Jean-Claude LE MOAL, Bernard HIDOINE, Lisette CALDERAN, *La recherche d'information sur les réseaux (cours Inria, 30 septembre-4 octobre 2002)*

Paris, ADBS Éd., 2002, 322 p.

Dominique Nauroy

Édition électronique

URL : <https://journals.openedition.org/questionsdecommunication/7145>

DOI : 10.4000/questionsdecommunication.7145

ISSN : 2259-8901

Éditeur

Presses universitaires de Lorraine

Édition imprimée

Date de publication : 1 juillet 2004

ISBN : 978-2-86480-838-1

ISSN : 1633-5961

Référence électronique

Dominique Nauroy, « Jean-Claude LE MOAL, Bernard HIDOINE, Lisette CALDERAN, *La recherche d'information sur les réseaux (cours Inria, 30 septembre-4 octobre 2002)* », *Questions de communication* [En ligne], 5 | 2004, mis en ligne le 19 juillet 2013, consulté le 03 janvier 2023. URL : <http://journals.openedition.org/questionsdecommunication/7145> ; DOI : <https://doi.org/10.4000/questionsdecommunication.7145>

Creative Commons - Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International
- CC BY-NC-ND 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Jean-Claude LE MOAL, Bernard HIDOINE, Lisette CALDERAN, *La recherche d'information sur les réseaux* (cours Inria, 30 septembre-4 octobre 2002). Paris, ADBS Éd., 2002, 322 p.

Comment rechercher avec pertinence une information contenue dans l'une des pages disséminées par milliards sur le Web ? Comment les moteurs de recherche indexent-ils les données ? Comment se faire comprendre par eux et formuler sa requête ? Comment, enfin, se faire référencer par ces moteurs avec pertinence ? Pour la deuxième fois (cf. *La recherche d'information sur les réseaux. Internet : pour en savoir plus* (cours INRIA, 30 septembre-4 octobre 1996, Trégastel), Paris, ADBS Éd., 1996), l'Inria décrit les multiples facettes d'un monde qui n'a pas encore terminé ses mutations et qui propose, face à la croissance vertigineuse du volume d'informations disponibles sur le Web, des standards destinés à les structurer, des outils et techniques pour les exploiter. Dans ce substantiel cours de plus de 300 pages, huit interventions traitent des problématiques du fonctionnement des systèmes de recherche, d'une part, et de l'organisation des documents susceptible de les rendre facilement accessibles sur l'internet, d'autre part.

La recherche du document numérique est actuellement privilégiée par rapport à celle de documents imprimés, ce qui suscite différentes questions. La première porte sur la nature des ressources qui se révèlent hétérogènes du point de vue de leurs formats, de la qualité des auteurs et des éditeurs, des objectifs et des publics visés. Dès lors, deux autres questions affluent : les documents trouvés sont-ils pertinents, leur source est-elle fiable ? Se pose également le problème de la désintermédiation, utopie en perte de vitesse face à l'adaptation numérique des repères traditionnels propres au document imprimé, tels les catalogues des bibliothèques.

L'objectif de l'ouvrage est, avant tout, d'offrir une mise à jour des connaissances sur le sujet, à destination d'un public de professionnels ou d'initiés. Trois chapitres sont consacrés à un panorama des sources et des outils du Web : Sylvie Dalbin, consultante au GIE

Desybel, une institution spécialisée dans les systèmes d'organisation, de gestion et de recherche d'informations et de documents, fait le point sur les outils de recherche sur le Web en évoquant d'abord les moteurs et le travail de collecte effectué par leurs robots – voire par des professionnels qui offrent un service de réponses aux questions des internautes –, puis les annuaires, catalogues de ressources classés par catégories. Elle évoque aussi les outils collaboratifs, dont l'Open Directory Project, annuaire de ressources alimenté par les internautes eux-mêmes, ou encore les anneaux qui lient, grâce au travail des administrateurs, des sites Internet traitant de sujets proches. Sylvie Dalbin note la complémentarité de tous ces outils, en étudie les fonctionnalités et propose une synthèse des recherches récentes sur ce sujet, en particulier sur les différentes dimensions de la pertinence de l'information sur le Web. Le chapitre 4 est centré sur l'interprétation des requêtes par les moteurs de recherche. La nature de ces requêtes est souvent peu précise ou de mauvaise qualité. Pascale Sébillot (Irisa) montre comment certains moteurs rectifient les fautes de frappe, proposent à l'utilisateur des extensions possibles, voire des questions associées. Le dernier chapitre fait écho au premier en se concentrant sur la veille stratégique. Armelle Thomas (Info-rizon) étudie les outils de surveillance sur les réseaux, fait le point sur la réalité des agents intelligents et traite des technologies qui donnent du sens à de vastes corpus de données.

Trois chapitres entrent dans le cadre de la deuxième problématique de l'ouvrage : pour que l'information puisse être répertoriée par un moteur de recherche ou hiérarchisée dans un annuaire, sa représentation doit être rigoureuse, structurée. C'est le rôle d'XML (*eXtensible Markup Language*) que présente, au chapitre 2, François Role, du ministère de la Recherche. XML n'est pas un simple langage de balisage comme HTML (*Hyper-text Markup Language*), mais une méthode normalisée pour baliser la structure logique d'un document. L'exposé est technique, mais d'une grande clarté didactique. Les exemples que donne l'auteur, étape par étape, sont accessibles à tous ceux qui ont de

vagues connaissances en HTML. Dans ce cours d'initiation, François Role nous invite à manipuler des structures au sein d'un document XML et conclut en présentant des exemples d'applications, notamment dans le domaine de la gestion documentaire. Le cours trouve son prolongement au chapitre 3, dédié aux métadonnées. Marie-Élise Fréon (Jouve) en donne une typologie illustrée d'exemples détaillés, en utilisant différents cadres de descriptions interprétables. Ces « données ajoutées au document original pour expliciter ses caractéristiques et représenter son contenu de façon conventionnelle, condensée et non ambiguë » – selon la définition d'Éric Sutter (*Documentation, information, connaissances : la gestion de la qualité*, Paris, ADBS Éd., 2002) – véritable « valeur ajoutée à l'information », en permettent la compilation et le repérage. Après la mise en jambe des chapitres 2 et 3, on peut aborder le chapitre 6 où Laurent Romary (Inria Lorraine) conduit le lecteur de la sémantique des contenus à la sémantique des structures. Ce cours se présente comme une révision des deux précédents, mais le propos se concentre sur la nécessité de la normalisation pour permettre des échanges d'information efficaces. Documenter une structure, c'est aussi la modéliser de manière à ce que l'information puisse être facilement repérée par un moteur ou catégorisée par l'homme. Il est nécessaire de s'appuyer sur une base conceptuelle unifiée pour décrire la variété des métadonnées susceptibles d'être associées à des bases documentaires. Enfin, dans le chapitre 5, Dominique Lahary (BDP du Val d'Oise) dresse une typologie des bibliothèques virtuelles et montre comment celles-ci continuent de fournir les services assurés par les bibliothèques traditionnelles. Il évoque notamment les perspectives apportées par XML dans la structuration des données.

Soucieux d'exhaustivité, cet ouvrage n'aborde pas exclusivement le traitement de l'information textuelle, mais nous renseigne sur la complexité de la recherche dans des documents multimédias. Au chapitre 7, Nozha Boujemaa, de l'Inria Rocquencourt, présente à titre d'illustration, l'indexation et la recherche d'images par le contenu visuel. Ainsi comprend-on comment il est aujourd'hui

d'hui possible de spécifier des régions d'intérêt d'une image pour n'en extraire, par exemple, que les visages qui s'y trouvent.

Ces interventions, quelque peu éclatées, auraient pu être davantage fusionnées afin d'éviter toute répétition et proposer un déroulement plus limpide. La matière passionnante dont il est ici question reste assez technique : présentée avec un souci plus grand des non-spécialistes, avec un découpage plus clair qui aurait inséré les éléments des huit cours dans un plan synthétique, l'ouvrage aurait pu être encore plus accessible. Pourtant, en décortiquant le principe de fonctionnement des outils de recherche, ce cours aide le chercheur d'informations à être plus efficace dans sa recherche sur l'internet. En outre, il décrit les stratégies actuelles d'organisation de l'information pour qu'une recherche d'information sur les pages Web dont on est l'auteur soit rendue possible par quelque agent que ce soit.

Dominique Nauroy
CREM, université de Metz

Hans-Jürgen LÜSEBRINK, *La Conquête de l'espace public colonial. Prises de parole et formes de participation d'écrivains et d'intellectuels africains dans la presse à l'époque coloniale (1900-1960)*.

Québec/Frankfurt am Main/London,
Éd. Nota bene/IKO-Verlag, coll. Studien zu
den frankophonen Literaturen ausserhalb
Europas, Bd. 7, 2003, 272 p.

Dans sa préface, Bernard Mouralis écrit que cet ouvrage « marque un moment nouveau dans le domaine des études concernant la littérature africaine d'expression française ». Il s'inscrit néanmoins dans une lignée. Parmi les « antécédents » cités par Hans-Jürgen Lüsebrink, il y a bien entendu les études de littérature nationale (Adrien Huannou, Richard Bjornson, Mukala Kadima-Nzujii), mais aussi certains travaux à caractère général, portant sur la période d'avant 1960, dont ceux de Robert Comevin, auxquels l'auteur a souvent recours et à qui il rend ainsi justice, et ceux de János Riesz et de Bernard Mouralis. Nonobstant leurs différences de propos, les uns et les autres ont délibérément placé l'accent sur une compré-