

L'utilisation des langages documentaires pour la recherche d'information

The Use of Index Languages for Information Retrieval

La utilización de los lenguajes documentales para la búsqueda de información

Clément Arsenault

Volume 52, Number 2, April–June 2006

Les langages documentaires

URI: <https://id.erudit.org/iderudit/1030017ar>

DOI: <https://doi.org/10.7202/1030017ar>

[See table of contents](#)

Article abstract

This article describes the use of index languages for information retrieval. The author places controlled languages in the context of a global model of information retrieval and compares their strengths and weaknesses to those systems using free vocabulary. The issues of pre- and post-coordination are examined with respect to their efficiency. The issue of controlled vocabulary, namely in the form of ontology, for automatic retrieval and the interoperational semantics of systems is also discussed.

Publisher(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (print)

2291-8949 (digital)

[Explore this journal](#)

Cite this article

Arsenault, C. (2006). L'utilisation des langages documentaires pour la recherche d'information. *Documentation et bibliothèques*, 52(2), 139–148. <https://doi.org/10.7202/1030017ar>

Tous droits réservés © Association pour l'avancement des sciences et des techniques de la documentation (ASTED), 2006

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

L'utilisation des langages documentaires pour la recherche d'information

CLÉMENT ARSENAULT

École de bibliothéconomie et des sciences de l'information
Université de Montréal
clement.arsenault@umontreal.ca

RÉSUMÉ | ABSTRACTS | RESUMEN

Cet article présente un état de la question sur l'utilisation des langages documentaires pour la recherche d'information. L'auteur situe les langages contrôlés à l'intérieur du modèle global de la recherche d'information et présente les forces et les faiblesses de ceux-ci par rapport au vocabulaire libre. Les questions de pré- et de post-coordination sont examinées en fonction de leur performance respective. La question de l'utilisation du vocabulaire contrôlé, notamment sous forme d'ontologie, pour les recherches automatiques et l'interopérabilité sémantique des systèmes est également abordée.

The Use of Index Languages for Information Retrieval

This article describes the use of index languages for information retrieval. The author places controlled languages in the context of a global model of information retrieval and compares their strengths and weaknesses to those systems using free vocabulary. The issues of pre- and post-coordination are examined with respect to their efficiency. The issue of controlled vocabulary, namely in the form of ontology, for automatic retrieval and the interoperational semantics of systems is also discussed.

La utilización de los lenguajes documentales para la búsqueda de información

Este artículo presenta el estado de la cuestión sobre la utilización de lenguajes documentales para la búsqueda de información. El autor sitúa los lenguajes controlados al interior del modelo global de búsqueda de información y presenta las fortalezas y debilidades de estos lenguajes en relación con el vocabulario libre. Las cuestiones de pre y post coordinación se examinan en función de su respectivo desempeño. También aborda la cuestión de la utilización del vocabulario controlado, principalmente bajo la forma de ontología, para las búsquedas automáticas y la interoperabilidad semántica de los sistemas.

LA RECHERCHE D'INFORMATION dans les systèmes de repérage informatisés peut se définir comme un processus interactif entre un système et un usager au cours duquel ce dernier tente, par l'intermédiaire d'une interface, de trouver des réponses à des questions précises, ou encore de trouver des documents pertinents sur un sujet (ou des références bibliographiques permettant de les localiser) dans une ou plusieurs bases de données. Les documents repérés devraient permettre à l'utilisateur de satisfaire, en totalité ou en partie, un besoin informationnel. Le type d'interaction usager-système dépendra à la fois du type et de la nature du besoin informationnel, ainsi que des fonctionnalités disponibles sur l'interface du système. De façon générale, deux modes d'interaction sont possibles : 1) la sélection d'éléments prédéfinis reposant généralement sur une classification hiérarchique par sujet (taxinomie) qui permet la recherche par navigation, de catégorie en sous-catégorie ; 2) l'utilisation de requêtes (généralement textuelles) reposant sur l'utilisation d'un index et d'un langage d'interrogation plus ou moins élaboré (Baeza-Yates et Ribeiro-Neto, 1999 ; Chu, 2003). C'est de ce dernier type d'interaction qu'il sera principalement question dans cet article.

Nous verrons également de quelles façons les langages documentaires, que l'on désignera également par l'expression « vocabulaires contrôlés », peuvent contribuer au processus de recherche d'information, quels sont les facteurs qui déterminent leur utilisation et quels sont les outils disponibles pour faciliter leur intégration dans ce processus.

La recherche d'information

Il n'existe pas encore de théorie satisfaisante qui explique de façon globale le processus complexe qu'est la recherche d'information. Plusieurs auteurs ont toutefois développé des modèles permettant de le schématiser (Harter, 1986 ; Hartley *et al.*, 1990 ; Baeza-Yates et Ribeiro-Neto, 1999 ; Chu, 2003). Dans la majorité de ces modèles, les différentes composantes du processus de recherche d'information seraient, du point de vue de l'utilisateur : 1) la définition du besoin d'information et la formulation des objectifs ; 2) la

Si les bénéfices du vocabulaire contrôlé semblent en théorie indubitables, en pratique, il n'est pas assuré que l'efficacité du système de repérage en soit améliorée, d'autant plus que la performance des systèmes n'est pas déterminée uniquement par les caractéristiques d'indexation, mais également par les

L'éternel débat sur l'avantage du vocabulaire contrôlé par opposition au vocabulaire libre dans les systèmes de recherche d'information persiste.



- ▷ permet de regrouper des éléments disparates mais ayant des caractéristiques communes.

Il est donc nécessaire de déterminer, selon un environnement ou une situation donnés, quelle sera la meilleure solution à utiliser pour maximiser la performance. Traditionnellement, cette performance s'établit en termes de précision et de rappel.

Les études comparatives

Quelques études ont été menées en vue d'évaluer s'il est avantageux de faire usage du vocabulaire contrôlé en recherche d'information. Les résultats de ces études sont mitigés et souvent contradictoires, étant donné que le bénéfice potentiel du vocabulaire contrôlé dépend largement de facteurs environnementaux et contextuels. Ainsi, l'éternel débat sur l'avantage du vocabulaire contrôlé par opposition au vocabulaire libre dans les systèmes de recherche d'information persiste, mais il est clair qu'on reconnaît de plus en plus l'utilité de l'un et de l'autre selon les circonstances. Dans son étude comparative sur l'efficacité du vocabulaire contrôlé dans les bases de données bibliographiques, Jacques Savoy (2005) démontre d'ailleurs que les systèmes combinant termes contrôlés et non contrôlés sont ceux qui offrent la meilleure performance de repérage, en particulier du point de vue de la précision.

La première grande étude portant sur l'intérêt des vocabulaires contrôlés en recherche d'information fut menée par Cyril Cleverdon de la fin des années 1950 au début des années 1960. Cette étude, menée à Cranfield (Royaume-Uni), comporte deux parties désignées sous le nom de Cranfield I et Cranfield II. Bien que controversés, en raison des approches méthodologiques employées dans l'étude, les résultats obtenus par Cleverdon sont venus ébranler les convictions *a priori* sur le repérage d'information et en particulier sur l'avantage du vocabulaire contrôlé (Chaudiron, 2004). Les résultats de Cranfield I et II indiquent que, en termes de repérage, les unitermes en vocabulaire libre offrent un rendement supérieur aux termes d'un langage contrôlé (Savoy, 2005). Certaines études plus récentes et à plus petite échelle tendent à confirmer les résultats de Cranfield. Dans une étude de cas menée sur la littérature en ingénierie du sol, Manikya Rao Muddamalle (1998) conclut que la recherche en langage naturel (c'est-à-dire en vocabulaire libre) est

légèrement plus performante que la recherche en vocabulaire contrôlé.

D'autres études tendent plutôt à démontrer le bénéfice des vocabulaires contrôlés, en particulier pour améliorer la précision des résultats. Ainsi, l'étude de Karen Markey *et al.* (1980) indique que l'utilisation du thésaurus dans la base ERIC (*Educational Resources Information Center*) par des chercheurs expérimentés favorise la précision, alors que les recherches en vocabulaire libre favorisent le rappel. Carol Tenopir (1985) arrive sensiblement aux mêmes conclusions. R. Betts et D. Marrable (1991) concluent que le thésaurus aide à obtenir une meilleure précision et indiquent que le classement des résultats favorise le rappel. L'étude de William Hersh *et al.* (1994) sur l'utilité des vedettes-matière MeSH (*Medical Subject Headings*) ne révèle pas d'avantage marqué de ce vocabulaire contrôlé sur le vocabulaire libre, sauf pour les chercheurs experts qui sont les seuls à avoir réussi à augmenter les taux de rappel de leurs recherches. Par ailleurs, James French *et al.* (2002) démontrent que l'ajout automatique d'un certain nombre de termes MeSH aux requêtes des usagers (technique d'expansion de requêtes) permet une meilleure sélection des résultats de recherche en vue du classement par pertinence des documents. L'étude de Jaap Kamps (2004) démontre aussi l'intérêt du vocabulaire contrôlé, mais cette fois pour le reclassement des résultats de recherche dans les systèmes de repérage basés sur la pertinence. Cet ajout améliorerait la performance des systèmes traitant des collections pour des domaines spécifiques.

On peut conclure de ce bref survol que le vocabulaire contrôlé peut, dans certaines circonstances, contribuer à améliorer la performance des systèmes de recherche d'information, en particulier sur le plan de la précision des recherches et aussi du rappel, pour les chercheurs expérimentés. Toutefois, l'usage de l'un ou de l'autre vocabulaire, ou même des deux concurremment, reste valide selon les circonstances.

Utilisation du vocabulaire contrôlé pour la recherche d'information

Approches

Le vocabulaire contrôlé peut être utilisé au moment de l'indexation (représentation) et au moment de la formulation des requêtes (recherche). Ainsi, quatre situations sont possibles (Lancaster et Warner, 1993) :

- ▷ Vocabulaire libre pour la représentation et la recherche
- ▷ Vocabulaire contrôlé pour la représentation et la recherche

Figure 1 : Affichage du thésaurus ERIC dans le système CSA

Figure 2 : Affichage du thésaurus ERIC dans le système Dialog

Related Terms	Type of Term	Records
☒ DATA ANALYSIS (PREPARATION OF FACTUAL INFORMATION ITEMS FOR...)		8677
☐ DATA COLLECTION (GENERATING OR BRINGING TOGETHER INFORMATION ...)	Narrower	8428
☐ DATA INTERPRETATION (EXPLANATION OF THE MEANING, IMPLICATIONS, OR...)	Narrower	957
☐ STATISTICAL ANALYSIS (APPLICATION OF STATISTICAL PROCESSES AND THE...)	Narrower	11238
☐ TREND ANALYSIS (USE OF A SERIES OF MEASUREMENTS OF A VARIABLE...)	Narrower	12770
☐ EVALUATION METHODS (OBJECTIVE OR SUBJECTIVE PROCEDURES USED TO O...)	Broader	33094
☐ CLASSIFICATION (ORDERING OF RELATED PHENOMENA INTO CATEGORIES...)	Related	15666
☐ CONTENT ANALYSIS (SYSTEMATIC, OBJECTIVE, AND QUANTITATIVE DESC...)	Related	5724
☐ DATA (INFORMATION, OFTEN NUMERICAL AND ESPECIALLY ...)	Related	165178
☐ DATA PROCESSING (SYSTEMATIC HANDLING, MANIPULATION, AND COMPU...)	Related	3376
☐ HYPOTHESIS TESTING (PROCESSES BY WHICH HYPOTHESES ARE ACCEPTED O...)	Related	3200
☐ RESEARCH (SYSTEMATIC INVESTIGATION, COLLECTION, AND AN...)	Related	309479
☐ RESEARCH METHODOLOGY (PROCEDURES USED IN MAKING SYSTEMATIC OBSERVA...)	Related	25623

- ▷ Vocabulaire contrôlé uniquement pour la représentation
- ▷ Vocabulaire contrôlé uniquement pour la recherche

La première approche est de toute évidence la plus aisée. C'est celle à laquelle recourent les grands moteurs de recherche sur le Web. L'approche la plus complexe est certainement la deuxième, puisque l'outil de contrôle utilisé par les indexeurs doit également être fourni à l'utilisateur — du moins en partie ou encore sous une forme adaptée — pour la formulation des requêtes. La troisième, de moins en moins courante, est celle des systèmes dans lesquels le traitement a été fait à l'aide d'un vocabulaire contrôlé, mais où celui-ci est difficilement ou pas accessible à l'utilisateur au moment de la formulation de ses requêtes. La quatrième approche, connue sous le nom de recherche en vocabulaire post-contrôlé (Chu, 2003), moins coûteuse à mettre en œuvre que la précédente, consiste

simplement à fournir un outil d'aide, par exemple un thésaurus, pour permettre à l'utilisateur de construire ses requêtes et duquel il peut extraire des synonymes, des termes associés ou autres. Cette dernière approche est même quelquefois utilisée à l'insu de l'utilisateur dans les systèmes qui utilisent les techniques d'expansion automatique des requêtes. Rappelons qu'il n'est pas garanti d'obtenir des rendements supérieurs avec l'un ou l'autre de ces quatre modèles. Rappelons également qu'il existe des approches hybrides où les vocabulaires libre et contrôlé sont utilisés de concert.

Outils, systèmes et interfaces

Les approches mentionnées précédemment trouvent leur application dans les systèmes de repérage par l'utilisation de divers outils spécialisés qui peuvent être intégrés à différents niveaux de l'interface. Ces outils prennent la plupart du temps la forme d'un

Figure 3a : Mécanisme d'interception des termes dans OVID

The screenshot shows the OVID search interface. At the top, there's a header with 'ERIC' and '1966 to January 2006'. Below this is a 'Search History' table with columns for '#', 'Search History', and 'Delete Searches'. The table is currently empty. Below the table are buttons for 'Combine Searches', 'Delete Searches', and 'Save Search/Alert'. The main search area has tabs for 'Advanced Search', 'Basic Search', and 'Find Citation'. The 'Basic Search' tab is selected. Below the tabs is a search input field with the text 'sight'. To the right of the input field is a checkbox labeled 'Map Term to Subject Heading' which is checked. Below the checkbox is a 'SEARCH' button.

Figure 3b : Contrôle effectué avant d'exécuter la requête dans OVID

The screenshot shows the 'Mapping Display' screen. It displays the text 'Your term mapped to the following Subject Headings:'. Below this is a table with two columns: 'Select' and 'Subject Heading'. The table contains two rows: one with 'Vision' and another with 'sight .mp. search as Keyword'. The 'Select' column has checkboxes, with the first one checked. Below the table is a 'CONTINUE' button.

thésaurus mais, dans certains cas, il peut s'agir tout simplement d'une liste de termes contrôlés. Certains producteurs de bases de données développent eux-mêmes leur outil de contrôle de vocabulaire. C'est le cas par exemple des bases LISA (*Library and Information Science Abstracts*), PsycINFO ou ERIC, qui ont chacune leur propre thésaurus qui a servi au départ à l'indexation des documents. Toutefois, ces thésaurus peuvent également être consultés par le chercheur pour l'aider à déterminer les termes à utiliser dans sa requête.

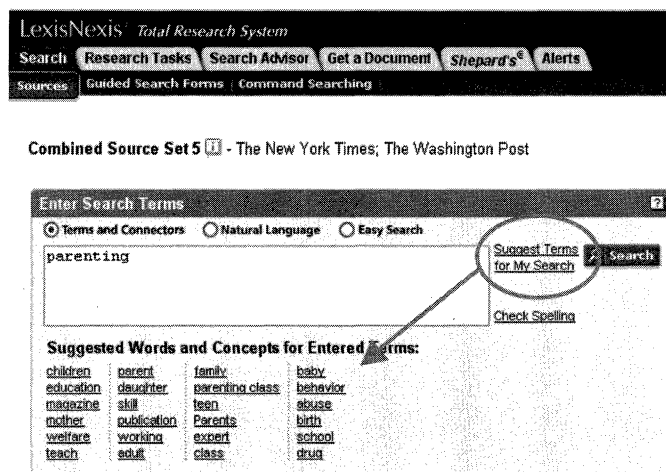
De plus en plus, les thésaurus sont disponibles à même l'interface de recherche développée par le distributeur (par exemple sur CSA ou Dialog), ce qui permet à l'utilisateur de sélectionner les termes et de les insérer directement dans son équation de recherche (voir figures 1 et 2). La consultation du thésaurus à même le système de recherche permet à l'utilisateur non seulement de tirer avantage des renvois, des relations, des définitions et des notes d'application contenus dans le thésaurus, mais également de bénéficier de fonctionnalités de recherche plus poussées, comme la fonction « explosion » (« *explode* » en anglais). Cette fonction de recherche, disponible sur certains systèmes, permet de tirer profit de la structure relationnelle et de navigation développée dans le thésaurus. Avec cette fonction, il est possible, en sélectionnant un terme X du thésaurus, de rechercher instantanément l'ensemble des spécifiques de ce

terme X, ce qui s'avère fort utile dans certains cas. Par exemple, en demandant « explosion « parents » », la requête sera étendue à tous les termes spécifiques associés à ce terme (par exemple: « mères », « pères », « parents adoptifs », et ainsi de suite).

L'architecture de certains systèmes prévoit également des mécanismes pour intercepter les requêtes formulées par l'utilisateur, en extraire les termes et les faire « parcourir » la structure syndétique du thésaurus avant même de lancer la recherche dans l'index. C'est ce que propose l'interface du fournisseur OVID. Ainsi, si l'utilisateur utilise un terme rejeté, le système lui proposera de chercher avec le terme accepté. Par exemple, si l'utilisateur cherche avec le mot « *sight* » et que « *vision* » est plutôt le terme qui a été utilisé, il en sera averti (voir figure 3a) et pourra décider de chercher soit avec le terme contrôlé « *vision* », soit de tout de même chercher « *sight* », mais en vocabulaire libre (voir figure 3b).

Les outils utilisés pour le contrôle du vocabulaire seront quelquefois développés par les fournisseurs de service plutôt que par les producteurs de bases de données. C'est souvent le cas pour les collections documentaires rassemblées par les fournisseurs (par exemple, les bases de journaux ou de périodiques divers) et pour lesquelles l'indexation est faite *a posteriori*. On trouve sur les systèmes ProQuest et Gale, par exemple, des thésaurus plus généraux qui sont quelquefois utilisés dans plusieurs bases. Cette unifor-

Figure 4 : Fonction de suggestion de termes associés dans LexisNexis



misation facilite la métarecherche (recherche multiple) en vocabulaire contrôlé, ce qui est difficilement réalisable lorsqu'on interroge simultanément des bases pour lesquelles un thésaurus spécifique a été utilisé. Enfin, dans certains systèmes, par exemple LexisNexis, on propose au chercheur des outils permettant d'exercer un certain contrôle sur le vocabulaire libre en lui suggérant des termes associés (généralement déterminés statistiquement de façon dynamique) au terme de l'équation (voir figure 4). Cette fonction est également offerte sous diverses formes par certains moteurs et métamoteurs de recherche Web (Exalead, Wisenut, Kartoo et Mamma, par exemple).

Types de vocabulaires contrôlés

Nonobstant les politiques d'indexation qui peuvent varier considérablement d'un système à un autre, les vocabulaires contrôlés sont développés de diverses façons. On retrouve trois types de termes contrôlés: 1) les unitermes; 2) les termes composés (expressions, syntagmes); 3) les termes complexes (composés syntaxiques). Prenons par exemple le sujet suivant: «l'emploi en thérapeutique des métaux alcalins». L'indexation de ce sujet donnerait les résultats suivants:

Unitermes: Alcalins / Emploi / Métaux / Thérapeutique

Composés: Emploi en thérapeutique / Métaux alcalins

Complexes: Métaux alcalins-Emploi en thérapeutique

Il est plus courant de retrouver des unitermes dans les systèmes où l'indexation est faite en vocabulaire libre, par extraction automatique à partir du texte intégral ou des éléments de métadonnées. Les termes composés sont en général issus des thésaurus documentaires et se retrouvent principalement dans les bases de données bibliographiques sous l'appel-

lation «descripteurs». Les termes complexes sont, quant à eux, régis par une syntaxe plus ou moins stricte déterminant la forme et l'ordre de présentation des composantes du terme; cette catégorie de termes se rencontre le plus souvent dans les catalogues de bibliothèques, sous forme de vedettes-matière. Comme on peut le voir, le degré de coordination des concepts représentés par les termes diffère d'un type de vocabulaire contrôlé à un autre. Dans le cas des vedettes-matière, le degré de coordination est plus élevé, ce qui permet à l'indexeur une représentation plus complète et plus précise par l'utilisation d'éléments de syntaxe.

Pré- ou post-coordination

La pré-coordination correspond à la combinaison des concepts durant la phase d'indexation. Les termes d'indexation résultant de cette action peuvent ainsi représenter des sujets complexes. Si l'analyse lexicale pour construire le fichier inversé (l'index) se fait par expression, alors le chercheur sera contraint d'accepter ces termes dans leur entièreté. À l'inverse, si les termes d'indexation représentent des concepts simples, la coordination des termes et des concepts doit être faite par le chercheur au moment de l'interrogation et à l'aide de diverses techniques (union, intersection, proximité, adjacence...). On parle alors de repérage en post-coordination. La différence entre les deux méthodes en est une de temps — avant ou après la recherche (Miller et Teitelbaum, 2002) — et déterminera qui de l'indexeur ou du chercheur (et même, en partie, du système de repérage) aura la tâche de coordonner les termes.

Certaines études ont tenté de mesurer la performance de l'une ou l'autre de ces approches en repérage d'information. Encore ici, les conclusions sont mitigées et certains auteurs prônent même le développement de systèmes hybrides en coordination partielle (Kambil et Bodoff, 1998). Dans leur étude, Uri

Figure 5a : Classement alphabétique des vedettes-matière

Noirs américains –Acculturation
 Noirs américains –Alabama
 Noirs américains –Associations –Bibliographie
 Noirs américains –Attitudes
 Noirs américains au cinéma
 Noirs américains au cinéma –Bibliographie
 Noirs américains au cinéma –Catalogues
 Noirs américains au cinéma –Histoire –20e siècle
 Noirs américains –Bibliographie
 Noirs américains –Biographies
 Noirs américains –Caroline du Nord

Figure 5b : Classement des vedettes-matière selon les règles

Noirs américains –Acculturation
 Noirs américains –Associations –Bibliographie
 Noirs américains –Attitudes
 Noirs américains –Bibliographie
 Noirs américains –Biographies
 ...
 Noirs américains –Alabama
 Noirs américains –Caroline du Nord
 ...
 Noirs américains au cinéma
 Noirs américains au cinéma –Bibliographie
 Noirs américains au cinéma –Catalogues
 Noirs américains au cinéma –Histoire –20e siècle

leur contenu que de leur degré de structuration, les systèmes de repérage mis en place sur le Web sont en général conçus pour déterminer *a posteriori* les relations terminologiques permettant, d'une part, de regrouper les documents similaires par des techniques de classification automatique et de mise en grappe (*clustering*) et, d'autre part, d'assurer un contrôle de la polysémie par des techniques de désambiguïsation sémantique. Ces techniques s'appliquent de façon automatique, avec des taux de succès variables selon les cas et les environnements, et contribuent à l'avancement et au développement du Web sémantique.

Cette vision du Web vise à établir des éléments de contrôle permettant aux systèmes de repérage d'être sémantiquement compatibles les uns avec les autres. L'utilisation de vocabulaires contrôlés, en particulier ceux qui se développent sous forme d'ontologies, est nécessaire pour assurer l'interopérabilité entre les systèmes (Franklin, 2003 : 99). Les ontologies, en tant que représentations formelles d'un domaine de la connaissance, contiennent des informations sémantiques permettant aux moteurs de recherche et aux agents intelligents d'établir des inférences logiques, et de les utiliser dans les techniques de recherche automatique telles que l'expansion et la traduction de requêtes (Nie, 2003). À l'instar des initiatives de mise en correspondance (*mapping*) terminologique entre les divers langages documentaires, notamment dans

le domaine de l'accès multilingue (McCulloch, 2004), ces initiatives d'interopérabilité sémantique permettront, espérons-le, de rendre la recherche d'information plus aisée et plus pertinente dans le contexte global du Web. ☉

Sources consultées

- Baeza-Yates, Ricardo et Berthier Ribeiro-Neto. 1999. *Modern information retrieval*, Don Mills, Ontario : Addison-Wesley.
- Betts, R. et D. Marrable. 1991. Free text vs controlled vocabulary: retrieval precision and recall over large databases. In *Online Information 91: Proceedings of the Fifteenth International Online Information Meeting*, sous la direction de David I. Raitt. Oxford, New Jersey : Learned Information, 153-165.
- Blair, David C. 2002. The challenge of commercial document retrieval, Part I: Major issues, and a framework based on search exhaustivity, determinacy of representation and document collection size. *Information Processing and Management* 38 (2): 273-291.
- Bodoff, David et Ajit Kambil. 1998. Partial coordination, I: The best of pre-coordination and post-coordination. *Journal of the American Society of Information Science* 49 (14): 1254-1269.
- Chaudiron, Stéphane. 2004. L'évaluation des systèmes de recherche d'informations. In *Les systèmes de recherche d'informations: modèles conceptuels*, sous la direction de Madjid Ihadjadene. Paris : Hermès Science, 185-207.
- Chu, Heting. 2003. *Information representation and retrieval in the digital age*. Medford : Information Today.

- Dillon, Andrew et Min Song. 1997. An empirical comparison of the usability for novice and expert searchers of a textual and a graphic interface to an art-resource database. *Journal of Digital Information* 1 (1).
- Drabenstott, Karen M. et Marjorie S. Weller. 1996. Failure analysis of subject searches in a test of a new design for subject access to online catalogs. *Journal of the American Society for Information Science* 47 (7): 519-537.
- Franklin, Rosemary Aud. 2003. Re-inventing subject access for the semantic Web. *Online Information Review* 27 (2): 94-101.
- French, James C., Allison L. Powell, Frederic Gey et Natalia Perelman. 2002. Exploiting manual indexing to improve collection selection and retrieval effectiveness. *Information Retrieval* 5 (4): 323-351.
- Harter, Stephen P. 1986. *Online information retrieval: Concepts, principles, and techniques*. San Diego: Academic Press.
- Hartley, Richard J., E. M. Keen, J. Andrew Large et Lucy A. Tedd. 1990. *Online searching: principles and practice*. London: Bowker-Saur.
- Harvey, Ross et Philip Hider. 2004. *Organising knowledge in a global society: principles and practice in libraries and information centres*. Wagga Wagga, NGS: Centre for Information Studies.
- Hersh, William, Chris Buckley, T. J. Leone et David Hickam. 1994. OHSUMED: an interactive retrieval evaluation and new large test collection for research. In *Proceedings of the 17th International Conference of the ACM-SIGIR'94*, sous la direction de W. Bruce Croft et C. J. van Rijsbergen. London: Springer, 192-201.
- Kambil, Ajit et David Bodoff. 1998. Partial coordination, II: A preliminary evaluation and failure analysis. *Journal of the American Society of Information Science* 49 (14): 1270-1282.
- Kamps, Jaap. 2004. Improving retrieval effectiveness by reranking documents based on controlled vocabulary. In *Advances in Information Retrieval: 26th European Conference on IR Research, ECIR 2004*, sous la direction de Sharon McDonald et John Tait. Berlin: Springer, 283-295.
- Kuhlthau, Carol Collier. 2004. *Seeking meaning: a process approach to library and information services*, 2nd ed. Westport, Connecticut: Libraries Unlimited.
- Lancaster, F. Wilfrid et Amy J. Warner. 1993. *Information retrieval today*. Arlington, Va.: Information Resources Press.
- Large, Andrew, Lucy A. Tedd et Richard J. Hartley. 1999. *Information seeking in the online age: principles and practice*. London: Bowker-Saur.
- Mann, Thomas. 2000. Teaching Library of Congress Subject Headings. *Cataloging and Classification Quarterly* 29 (1-2): 117-126.
- _____. 2003. Why LC Subject Headings are more important than ever. *American Libraries*, October 2003, 52-54.
- Markey, Karen, Pauline Atherton et Claudia Newton. 1980. An analysis of controlled vocabulary and free text search statements in online searches. *Online Review* 4 (3): 225-236.
- Markey Drabenstott, Karen et Diane Vizine-Goetz. 1994. *Using subject headings for online retrieval: Theory, practice, and potential*. San Diego: Academic Press.
- McCulloch, Emma. 2004. Multiple terminologies: an obstacle to information retrieval. *Library Review* 53 (6): 297-300.
- Miller, Uri et Ruth Teitelbaum. 2002. Pre-coordination and post-coordination, past and future. *Knowledge Organization* 29 (2): 87-93.
- Mizzaro, Stefano. 1997. Relevance: the whole history. *Journal of the American Society for Information Science* 48 (9): 810-832.
- Nie, Jian-Yun. 2003. Query expansion and query translation as logical inference. *Journal of the American Society of Information Science and Technology* 54 (4): 335-346.
- Olson, Hope et John Boll. 2001. *Subject analysis in online catalogs*. Englewood, Colorado: Libraries Unlimited.
- Ozmutlu, Seda. 2005. Performance of question-based vs keyword-based search engines and effect of Web user characteristics on search engine performance. *Online Information Review* 29 (6): 656-675.
- Rao Muddamalle, Manikya. 1998. Natural language versus controlled vocabulary in information retrieval: A case study in soil mechanics. *Journal of the American Society of Information Science* 49 (10): 881-887.
- Rowley, Jennifer et John Farrow. 2000. *Organizing knowledge: An introduction to managing access to information*, 3^e éd. Aldershot, England: Gower.
- Saracevic, Tefko. 1996. Relevance reconsidered '96. In *CoLIS2, Second International Conference on Conceptions of Library and Information Science*, sous la direction de Peter Ingwersen et Niels Ole Pors. Copenhagen: Royal School of Librarianship, 201-218.
- Savoy, Jacques. 2005. Bibliographic database access using free-text and controlled vocabulary: an evaluation. *Information Processing and Management* 41 (4): 873-890.
- Soergel, Dagobert. 1994. Indexing and retrieval performance: the logical evidence. *Journal of the American Society of Information Science* 45 (8): 589-599.
- Tenopir, Carol. 1985. Full text database retrieval performance. *Online Review* 9 (2): 149-164.
- Yee, Martha M. 2004. New perspective on the shared cataloging environment and a MARC 21 shopping list. *Library Resources & Technical Services* 48 (3): 165-178.