemple, moins puissantes sur le plan de la communication d'information mais davantage adaptables à un environnement et à des besoins particuliers.

Contrairement à la structure hiérarchique pure, l'arborescence ne repose pas sur l'unique relation de genre à espèce entre entités. La relation générique est utilisée à l'occasion dans l'arborescence, mais elle y côtoie d'autres principes de division et d'autres relations, notamment la relation partitive (méronymique), la relation causale, la relation sociale ou administrative conventionnelle (par exemple, les grades dans une armée ou la structure d'une administration publique) et la relation basée sur une caractéristique non essentielle de l'entité à classer (par exemple, la profession exercée, pour créer des sous-ensembles de personnes). Dans une arborescence, les caractéristiques de la classe inclusive ne sont pas automatiquement transmises aux classes incluses ; la roue et la bicyclette, deux entités distinctes liées par une relation partitive, ne partagent ni caractéristiques essentielles, ni même accidentelles. Par le fait même, l'arborescence est beaucoup moins riche sur le plan sémantique que la hiérarchie pure et elle ne permet pas de définir et d'inférer ; la relation entre roue et bicyclette est évidente certes, mais elle offre bien peu d'information sur ce qu'est une roue, sur ce qu'est une bicyclette, et sur ce que sont les entités que l'on trouvera au même niveau de la hiérarchie que la roue, le guidon et la selle, par exemple.

Nettement plus souple que la structure hiérarchique générique et présentée comme une alternative intéressante permettant la représentation de points de vue différents au sein d'une même structure classificatoire, l'arborescence n'est pas sans faiblesses. Dans un domaine de la connaissance particulier, l'arborescence ne présente qu'une seule perspective, au risque de masquer d'autres relations intéressantes entre entités, d'autres façons d'ordonner les

composantes d'un univers. Tout comme la structure hiérarchique pure, elle ne laisse que peu de place à la polyhiérarchie, c'est-à-dire à la multiplication des possibilités de navigation logique entre une entité spécifique et l'entité inclusive qui occupe la tête d'une hiérarchie. En se développant, l'arborescence tend à devenir de plus en plus rigide, forçant l'intégration d'entités récemment créées ou de concepts nouvellement définis dans des classes existantes pour éviter un travail de restructuration.

La mise en ordre ponctuelle, en réponse à un besoin précis, est une autre forme d'organisation facilement réalisable et relativement efficace qu'autorise également l'élargissement de la définition de la notion de classe. Elle résulte en la constitution de sous-ensembles d'entités n'ayant pas d'attributs essentiels en commun mais que lie temporairement un contexte, un événement, etc. (par exemple, le sous-ensemble des objets à mettre dans une valise avant un départ en croisière, ou l'ensemble des instruments à disposer sur un plateau stérile avant une intervention chirurgicale). En sciences de l'information, on peut associer à ce modèle souple d'organisation des connaissances les regroupements documentaires basés sur les « centres d'intérêts », présents en bibliothèques municipales surtout, et de plus en plus sur le web. Dans cette structure, les classes à contenu plus ou moins fixe et à durée de vie limitée sont d'utilité immédiate et certaine, mais trop pauvres sur le plan sémantique et trop subjectives pour constituer un modèle d'organisation stable.

Une alternative : l'organisation à base de facettes

En 2001, un groupe d'architectes de l'information chargés d'organiser les éléments de contenu d'un site web et d'améliorer l'accès à une base de connaissances en milieu corporatif adopte un modèle de mise en ordre conçu 70 ans plus tôt par le mathématicien et bibliothécaire indien S.R. Ranganathan : l'analyse, la description, et la classification à base de facettes (La Barre, 2007, 129). Ce modèle d'organisation alternatif permet de se libérer de structures hiérarchiques encombrantes et toujours contestées, tout en préservant la logique de navigation dans un domaine de la connaissance ou une collection de ressources documentaires. C'est ainsi que les facettes sortent des laboratoires où elles ont été manipulées jusque-là par quelques chercheurs et deviennent très populaires dans les catalogues et les bases de données numériques ainsi que sur le web, où on les utilisera désormais sans vraiment se préoccuper des fondements théoriques, philosophiques et bibliothéconomiques qui en justifient l'application.

Il n'existe pas de réalité, de situation, d'objet, de concept ou de sujet qui ne soit multidimensionnel. Les concepteurs des modèles hiérarchiques d'organisation des connaissances n'ont jamais nié ce phénomène qu'ils n'ont

pu, ou n'ont pas voulu, refléter dans les structures classificatoires traditionnelles. Ranganathan a osé rompre avec la tradition en tenant compte de facteurs tels la dépendance conceptuelle et la multiplicité des points de vue d'abord, et en installant ensuite le point de vue, ou la facette, à la base et au cœur du modèle d'organisation des connaissances et du schéma de classification proposé en 1933 à la communauté bibliothéconomique de l'époque sous le nom de classification Colon (CC)³. Ranganathan reconnaissait ainsi l'existence et la validité de points de vue distincts lorsqu'il s'agit d'organiser l'univers des connaissances, en même temps que l'inexistence d'un modèle d'organisation idéal, valide en toutes circonstances.

Seule structure classificatoire qui puisse être qualifiée d'analytico-synthétique, la CC est le dernier des schémas encyclopédiques à avoir vu le jour. L'exploitation de la CC exige une analyse en deux étapes : la déconstruction d'un sujet en ses diverses composantes à partir de ses différentes facettes (étape analytique) et la génération à l'aide de règles syntaxiques complexes d'un indice significatif intégrant les composantes du sujet (étape de la synthèse). Contrairement aux schémas hiérarchiques décrits dans les sections précédentes, la CC n'est pas énumérative et ne propose aucun indice de classification préconstruit. Selon Ranganathan, cinq facettes sont nécessaires et suffisantes à l'analyse et à la représentation de tout sujet : la personnalité (P) ou l'essence du sujet (par exemple « Barres » dans Moulage de barres d'acier) ; la matière (M) (par exemple « Acier » dans Moulage de barres d'acier) ; l'énergie (E) ou l'action décrite (par exemple « Moulage » dans Moulage de barres d'acier), l'espace (S) et le temps (T).

La classification Colon est un instrument complexe dont l'utilisation exige des compétences que bien peu de professionnels de l'information possèdent. En outre, la CC ne peut remplir la fonction bibliothéconomique (de rangement) attribuée à la classification documentaire. Pour ces raisons, elle n'a jamais été utilisée que dans quelques institutions spécialisées, en Inde surtout. Mais le modèle d'organisation imaginé par Ranganathan continue d'être admiré pour la richesse et la précision des représentations qu'il permet de générer, pour sa grande souplesse et surtout pour la possibilité qu'il offre de représenter un domaine de la connaissance de façon différente selon le point de vue adopté, l'objectif poursuivi, l'auditoire auquel on s'adresse, etc. Svenonius suggère que l'innovation chez Ranganathan a consisté à appliquer une approche-système et à adopter un point de vue structuraliste dans ses réflexions sur l'organisation du savoir et des documents (2000, 14).

3. Le nom de la classification vient de l'utilisation qu'elle fait du point-virgule (en anglais *colon*) comme indicateur de facette Énergie (E), la plus importante selon Ranganathan.

Les cinq facettes proposées par Ranganathan ont été jugées insuffisantes et inappropriées dans certains domaines de la connaissance (les arts et les sciences sociales notamment) ; de plus, la difficulté de définir de manière fonctionnelle la facette « personnalité » a certes contribué à réduire l'intérêt pour la CC. Mais la pratique de l'analyse et de la représentation d'un sujet à l'aide de facettes prédéfinies a trouvé un environnement idéal d'application lorsque la technologie a permis le développement de systèmes d'information dynamiques, ouverts, interreliés et interopérables. L'explosion de la production de ressources numériques et le développement de collections virtuelles, au sein desquelles il n'est plus besoin d'assigner une ressource à une seule classe pour en permettre le rangement, ont propulsé à l'avant-plan l'organisation à base de facettes, exploitable par les moteurs de recherche. Si la CC elle-même est en passe de devenir un objet de recherche plutôt qu'un outil de travail, les principes inhérents à la classification à facettes proposés par Ranganathan sont de plus en plus présents sur le web. Quelques administrations publiques, par exemple, font usage de facettes pour structurer leurs intranets ; on y accède à des listes (souvent hiérarchisées) de Processus d'affaires (par exemple, Gestion du personnel), d'Activités (par exemple, Recrutement), d'Agents (par exemple, Conseiller au recrutement), de Sujets (par exemple, Concours), de lieux géographiques (par exemple, Québec), de données temporelles (par exemple, 2006), de Types de document (par exemple, Lettre), etc. auxquelles on a recours pour affiner une requête ou localiser un document particulier. On note aussi l'utilisation de « facettes » telles Auteur, Langue, Sujet, Date, Collection, etc. dans les interfaces d'accès aux catalogues de bibliothèques⁴ ; ces facettes y facilitent la poursuite de la navigation suite à l'obtention d'un premier ensemble de résultats.

Lorsqu'on la compare aux modèles hiérarchiques traditionnels, l'organisation à base de facettes paraît particulièrement avantageuse : 1) elle établit des liens ponctuels, fonctionnels et pragmatiques entre concepts et/ou sujets, plutôt que des relations strictement taxinomiques permanentes ; 2) les composantes de la structure restent indépendantes les unes des autres jusqu'au moment de la synthèse d'un sujet et de l'indice qui le représente ; 3) il n'existe théoriquement aucune limite quant au nombre et à la spécificité des sujets qui peuvent être représentés ; 4) la structure classificatoire est lâche et expansible à l'infini ; 5) l'ordre de présentation des facettes d'un sujet est ajustable au contexte et aux besoins ; 6) le modèle est efficace dans les disciplines en émergence et la structure de référence (définissant les facettes, énumérant les concepts et proposant les termes) peut être établie au fur et à mesure que se développe une collection de ressources ; 7) l'évolution des connaissances, la

4. Voir par exemple, ariane2.bibl.ulaval.ca/, et aqua.esc-lille.eu/

croissance ou au contraire la diminution de l'importance accordée à un domaine ou à un sujet, etc., ont peu d'impact sur les composantes de base du système et sur sa structure globale.

L'une des faiblesses du modèle tient à la difficulté d'identifier et de définir les facettes les plus appropriées à la description d'un domaine de la connaissance ou d'un champ d'activités. L'ambiguïté entretenue par Ranganathan lui-même quant à la nature et à la fonction de la facette (voir à ce sujet Maniez, 1999) n'est pas étrangère à l'absence de principes directeurs concrets pour le choix, la définition et l'application de facettes. Le discours des spécialistes contemporains révèle des divergences profondes qui se répercutent dans leurs pratiques. Il en résulte des différences importantes dans la structure d'organisation et dans la terminologie utilisée pour accéder à des réservoirs qui offrent le même type d'information et visent les mêmes objectifs (c'est le cas dans les nombreux catalogues de bibliothèques offerts sur le web, par exemple). En favorisant la flexibilité et la diversité des points de vue, le risque s'accroît de mettre de l'avant, là encore, une perspective et une vision locales et de faire obstacle à l'interopérabilité sémantique des systèmes d'information. Et au final, le chercheur d'information ne sera pas nécessairement mieux servi par une organisation à base de facettes que par une organisation hiérarchique stricte.

Ranganathan aurait été le premier à admettre que le modèle d'organisation qu'il proposait ne pourrait répondre à tous les besoins, dans toutes les circonstances ; c'est le lot de tout système d'organisation. Mais parions qu'il aurait été intéressé aussi par l'intégration à deux schémas de classification documentaire plus que centenaires, la CDD et la CDU, des principes de l'organisation sur la base de facettes. De fait, la CDU a toujours autorisé la synthèse d'indices de classification pour représenter des combinaisons de sujets non prévues par les concepteurs du système. Et les deux schémas de classification regroupent depuis longtemps dans des tables auxiliaires les subdivisions, géographiques ou chronologiques par exemple, applicables à l'ensemble des sujets traités et assimilables aux facettes proposées par le maître indien. Il est à prévoir, la technologie aidant, que d'autres éléments de l'analyse par facettes viendront enrichir les structures et accroître la convivialité de ces deux schémas de classification au cours des prochaines années.

On peut imaginer qu'un nouveau modèle d'organisation des connaissances, intégrant de façon intelligente plutôt que strictement mécanique les avantages respectifs du modèle hiérarchique et du modèle à facettes, pourrait améliorer encore l'efficacité de la navigation dans un domaine de la connaissance et dans les immenses collections de ressources numériques accessibles au chercheur d'information.

Organisation et nouveaux environnements informationnels

Sur le web comme en bibliothèque traditionnelle, la recherche par sujet reste la plus difficile et ses résultats sont souvent décevants. Certains gestionnaires du web ont rapidement perçu l'intérêt d'organiser les ressources mises à disposition et proposent dans des répertoires thématiques (par exemple, le Open Directory Project à www.dmoz.org), des portails (par exemple, Science.gouv.fr, le portail de l'Internet scientifique), des encyclopédies (par exemple, Wikipédia) et des bibliothèques virtuelles (par exemple, la Bibliothèque de référence virtuelle à www.bibliovirtuelle.ca/) des structures de navigation plus ou moins flexibles qui permettent d'accéder plus aisément à des ressources ayant des caractéristiques communes. Qu'on les désigne sous le terme de plans de classement ou de classification, listes de catégories ou de « facettes », folksonomies, taxinomies ou parfois même ontologies, ces aides à la navigation dirigent le chercheur progressivement, et efficacement, vers des ressources pertinentes dont il ne connaît pas l'existence et qu'il n'aurait pu localiser autrement.

L'utilisation d'un mode d'organisation structuré pour accéder à une collection virtuelle présente plusieurs avantages : 1) elle autorise le furetage et la navigation ; 2) elle facilite l'élargissement ou le rétrécissement de la portée d'une recherche et permet d'atteindre un niveau optimal de précision ou de rappel ; 3) elle fournit un contexte permettant l'interprétation des termes de recherche utilisés et des ressources repérées ; 4) elle améliore l'efficacité de la recherche simultanée dans plusieurs bases d'information ; 5) elle accroît l'interopérabilité sémantique des systèmes de recherche d'information ; 6) elle garantit une assistance appréciée aux chercheurs inexpérimentés ou non spécialistes d'un domaine.

On observe que les schémas de classification documentaire traditionnels sont peu présents sur le web, et ce bien qu'ils répondent à la majorité des exigences rappelées par Chan (2000) pour qu'un système d'organisation et d'accès aux ressources numériques soit performant : 1) la capacité à gérer efficacement un très grand nombre de ressources ; 2) la flexibilité et la capacité d'adaptation à diverses communautés informationnelles (musées, archives, etc.) ; 3) l'extensibilité et la capacité d'adaptation à différents niveaux de spécificité selon les besoins ; 4) la polyvalence (*i.e.* la capacité à remplir plusieurs fonctions) ; 5) la potentialité d'une utilisation efficace et cohérente par un moteur de recherche.

Aussi efficace soit-il dans un environnement particulier, le transfert d'un outil à un nouvel environnement n'est donc pas si facile à effectuer. Conçues à la fin du XIX^e siècle pour organiser des collections physiques volumineuses et

diversifiées, les classifications documentaires encyclopédiques (la CDD, la CDU et la LCC surtout) qui ont traversé le XX^e siècle sans trop de mal sont projetées aujourd'hui dans un environnement complètement différent auquel elles s'adaptent avec plus ou moins de succès. Ces nouveaux environnements d'utilisation potentielle des classifications documentaires incluent l'ensemble du web, mais également les bases de données multidisciplinaires et multilingues qu'on peut y explorer. Quant aux nouveaux utilisateurs potentiels, ils incluent tous les chercheurs d'information utilisant un poste de travail de moins en moins souvent localisé dans une institution documentaire.

L'intérêt des professionnels pour la classification documentaire a faibli au cours des décennies 1980 et 1990, sous l'influence du développement de moteurs de recherche de plus en plus performants, qui ont pu faire croire que la recherche non contrôlée au moyen de mots-clés, par exemple, était efficace dans toutes les situations. Pour les schémas de classification documentaire, le passage de l'utilisation locale en institution à l'utilisation globale et partagée sur des réseaux ouverts constitue un défi de taille. Ce défi est plus grand encore en Amérique du nord où les professionnels de la documentation ont priorisé la fonction de rangement de la classification plutôt que sa mission cognitive. À partir du moment où l'ensemble d'une collection, ou une grande partie de celle-ci, n'existe plus que sous forme numérique virtuelle, l'utilisation d'un schéma de classification considéré d'abord comme outil de classement apparaît bien évidemment inutile. Les classifications documentaires pourraient pourtant jouer, dans les nouveaux environnements informationnels, un rôle important pour la localisation, le furetage, la navigation, le repérage, l'identification, la sélection et le filtrage d'information et de ressources documentaires pertinentes.

Puisque les forces et les faiblesses des schémas de classification documentaires sont connues, il est possible de décrire assez précisément les modifications à leur contenu et à leur structure qui pourraient leur assurer un rôle de premier plan sur les réseaux d'information. Il faut développer des mécanismes qui permettront de suivre de plus près l'évolution des savoirs et d'intégrer rapidement non seulement les nouveaux sujets, mais également les nouvelles façons de présenter les sujets existants. Il faut moderniser et populariser les termes utilisés pour représenter les sujets. Il faut prévoir plus de flexibilité et des structures alternatives qui permettent la représentation des besoins particuliers de certains usagers, en fonction de différences culturelles, notamment. Sur le plan des applications, il faut développer des outils qui favorisent l'exploitation automatique du schéma de classification documentaire pour le repérage d'information et concevoir des modalités d'utilisation en environnement multilingue, comme langage de commutation, par exemple. La technologie aidant, l'intégration de divers modes d'organisation et de

présentation des savoirs (par exemple, le mode hiérarchique, tel qu'actualisé dans le schéma de classification, et le mode alphabétique, le plus pratiqué sur le web) ne peut qu'être utile au chercheur d'information ; une telle intégration assurerait une meilleure description des contenus thématiques et un rappel plus conséquent de ressources pertinentes.

Les gestionnaires du web connaissent mal et entretiennent de nombreux préjugés face aux structures traditionnelles d'organisation des connaissances et aux schémas de classification documentaire qui s'en inspirent. On les croit totalement inflexibles, obsolètes, peu dynamiques, lourds, complexes, coûteux, et non appropriés pour la mise en ordre des nouvelles connaissances et de nouveaux types de ressources. Aux schémas de classification documentaire traditionnels, on a donc préféré sur le web des structures souples, non disciplinaires, conviviales et intuitives, développées autour de grandes catégories représentant des thématiques (par exemple, Enseignement, Tourisme, Loisirs), des types de documents (Dictionnaires, Référence), des auditoires (Enfants, Immigrants), des lieux (Québec, Californie), etc. Ces structures classificatoires sont plus à même que les schémas traditionnels d'intégrer temporairement un centre d'intérêt populaire, mais leurs faiblesses ne sont pas négligeables. Ces nouvelles structures se développent rarement sur plus de trois ou quatre niveaux de profondeur et elles n'offrent 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 grand. On ne s'y inquiète guère de normalisation conceptuelle, terminologique ou structurelle. Les nouveaux schémas d'organisation ne sont pas toujours cohérents et l'absence de relations logiques entre les différentes composantes de la structure génère des hiérarchies peu informatives qui rendent la navigation difficile. De plus, le manque d'uniformité entre structures couvrant un même domaine du savoir ou un même champ d'activités 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.

Là où les schémas de classification traditionnels se voulaient assimilés à des classifications du savoir et se réclamaient de la philosophie, ces nouvelles structures sont résolument utilitaires et se soucient peu de normalisation, de stabilité et même d'efficacité. Délestés de l'obligation de collocation physique de documents, les structures d'organisation du web font la part belle à une polyhiérarchie libératrice, permettant notamment la coexistence de diverses visions du monde naturel et de l'univers des connaissances.

Web 2.0 et organisation des connaissances : nouveaux outils, nouvelles pratiques

L'Internet représente une source incontournable pour le traitement des ressources en ligne par des professionnels de l'information (des bibliothécaires, des documentalistes, des veilleurs). À ses débuts les bibliothécaires et autres spécialistes de l'information ont tenté d'étendre leur mission d'organisation des connaissances à ce nouvel environnement (Voss, 2007 cité par Durieux, 2010). Mais ils se sont rendu compte très vite de leurs limites face à l'expansion d'Internet et à l'augmentation constante de ressources publiées.

D'autres évolutions ont accompagné ces changements, à savoir le rôle et la place de l'utilisateur. Depuis l'avènement du web 2.0, l'utilisateur est passé du rôle de consommateur passif à celui d'acteur à part entière (Durieux, 2010). Ces possibilités d'interactions ont mené à l'émergence du principe de collaboration sur le réseau dont l'exemple le plus marquant est *Wikipédia* (Durieux, 2010). Ce nouveau mode d'écriture a eu un impact important sur les modes d'indexation et par conséquent sur les systèmes d'organisation de connaissances.

Face à cette dynamique de production pour classer les quantités massives de contenus, les éditeurs offrent aux utilisateurs une panoplie d'outils d'aide à la navigation. Parmi ces outils, le tagging collaboratif s'est progressivement diffusé et représente aujourd'hui un mode de catégorisation de l'information typique des sites du web 2.0.

En raison de l'expansion d'Internet, les professionnels de l'information se trouvent devant l'impossibilité d'indexer l'ensemble des ressources accessibles en ligne. La pratique d'indexation collaborative est apparue comme une des possibilités. Depuis l'intégration du premier système de tagging collaboratif sur le site Delicious⁵ en 2003, le tagging a connu d'importantes évolutions sur un plan technique et s'est progressivement diffusé à d'autres sites de partage de contenus tels que Flickr en 2004, Youtube et Dailymotion en 2005 (Crepel, 2008).

Le mode d'indexation pratiqué dans ces sites est une indexation collaborative produite sur la base de l'indexation libre au moyen de tags, appelé communément « folksonomie ». Les folksonomies désignent un processus d'organisation collaborative à l'aide de mots-clés librement choisis. Ce mode d'indexation dite sociale, par opposition à l'indexation normée (métadonnées normées, listes d'autorité, thésaurus...) trouve ses origines dans des pratiques classiques (organisation et classification des connaissances) mais aussi dans des

5. <http://del.icio.us>

pratiques métier d'ordre techniciste (conservation et accès aux documents, métadonnées, Dublin Core). Il constitue une des bases de développements actuels sur un web sans cesse en évolution (web 2.0, ontologies sociales, web sémantique et socio-sémantique). Ces pratiques disposent aujourd'hui d'outils et d'interfaces permettant un « balisage »⁶ entièrement subjectivé (par opposition à des balises normées, émanant d'une autorité et ayant été validées).

L'intérêt des tags paraît indéniable : une fois assignés, les tags sont immédiatement accessibles aux internautes et exploitables dans le cadre de recherches d'information. Un nombre croissant de sites web permet aux utilisateurs de taguer des photographies, des vidéos ou encore de la musique. Le choix des termes est totalement libre, il ne repose sur aucun vocabulaire contrôlé (Goldes *et al*, 2006 cité par Durieux, 2010).

Un autre concept associé au processus d'indexation collaborative est le *tagcloud* (ou nuage de mots clés). Il permet de visualiser l'ensemble des tags attribués à une ressource donnée ou par un utilisateur déterminé. Plus un tag est utilisé, plus sa visibilité à l'intérieur du nuage augmente. La sélection d'un mot clé du nuage mène à l'ensemble des ressources associées à ce dernier (Durieux, 2010). Un *tagcloud* permet dès lors la navigation à l'intérieur de la folksonomie et la découverte des ressources susceptibles de rencontrer l'intérêt de l'utilisateur.

Certains auteurs mettent en évidence l'amalgame fréquent entre tagging collaboratif et social bookmarking (ou partage de signets). Ce dernier consiste en la sauvegarde et l'organisation de signets de pages web (Durieux, 2010). Par ailleurs, le collaborative tagging peut s'appliquer à d'autres ressources que des signets de pages web.

Ce mode d'indexation a fait l'objet d'évaluations et d'appréciations entreprises par de nombreux chercheurs. Certains soulignent cependant ses limites par rapport aux langages d'indexation contrôlés (Durieux, 2010). Parmi les reproches faits à cette pratique d'indexation, on peut relever l'absence de structure et d'organisation des folksonomies comparativement aux langages d'indexation traditionnels. Contrairement à l'indexation réalisée à l'aide d'un langage documentaire, l'indexation collaborative ne gère pas les problèmes liés à la polysémie et à la synonymie, ni les autres problèmes linguistiques habituellement pris en charge par les thésaurus documentaires.

D'autres évaluations ont été plus positives. Les folksonomies et d'autres techniques d'indexation sociale seraient pérennes d'après Ertzscheid et Galzot

6. On parle de « tags » pour les balises et « tagging » pour le balisage de l'information. On trouve d'autres termes comme « signets » ou *social bookmarking*.

(2006, 299-301) pour plusieurs raisons dont nous relevons ici les plus importantes :

– Raisons économiques : les outils de recherche sont intégrés de manière plus ou moins naturelle et intuitive, la gestion de ces communautés d'intérêt et la possibilité de soumettre tout type de document ;

– Raisons techniques : les interfaces souvent très intuitives, accessibles à tous en ajoutant une dimension de partage, de capitalisation et de diffusion en temps réel ;

– Faible coût cognitif. Ertzscheid et Galzot (2006) expliquent qu'en raison de leur complexité, les activités de catégorisation ou d'indexation, dans un contexte documentaire professionnel, nécessitent l'usage de métadonnées normalisées alors que l'indexation dite sociale n'engendre qu'un faible coût cognitif.

Pour une revue complète des forces et faiblesses de ces pratiques d'indexation, nous renvoyons les lecteurs à Durieux (2010).

Politique informationnelle croissante depuis 2001 et l'article fondateur du web sémantique écrit par Tim Berners-Lee (Berners-Lee *et al.*, 2001), la nécessité de se tourner vers un web plus structuré et dans lequel les données priment sur le document se fait sentir au sein de la communauté des utilisateurs actifs du web, qu'il s'agisse de scientifiques de l'information, de journalistes ou encore d'informaticiens. Certes, des initiatives d'organisation ont déjà été instaurées, telle l'arrivée massive du XHTML, permettant une structuration du web plus sémantisée qu'avec des pages construites en HTML pur, grâce à la séparation désormais créée entre contenu et représentation.

Néanmoins, l'utilisation des microformats et de la surcouche sémantisante basée sur le RDF va permettre d'organiser l'information selon une architecture propre à rendre l'information accessible de manière structurée à la machine, et par là-même à en automatiser le traitement.

Cette translation vers une toile de plus en plus organisée n'en est donc, malgré tout, pas à ses débuts. Toutefois, de par l'intégration progressive de documents web au format RDF/XML, la restructuration des données n'est pas sans engendrer des contraintes et des questionnements : comment intégrer ce nouveau moyen de construction du web sans perturber les usages courants ? Comment assurer la croissance du *Linked Open Data Cloud*?⁷ Comment toucher le « grand public » et l'impliquer dans les démarches de sémantisation ?

7. LOD Cloud, le LOD se base sur les métadonnées collectées et traitées par les contributions des membres du *CKAN Directory*. <http://richard.cyganiak.de/2007/10/>

La situation se doit d'être étudiée sous différentes facettes, afin de parvenir à un équilibre entre les bases sociales acquises et les fondements sémantiques espérés. Des contradictions importantes sont à prévoir. Ainsi, le travail de reconstruction sémantisée de l'information en ligne doit se faire à la fois de façon automatisée (*data mining*), mais également humaine, ce qui rejoint les propos tenus par Berners-Lee (2001) : le web sémantique est « une extension du web actuel dans laquelle l'information se voit associée à un sens bien défini améliorant la capacité des ordinateurs et des hommes à travailler en coopération »⁸. De ce fait, nos efforts doivent être centrés sur l'amélioration des outils d'indexation existants, tout en nous focalisant sur l'acquisition de nouvelles données basées sur les schémas RDF.

Il est toutefois nécessaire de promouvoir les intérêts d'une implication sociale dans ce remaniement de l'indexation des documents avec des arguments convaincants et propres à donner aux internautes l'envie de travailler à l'amélioration du système. Il convient de rappeler que les écarts perçus entre le web social actuel et le web sémantisé, attendu dans les années à venir, ne se cantonnent qu'à un changement, certes important mais discret, dans l'organisation du web.

Ainsi, la question de l'indexation des informations se doit d'être analysée et devra offrir des moyens concrets. Néanmoins, comme nous avons pu l'observer, des pistes sont déjà explorées, telle l'élaboration de *folksontologies* ou *ontologies collaboratives*. La construction d'une folksontologie implique de tourner les folksonomies en ontologies ou folksontologie. Plusieurs approches complémentaires existent afin de rapprocher les folksonomies et les ontologies pour la gestion des connaissances partagées.

Finalement, il ressort de toutes ces considérations que les modifications apportées au contenu du web seront essentiellement axées sur l'automatisation et le traitement purement informatique des données.

Web 2.0, social et web sémantique ne sont pas deux conceptions antinomiques, pas plus que deux étapes successives d'un développement technologique (Gruber, 2008). Web participatif et web sémantique ne s'excluent pas et peuvent nourrir des points de convergence comme l'illustre déjà l'application *DBpedia*⁹, base de données sémantique puisant ses informations

lod. Ce graphique montre les bases de données publiées aux formats associés aux *Linked Data*, par les contributeurs de la communauté du projet *Linked Open Data* et d'autres institutions, ou individus à titre personnel.

8. Passage traduit du texte de l'article de Berners-Lee, J. Hendler, and O Lassila, « The semantic web », *Scientific American*, v.284(5), pp. 34-43, 2001.

9. <http://dbpedia.org>

dans l'encyclopédie collaborative *Wikipedia*. Le web social a contribué à faire évoluer le web vers le modèle initialement imaginé par ses concepteurs en plaçant les utilisateurs et leurs échanges informationnels au centre des processus de publication et d'échange de contenus. Zacklad *et al.* (2003) ont forgé le terme *web socio-sémantique* pour exprimer cette coopération souhaitable entre deux types d'organisations.

Les sept textes qui constituent ce numéro thématique présentent des perspectives différentes sur l'organisation des connaissances telle qu'elle se reflète dans les institutions documentaires traditionnelles et telle qu'elle émerge dans les environnements numériques, notamment sur les réseaux sociaux et dans les sites de partage d'information et de ressources. Qu'ils soient davantage théoriques ou qu'ils décrivent un projet particulier, ces articles offrent matière à réflexion ainsi que des éléments de réponses aux questions de départ : d'où venons-nous ? où en sommes-nous ? où allons-nous ?

Sous la plume, métaphorique, de Pierre Carbone, organisation du savoir et bibliothèques paraissent se confondre, la seconde ne pouvant exister sans l'apport de la première. Dans *Les bibliothèques, ou les mémoires mobilisées*, l'auteur présente la bibliothèque du XXI^e siècle, la bibliothèque apprenante, où « le partage du savoir, la valorisation de la mémoire collective, le développement des compétences sont au cœur des projets ». La bibliothèque est en mutation, et s'il est indéniable qu'en un tel contexte l'organisation reste essentielle, il faut se demander quel type et quel système d'organisation seront le plus à même de faciliter l'atteinte d'objectifs très ambitieux.

Vanda Broughton examine un phénomène beaucoup plus récent, celui de l'assignation par l'utilisateur de termes représentant le sujet des ressources documentaires (communément appelé *tagging*). Sous prétexte de décrire les éléments de contrôle du vocabulaire qui se glissent dans les folksonomies et autres listes terminologiques ainsi créées, l'auteure rappelle d'abord les origines des différentes formes de contrôle du vocabulaire bien connues des spécialistes du traitement documentaire, et les avantages de ce contrôle. Dans *Emergent vocabulary control in web 2.0: comparisons with conventional LIS theory and practice*, Broughton conclut que les systèmes de *tagging* possèdent des caractéristiques utiles, tout particulièrement pour la description et le repérage dans le web non organisé, mais sans offrir la précision et la puissance sémantique des langages documentaires préconstruits. C'est la structure relationnelle qui confère précision et puissance à un langage documentaire et l'élaboration de cette structure relationnelle ne peut être pour le moment complètement automatisée.

La logique de la hiérarchie est-elle essentielle à l'efficacité d'une structure d'organisation des connaissances ou aux systèmes de classification de

l'information et des documents qui en dérivent ? Dans *Hiérarchie et anarchie : dépasser l'opposition entre organisation centralisée et distribuée*, Isabelle Boydens montre que la coexistence des organisations centralisées et distribuées du savoir ne date pas d'hier et qu'elle rend inévitable l'interaction entre hiérarchie et anarchie ; le degré d'interaction augmente nécessairement avec le développement des technologies de l'information et la participation de l'utilisateur à la description et à la mise en ordre des ressources. Le phénomène n'a pas que des conséquences négatives : l'auteure présente des exemples concrets d'application de modèles reposant sur une approche herméneutique et sur la théorie de la modélisation conceptuelle qui permettent de maximiser l'efficacité de méthodes mixtes d'organisation.

La classification à facettes suscite de plus en plus d'intérêt sur le web, et plus particulièrement dans le cadre d'un web que l'on veut sémantique. Le système de description et d'organisation du savoir et de ses représentations conçu par Ranganathan, aussi génial soit-il, ne pouvait s'implanter à l'époque de sa création dans un milieu traditionnel où l'organisation hiérarchique était la seule qui soit viable en raison des limites imposées par les systèmes de recherche ; à l'ère du numérique, de la flexibilité presque totale, de la disponibilité de technologies capable de « calculer » et de réorganiser infiniment plus vite que l'esprit humain, la classification à facettes trouve un terrain idéal pour s'implanter, croître, se transformer, s'améliorer. Dans *A semantic (faceted) web?*, Kathryn La Barre présente les pratiques communément connues sous le nom de théorie des facettes, telles qu'elles sont appliquées dans le contexte du web sémantique. La Barre décrit des applications du web qui utilisent, adaptent et interprètent, parfois de façon erronée, la pratique de l'analyse par facettes. Elle conclut que l'utilisation des facettes est loin d'être optimale et suggère la création de définitions et de spécificités fonctionnelles pour éviter les dérives fréquemment observées lorsque la facette acquiert de multiples identités ; cette dérive avait été relevée aussi, rappelons-le, par Maniez (1999).

Manuel Zacklad souligne que les enjeux liés à la conception et à l'usage des systèmes d'organisation des connaissances (SOC), ne concernant jusqu'alors qu'un petit nombre de spécialistes, tendent à devenir majeurs dans la gestion de l'information numérique. Dans *Cinq critères d'évaluation des systèmes d'organisation des connaissances et de leur sémantique*, il montre que la diversité des SOC, intégrés dans des dispositifs techno-socio-expressifs très divers, est telle qu'il est très difficile de bien comprendre leurs propriétés et leur complémentarité. Après un rappel des six grandes familles de SOC qui semblent représentatives des évolutions actuelles, l'auteur présente cinq critères d'analyse et de comparaison des SOC.

Dans *L'organisation des connaissances dans les habitats numériques de communauté*, Diane Mercier présente les résultats d'une étude exploratoire qui vise à mieux comprendre la relation entre communauté de pratique et habitat numérique au regard de l'organisation des connaissances. Cette étude met en évidence l'organisation en facettes des connaissances. L'auteure démontre aussi que la coconstruction de base de connaissances avec des outils accessibles et conviviaux pour des équipes ou des groupes multidisciplinaires contribue à centrer leurs activités sur la satisfaction des besoins des utilisateurs-contributeurs et encourage la compréhension mutuelle.

Le dernier article, proposé par Jean-Max Noyer, définit un certain nombre de critères conceptuels qui permettent d'orienter de manière critique le développement des nouvelles technologies intellectuelles. Ces critères sont définis en fonction des rapports entre connaissance/pensée/réseaux dans le contexte des écritures et mémoires numériques. Cette réflexion est mise en perspective selon divers niveaux d'échelle qui prennent en compte les différentes dimensions de la crise écologique en cours. Dans *Connaissance, pensée, réseaux à l'heure numérique. Pour une nouvelle Renaissance*, l'auteur souligne que sa réflexion s'inscrit dans le cadre d'une approche critique des nouvelles écologies de l'intelligence et de la connaissance, y compris les dimensions cognitives, sémiotiques et politiques.

Bibliographie

- Berners-Lee J., Hendler, and O. Lassila, « The semantic web », *Scientific American*, vol. 284, n° 5, 2001, p. 34-43.
- Béthery A., « Liberté bien ordonnée... les classifications encyclopédiques revues et corrigées », *Bulletin des bibliothèques de France*, vol. 33, n° 6, 1988, p. 450-455.
- Broudoux E., « Indexation collaborative : entre gain informationnel et déperdition conceptuelle ? », *Document numérique et société*, 17-18 novembre 2008, Paris, Cnam (http://archivesic.ccsd.cnrs.fr/sic_00364626/en/)
- Bowker G.C., Star S. L., *Sorting Things Out: Classification and Its Consequences*, Cambridge, MA, MIT Press, 1999.
- Chan L.M., *Exploiting LCSH, LCC, and DDC to retrieve networked resources: issues and challenges*, 2000, en ligne à lcweb.loc.gov/catdir/bibcontrol/chan_paper.html.
- Crepel M., « Les folksonomies comme support émergent de navigation sociale et de structuration de l'information sur le web », In Cardon, D. (dir) *Les réseaux sociaux de l'Internet*, vol 26, n° 152, 2008, p. 169-2004.
- De Grolier E., « Taxilogie et classification : un essai de mise au point et quelques notes de prospectives », *Bulletin des bibliothèques de France*, vol. 33, n° 6, 1988, p. 468-489.

- Durieux V., (2010), Collaborative tagging et folksonomies, l'Organisation du web par les internautes. In Quoniam L., Zimbardo P. (dir), *Du web 2.0 au concept 2.0*, Revue *Les cahiers du numérique*, vol. 6, n° 1, Hermès-Lavoisier, Paris, 2010, p. 69-80.
- Gandon F., Le web sémantique n'est pas antisocial, Proceedings Conférence Ingénierie des Connaissances, Semaine de la Connaissance, Nantes, pp. 131-140, 20-30 Juin 2006
- Ertzscheid O., Galzot (2006), « Indexation sociale et continents documentaires », in Broudoux Y., Chartron G., *Document numérique et société*, Actes de la Conférence, DocSoc-2006, Semaine du Document numérique, Fribourg, 20-21 septembre 2006, p. 291-306.
- Gruber T., « Collective knowledge systems: where the social web meets the semantic web », *Journal of web semantics*, vol. 6, 2008, p. 4-13.
- Le Deuff O., « Folksonomies : les usagers indexent le web », *Bulletin des bibliothèques de France*, vol. 51, n° 4, 2006, p. 66-70.
- Hudon M., « Structuration des savoirs et organisation des collections dans les répertoires du web », *Bulletin des bibliothèques de France*, vol. 46, n° 1, 2001, p. 57-62.
- Kwasnik B. H., « The role of classification in knowledge representation and discovery », *Library Trends*, vol. 48, n° 1, 1999, p. 22-47.
- LaBarre K., « The heritage of early FC in document reference retrieval systems, 1920-1969 », *Library History*, vol. 23, 2007, p. 129-149.
- Mai J.E., « The Future of General Classification », *Cataloging & Classification Quarterly*, vol. 37, n° 1-2, 2003, p. 3-12.
- Maniez J., *Actualité des langages documentaires : fondements théoriques de la recherche d'information*, Paris, ADBS, 2002.
- Maniez J., « Du bon usage des facettes », *Documentaliste – Sciences de l'information*, vol. 36, n° 4/5, 1999, p. 249-262.
- Miksa F.L., « The Concept of the Universe of Knowledge and the Purpose of LIS Classification », *Classification Research for Knowledge Representation and Organization*, Proceedings of the 5th International Study Conference on Classification Research, Toronto, Canada, June 24-28, 1991, Amsterdam, Elsevier, 1992, p. 101-126.
- Quoniam L., Zimbardo P. (dir), *Du web 2.0 au concept 2.0*, Revue *Les cahiers du numérique*, vol. 6, n° 1, Hermès-Lavoisier, Paris, 2010.
- Satija M.P., « Classification: some fundamentals, some myths, some realities », *Knowledge Organization*, vol. 25, n° 1-2, 1998, p. 32-35.
- Satija M.P., « The revision and future of Colon Classification », *Knowledge Organization*, vol. 24, n° 1, 1997, p. 18-23.
- Svenonius E., « Classification: Prospects, problems and possibilities », *Classification Research for Knowledge Representation and Organization*, Proceedings of the 5th

International Study Conference on Classification Research, Toronto, Canada, June 24-28, 1991, Amsterdam, Elsevier, 1992, p. 5-25.

Voss J., « tagging, folksonomy & co-renaissance of manual indexing », 10th international *symposium for information science*, cologne, 2007, (http://arxiv.org/ps_cache/cs/pdf/0701/0701072v2.pdf)

Weihls J., « A brief history of classification, Part 1-5 », *Technicalities*, 30, 2010, n° 1-5.

Weinberger D., *Everything is miscellaneous: the power of the new digital disorder*, New York, NY, Times Books, 2007.

Zacklad et Cahier J-P., Pétard X., (2003), *Du web cognitivement sémantique au web socio sémantique : Exigences représentationnelles de la coopération*, http://www.zacklad.org/articles_web_socio_semantique/diapo%20web%20semantique%20et%20SHS.pdf