

CE QUE L'ANNUAIRE FAIT À INTERNET

Sociologie des épreuves documentaires

Christophe LEJEUNE

Chercheur en sociologie, Institut Charles Delaunay

Internet regorge de communautés de passionnés en tous genres. C'est de l'une d'elles que traite cette contribution. Lancée en 1998, l'initiative en question regroupe des volontaires construisant un annuaire de sites Internet. Bien avant le succès de Wikipédia, cette communauté est loin d'avoir laissé indifférents les géants de l'informatique puisque Netscape l'a sponsorisée depuis ses débuts et que le service d'annuaire de Google utilise sa base de données. Apparemment éloigné des bibliothèques, ce groupe de passionnés n'en est pas moins en train de réaliser le rêve de Paul Otlet, pionnier (belge) de la documentation. En effet, en collectant les adresses de sites Internet, en les répertoriant, les classant, ces personnes reproduisent, un siècle plus tard, ce qui fut le travail des premiers documentalistes. Ils se confrontent à des problèmes similaires. Ils y trouvent souvent des solutions analogues. Et, parfois, leurs solutions divergent.

Après quatre années d'observation fouillée de cette communauté virtuelle, l'auteur propose une mise en perspective de cette activité classificatoire avec les usages et les outils de la documentation.

Internet groepeer tal van liefhebbers allerlei genre. Een van die gepassioneerden levert deze bijdrage over het in 1998 ontstane initiatief om louter vrijwillig een gids van Internetsites op te stellen. Lang voor het succes van Wikipedia trok dit initiatief de aandacht van enkele hoofdrolspelers op de informaticamarkt: Netscape was van in het begin bereid tot sponsoring en Google gebruikt geregeld de data. Op het eerste zicht ver verwijderd van het klassieke bibliotheekgebeuren, is deze belangengroep toch onrechtstreeks bezig met het verwezenlijken van de droom van een van de (Belgische) pioniers op het vlak van het documentatiebeheer, Paul Otlet. Een van de primaire taken van een documentalist was immers het repertoriëren, het klasseren van documentatie. Zowat een eeuw later spitst zich dit toe op het Internet, doen zich gelijkaardige problemen voor als bij het ontsluiten van klassieke documentatie met dikwijls bijna analoge oplossingen en op andere momenten toch afwijkend. Na een periode van ongeveer vier jaar observatie van deze virtuele gemeenschap formuleert de auteur een perspectief voor het classificatiewerk gebruikmakend van wat we eerder klassieke documentatietools kunnen noemen.

Les constructeurs d'annuaires témoignent d'une culture et d'un univers référentiel composites que cet article s'emploie à expliciter¹. Je présente tout d'abord l'organisation thématique des annuaires en tant qu'outils de recherche d'information. J'évoque également la structure sociale de la communauté des contributeurs à l'*Open Directory Project*² du point de vue de sa mise sur pied. Suit une discussion du choix terminologique du terme "annuaire" ce qui m'amène à évoquer l'histoire de ce concept. Celle-ci explicite une partie des références professionnelles des annuaires de sites Internet (à la rencontre des techniques informatiques et des outils de classement). J'examine ensuite la structure hiérarchique de l'outil et met en évidence la coexistence de plusieurs types de hiérarchies. Après avoir évoqué la diversité d'usages des relations verticales, je passe en revue trois types de renvois (horizontaux) utilisés au sein d'une arborescence.

d'un annuaire (aussi appelé répertoire) d'adresses de sites Internet. Cet annuaire est un outil qui peut être consulté par toute personne qui recherche de l'information sur Internet (c'est, en quelque sorte, l'équivalent des pages jaunes). Contrairement aux célèbres moteurs de recherche, l'annuaire est structuré de manière arborescente : les catégories s'organisent dans une logique d'inclusion qui va de la plus générale à la plus spécifique. Chaque catégorie accueille des références de sites Internet (c'est-à-dire une adresse, un titre et une description). Je montrerai que cette organisation hiérarchique se complète de renvois transversaux de différents types (les catégories apparentées, les équivalences linguistiques et les liens symboliques).

Présentation de l'annuaire

Cette section liminaire présente brièvement la communauté virtuelle qui a fait l'objet de mes investigations. Les acteurs qui la composent se désignent comme des éditeurs ou des indexeurs. Leur travail consiste en l'élaboration

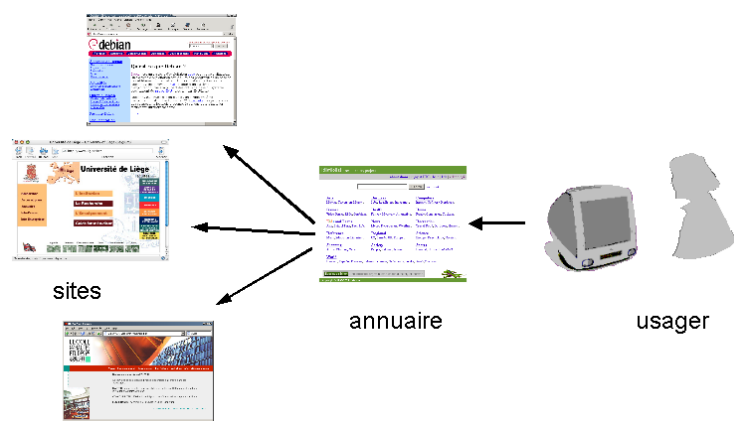


Fig. 1 : Figure de l'utilisateur

L'annuaire peut être abordé de deux manières. Pour ses utilisateurs (les internautes), il est envisagé comme un outil de recherche d'information (au même titre qu'un moteur de recherche). Cet outil est alors envisagé comme une boîte noire ; c'est un intermédiaire dont seule l'efficacité compte (figure 1). Cette première perspective recouvre le point de vue de l'utilisateur.

Une deuxième façon d'envisager le répertoire est d'entrer, comme les anthropologues des sciences, dans l'annuaire en train de se faire [11]. En étudiant l'élaboration de cet outil, on ouvre la boîte noire, et le travail des concepteurs (les éditeurs-indexeurs) devient tangible. L'annuaire est alors conçu comme une réalisation collective continue (à laquelle chacun contribue), il est une fin, et non plus un intermédiaire (figure 2). Cette perspective rencontre celle du contributeur.

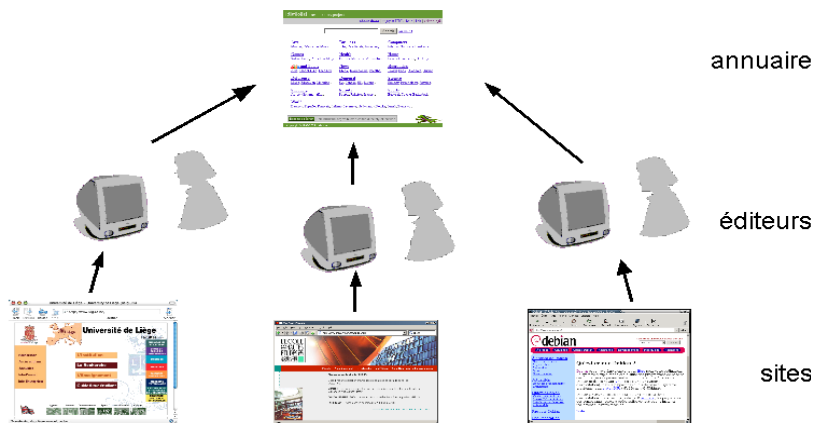


Fig. 2 : Figure de l'éditeur

Les figures de l'utilisateur et de l'indexeur permettent de clarifier la différence entre l'usage et la constitution de l'annuaire. Cette distinction ne doit cependant pas faire oublier que, dans le cas de l'*Open Directory Project*, tout utilisateur est invité à contribuer à l'enrichissement du répertoire. Cette participation peut prendre la forme du commentaire ponctuel, de la suggestion d'une référence voire de l'accès au statut d'éditeur. Le commentaire s'effectue au moyen d'un des formulaires prévus à cet effet³ ; il s'agit le plus souvent de suggérer la rectification ou la mise à jour⁴ d'une référence de site Internet déjà présente dans l'annuaire. Le visiteur peut également soumettre la référence d'un site Internet qui ne serait pas encore listée⁵. Il est enfin également possible de poser sa candidature comme éditeur d'une des catégories du répertoire⁶. Ces trois modes de participation constituent un gradient entre les deux figures (de l'usage et de la constitution) dégagées ci-dessus.

Après cette présentation de l'annuaire et de ses usages, je discute dans la section suivante le choix de la qualification de l'annuaire. Cette section terminologique permet déjà d'entrevoir les

différents univers référentiels qui ont historiquement marqué les protagonistes qui m'occupent.

Question de terminologie

En français, le mot "annuaire" renvoie à deux composantes. Ces deux dimensions sont celle de la liste de coordonnées d'humains (on emploie également le terme "répertoire" dans cette acception) et celle de la mise à jour sur base annuelle (l'annuaire est par définition un répertoire qui est revu chaque année). Vu que les annuaires de sites Internet ne regroupent pas de liste de personnes et qu'ils ne connaissent pas de cycles annuels, il est autorisé de se questionner sur la raison d'être de cette dénomination.

Dans les lignes qui suivent, j'avance néanmoins l'idée que ces deux dimensions ont convergé pour que le terme annuaire s'impose. Une influence de l'informatique est probablement à l'origine de ce choix (même si les répertoires informatiques ne s'appellent pas des annuaires). Celle-ci passe par l'anglais – via l'assimilation au système de fichier que cette langue rend possible. L'anglais qualifie en effet de "directories" les répertoires (ou "dossiers") dans lesquels les utilisateurs répartissent leurs fichiers informatiques. Les projets d'annuaire ayant été initialement fondés aux États-Unis, la référence à l'anglais (langue prégnante dans le monde informatique) ainsi que – à travers elle – l'homologie entre les répertoires de sites Internet et le système de fichiers des disques durs permettent d'amorcer la compréhension de ce choix terminologique.

Ainsi, l'informatique recourt effectivement au vocable "directory" pour désigner les dossiers ou répertoires de fichiers sur le disque dur, et c'est en référence à ce monde professionnel que les créateurs d'annuaires ont effectués leur choix terminologique. Ce constat renseigne sur la culture des fondations de ces outils de recherche d'information. Mais la question terminologique n'est guère plus avancée. En effet, avant d'être importé en informatique, le terme anglais "directory" recouvre lui aussi un des sens de l'annuaire en français, celui d'une liste de personnes (ou de leurs coordonnées) [14]. La dimension annuelle lui est par contre étrangère. La compréhension de cet emploi particulier de "directory" exige donc une remontée dans le temps de quelques décen-

nies... afin de comprendre l'importation de ce terme en informatique.

Pourquoi les concepteurs de ces fichiers informatiques d'un type particulier que sont les dossiers [17] ont-ils décidé de les baptiser "directories" ? Une explication possible m'a été suggérée par l'informaticien Aurélien Bénel [5]. Comme les comptes-rendus historiques des systèmes d'exploitation semblent le suggérer, les premiers systèmes de fichiers à connaître une hiérarchie (et donc des dossiers) ont introduit un seul niveau de répertoire afin de fournir un espace séparé à chaque utilisateur d'un même ordinateur [19]. Les premiers répertoires informatiques auraient donc été des conteneurs personnalisés, chacun d'entre eux étant réservé à une personne différente. Dès lors, la liste des dossiers aurait correspondu à une liste de personnes. L'analogie avec l'annuaire (compris comme une liste de noms de personnes) n'est donc pas fortuite, et c'est sans doute elle qui a incité les informaticiens des années soixante à opter pour ce terme.

Voici donc une première explication du choix terminologique : celui-ci a été opéré en référence à l'informatique et l'histoire technique de cette discipline permet *in fine* de comprendre pourquoi le terme s'est imposé. Le français aurait alors traduit le terme "directory" par son équivalent "annuaire". Bien qu'elle soit également mobilisée par les éditeurs, la traduction alternative - "répertoire" - n'a pas été retenue comme désignation officielle de l'ensemble des annuaires de sites Internet, peut-être pour la raison suivante.

L'influence sociolinguistique de l'anglais s'est sans doute combinée au rapprochement de l'annuaire aux pages jaunes. De nombreux documents - tant de vulgarisation que rédigés par des analystes reconnus [4] - filent cette comparaison. Si elle peut être commentée (voire critiquée) d'un point de vue analytique, cette assimilation de l'usage de l'annuaire au bottin de téléphone est particulièrement habile d'un point de vue sociologique. Elle opère en effet une traduction [6] entre un usage largement partagé et intégré à la vie de tous les jours, usage découlant d'ailleurs de l'intrusion dans les foyers, il n'y a guère plus d'une génération, d'une nouvelle technologie de la communication, à savoir : le téléphone. En rapprochant l'annuaire aux pages jaunes, on fait de même entre Internet et le téléphone. À l'heure où ces lignes sont écrites, la pénétration du réseau dans les foyers est patente. Je me garderai de trancher entre le fait que cette traduction était visionnaire ou si elle fut efficace (ce qui signifierait qu'elle n'a pas seulement anticipé le mouvement mais qu'elle y a contribué). Sans doute un peu des deux. Ce qui importe ici, c'est que le choix du terme procède d'une économie cognitive (et sociale) importante

car elle mobilise le nom d'un outil quotidien connu du grand nombre pour désigner un outil qui, de son côté, inspire la nouveauté. Le terme "annuaire" a donc une dimension politique (il contribue à l'acceptation de cet outil). En s'appuyant sur un existant, il témoigne également d'un attrait cognitif. Le contenu sémantique de cette homologie désigne l'usage d'un outil assurant la médiation entre l'utilisateur et ce qu'il cherche à atteindre, et présentant une structuration de l'information en rubriques.

À l'aune de son rôle de passeur, la qualification des annuaires devient un choix non seulement compréhensible mais (replacé dans son contexte d'émergence) stratégique et (rétrospectivement) fructueux⁷. Toutefois, l'emploi de ce terme draine avec lui une seconde dimension qui semble belle et bien étrangère : l'annualisation. En effet, le choix du terme annuaire semble faire converger la dimension de liste de personne avec la dimension annuelle, sans autre raison que l'arbitraire de la langue [7]. Qu'elle soit fortuite ou non, cette convergence recèle quelque bonheur. Selon sa définition originelle, un annuaire est mis à jour chaque année. Souvent, cela se concrétise matériellement par la publication d'une nouvelle édition. Contrairement à leurs cousins de papier, les annuaires de sites Internet ne connaissent pas de cycles annuels de ce type. Par contre, ils sont mis à jour de manière régulière, voire continue. Bien que cette proximité soit fortuite, comme je l'ai indiqué plus haut, les "web directories" rejoignent (du moins en partie) l'esprit de ce second axe.

Si ce choix terminologique peut toujours étonner certains, les lignes précédentes permettent maintenant de comprendre pourquoi les annuaires continuent à s'appeler ainsi. Outre son apport terminologique, cette section permet donc de saisir non seulement la dynamique des annuaires (mise à jour continue) mais également les mondes de référence (documentaires et informatiques) convoqués par cette dénomination. La suite de cet article poursuit cette investigation à travers l'étude de la forme (arborescente) de ces outils.

Une ou plusieurs hiérarchies ?

Comme je l'ai précisé dès l'introduction de cet article, les catégories qui accueillent les références de sites Internet sont organisées de manière hiérarchique. Les catégories de plus haut niveau (appelées "majeures") couvrent les grandes thématiques comme les arts ou les sciences et contiennent des sous-catégories plus spécifiques. Chaque niveau peut contenir des références de sites Internet et des sous-catégories. Le défi de

L'*Open Directory* repose sur sa nature volontaire. Insatisfaits des annuaires proposés par des entreprises commerciales, Rich Skrenta et Bob Truel se sont inspirés de l'alternative des logiciels libres pour lancer leur projet. Vu que la qualité et la complétude d'un annuaire sont tributaires des experts qui, sur chaque thématique spécifique, recensent les sites pertinents, un groupe de volontaires est à même – s'il réussit à mobiliser suffisamment de personnes – de surpasser une entreprise qui ne peut engager qu'un nombre limité d'employés.

L'idée était séduisante mais rien ne disait, en juin 1998, si elle serait fructueuse. Année après année, des volontaires ont rejoint le projet. Six mois après la naissance du projet, *Netscape* offrit son soutien via la mise à disposition de serveurs. Alors que la communauté n'a pas deux ans, *Google* décide d'adopter pour son propre annuaire la base de données mise librement à disposition par l'*Open Directory*. Actuellement, plusieurs dizaines de milliers de bénévoles colligent des références dans des spécialités toutes plus exotiques les unes que les autres. Ce mode d'organisation (basé sur le volontariat) serait, de l'avis des spécialistes [3], la seule organisation viable pour un annuaire (qui, par définition, confronte un travail manuel de catalogage à une masse de sites Internet dont la croissance est exponentielle).

Ce diagnostic ne renseigne cependant pas sur la façon dont ces immenses communautés se débrouillent pour passer en revue une telle quantité de sites Internet. Ou plutôt (vu que le nombre de bras n'est vraisemblablement pas le problème du projet) sur comment cette communauté s'organise. En outre, les pages de présentation du projet ont de quoi interpellier. Selon les versions, elles parlent de "démocratie" ou de "république du Web". Tout un programme !

Vu son ampleur, la tâche qui incombe aux membres du projet nécessite une division du travail et une répartition des responsabilités. Chaque candidat à l'entrée dans le répertoire est ainsi invité à faire valoir une compétence thématique. Si sa candidature est acceptée, il devient responsable d'une catégorie. La catégorie dans laquelle un nouvel arrivant est affecté n'est jamais de très grande généralité ; on ne commence pas par la catégorie "arts" ou "sciences" mais plutôt en "peinture sur bois" ou en "sociologie de la famille", c'est-à-dire dans une catégorie qui ne contient pas elle-même de sous-catégorie. Une fois accepté, l'éditeur procède à l'alimentation de la catégorie (avec de nouvelles références) et corrige au besoin les références qui y sont déjà à son arrivée. C'est sa compétence thématique qui lui permet d'opérer une veille efficace sur le domaine dont il est responsable et, par conséquent, d'être informé de l'apparition de nouveaux sites

Internet sur le sujet. De nouveaux enregistrements lui sont également soumis par les visiteurs de l'annuaire ou par ses collègues éditeurs. Il lui revient d'en jauger la pertinence et l'adéquation par rapport au thème dont il s'occupe. Cette période d'apprentissage permet à la communauté d'évaluer le sérieux et le talent du nouvel arrivant durant une période d'écolage (dont la durée n'est pas prédéterminée). Elle est également un barrage contre les éventuelles velléités de sabotage d'une personne mal intentionnée. Lorsqu'il a fait ses preuves, un membre de l'annuaire peut (s'il le souhaite) poser sa candidature pour une autre rubrique (éventuellement d'un rang supérieur) ; celle-ci est à chaque fois examinée par ses pairs. Le responsable d'une catégorie peut non seulement intervenir sur les références de sites qu'elle contient mais également sur le contenu de toutes ses sous-catégories. Il existe donc un enjeu conséquent à être responsable d'une catégorie d'un haut niveau de généralité. Les éditeurs des catégories très spécifiques vivent parfois comme des intrusions l'intervention d'indexeurs des rubriques plus générales ; ce mode d'organisation nécessite donc des dispositifs de coordination et d'arbitrage [12]. Cette asymétrie entre le champ d'action des uns et des autres m'amène à avancer qu'il n'existe pas une hiérarchie mais deux : l'arbre thématique se double d'un arbre social [13].

Dimensions verticales

Dans cette section, j'affine l'affirmation selon laquelle l'annuaire est organisé selon une hiérarchie thématique. Je montre, d'une part, que plusieurs hiérarchies cohabitent et, d'autre part, que les relations verticales qui composent ces arbres ne sont pas toutes du même type.

À la naissance du répertoire, les fondateurs ont proposé une seule arborescence thématique destinée à accueillir toutes les catégories. Pour construire cette hiérarchie, ils ont tout d'abord envisagé d'employer la classification de Melvil Dewey, l'homologue américain du père (belge) de la classification décimale universelle, Paul Otlet [16, 18, 9, 2, 8]. Cet outil documentaire leur a cependant semblé peu adapté aux sites Internet. Leur deuxième tentative consista à calquer la structure des groupes de discussion *Usenet*. Bien entendu, celle-ci fournissait une simple première ébauche. Aucune limitation n'a été formulée a priori sur le nombre de branches et de niveaux taxinomiques acceptables de sorte que, dans une prolongation des Lumières, l'arbre pouvait en effet prétendre recouvrir toutes les aires du savoir humain. Un arbre à géométrie variable, en quelque sorte. Bien que la structure de départ ait été largement abrogée depuis le lancement du projet,

les majeures correspondent encore actuellement aux premiers niveaux de *Usenet*. Néanmoins, avec le temps, certaines catégories ont acquis un statut particulier. C'est ce qu'expliquent les sous-sections suivantes.

L'internationalisation

Conçu aux États-Unis, l'annuaire regroupe, à sa conception, des rubriques libellées exclusivement en anglais. Quelle que soit sa langue d'expression originale, le site Internet était inscrit sous une thématique exprimée en anglais. Avec l'arrivée chaque jour de volontaires de tous les horizons, cette configuration – non problématique à l'origine – est devenue de moins en moins évidente. Des communautés linguistiques non anglophones se constituent peu à peu via les forums de discussion internes de la communauté. La proposition de démultiplier l'annuaire selon la langue d'expression émerge dans ces groupes, elle est ensuite proposée à l'ensemble de la communauté, débattue, acceptée puis mise en pratique. La forme arborescente est néanmoins conservée. L'outil étant jugé opérant en l'état, les branches linguistiques sont créées comme des catégories particulières. Dès juillet 1998, une nouvelle majeure ("world") accueille ainsi les rubriques dédiées à chaque langue, puis les références à des sites Internet d'expression non anglaise sont progressivement renvoyées vers ces branches. Qualifiée d'internationalisation, cette ventilation prend du temps et s'accompagne de la création d'une arborescence propre à chaque langue. Dans la mesure du possible, celle-ci emprunte sa morphologie à la hiérarchie anglophone. La disponibilité de sites Internet dans une thématique prévaut néanmoins sur la contrainte d'homogénéité, de sorte que certaines catégories sont créées uniquement pour une branche linguistique déterminée. Assez logiquement, la disparité entre branches linguistiques s'accroît à mesure que l'on s'enfonce dans des rubriques spécifiques.

La régionalisation

Après la création des branches linguistiques, un deuxième mouvement – similaire – a été initié à propos de la couverture géographique, et non plus de la langue d'expression. Qualifié de régionalisation, ce deuxième mouvement a été motivé par le constat que le site des habitants de la petite ville de Montréal, en Bourgogne, comportant notamment l'horaire des offices paroissiaux, n'intéresse pas les personnes cherchant des informations sur l'organisation paroissiale de la France dans son ensemble. Ce constat ne traduit aucun jugement moral sur l'intérêt des sites dont la pertinence est circonscrite géographiquement. La seule considération en jeu est le service aux

usagers de l'outil : les éditeurs ont jugé efficace de rassembler les sites concernant des territoires limités plutôt que de les laisser éparpillés dans l'ensemble du répertoire. Au sein de chaque branche linguistique a donc été entreprise la ventilation vers une sous-catégorie intitulée "régional" des fiches descriptives de sites à portée géographique restreinte. Cette régionalisation comporte également plusieurs degrés. Les sites susceptibles d'intéresser tous les francophones, indépendamment de leur nationalité sont listés dans l'arborescence linguistique (donc ne sont pas envoyés sous la catégorie régionale). Les fiches concernant des sites réservés à des visiteurs d'une nationalité déterminée sont renvoyées, sous régional, dans une catégorie correspondant au pays en question (le portail du gouvernement fédéral belge se trouve donc sous world, français, régional, Belgique). Enfin, une subdivision en régions et villes permet encore de répartir les sites dans des catégories plus spécifiques si nécessaires. Pour autant que le nombre de références le justifie, une arborescence similaire à celle créée pour la branche thématique est agencée sous chacune de ces catégories.

Les enfants et les adolescents

Enfin, une arborescence parallèle a été créée pour un projet spécifique (et, avec elle, une nouvelle catégorie majeure). Ce projet consiste à recenser l'ensemble des sites Internet à portée didactique qui sont adéquats à une consultation par les enfants et les adolescents. Initié dans la partie anglophone de l'annuaire le 14 novembre 2000, ce projet apparaît donc dès la page d'accueil de l'annuaire (sous le nom de "kids and teens"). Il est également poursuivi par des indexeurs non anglophones de sorte qu'une telle catégorie de haut niveau existe en français (sous le nom de "enfants et adolescents"). Cette catégorie inclut également une arborescence thématique similaire à celle développée initialement de manière thématique. Il est intéressant de noter que l'internationalisation de la catégorie "kids and teens" opère d'une manière différente de toutes les autres majeures (comme "sciences" et "arts"). En effet, les langues autres que l'anglais ne sont pas sous la majeure "world" mais empruntent un proxy spécifique, directement sous "kids and teens". Intitulé "international", ce nœud remplit une fonction analogue à "world" (il regroupe les différentes branches linguistiques) mais, contrairement au reste de l'annuaire, la majeure "kids and teens" ne connaît pas de ventilation. Cette construction implique une répartition des champs d'actions différente : les responsables de la majeure "kids and teens" sont en mesure d'intervenir dans toutes les branches linguistiques couvrant ce sujet (puisque ces dernières sont – techniquement – des sous-catégories). Par

contre, un éditeur de la rubrique anglophone des "arts" qui souhaiterait intervenir dans la branche francophone correspondante devrait au préalable poser sa candidature à cette catégorie, comme tout autre éditeur (puisque, comme l'illustre la figure 3, la majeure "arts" est une branche parallèle à la branche francophone inscrite sous la majeure "world").

À proprement parler, il n'y a donc pas une arborescence mais plusieurs. Bien que l'annuaire se compose formellement d'un seul arbre, celui-ci regroupe, grâce aux catégories "world", "régional" ainsi que "enfants et adolescents", plusieurs hiérarchies thématiques parallèles. Cette multiplication parallèle des arborescences est représentée à la figure 3.

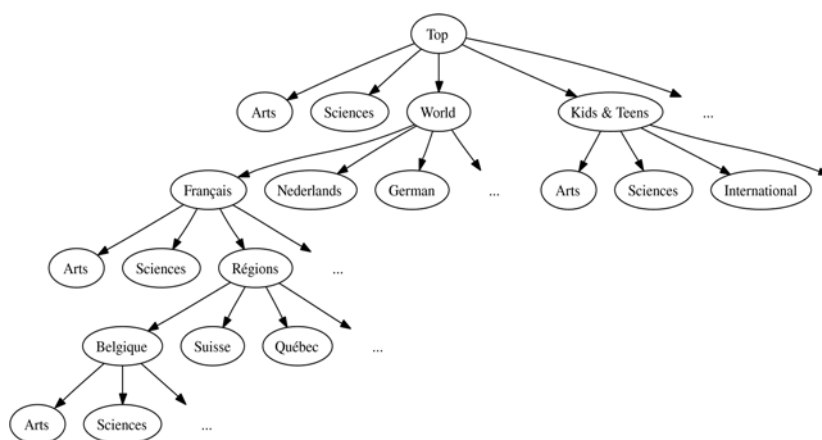


Fig. 3 : Plusieurs arborescences parallèles

Ces différentes arborescences se sont constituées successivement (ainsi que les ventilations nécessaires) au cours de l'histoire de la communauté, elles ont été débattues et négociées collectivement puis mises en œuvre par l'ensemble des éditeurs, chacun à leur niveau. Chacune de ces "étapes" est un processus ; loin d'être ponctuelle, chaque transformation a elle-même une durée. Si la répartition linguistique est désormais acquise, la vitesse à laquelle la régionalisation s'opère est variable d'une branche à l'autre. Les raisons peuvent être variées. Tel indexeur estime que maintenir sa catégorie à jour est un objectif premier par rapport à la régionalisation ; tel autre considère qu'il s'agit d'une décision absurde et, par conséquent, ne l'applique pas. Une grande catégorie (dont la nature la prédestine à être largement dispersée dans les différentes branches régionales) peut être laissée en l'état dans la mesure où son responsable, très actif (et donc précieux pour le projet global), est réfractaire à cette ventilation. Bien que le groupe argumente abondamment sur les forums internes en faveur de la cohérence de l'application des règles communes au projet, un régime d'exception pourra être toléré (mais régulièrement renégocié) par rapport au maintien d'une telle catégorie. Cet équilibre (dit "de transition" par le plus grand

nombre) peut se prolonger dans le temps et cela d'autant plus longtemps que le responsable est actif : la communauté préfère conserver une précieuse cheville ouvrière plutôt que de risquer, au nom de la morphologie du répertoire, de perdre une ressource qui peut s'avérer difficile (voire impossible) à remplacer.

Outre la coexistence de ces arborescences parallèles, je vais montrer dans la suite de cet article que la nature des liens verticaux de filiation est différente selon le nœud considéré.

Les significations de l'emboîtement

Parmi les différentes sémantiques qu'empruntent ces relations verticales figure la relation méronymique. Celle-ci unit le tout à

ses parties. Ainsi, la sous-catégorie "régions" sous "Belgique" témoigne-t-elle d'une signification de ce type : la Belgique comporte plusieurs régions et chaque région appartient à la Belgique. Une deuxième relation est celle de la thématique ou du point de vue. Les catégories thématiques – comme "art" ou "science" – sous "Belgique" ou "enfants et adolescents" ne signifient pas que l'art est une partie de la Belgique ou que les enfants sont constitués de

science mais que ces rubriques regrouperont des références à l'art en Belgique (thème) ou des introductions scientifiques susceptibles d'être consultées par des mineurs d'âge (point de vue). La relation hiérarchique peut comporter une troisième signification : celle de la relation générique. En insérant par exemple, sous la rubrique des sciences, la catégorie des sciences humaines et sociales, les éditeurs opèrent une relation du genre à l'espèce [1]. Enfin, la célèbre relation "is a" ("est un") est également représentée dans l'annuaire [15, 10]. Sous la catégorie régions (belges) se trouve notamment Bruxelles-Capitale. Or Bruxelles-Capitale n'est ni une partie de "région", ni un domaine thématique de "région". Tout comme la sculpture est un art ou la sociologie une science humaine, Bruxelles-Capitale est une région.

Cette multiplication des sémantiques que recouvrent les relations d'inclusion catégorielle témoigne aussi de l'indétermination d'une telle relation pour le sens commun. En effet, les éditeurs construisent un outil qu'ils veulent intuitif pour les usagers (ce faisant, ils mobilisent ce que représente pour eux le sens commun des internautes). En outre, les éditeurs sont recrutés, en grande majorité, sur base de leur compétence thémati-

que (plus, en tout cas, que pour leur goût envers les techniques documentaires). Loin de constituer une limite ou une faiblesse de l'outil, cette indétermination garantit au contraire flexibilité et adéquation de l'organisation catégorielle au contenu des sites Internet recensés.

Dimensions horizontales

Loin d'emprunter une sémantique unifiée, les relations verticales procèdent de logiques variées. En outre, comme je l'ai montré, les arborescences parallèles coexistent. Ces complexités peuvent amener le lecteur à se demander com-

lement". Chacun de ces renvois se présente comme un lien vers une autre catégorie, dont le chemin est visible in extenso. Cela signifie que le nom de ce lien correspond au nom de la catégorie cible. L'éditeur qui installe un tel lien ne peut donc pas introduire d'élément supplémentaire (ni de commentaire) au simple renvoi vers cette catégorie apparentée. Ce type de lien sert de passerelle soit entre deux thématiques connexes, soit entre deux arborescences parallèles.

Au niveau des thématiques connexes, par exemple, la catégorie dédiée à l'art de l'"illustration" contient un lien vers la catégorie concernant la "bande dessinée" ainsi que vers la rubrique dé-

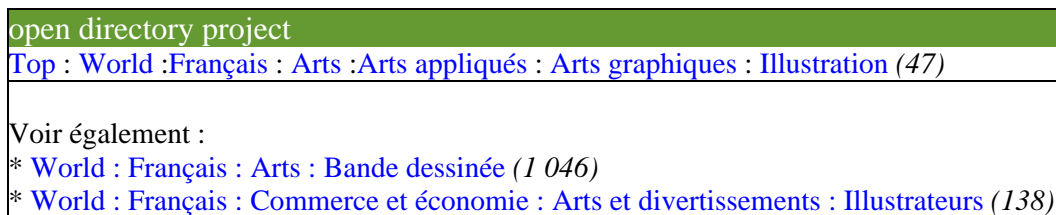


Fig. 4 : Catégories apparentées à l'illustration.

ment de tels outils – qui prétendent pourtant assister l'internaute dans son recueil d'information – peuvent tout simplement rester utilisables. Afin de ne pas perdre les utilisateurs dans les méandres de ces arborescences parallèles et concurrentes, les concepteurs de l'annuaire ont conçu des relations transversales qui tracent des voies latérales entre les différentes branches du répertoire. Ces relations horizontales sont de trois types : les catégories apparentées, les équivalences inter-linguistiques et les liens symboliques.

Les liens vers une catégorie apparentée

Les liens vers des catégories apparentées (aussi qualifiés de "liens connexes") sont des liens transversaux placés au sein d'une rubrique afin

diée à la profession d'"illustrateur". Bien qu'appartenant tout deux à la hiérarchie francophone, ces nœuds appartiennent à des branches distinctes, puisque la bande dessinée et l'illustration appartiennent à la catégorie des arts alors que la rubrique des illustrateurs s'inscrit dans la présentation de prestations rémunérées typiques de la branche "commerce et économie".

Les mêmes catégories fournissent un exemple de lien vers une catégorie apparentée qui opère la jonction entre deux arborescences. La catégorie des bandes dessinées belges comporte ainsi un renvoi de ce type vers la catégorie générale de la bande dessinée, le lien établissant ce faisant une jonction transversale entre la hiérarchie "régionale" belge et le répertoire thématique francophone (Figure 5).

Les deux utilisations renvoient au même disposi-



Fig. 5 : Catégories apparentées à la bande dessinée belge

d'indiquer au visiteur l'existence d'autres catégories dignes d'intérêt. Au niveau de l'interface, ces liens vers des catégories apparentées figurent en dessous de la liste des sous-catégories et avant la liste des références de sites Internet (figure 4). Ils sont précédés d'une entête intitulée "voir éga-

tif. Il s'agit, dans les deux cas, de catégories apparentées. Bien que la communauté des éditeurs en préconise un usage pondéré, leur nombre n'est pas limité. Les liens entre catégories ne sont pas symétriques : ils sont orientés (d'une source à une cible) et rien n'oblige à ce qu'un

renvoi inverse soit inséré dans la catégorie cible. Au niveau de l'interface, il n'est donc pas possible de "remonter" un lien de ce type (c'est à dire d'accéder aux catégories qui comportent un lien vers la catégorie courante). Bien entendu, lorsque les éditeurs les trouvent pertinents, de tels liens réciproques sont apposés. Ainsi, la catégorie des bandes dessinées comporte-t-elle un lien vers la catégorie "illustration" (qui répond au lien établissant la relation dans l'autre sens).

Les liens connexes peuvent a priori être orientés vers n'importe quelle catégorie de destination.

d'ailleurs dans d'autres dispositifs de l'annuaire ; elle est particulièrement explicite dans l'écran de mise en garde qui accueille le formulaire de création d'une nouvelle catégorie puisque celui-ci fait explicitement référence à Melvil Dewey (voir figure 6).

Les liens interlinguistiques

Les liens entre langues représentent un deuxième type de lien horizontal. Ces liens ont découlé de l'internationalisation évoquée ci-

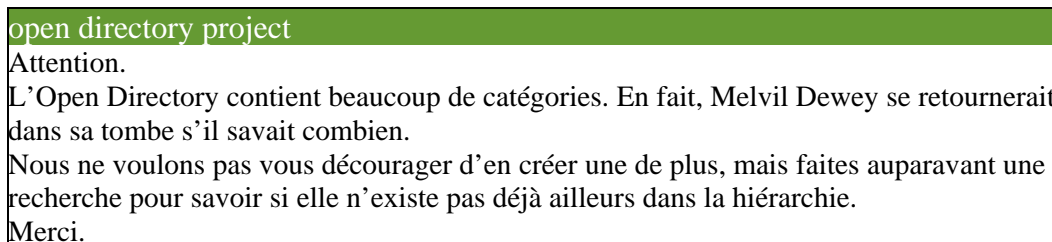


Fig. 6 : Écran de mise en garde lors de la création d'une catégorie

Cette absence de contrainte connaît une exception pour l'arborescence réservée aux mineurs ("kids and teens"). Il est en effet interdit de placer des renvois transversaux à partir d'une catégorie de cette branche vers une catégorie qui lui serait extérieure. Autrement dit, ces renvois peuvent exister au sein du répertoire "enfants et adolescents" mais ne peuvent en sortir. Cette mesure autarcique vise à éviter que les visiteurs (particuliers) de cette branche ne soient heurtés par un site listé dans une autre branche. Il est, par contre, tout à fait envisageable que des catégories thématiques comportent un tel renvoi vers une catégorie de la branches des enfants et des adolescents (comme l'illustre le milieu de la figure 7).

Le nom de l'entête qui figure au dessus des liens connexes, "voir également", évoque les renvois connus des documentalistes. De fait, cet emprunt aux sciences de la documentation procède de la même logique que les liens transversaux dans les thesauri [1]. Cette référence à la documentation – dont j'ai montré au début de la section "Dimensions verticales" qu'elle était présente dès la fondation de l'annuaire – apparaît

dessus. Dans un premier temps, les liens entre différentes branches linguistiques étaient tout simplement indiqués comme des liens de catégories apparentées. En octobre 2000, les concepteurs de l'annuaire introduisent de nouveaux liens transversaux dédiés spécifiquement aux équivalences linguistiques (les liens connexes



Fig. 7 : La catégorie dédiée à la bande dessinée (fragment)

retrouvant dès lors leur spécificité, décrite à la sous-section précédente).

Dans l'interface publique de l'annuaire, ces liens vers des catégories équivalentes en d'autres langues figurent sous les liens connexes, juste avant la liste des références aux sites Internet sur le sujet. Ils sont précédés du titre "cette catégorie en d'autres langues". Contrairement aux liens précédents, l'intitulé de chacun de ces renvois ne comporte pas son chemin. Il n'inclut pas non plus

le nom de la catégorie cible. Le nom de ces liens est celui de la langue vers laquelle il renvoie. Ce nom de langue est soit le nom de la sous-catégorie de "world" si la catégorie cible appartient à une branche de cette "majeure", soit "anglais" dans le cas contraire. Une catégorie déterminée ne peut avoir qu'une seule homologue par langue existant au sein du répertoire (c'est là une contrainte forte : le système ne permet pas aux responsables de catégories de renseigner plus d'un "équivalent" par branche linguistique). Ces liens sont censés être symétriques. Cette propriété n'est cependant pas assurée par le système : il revient à chaque éditeur de placer les équivalences avec d'autres langues.

Le bas de la figure 7 montre des exemples d'équivalences linguistiques. À la catégorie (francophone) "bande dessinée" correspondent ainsi la catégorie en chinois traditionnel (qui comporte, dans son chemin, le proxy "world") et la catégorie anglophone "comics" qui, langue des fondateurs oblige, dépend directement de la majeure "arts" (sans proxy ni précision de la langue "English").

Les liens symboliques

Les liens symboliques témoignent d'une troisième et dernière manière de traverser horizontalement le répertoire. Ceux-ci empruntent néanmoins une sémantique particulière. La catégorie cible qu'ils désignent est considérée comme une sous-catégorie de la catégorie courante. La catégorie francophone destinée à la bande dessinée contient un lien de ce type vers la branche régionale belge (ce lien est visible en haut de la figure 7). Les bandes dessinées belges constituent en effet bien un sous-ensemble des bandes dessinées, c'est pour cette raison que le lien se situe parmi les autres sous-catégories (auteurs, éditeurs, etc.) tout en étant signalé (grâce à l'arobase) comme étant un lien vers une catégorie "distante".

Cet usage du lien symbolique correspond à son acception horizontale. Toutefois, il existe des configurations qui font un usage vertical des liens symboliques. Dans ce type d'usage, la catégorie dans laquelle le lien symbolique est placé appartient à la même branche que la catégorie cible. Dans cette configuration, le lien symbolique est qualifié de raccourcis. Bien qu'elle puisse apparaître singulière, cette situation semble bien être celle qui fut le lieu d'émergence des liens symboliques. Tant les membres que j'ai rencontrés que les analystes anglo-saxons [20] en attribuent d'ailleurs la paternité à l'annuaire commercial qui

préexistait à l'*Open Directory Project : Yahoo!*. On doit donc aux concepteurs de cet autre annuaire la notation dotée d'un arobase final. Par contre, la qualification de "lien symbolique" revient à la communauté anglophone de l'*Open Directory*. Ce terme fait référence à des liens utilisés sur les systèmes d'exploitations Unix et Gnu/Linux [13]. La figure 8 reprend l'illustration graphique pour illustrer les trois types de relations horizontales décrites dans les sous-sections précédentes.

Conclusion

À travers l'examen de la structure de l'*Open Directory*, j'ai montré comment une communauté a construit un outil riche et complexe. J'ai ainsi explicité comment la communauté avait négocié l'internationalisation de ses membres à travers la

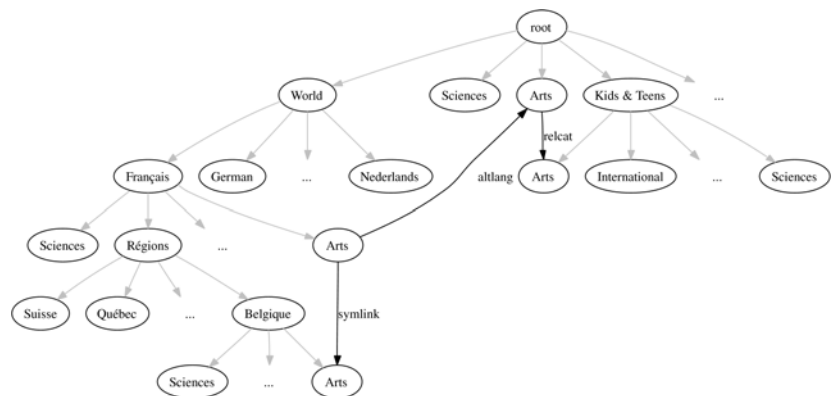


Fig. 8 : Les trois relations horizontales

duplication de l'arborescence. Les éditeurs ont témoigné d'une belle inventivité en tirant parti d'une architecture robuste plutôt qu'en la réformant. Ces éditeurs manifestent en outre une culture du débat et de la négociation, rendus nécessaires par le nombre de membres impliqués. J'ai ainsi montré combien la question de la ventilation des catégories recèle d'enjeux sociaux.

À cette organisation de la vie sociale, s'ajoute une culture professionnelle (que les sociologues qualifieront, selon leur orientation, d'*ethos* ou de monde vécu). J'ai montré que celle-ci se compose d'emprunts à l'informatique, aux sciences documentaires ainsi qu'aux techniques propres aux annuaires de sites Internet.

Christophe LEJEUNE

Université de Technologie de Troyes
12, Rue Marie Curie - BP2060
10000 Troyes
France
christophe.lejeune@utt.fr

4 septembre 2006

Bibliographie

1. Aitchison, Jean et Gilchrist, Alan. *Construire un thésaurus : manuel pratique*. Paris : ADBS Éditions, 1992.
2. Bay, Christine ; Boitte, Marguerite ; Cerisier, Bettina ; Demoustiez, Christelle ; Füg, Jean-François ; Lecocq, Suzanne ; Lefebvre, Daniel ; Manfroid, Stéphanie ; Mascolo, Tony (ed.) *Un Internet de papier*. Mundaneum, Mons, 1998.
3. Beauvisage, Thomas ; Assadi, Houssein. Les annuaires du Web. Entre intermédiation neutre et choix éditorial marqué. *Réseaux*, 2002, Vol. 20, n° 116, pp. 141-170.
4. Bourdoncle, François. Livetopics : Recherche visuelle d'information sur l'Internet. *La documentation française, Dossiers de l'audiovisuel*, 1997, Vol. 74, pp. 36-38.
5. Bénel, Aurélien ; Zaher, L'Hédi ; Cahier, Jean-Pierre ; Zacklad, Manuel. Gestion des identités pour un web socio-sémantique, document de travail, Laboratoire Tech-CICO, Troyes : UTT.
6. Callon, Michel. Éléments pour une sociologie de la traduction. La domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc. *L'Année sociologique*, 1986, Vol. 36, pp. 169-208.
7. de Saussure, Ferdinand. *Cours de linguistique générale*. Paris : Payot, 1995.
8. Fayet-Scribe, Sylvie. *Histoire de la documentation en France. Culture, science et technologie de l'information : 1895 - 1937*. Paris : CNRS, 2000.
9. Füg, Jean-François. Le mondaneum d'Otlet à Internet, une machine à faire la paix. in *Information et Documentation : du réel au virtuel*, Roelants-Abraham, Josiane (ed.), ULB, 1998.
10. Gilchrist, Alan. Thesauri, taxonomies and ontologies - an etymological note. *Journal of Documentation*, 2003, Vol. 59, n° 1, pp. 7-18.
11. Latour, Bruno. *La Science en action*. Paris : Gallimard, 1995.
12. Lejeune, Christophe. Indexation et organisation de la connaissance. La régulation des décisions sur un forum de discussion. *Les cahiers du numérique*, 2002, Vol. 3, n° 2, pp. 129-144.
13. Lejeune, Christophe. Symlinks as boundary objects. *TeamEthno Online*, 2006, Vol. °2, 108-111.
14. G & C. Merriam Co. *Webster's Revised Unabridged Dictionary*. 1913.
15. Michiels, Archibald. *Traitement du langage naturel et Prolog. Une introduction*. Paris : Hermès, 1991.
16. Otlet, Paul. *Traité de documentation. Le livre sur le livre. Théorie et pratique*. Bruxelles : Mundaneum, 1934.
17. Schwartz, Randal ; Phoenix, Tom. *Introduction à Perl*. O'Reilly, Paris, 2002.
18. Van Slype, Georges. La genèse d'une science de l'information et d'une documentologie. in *Cent ans de l'Office International de Bibliographie*. Mons : Mundaneum, 1995, pp. 161-179.
19. Tanenbaum, Andrew. *Systèmes d'exploitation*. Paris : Pearson Education France, 2003.
20. Thomas, Alan ; Shearer, James(ed.) *Internet Searching and Indexing. The Subject Approach*. New York : The Haworth Press, 2000.

Notes

- ¹ Pour leurs précieux commentaires, je tiens à remercier Diéyi Diouf, documentaliste à l'université de Dakar ainsi que

le comité de l'Association Belge de Documentation. Pour leur soutien, à la fois logistique et intellectuel, merci à Alexandre Serres (unité régionale de formation à l'information scientifique et technique à Rennes) et Didier Vrancken (président de l'institut des Sciences Humaines et Sociales de l'Université de Liège). Je tiens enfin à remercier chaleureusement les membres de la branche francophone de l'*Open Directory* qui m'ont accueilli, quatre ans durant, de manière particulièrement conviviale. Je reste évidemment seul responsable des erreurs qui pourraient subsister dans cet article.

² <<http://www.dmoz.org/>>

³ Par exemple, <<http://dmoz.org/cgi-bin/send.cgi?toeditor=freezincertain>>

⁴ Par exemple, <<http://dmoz.org/cgi-bin/update.cgi?where=Reference/Libraries/Public>>

⁵ Par exemple, <<http://dmoz.org/cgi-bin/add.cgi?where=Reference/Libraries/Public>>

⁶ Par exemple, <<http://dmoz.org/cgi-bin/apply.cgi?where=Reference/Libraries/Public>>

⁷ Le succès des annuaires demande à être relativisé. En effet, même aujourd'hui, la majorité des utilisateurs d'outils de recherche ne connaissent pas leur existence (même lorsque les outils auxquels ils recourent reposent souvent en partie sur la base de données d'un annuaire. Pour une revue critique des usages des différents outils de recherche sur Internet, voir [3].