

Index, Google et bibliothèques

DOMINIQUE MANIEZ

dominique.maniez@wanadoo.fr

Après des études de philosophie, Dominique Maniez s'est reconverti dans l'informatique. Il a été développeur de bases de données, journaliste spécialisé, auteur et traducteur d'ouvrages sur les nouvelles technologies, professeur associé à l'Enssib et à l'université Lyon 2. Il enseigne actuellement les TICE aux étudiants qui préparent les concours de recrutement de l'Éducation nationale.

Quand il entre dans le hall de l'Enssib, le flâneur observateur peut apercevoir, en levant les yeux sur sa gauche, un assemblage de néons multicolores qui sont du plus bel effet à la tombée de la nuit. Cette œuvre, que l'on doit à Maurizio Nannucci¹, laisse souvent perplexes les visiteurs et rares sont ceux qui découvrent immédiatement que l'artiste a tout simplement imbriqué les lettres du mot INDEX. Que l'école qui est chargée de la formation du personnel des bibliothèques soit, dès que l'on en franchit le seuil, placée sous le signe de l'index est un symbole fort qui, à ma connaissance, n'a jamais été questionné.

Il faut préciser d'emblée qu'il y a quelque péril à parler des « bibliothèques » en général, tant il y en a de modèles différents. Cette précaution terminologique s'adresse également au deuxième objet d'étude, l'index, et on évoquera dans une rapide perspective historique des réalités très diverses qui partagent cependant le même nom. Il est important d'étudier les phénomènes techniques sur des temporalités longues, au rebours de la propension actuelle qui, quand on évoque des sujets touchant au numérique, fait fi des périodes antérieures dès qu'elles dépassent la vingtaine d'années, comme si la technologie pouvait être analysée intrinsèquement en éludant ses déterminismes socioculturels².

1. Voir photo en page suivante.

2. Même si les livres numériques n'existent que depuis une dizaine d'années, on notera que l'on convoque rarement l'histoire du livre pour en parler. De la même manière, le droit d'auteur est sommé de s'adapter aux technologies numériques comme si l'on pouvait balayer d'un revers de main tous les acquis en ce domaine depuis plus de deux cents ans. Cette perspective anhistorique

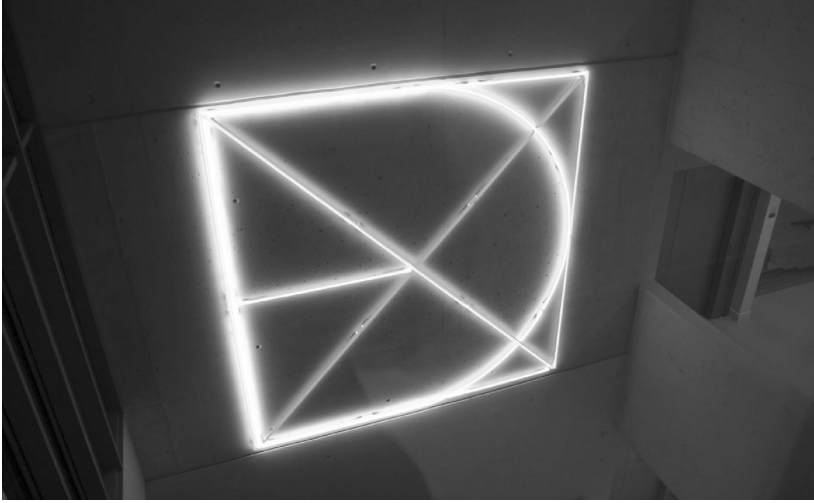
Même si cet article évoque des techniques documentaires très précises, il a pour ambition de rester dans un registre épistémologique et a pour objectif de montrer comment les domaines du livre et du document tissent un rapport dialectique avec cette technologie de l'intellect que l'on appelle index.

Des premiers index de livres au web sémantique

Il ne s'agit pas ici de retracer à grands traits l'histoire des index de livres³, mais d'en extraire certaines caractéristiques qui sont susceptibles d'éclairer le propos. À l'échelle de l'histoire du livre, les index sont finalement apparus relativement tard, car l'ordre alphabétique a mis du temps à s'imposer et les pratiques de lecture médiévales ne permettaient pas d'envisager cet outil de repérage comme indispensable. Si les index de livres ont pu se développer à partir du xvi^e siècle, c'est bien entendu à cause de l'essor de l'imprimerie, mais il serait illusoire d'imaginer que cette invention en soit l'unique raison. C'est en effet une conjonction de facteurs techniques, linguistiques et culturels qui a permis la modélisation de cet auxiliaire de lecture qui a encore sa place de nos jours à la fin de certains ouvrages. En perdant de vue cette origine multifactorielle et en oubliant la longue his-

du numérique est sans doute un indice des difficultés qu'entretiennent les nouvelles technologies avec le temps.

3. Le lecteur intéressé par cette question pourra se reporter à la deuxième partie de notre ouvrage écrit avec Jacques Maniez, *Concevoir l'Index d'un livre – Histoire, actualités et perspectives*, ADBS, 2009.



L'œuvre de Maurizio Nannucci, *Index*, allumée à la tombée de la nuit.

toire des index de livres, nous pensons que les informaticiens qui abordent aujourd'hui les problématiques de l'indexation du web vont dans la mauvaise direction, et laissent de côté des considérations épistémiques qui méritent d'être approfondies.

Il ne s'agit pas pour autant de reléguer aux oubliettes la technologie, et chacun, sans être spécialiste de la question, peut aisément constater que des progrès considérables ont été réalisés en matière d'indexation depuis une dizaine d'années. Dans cette optique, il est toujours intéressant de se replonger dans le passé, et nous renvoyons le lecteur à deux excellents articles de Dominique Lahary et de Jean-Claude Le Moal qui ont été publiés en 2002 dans le *BBF*⁴. À la lecture de ces articles, qui ne sont finalement pas si anciens que cela, chacun voit bien que les choses ont évolué à grande vitesse. Dominique Lahary tenait, par exemple, les propos suivants : « *Il y a enfin, ce qui agite les bibliothécaires qui peuvent y voir une confirmation ou au contraire une négation de leurs traditions, la notion de métadonnées (metadata), ces données sur les données qui peuvent être contenues dans la ressource électronique elle-même ou dans un enregistrement séparé, et qui font l'objet de formalisation comme le RDF (Resource Description Framework).* »

Aujourd'hui, RDF et OWL⁵ sont à la base du web de données, que l'on

4. Dominique Lahary, « Informatique et bibliothèques », *BBF*, 2002, n° 1, p. 60-67. En ligne : <http://bbf.enssib.fr/consulter/bbf-2002-01-0060-006>; Jean-Claude Le Moal, « La documentation numérique », *BBF*, 2002, n° 1, p. 68-72. En ligne : <http://bbf.enssib.fr/consulter/bbf-2002-01-0068-007>

5. OWL pour Web Ontology Language : extension de RDF qui sert à décrire ce que les informaticiens appellent des ontologies et qui, malgré ce que certains en pensent, n'ont pas

appelle également web sémantique⁶ ou web 3.0. Et l'on est bien obligé de se rendre à l'évidence : ces technologies sont aujourd'hui matures et implantées dans des outils que nous employons couramment. Par exemple, c'est le projet Isidore⁷ qui permet aux ressources de l'université ouverte des humanités (UOH)⁸ d'être moissonnées automatiquement. Projet d'envergure piloté par le TGE Adonis, Isidore est sans doute la réalisation française la plus emblématique du web sémantique à l'heure actuelle, mais il existe une myriade d'autres initiatives en cours dans ce secteur qui est foisonnant⁹.

Le tour de force du web sémantique est sans doute d'arriver à extraire des informations pertinentes de données qui à la base ne le sont pas forcé-

grand rapport avec la science de l'être en tant qu'étant...

6. Le lecteur peu au courant de la terminologie et des enjeux du web sémantique trouvera une bonne synthèse dans un dossier récent de la revue *Documentaliste - Sciences de l'information*, volume 48, n° 4, décembre 2011, dossier : « Web sémantique, web de données : quelle nouvelle donne ? » : www.adbs.fr/revue-docsi-volume-48-n-4-decembre-2011-dossier-web-semantique-web-de-donnees-quelle-nouvelle-donne-111766.htm

7. www.rechercheisidore.fr et <http://blog.antidot.net/tag/isidore> pour une description technique du projet.

8. L'UOH est l'université numérique thématique (UNT) qui est consacrée aux sciences humaines et sociales : www.uoh.fr

9. Le programme de l'édition 2012 de SemWeb.Pro (journée de conférences pour les professionnels du web sémantique) disponible à www.semweb.pro/blogentry/2713 donne une idée des principaux acteurs du domaine ainsi que des orientations de recherche. Bien entendu, le web sémantique n'attire pas que la recherche universitaire, et on trouvera quelques exemples d'utilisation en entreprise à : <http://dossierdoc.typepad.com/descripteurs/2012/02/le-web-s%C3%A9mantique-en-entreprise-quelques-cas-dusage.html>

→ Index, Google et bibliothèques :

ment. Par exemple, un projet comme DBpedia¹⁰, qui vise à sémantiser Wikipédia, arriverait presque à vous réconcilier avec cette encyclopédie collaborative.

Toujours plus vite

Pour résumer, grâce au web sémantique, on dispose aujourd'hui d'index construits de manière automatique qui permettent d'accéder à des données pertinentes si l'on emploie les bons outils. Il demeure cependant une question embarrassante : combien sommes-nous à utiliser ces bons outils ? Il ne s'agit pas ici de reprendre un procès contre Google qui a déjà été instruit¹¹, mais bien de constater qu'une immense majorité des étudiants, des enseignants et des personnels de bibliothèque (pour ne citer que ces catégories socioprofessionnelles qui sont censées avoir des rapports privilégiés avec l'information scientifique et technique) utilisent systématiquement le moteur de recherche de la firme de Mountain View en première intention dès qu'il s'agit de trouver la moindre information, qu'elle soit factuelle, anecdotique, savante ou technique. C'est un fait établi que peu de gens de bonne foi osent aujourd'hui contester, et le recours à Google est

10. <http://fr.dbpedia.org> et http://cblog.culture.fr/2012/02/23/linkeddata_dbpedia_webdedonnees pour une présentation du projet.

11. L'honnêteté nous impose d'ailleurs d'avouer que malgré un certain succès d'estime, ce combat n'a eu aucun impact ni aucune conséquence pratique, les utilisateurs des services de Google étant toujours de plus en plus nombreux. Il semblerait également que la modification récente des règles de confidentialité n'ait pas plus éveillé les consciences, malgré les protestations de la Commission nationale Informatique et libertés. Les lecteurs qui ignoreraient nos thèses pourront se reporter au premier chapitre de notre ouvrage, « Les 10 plaies d'Internet », qui est disponible en ligne à : <http://dunod.ebrochure.fr/10-plaies-internet/page1.html>

On trouvera également un réquisitoire brillant et solidement argumenté contre cette tendance à tout envisager sous le prisme de Google dans l'essai récent de Siva Vaidyanathan, *The Googolization of Everything: And Why We Should Worry*, University of California Press, 2012.

devenu un réflexe conditionné contre lequel il est sans doute vain de vouloir lutter.

Il y a cependant une difficulté supplémentaire : dans l'esprit de bon nombre d'internautes, il existe une équation plus ou moins consciente que l'on peut formuler en ces termes : Google = internet = bibliothèque. Même si la métaphore d'internet vue comme une immense bibliothèque a été employée assez tôt, au fur et à mesure que l'on s'est rendu compte que le web engrangeait des documents à une vitesse exponentielle, l'équivalence symbolique entre le moteur de recherche et le contenu de ce qu'il est censé indexer a été relativement peu étudiée. Chacun voit bien cependant que le fait que Google soit la page d'accueil d'un grand nombre de navigateurs ou que la fenêtre de recherche du navigateur soit quasi systématiquement celle de Google entretient l'illusion que le navigateur Google est le seul moyen d'accéder au web. Et, dans les faits, c'est bien ce qui se passe. Si c'est le rôle du navigateur d'interpréter le code des pages web pour l'afficher, le recours à Google est devenu indispensable pour accéder aux sites web, car plus grand monde ne prend la peine de saisir directement une adresse dans la barre d'adresses du navigateur. Par métonymie, Google devient le web, ce qui, si l'on met à part les données du web invisible, n'est pas tout à fait inexact, puisqu'il a pris soin d'inclure dans sa base de données toutes les pages d'internet qu'il a pu télécharger.

Si l'on admet que Google se résume à son index, il en découle une autre équation : index = internet = bibliothèque. Ce renversement dialectique est saisissant, car un bibliothécaire a sans doute du mal à imaginer que le catalogue de sa bibliothèque soit équivalent à la somme des livres qu'il inventorie. Cette confusion se fait bien évidemment au détriment de la bibliothèque, que l'on n'a plus besoin de fréquenter puisque l'on a accès à internet. Dans cette dérive sémantique, l'index est confondu avec la bibliothèque, comme si les métadonnées étaient devenues aussi importantes que les données. On serait presque tenté de résumer cette situa-

tion par la formule suivante : « *Quand le bibliothécaire montre la lune, l'imbécile regarde l'index.* »

Cette hérésie se double d'une autre plaie du monde contemporain, la dictature du temps réel. Si Google a arrêté depuis longtemps de communiquer sur la taille de son index, il continue à afficher le temps de chaque requête effectuée sur le moteur de recherche. Plus la requête est rapide, plus les liens commerciaux s'afficheront vite. De nombreux théoriciens, notamment Paul Virilio, ont réfléchi à cette fuite en avant qui consiste à vouloir toujours aller plus vite, et il faut bien reconnaître que l'usage omniprésent des nouvelles technologies de la communication a contribué à renforcer cette pente dangereuse. Cette quête de l'urgence qui est caractéristique des sociétés capitalistes modernes¹² participe à l'aliénation de l'internaute, qui est avant tout considéré comme un client potentiel, plutôt que comme un citoyen souhaitant s'informer. Ce désir d'instantanéité, qui frise parfois la névrose, est bien entendu un frein à la formation scientifique et technique, mais tous ceux qui fréquentent de près des étudiants savent pertinemment que la concurrence est devenue déloyale entre internet et la bibliothèque universitaire.

Pour être tout à fait franc, nous ne croyons pas à un retour en arrière. Faut-il pour autant baisser les bras et se résigner, en constatant tous les jours les lacunes produites par ces nouvelles habitudes de travail dans la formation intellectuelle des étudiants¹³? Assurément non! Il faut alors se battre sur le seul terrain qui vaille la peine et sur lequel nous pouvons exercer notre supériorité : l'humain. Même si cette formule peut paraître totalement galvaudée, il faut

remettre l'homme au centre de la bibliothèque et faire en sorte que le bibliothécaire redevienne un médiateur humaniste.

Face au web qui est un monstre froid et secret¹⁴ voyant en chaque internaute un consommateur potentiel¹⁵, le bibliothécaire doit jouer un rôle citoyen de premier plan si l'on croit encore aux vertus du service public ainsi qu'à la fonction émancipatrice de l'éducation à l'*Information Literacy*. Par le jeu de dérives technicistes, un certain nombre de professionnels de l'information-documentation ont réussi à mettre l'être humain au second plan, et il convient aujourd'hui de reconquérir le terrain perdu. Il est sans nul doute plus facile de dresser un constat d'échec que de proposer une solution miracle à ce problème, mais nous avons le sentiment que le monde des bibliothèques doit avant tout miser sur ses atouts et ne pas courir après Google en multipliant les partenariats, même si cela fait augmenter dans des proportions importantes les statistiques de consultation des bibliothèques numériques. Il ne s'agit à ce stade-là même plus d'une question d'éthique, mais de stratégie. Si l'on veut que nos concitoyens retournent dans nos bibliothèques pour chercher de l'information et non plus consommer des données, il faut miser sur l'humain, sinon les bibliothèques seront mises à l'index. ●

Juin 2012

12. On lira avec profit à ce sujet l'analyse que livre le philosophe Hartmut Rosa, notamment dans son dernier ouvrage, *Aliénation et accélération : vers une théorie critique de la modernité tardive*, La Découverte, 2012.

13. Même si nous ne partageons pas tous les points de vue de Nicholas Carr, il faut reconnaître au journaliste américain le mérite d'avoir mis les pieds dans le plat avec son dernier livre : *Internet rend-il bête? Réapprendre à lire et à penser dans un monde fragmenté*, Robert Laffont, 2011.

14. La transparence et l'*open data* sont des thèmes très à la mode, mais, à notre connaissance, Google n'a toujours pas publié les algorithmes de son moteur de recherche...

15. Si l'on reprend l'aphorisme d'Andrew Lewis, il arrive d'ailleurs parfois que le consommateur change de statut : « *If you are not paying for it, you're not the customer; you're the product being sold.* »