

## **Quand le lézard s'en mêle... Ethnographie de l'indexation collective de sites internet**

Christophe LEJEUNE\*

Les utilisateurs du réseau Internet partent souvent d'outils de recherche d'information pour se rendre *in fine* aux documents qui les intéressent (Zaher & al, 2006). Parmi les outils disponibles, figurent les moteurs de recherche et les annuaires de sites Internet (Nicholson, 2000). A la différence des moteurs, les annuaires sont organisés en catégories thématiques s'emboîtant, de la plus générale à la plus particulière, dans une logique arborescente. En outre, là où les données des moteurs sont indexées automatiquement, la base de données des annuaires est le fruit d'une indexation patiente, entrée après entrée, d'indexeurs humains (Beauvisage, Assadi, 2002).

Identifier les choix (collectivement négociés) présidant à l'élaboration de ces annuaires paraît crucial pour qui veut comprendre ces outils. À cette fin, nous avons mené une enquête ethnographique visant à observer la construction d'un annuaire de sites Internet : l'Open Directory Project<sup>1</sup>. Fondé il y a plus de dix ans (le 5 juin 1998), l'annuaire en question est constitué par des bénévoles qui inscrivent les références de sites Internet

<sup>0</sup> Docteur en sociologie, Centre d'Expertise en Méthodologie et Analyse de Données, Institut des sciences humaines et sociales, Université de Liège, Bd du Rectorat, 7, B31 Bte 47, B4000 Liège, Belgique. [christophe.lejeune@ulg.ac.be](mailto:christophe.lejeune@ulg.ac.be)

<sup>1</sup> <http://dmoz.org/>

dans les différentes catégories d'un répertoire arborescent (Gilchrist, 2003). Ces volontaires se qualifient eux-mêmes d'indexeurs, d'évaluateurs, d'éditeurs ou de rédacteurs (Skrenta, 2001). Notre investigation a été conduite en deux temps. De 1999 à 2003, nous avons intégré l'équipe francophone des constructeurs de l'annuaire. En position d'observation participante, nous avons saisi l'invention collective de pratiques d'indexation en cours d'élaboration ; nous avons veillé à intégrer les pratiques du collectif en question afin d'en être reconnu comme un membre à part entière<sup>2</sup>. De 2003 à 2007, l'observation (cette fois, de manière distanciée) a cherché à comprendre ce que les pratiques inventées sont advenues. Ce suivi a principalement reposé sur l'analyse textuelle des forums de discussion internes à cette communauté au moyen de *Cassandra*<sup>3</sup>, un logiciel libre d'analyse qualitative<sup>4</sup> (Lejeune, 2008a, 2008 b). La comparaison des deux périodes met au jour que certaines pratiques, jusque-là informellement partagées, se sont institutionnalisées. D'autres ont été abandonnées, voire oubliées. Enfin, de nouvelles procédures ont été inventées afin de suppléer le manque consécutif à l'oubli ou à l'abandon des précédentes.

Nous rendons compte ci-dessous d'un aspect de nos observations, à savoir l'explicitation des pratiques d'évaluation et d'indexation des sites Internet à l'œuvre dans ce collectif. Celles-ci sont livrées sur un mode descriptif, afin de contribuer au débat méthodologique sur l'articulation entre les niveaux micro et macro en sciences sociales (Knorr-Cetina, Cicourel, 1981) : un des enjeux de cette description est d'illustrer la généralisation (au niveau « global ») de pratiques sociales à travers une description essentiellement locale et située.

La présente description voudrait également contribuer aux théories portant sur la particularité d'Internet comme milieu social. Elle entend tout d'abord contrer l'argument relativiste selon lequel, sur Internet, tous les sites se valent. Elle rappelle également que, dès 1998, bien avant le succès de l'encyclopédie libre Wikipédia, le collectif de l'Open Directory s'était organisé sur le modèle des communautés Open Source<sup>5</sup> et, compte tenu que les utilisateurs y produisaient les contenus du service, préfigurait déjà le mouvement du Web 2.0.

Mais, avant tout, elle propose de saisir, dans le détail de leur élaboration, des pratiques d'indexation (partagées mais implicites) régissant un annuaire de sites Internet. Dans une première section, nous nous intéressons aux parcours qu'empruntent ces évaluateurs de sites. Dans les sections suivantes, nous explicitons les critères d'évaluation des sites, puis les normes de mise en forme de leur enregistrement dans la base de données de l'annuaire.

<sup>2</sup> Cette première phase a constitué le terrain de notre thèse de doctorat (Lejeune, 2004).

<sup>3</sup> <http://cassandra-qda.sourceforge.net/>

<sup>4</sup> Ce logiciel a été conçu par nous même.

<sup>5</sup> Cet intérêt n'avait alors pas échappé à la société Nescape qui s'associa au projet dès 1998, lui fournissant ainsi sa mascotte (le lézard auquel le titre du présent article fait allusion).

## Collecte des indices

Contrairement aux parcours centrifuges qu'empruntent les utilisateurs des annuaires et des moteurs de recherche, les trajectoires des constructeurs de l'annuaire sont centripètes, puisqu'elles consistent à ramener, des sites visités, une série d'informations à intégrer dans l'annuaire. Comme l'explorateur des mondes distants (Latour, 1993 : 173), l'éditeur quitte son laboratoire (ici l'annuaire) pour visiter des contrées inconnues (ici Internet); lors de son expédition, il observe et évalue<sup>6</sup> des sites Internet. Il collecte des informations et des échantillons qu'il ramène au laboratoire (l'annuaire). Il y a donc eu déplacement d'information des sites vers l'annuaire (Latour, 1995 : 163; Latour, 1996 : 24). Une fois ces éléments collectés, l'éditeur extrait les informations jugées pertinentes pour l'utilisateur, afin de ne conserver qu'un compte rendu synthétique qu'il enregistre dans l'annuaire; il y a donc également eu réduction de l'information<sup>7</sup>.

Des nombreux textes, images, couleurs et sons présents sur les sites Internet, seules des coordonnées (l'adresse du site) et quelques lignes de description sont conservées. Néanmoins, cette réduction permet de regrouper des milliers d'inscriptions. Ce mouvement rend donc disponibles et accessibles la multitude des sites, visités et consignés par les explorateurs-indexeurs de manière synoptique. La réduction s'accompagne donc d'une amplification ; l'annuaire peut être qualifié de centre d'accumulation (Latour, 2006 ; Desrosières, 1993). Cette vertu lui vaut d'être choisi comme point de départ de bien des recherches d'information.

Afin de mieux comprendre comment est construit cet outil, il est nécessaire d' expliciter les procédures opérées par ses membres lorsqu'ils visitent, évaluent et indexent les sites. On distingue la collecte des indices (lors de la visite des sites candidats à l'indexation) et l'enregistrement dans la base de données (lors de la complétion de l'interface d'indexation).

Tous les sites présents sur Internet ne sont pas inclus dans l'annuaire. Les éditeurs en évalue préalablement l'apport, la fiabilité ainsi que l'adéquation avec une certaine politique éditoriale (par exemple la branche francophone de l'Open Directory Project refuse systématiquement les sites d'orientation raciste ou pédo-pornographique). Pour ce faire, ils collectent une série d'indices provenant du contenu du site, de son ergonomie et de son adresse (Ginzburg, 1980).

Le contenu du site, son apport et sa fiabilité sont donc déterminants. L'éditeur commence par se documenter au sujet des sites qu'il évalue. Selon l'ambiguïté de ces derniers et la détermination de l'indexeur, l'enquête peut prendre des extensions plus ou moins lointaines. La stratégie la plus courante consiste à chercher, au delà du site lui-même, des informations

<sup>6</sup> Les ressorts de cette évaluation sont détaillés dans la suite de cet article.

<sup>7</sup> Cette précision n'est pas gratuite. Il existe, en effet, des projets (comme [archive.org](http://archive.org)) qui visent bel et bien à rapatrier en un seul centre la totalité des sites visités.

corroborant son contenu ou permettant de se faire une idée sur son statut<sup>8</sup>. Il arrive parfois que l'investigation se prolonge hors d'Internet, comme en témoigne l'exemple suivant. Un éditeur de la catégorie « sciences humaines et sociales » rapporte son désarroi lorsque la candidature du site « swemorph.com » lui est soumise. Le site auquel le conduit cette adresse appartient à la « société morphologique suédoise » et porte sur une méthode qualifiée d'analyse morphologique générale permettant de construire des modèles « non quantifiés ». Bien que le site présente une apparence et des formats proches des autres sites scientifiques, l'éditeur est réservé. Tout d'abord, les références bibliographiques avancées lui sont inconnues. Ensuite, ni l'adresse, ni le courriel de la personne ayant soumis son adresse ne comportent le nom de domaine d'une institution comparable à une université. Enfin, la présentation du champ de recherche en question lui semble très vague. Il emprunte dès lors la stratégie classique et cherche sur Internet une trace de ce courant de recherche. Alors que les références citées sur le site suédois remontent aux années 1940, la discipline n'est représentée nulle part ailleurs. Vu que peu d'éléments concluants sont rassemblés, l'indexeur tente d'entrer en contact avec l'auteur du site. Une correspondance électronique est engagée. Un échange de documents s'amorce, fournissant des critères matériels à l'évaluation (via deux documents – l'un au format portable PDF – envoyé par courrier électronique puis l'autre par courrier postal). L'éditeur a pu alors juger sur pièces la pertinence du courant en question et finalement décider de l'enregistrement du site dans l'annuaire.

Le contenu du site n'est pas le seul élément pris en compte par les évaluateurs. Son ergonomie est jaugée, notamment à l'aune d'une barre de menus agissant comme une table des matières clairement intitulées, facilement accessibles et divisant le site en thématiques cohérentes<sup>9</sup>. Le « comportement » du site est également un facteur pris en compte. Certains sites ouvrent des fenêtres de navigateurs supplémentaires, le plus souvent pour inviter le visiteur à visiter un site ami ou, plus généralement, pour faire de la publicité. Cette pratique, jugée agressive, est tolérée lorsqu'elle concerne l'ouverture d'une ou deux de ces fenêtres (appelée *pop-up* ou console). Lorsque plus de deux *pop-up* apparaissent dans la page d'accueil, ou lorsque le visiteur est confronté à une chaîne de consoles (une nouvelle s'ouvrant automatiquement à chaque fois que l'on en ferme une), le site est rejeté. Cette mesure est particulièrement discriminante dans le cas des sites de jeu d'argent, de hasard ou pornographique.

Les indexeurs examinent également l'adresse du site Internet (URL) comme une ressource pour l'indexation. Accessible dans la barre d'adresse (le plus souvent au dessus de la fenêtre de visualisation du navigateur), celle-ci est rarement observée par la plupart des internautes. Les différentes composantes de l'adresse apportent chacune un renseignement. Le nom de domaine est considéré comme un indicateur de la thématique ou de

<sup>8</sup> Le site « hoaxbuster.com » s'est rendu célèbre en publiant des enquêtes du même type statuant sur les rumeurs qui circulent sur Internet.

<sup>9</sup> Nous verrons également que les menus servent de première approximation à l'éditeur pour la rédaction de la description du site Internet.

l'institutionnalisation du site. Son suffixe peut être un indice de sa nationalité ou du pays dans lequel est hébergé le serveur (.fr pour la France, par exemple) ; mais il peut également être un indicateur de son activité (.com indiquant une activité commerciale américaine alors que .org est – théoriquement du moins – réservé aux sites non commerciaux).

Les indexeurs tiennent aussi compte de la structure des répertoires à droite du nom de domaine. Dans le cas de sites d'hébergement, le premier niveau de la structure permet d'identifier les comptes des différents utilisateurs. Sur les sites de grandes sociétés, ces premiers répertoires peuvent correspondre aux sites des différents départements ou des membres du personnel. Les sous-répertoires suivants (c'est-à-dire ceux qui sont plus à droite encore) indiquent la division thématique du site. Ceux-ci varient (contrairement au reste de l'adresse) au cours de la navigation au sein du site Internet. Leur organisation peut donc offrir une aide à la visite du site.

Dans certains cas, les noms des répertoires et des pages du site permettent d'orienter le rédacteur vers la catégorie adaptée dans lequel inscrire l'adresse<sup>10</sup>. Le fait que les noms des pages et des répertoires fournissent des critères pour l'édition permet de soutenir, comme Nielsen (1999), que l'adresse (« URL ») fait partie intégrante de l'interface utilisateur (« UI ») du site Internet (Bénel, 2003 : 20).

## Conventions d'indexation

Pour les constructeurs de l'annuaire, la visite et l'évaluation d'un site Internet (à l'aune des critères explicités ci-avant) débouche sur son indexation. Celle-ci se réalise grâce à un masque de saisie informatique (voir figure 1) composé de différents champs. Leur complétion est régie par des normes de rédaction (1) d'une adresse (URL), (2) d'un titre, (3) d'une description, (4) d'une catégorie et (5) d'un commentaire qui sont respectivement traitées dans les sous-sections suivantes.

<b>Adresse:</b> [nouveau]	http://w3.scsoc.univ-tlse2.fr/
<b>Titre:</b>	Sciences de la Société
<b>Description:</b> <small>S.V.P. soyez clair, précis et concis.</small>	Revue scientifique interdisciplinaire. Présentation, comité, numéros parus et consignes aux auteurs.
<b>Catégorie:</b>	Sciences: <a href="#">Sciences humaines et sociales</a>
<b>Déplacer vers une autre catégorie:</b>	Entrez une autre catégorie ci-dessous ou sélectionnez-en une dans cette liste ▾
	Sciences/Sciences_humaines_et_sociales/Revue_et_publications/
<b>Note:</b> <small>Note interne destinée aux autres éditeurs.</small>	http://w3.univ-tlse2.fr/scsoc/ = http://w3.scsoc.univ-tlse2.fr/ <a href="#">Voir les notes</a> <a href="#">dernière note</a>
<input type="button" value="Mettre à jour"/> <input type="button" value="Annuler"/> <input type="button" value="Réinitialiser"/>	

<sup>10</sup> Comme les menus et les boutons, les noms de répertoires constituent également des ressources pour la description du site.

## **Normes de rédaction des adresses**

Les constructeurs de l'annuaire veillent à n'inscrire chaque site qu'une seule fois. Dès lors, afin d'identifier les adresses déjà enregistrées, ils se sont dotés de règles quant à la standardisation de leur écriture. Cette normalisation prescrit que, dans la mesure du possible, les adresses se terminent plutôt par le nom d'un répertoire que celui d'une page. L'épreuve mise en place est la suivante. Si une adresse candidate à l'inscription se termine par le nom d'une page (donc, par exemple, par les lettres php, htm ou html), l'éditeur tente de visiter le répertoire qui contient cette page. Il entre donc dans son navigateur une adresse tronquée du nom de la page. Pour ce faire, il soustrait la chaîne de caractères située à droite la dernière barre oblique « / » de l'adresse. Cette chaîne est, la plupart du temps, composée d'un nom puis d'un point suivi des lettres html ou htm. Le test vise à vérifier si le navigateur affiche quelque chose de différent selon que l'adresse demandée se termine par le nom d'une page ou par celui d'un répertoire. Si le serveur du site visité renvoie la même page, l'éditeur enregistre l'adresse tronquée. Cette pratique de troncature de l'adresse n'est pas idiosyncratique, elle est également attestée chez les internautes expérimentés. Un des spécialistes de l'Internet, Jakob Nielsen, l'évoque comme une pratique (non marginale) d'explorations des sites (Nielsen, 1999). S'adressant aux concepteurs de sites, Nielsen conseille d'ailleurs de configurer les serveurs de sorte que ces derniers renvoient bel et bien à une page lorsqu'une adresse tronquée est soumise (soit une page existante, soit une page générée par un script).

Les concepteurs de sites Internet (qui sont souvent des acteurs distincts des gestionnaires des serveurs) témoignent également d'une préférence envers ces adresses tronquées, lorsqu'il s'agit d'insérer sur leurs pages un lien vers un autre site Internet. Par exemple, la publication au 30 avril 1998 de la version 1.17 de la page d'accueil du site dédié à la diffusion du navigateur Mozilla comporte un commentaire qui atteste d'une normalisation de ce type. Le concepteur de cette page Internet recommande d'éviter de mentionner la page par défaut du répertoire visé par un tel lien. La norme mise en pratique par les éditeurs n'est donc pas une convention locale. Elle s'est diffusée, en 1998, comme une façon économique, persistante et durable d'indiquer les adresses. Elle rend transparentes pour l'utilisateur les modifications technologiques dans la gestion des pages. De telles modifications dépendent de l'hébergeur et du concepteur du site et peuvent avoir comme conséquence que la page d'accueil passe par exemple du nom `index.html` à `index.php`. Le fait de ne pas inscrire les pages permet d'éviter que, lorsqu'une modification technologique intervient, le lien pointe vers l'adresse d'une page qui n'existe plus – et génère donc une erreur pour l'utilisateur.

La proximité entre la pratique des indexeurs et celle des gestionnaires de serveurs n'est pas fortuite. En effet, si les indexeurs préfèrent enregistrer la racine d'un répertoire plutôt qu'une page, c'est précisément parce qu'ils tiennent compte des stratégies mises en œuvre, respectivement, par les

gestionnaires des serveurs (pour cacher le contenu de certains de leurs répertoires), par les concepteurs de sites (pour limiter les opérations de maintenance) et par les internautes (pour trouver l'information recherchée). Cette standardisation de l'écriture des adresses de sites Internet (URL) découle donc à la fois d'un chassé croisé entre gestionnaires, concepteurs et visiteurs de sites, et d'une intervention d'un tiers (l'indexeur) dans cette dynamique.

### **Normes de rédaction des titres**

L'indexation suppose également d'attribuer un titre au site Internet (enregistré dans le deuxième champ de l'interface, figure 1). Parmi les normes présidant à la rédaction du titre, on trouve la fidélité aux choix du concepteur du site, l'exigence d'adéquation avec le contenu du site, l'intelligibilité et le barrage opposé aux stratégies de positionnement. La première convention consiste à partir de la page d'accueil du site. En effet, le concepteur du site appose, souvent un nom de baptême à son site, soit relatif à un sujet particulier, soit correspondant au nom de l'organisation, du groupe, de la personne, de l'association ou de la société dont il est l'émanation (et souvent la vitrine). Le travail de l'éditeur consiste alors à rapprocher le titre figurant sur cette page d'accueil du contenu effectif de celui-ci. Lorsque le nom de baptême est adéquat et explicite, il est enregistré comme titre.

A ces principes (de fidélité à la volonté de l'auteur et d'adéquation au contenu) s'ajoutent des règles de formulation : les titres sont typiquement des nominalisations (et non des phrases) ; ils ne commencent normalement pas par un article mais par un nom dont seule la première lettre est capitalisée ; seuls les acronymes, les sigles et les noms propres ont droit à une capitale initiale (les noms communs ne sont capitalisés que s'ils sont les premiers mots du titre) ; aucun terme ne peut être entièrement capitalisé ; une série de termes – jugés inutiles – sont proscrits, comme « en ligne » ou « bienvenue ». Les contraintes de mise en forme typographique, de majuscule initiale et de suppression de l'article sont motivées par une caractéristique technique du mode d'affichage de la liste des adresses contenues dans une catégorie : c'est le traditionnel classement alphabétique qui s'est imposé.

### **Normes de rédaction des descriptions**

Contrairement au titre, la description se compose de phrases. Au niveau du fond, celles-ci correspondent au contenu effectif du site. Au niveau de la forme, les règles usuelles de ponctuation et de grammaire sont de rigueur. Une série (non exhaustive) de termes sont exclus de la description ; il s'agit de termes considérés comme non neutres, ou de termes n'apportant pas d'information au visiteur (cette liste comprend, par exemple, « bienvenue », « meilleur », « en ligne », « site »). La pratique veut que la description se

compose de deux phrases. La première phrase porte sur une présentation du sujet traité par le site. Cette présentation se situe à un niveau d'abstraction intermédiaire entre le titre du site et le nom de la catégorie. La seconde phrase détaille les ressources que le site propose (carte, vidéo, image, formulaire, répertoire...). La seconde phrase prend parfois appui sur le nom des différents menus de navigation du site. Les sites Internet proposent en effet souvent une répartition en menus des pages qu'ils contiennent. Comme nous l'avons exposé plus haut, l'évaluation du site tient compte de l'aide à la navigation que constituent les menus. A ce stade de l'indexation, les menus jouent un autre rôle. Ils constituent les indices de la description (structurée) du contenu<sup>11</sup>. Les indexeurs s'appuient par conséquent sur leurs intitulés pour rédiger la deuxième phrase de la description. Néanmoins toutes les différentes rubriques ne sont pas reprises telles quelles par l'éditeur. Comme pour le titre, elles doivent passer une épreuve de pertinence qui consiste à rapprocher le contenu effectif du site (établi sur la base de son exploration) de ce qu'annonce le menu. Ce rapprochement amène à supprimer les items ne correspondant pas à un contenu effectif, à utiliser un terme plus générique ou plus « parlant » que celui choisi par le concepteur du site pour désigner le contenu attesté par la visite ou encore à mentionner une ressource présente sur le site mais non nécessairement signalée dans sa table des matières. Une ressource du même type réside dans la répartition des pages dans plusieurs répertoires sur le même serveur, que l'éditeur a relevé lors de son exploration du site. Si elle n'est pas remarquée par la plupart des visiteurs du site, les éditeurs sont, quant à eux, attentifs à ce genre de détails techniques, qui peuvent les aider dans la détermination des différents thèmes d'un site.

L'ensemble de ces éléments (menus, rubriques, noms des pages et des répertoires) apporte un guide à l'édition, une première approximation qui est rapprochée du contenu effectif des pages correspondantes. Les éditeurs se transmettent ces pratiques informelles comme des astuces, ils les découvrent par eux-mêmes ou en observent l'usage chez d'autres membres de la communauté. Elles deviennent dès lors des conventions partagées. Il arrive également qu'elles s'institutionnalisent avec le temps. C'est le cas de la division de la description en deux phrases, qui a intégré la liste des directives officielles d'édition en février 2002.

### **Choix de la catégorie**

L'indexation d'un site Internet dans un annuaire implique le classement de sa référence dans une catégorie thématique. Chaque éditeur prend en charge une ou plusieurs catégories; L'évaluation qu'il produit porte également sur la catégorie dans laquelle il est pertinent de lister la référence du site. S'il lui apparaît que le site aurait plus sa place dans une catégorie autre que les siennes, il peut encoder la fiche d'indexation du site évalué et la transmettre à l'éditeur qui s'occupe du thème en question. Ce choix requiert de l'éditeur

<sup>11</sup> Dans un autre contexte (celui des journaux en ligne), les rubriques peuvent procéder de l'explicitation des caractérisations professionnelles de leur « fabriquant » (Cotte, 2006).

qu'il connaisse les catégories thématiquement proches de celles qu'il maintient, susceptibles d'appartenir à une autre branche, ou gérées par une autre équipe d'indexeurs.

L'arborescence générale de l'annuaire recouvre en effet une catégorisation relevant de dimensions hétérogènes (le thème, la portée géographique, la langue et le public cible)<sup>12</sup>. Cette morphologie particulière résulte notamment du nombre de sites Internet recensés dans une langue donnée. Dès lors, le fait « que le découpage régional diffère selon la langue d'accès » et que « en déroulant ces arborescences, [...] cette différence s'accentue », n'apparaît « labyrinthique » que si on l'évalue à l'aune de standards classificatoires (Donze, 2006). Si, au contraire, on la rapporte aux pratiques sociales décrites jusqu'ici, cette différence entre les branches du répertoire se comprend aisément. Vu que plus de sites portant sur des localités américaines sont rédigés en langue anglaise qu'en français, la subdivision régionale des Etats-Unis dans l'annuaire est plus élaborée en anglais qu'en français.

### **Commentaires et mémoires**

Comme nous venons de le mentionner, il arrive que la fiche du site à évaluer soit passée par les mains d'autres éditeurs avant de parvenir à l'emplacement dans lequel elle est passée en revue. Afin de se coordonner, les éditeurs peuvent saisir un commentaire dans le quatrième champ de l'interface d'indexation (figure 1). Ces notes vont rejoindre la série de traces générées automatiquement par le système et accessibles aux éditeurs via l'historique de chacune des fiches. Les traces automatiques comportent l'identité de celui qui a procédé à une opération ainsi que la nature de cette opération. La mention de l'identité se résume par le pseudonyme de l'éditeur. L'opération effectuée peut procéder de la modification du titre, de la description ou de l'adresse, du déplacement d'une catégorie à une autre ou du déplacement de la face cachée à la face publique de l'annuaire (cette dernière opération est qualifiée de publication). Ces traces automatiques sont complétées par l'éventuel commentaire de l'éditeur. Ce commentaire documente l'évaluation déjà opérée, les enquêtes menées, des éventuels contacts avec la personne qui a soumis (ou conçu) le site ou le motif de la suppression de l'adresse. Les indexeurs y mentionnent leurs doutes, leurs impressions, la qualité du site, son accessibilité ou les intentions de la personne ayant soumis l'adresse (que celles-ci soient ouvertement déclarées – via une correspondance avec l'éditeur – ou suspectées – à cause, par exemple, de soumissions répétées). Au niveau individuel, elles permettent une économie cognitive (Rosch, 1978; Chomsky, 1969) typique de la cognition distribuée dans un artefact (Norman, 1993). Collectivement, elles assurent la coordination du travail collaboratif (CSCW) et correspondent à une cognition partagée entre acteurs (Conein, 2004).

<sup>12</sup> LEJEUNE (2006) expose les raisons historiques qui ont présidé au développement de ces branches parallèles.

En tant que réceptacle des motifs de rejet (ou de méfiance), elles constituent enfin la mémoire de l'élaboration d'une politique éditoriale collective. La controverse publique initiée en février 2000 – et suivie d'un procès – à propos de la présence d'objets néo-nazis sur le site de vente aux enchères du portail Yahoo! (qui propose par ailleurs l'annuaire concurrent de l'Open Directory Project) constitue l'événement déclencheur d'une réflexion collective au sein de la branche francophone. Quelques mois plus tard, après de nombreuses discussions (attestées sur les forums de discussion), les volontaires francophones décident d'exclure les références néo-nazies ou pédophiles, s'écartant sur ce point de la ligne éditoriale de la branche anglophone, qui fait d'habitude autorité. Vu que les directives d'indexations officielles du projet constitue la traduction des *guidelines* anglophones, aucune documentation officielle ne relate cette convention (qui, bien qu'informelle, est partagée). Celle-ci transparait cependant dans les notes d'édition commentant le rejet de telle ou telle fiche. L'histoire de cette convention, sa délibération et, si l'on peut dire, le schisme entre les deux branches de l'annuaire rappellent le contexte historique, un siècle auparavant, de l'élaboration d'un autre outil documentaire : la classification décimale universelle (CDU). Dans l'entre-deux-guerres, Paul Otlet et Henri La Fontaine, pétris d'idéaux universalistes et pacifiques, fondent cet outil afin de rendre l'information accessible à tous (Bay, 1998; Füeg, 1998; Haesenne, 2002). De la même manière, le rejet de sites jugés indésirables procède indéniablement d'une construction d'une représentation idéalisée d'Internet (non plus pour donner accès à toute l'information, mais pour rendre les documents indésirables moins accessibles). Même si les disparités de choix éditoriaux entre les branches du répertoire ne sont accessibles que dans les notes d'édition (et donc absentes de la face publique de l'annuaire), on peut toutefois se demander si la ligne éditoriale se ressent également dans la morphologie des catégories de l'annuaire<sup>13</sup>.

En organisant, dès 1998, un collectif dédié à la production de contenus sur le modèle des groupes de programmeurs de logiciels libres, l'Open Directory fait figure de précurseur du Web 2.0. L'adoption, en 2000, de leur production par le géant Google atteste du succès de cette communauté de bénévoles. Au sein des pratiques d'indexation informelles développées chez ces évaluateurs de sites Internet le travail engage des procédures de nature variée reposant sur des percepts, des critères institutionnels et des mémoires externalisées. Cette ethnographie de la tâche recèle en outre nombre d'éléments empiriques permettant d'alimenter la réflexion sur les activités coordonnées par le réseau. Tout d'abord, les enquêtes en crédibilité menées par les indexeurs attestent qu'Internet ne fonctionne pas « en vase clôt », indépendamment des médiations traditionnelles. Les pratiques décrites attestent d'une métrique selon laquelle tout ne se vaut pas sur Internet. En outre, la description de la généralisation (voire de l'institutionnalisation) de pratiques initialement informelles (comme les deux phrases de la description d'un site Internet ou la troncature de leurs adresses) entend contribuer au controverse méthodologique autour de l'articulation des niveaux micro et

<sup>13</sup> L'étude des indices des trois versions abrégées de la CDU (de 1958, 1986 et 1998) a ainsi montré ce qu'il est advenu des idéaux d'Otlet et La Fontaine (Courbières, Couzinet, 2006).

macro de la description sociologique. Le relevé des pratiques locales de normalisation des adresses Internet atteste qu'elles sont partagées par différentes communautés (l'annuaire, le site Mozilla et le chercheur Jakob Nielsen). Sans changer d'échelle, la description donne à voir l'installation d'une pratique standardisée. Il est donc possible, sans quitter le niveau de la description locale, de mettre en perspective les traces – historiquement et localement – situées de la généralisation d'une pratique en train de devenir globale. Au delà de ces enjeux, notre description permet surtout de mieux comprendre comment sont construits les annuaires et, en définitive, comment se construit leur morphologie particulière.

Lejeune Christophe (2008), paru dans *Sciences de la société*, n° 75, pp. 101-113.

## Références bibliographiques

BEAUVISAGE (T.), ASSADI (H.), 2002, « Les annuaires du web. Entre intermédiation neutre et choix éditorial marqué », *Réseaux*, vol. 20, n° 116, p. 141-170.

BAY (C.) *et al.*, 1998, *Un Internet de papier*, Mundaneum, Mons.

BÉNEL (A.), 2003, *Consultation assistée par ordinateur de la documentation en Sciences Humaines : Considérations épistémologiques, solutions opératoires et applications à l'archéologie*, Thèse de doctorat en informatique, Institut National des Sciences Appliquées de Lyon.

CHOMSKY (N.), 1969, *Le langage et la pensée*, Payot [Harcourt, Brace & World, 1968], Paris.

CONEIN, (B.), 2004, « Cognition distribuée, groupe social et technologie cognitive », *Réseaux*, n° 124/2, p. 53-79.

COTTE (D.), 2006, « La structure comme index : le rôle des interfaces documentaires dans la construction du sens », In TIMIMI, I., KOVACS, S. (eds), *Indice, index, indexation*, ADBS, Paris, p. 105-114.

COURBIÈRES (C.), COUZINET (V.), 2006, « Du bleu horizon à l'horizon documentaire : représentation des savoirs à l'aube de la construction européenne », In TIMIMI (I.), KOVACS (S.) (eds), *Indice, index, indexation*, ADBS, Paris, p. 81-90.

DESROSIÈRES (A.), 1993, *La politique des grands nombres. Histoire de la raison statistique*, La Découverte, Paris.

DONZE (I.), 2006, « "Pékin" ou "Beijing" ? La construction d'un thésaurus de noms de lieux et la problématique de l'indexation des noms géographiques », In TIMIMI (I.), KOVACS (S.) (eds), *Indice, index, indexation*, ADBS, Paris, p. 177-189.

FÜEG (J.-F.), 1998, « Le Mundaneum d'Otlet à Internet, une machine à faire la Paix », in: ROELANTS-ABRAHAM (J.), ed, *Information et Documentation : du réel au virtuel*, ULB.

GILCHRIST (A.), 2003, « Thesauri, taxonomies and ontologies - an etymological note », *Journal of Documentation*, 59/1, p. 7-18.

GINZBURG (C.), 1980, « Signes, traces, pistes. Racines d'un paradigme de l'indice », *Le Débat*, 6, p. 3-44.

HAESSENNE (N.), 2002, « Un bibliographe passionné », In HASQUIN (H.), LECOCQ (S.), LEFEBVRE (D.) *et al.*, *Henri La Fontaine. Prix Nobel de la Paix. Tracé[s] d'une vie*, Mundaneum, Mons, p. 89-103.

KNORR-CETINA (K.), CICOUREL (A. V.), ed., 1981, *Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies*, Routledge & Kegan Paul, Boston.

LATOUR (B.), 1993, *La clef de Berlin et autres leçons d'un amateur de sciences*, La Découverte, Paris.

LATOUR (B.), 1995, *La Science en action*, Gallimard [La Découverte, 1989], Paris.

LATOUR (B.), 1996, « Ces réseaux que la raison ignore : laboratoires, bibliothèques, collections », In BARATIN (M.), JACOB (C.) (eds), *Le pouvoir des bibliothèques. La mémoire des livres en Occident*, Albin Michel, Paris, p. 23-46.

LEJEUNE (C.), 2004, *Sociologie d'un annuaire de sites Internet. Les sciences documentaires saisies par l'informatique libre*, Thèse de doctorat en sociologie, Université de Liège.

LEJEUNE (C.), 2006, « D'un annuaire de sites Internet à l'organisation documentaire. Une sociologie des relations sémantiques », *Cahiers de Documentation*, n° 3, p. 12-22.

Lejeune Christophe (2008), paru dans *Sciences de la société*, n° 75, pp. 101-113.

LEJEUNE, (C.), 2008a, « Software features for qualitative analysis. A methodological and epistemological experiment with *Cassandra* », *7th International Conference on Social Science Methodology*, ISA, RC33, Naples.

LEJEUNE, (C.), 2008b, « Au fil de l'interprétation. L'apport des registres aux logiciels d'analyse qualitative », *Revue Suisse de Sociologie*, Vol. 34, n° 3, p. 593-603.

NICHOLSON (S.), 2000, « A Proposal for Categorization and Nomenclature for Web Search Tools », In THOMAS (A.), SHEARER (J.) (eds), *Internet Searching and Indexing. The Subject Approach*, The Haworth Press, p. 9-28.

NIELSEN (J.), 1999, « URL as UI », *Alertbox: Current Issues in Web Usability*.

NORMAN (D.), 1993, « Les artefacts cognitifs », *Raisons pratiques*, 4, p. 15-34.

ROSCH (E.), 1978, « Principles of Categorization », In ROSCH, (E.), LLOYD, (B.) (eds), *Cognition and Categorization*, Erlbaum, p. 27-48.

SKRENTA (R.), 2001, « Genesis of the Open Directory Project », *The Internet Developer Group*.

ZAHER (L.), CAHIER (J.-P.), ZACKLAD (M.), 2007, « De la recherche d'information à une recherche ouverte d'information », In MS. Bouhlef and B. Solaiman (eds), *Proceedings of the 4th International Conference on Sciences of electronics, Technologies of Information an Telecommunication (SETIT'2007)*, IEEE.