

Traitement documentaire de l'image ordinaire : analyse de deux approches d'indexation

Cataloguing Ordinary Images: An Analysis of Two Indexation Approaches

Tratamiento documental de la imagen común: análisis de dos enfoques de indexación

Elaine Ménard

Volume 55, Number 1, January–March 2009

URI: <https://id.erudit.org/iderudit/1029049ar>

DOI: <https://doi.org/10.7202/1029049ar>

[See table of contents](#)

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

Ménard, E. (2009). Traitement documentaire de l'image ordinaire : analyse de deux approches d'indexation. *Documentation et bibliothèques*, 55(1), 29–38. <https://doi.org/10.7202/1029049ar>

Article abstract

This article presents the results of a research project whose purpose it was to describe the characteristics of two approaches used to index a collection of ordinary images of the objects of everyday life. The first approach uses terms from a visual dictionary, whereas the second approach favors an uncontrolled vocabulary. The analysis of the indexation terms reveals the tendencies regarding terminology, perception and structure. The factors that influence the attribution of indexation terms are also described.

Traitement documentaire de l'image ordinaire : analyse de deux approches d'indexation

ELAINE MÉNARD

Professeure-adjointe
School of Information Studies
Université McGill
elaine.menard@mcgill.ca

RÉSUMÉ | ABSTRACTS | RESUMEN

Cet article* présente les résultats** d'une recherche ayant pour objectif de recenser les caractéristiques de deux approches utilisées pour l'indexation d'un ensemble d'images ordinaires représentant des objets de la vie quotidienne. La première approche suppose l'attribution de termes d'indexation extraits d'un dictionnaire visuel, alors que la deuxième approche préconise l'utilisation du vocabulaire libre pour la description des images. L'analyse des termes d'indexation révèle les tendances observées sur le plan terminologique, perceptuel et structurel. Les facteurs ayant influencé l'attribution des termes d'indexation sont également décrits.

Cataloguing Ordinary Images : An Analysis of Two Indexation Approaches

This article[§] presents the results of a research project^{§§} whose purpose it was to describe the characteristics of two approaches used to index a collection of ordinary images of the objects of everyday life. The first approach uses terms from a visual dictionary, whereas the second approach favors an uncontrolled vocabulary. The analysis of the indexation terms reveals the tendencies

regarding terminology, perception and structure. The factors that influence the attribution of indexation terms are also described.

Tratamiento documental de la imagen común : análisis de dos enfoques de indexación

Este artículo[†] presenta los resultados de una investigación^{††} cuyo objetivo es inventariar las características de dos enfoques utilizados para la indexación de un conjunto de imágenes corrientes que representan objetos de la vida cotidiana. El primer abordaje supone la atribución de términos de indexación extraídos de un diccionario visual, mientras que el segundo preconiza la utilización de vocabulario libre para la descripción de las imágenes. El análisis de los términos de indexación revela las tendencias observadas en el plano terminológico, perceptual y estructural. Se describen asimismo los factores que han influenciado la atribución de los términos de indexación.

Introduction

LE PROCESSUS D'INDEXATION vise principalement à « décrire et à caractériser un document à l'aide de la représentation des concepts contenus dans ce document, c'est-à-dire à traduire en langage documentaire les concepts extraits du document au moment de l'analyse » (Association française de normalisation, 1996). Traditionnellement, deux approches sont utilisées pour l'indexation de l'image : l'approche basée sur le contenu ou les caractéristiques physiques de l'image et l'approche basée sur le contexte ou la description de l'image.

Inspirés du domaine de l'imagerie médicale et géophysique, et du développement de nouvelles technologies informatiques telles que le traitement d'images par ordinateur, la reconnaissance de formes et d'objets, la vision artificielle, etc., les systèmes basés sur le contenu (*Content based image retrieval* [CBIR]) utilisés pour l'indexation de l'image exploitent les caractéristiques physiques de l'image (Tsai, 2003). Ces caractéristiques, généralement extraites automatiquement des fichiers d'images, sont considérées de bas niveau, par opposition aux éléments dits de haut niveau (termes d'indexation) qui sont attribués à la suite de l'analyse de l'image effectuée par un indexeur humain ou par traitement automatique. Avec les systèmes de CBIR, la phase d'indexation des images n'implique à aucun moment l'utilisation du langage. Les images sont décrites non pas par des

* Cet article est extrait d'une thèse de doctorat intitulée *Étude sur l'influence du vocabulaire utilisé pour l'indexation des images en contexte de repérage multilingue*, soutenue à l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal en 2008. Le texte complet de la thèse est disponible à l'adresse <<https://papyrus.bib.umontreal.ca/jspui/bitstream/1866/2611/1/menard-e-these-indexation-reperage-images.pdf>>. L'auteure remercie les membres de son comité de recherche doctoral, les professeurs Clément Arsenault et Christine Dufour.

** La recherche a été réalisée grâce à une bourse doctorale décernée par le Conseil de recherches en sciences humaines du Canada (CRSH).

§ This article is based on a doctoral thesis titled *Étude sur l'influence du vocabulaire utilisé pour l'indexation des images en contexte de repérage multilingue*, submitted to the École de bibliothéconomie et des sciences de l'information de l'Université de Montréal en 2008. The complete text of the thesis is available at <https://papyrus.bib.umontreal.ca/jspui/bitstream/1866/2611/1/menard-e-these-indexation-reperage-images.pdf>. The author wished to acknowledge the members of her doctoral research committee, namely professors Clément Arsenault et Christine Dufour.

§§ This research project was made possible with a doctoral grant from the Social Sciences and Humanities Research Council of Canada (SSHRC).

† Este artículo ha sido extraído de una tesis de doctorado titulada *Estudio sobre la influencia del vocabulario utilizado para la indexación de las imágenes en contexto de localización multilingüe*, realizada en la Escuela de biblioteconomía y ciencias de la información de la Universidad de Montreal en 2008. El texto completo de la tesis se encuentra disponible en la siguiente dirección: <https://papyrus.bib.umontreal.ca/jspui/bitstream/1866/2611/1/menard-e-these-indexation-reperage-images.pdf>. La autora agradece a los miembros de su comité de investigación doctoral, los profesores Clément Arsenault y Christine Dufour.

†† La investigación ha sido realizada gracias a una beca de doctorado otorgada por el Consejo de investigaciones en ciencias humanas de Canadá (CRSH).

éléments textuels, mais plutôt par des valeurs associées à certains paramètres tels que la couleur, la texture et la forme (Boudry et Agostini, 2004). Pour le moment, les méthodes d'indexation utilisant les caractéristiques physiques de l'image présentent toujours de nombreuses lacunes et demeurent au stade expérimental (Jørgensen, 2003). Par conséquent, l'indexation à l'aide de la description textuelle de l'image constitue toujours la voie la plus utilisée menant à son repérage. Les termes d'indexation peuvent être tirés d'un vocabulaire contrôlé ou provenir de la langue naturelle.

Vocabulaire contrôlé

La première approche d'indexation de l'image, définie pour cette recherche comme la manière d'indexer caractérisée par le vocabulaire choisi pour le traitement, préconise l'utilisation du vocabulaire contrôlé. Le vocabulaire contrôlé est un langage comportant ses propres termes, sa propre syntaxe et sa propre sémantique. Le principal avantage du contrôle du vocabulaire est de favoriser la cohérence de l'indexation et d'accroître, grâce à un système de renvois, la probabilité d'adéquation entre les termes choisis par l'indexeur et le chercheur (Jørgensen, 2003 ; Arsenault, 2006). De prime abord, les vocabulaires contrôlés ont pour objectif de faciliter le processus d'indexation. En plus d'exercer une certaine forme de standardisation au cours de l'indexation, ces vocabulaires permettent également d'améliorer le repérage. Toutefois, les vocabulaires contrôlés présentent certaines faiblesses dont la principale est de représenter les concepts de manière artificielle (Macgregor et McCulloch, 2006). En effet, la plupart des vocabulaires contrôlés incluent des termes corrects sur le plan linguistique, mais peu souvent utilisés dans la vie quotidienne. Par exemple, dans un vocabulaire contrôlé on proposera le terme « brodequin de travail », alors que le terme « botte de travail » est beaucoup plus usuel. On remarque souvent une grande divergence entre les termes d'indexation offerts dans le vocabulaire contrôlé et les termes généralement utilisés dans les requêtes au moment du repérage (Furnas *et al.*, 1987). En outre, l'utilisation des vocabulaires contrôlés demeure une tâche complexe pour la plupart des indexeurs, particulièrement pour l'indexation des documents iconographiques (Goodrum, 2000). Enfin, la plupart des vocabulaires contrôlés habituellement utilisés pour l'indexation de l'image n'existent qu'en anglais. Par conséquent, un indexeur possédant peu de connaissances de la langue anglaise et désirant utiliser ces vocabulaires, risque de se heurter à un problème linguistique important, à moins qu'un mécanisme de traduction ne soit prévu.

En règle générale, les schémas de métadonnées développés pour l'indexation des images suggèrent l'utilisation de vocabulaires contrôlés pour certains éléments de métadonnées. Certains de ces vocabulaires sont de nature générale. Cependant, l'utilisation de vocabulaires contrôlés de nature générale pour l'indexation des

images n'est pas toujours couronnée de succès, puisque ceux-ci sont souvent peu étoffés et ne conviennent pas nécessairement à des collections diversifiées d'images. Dans la majorité des cas, ces vocabulaires contrôlés ne sont pas assez précis. De ce fait, d'autres vocabulaires contrôlés ont dû être élaborés pour l'indexation des ressources visuelles. Ces vocabulaires sont conçus pour la description d'éléments comme le matériel, le style ou le sujet de l'objet illustré. Parmi les vocabulaires contrôlés pouvant être utilisés pour l'indexation, on retrouve les schémas de classification, les listes de vedettes-matière et les thésaurus.

Schémas de classification

Un schéma de classification est un langage documentaire construit à partir de signes linguistiques ou de symboles. Ce langage est basé sur la structuration des sujets en classes. Les classes et les relations entre celles-ci sont représentées par des indices codés. Un schéma de classification possède une double fonction, puisqu'il permet la classification d'un domaine et le classement physique des documents, selon la nature du sujet (Maniez, 1987). Traditionnellement, les schémas de classification sont le plus souvent utilisés en vue de l'organisation de l'information monographique. Cependant, ces schémas peuvent aussi être utilisés pour l'indexation de documents tels que l'image numérique. Parmi les principaux schémas de classification documentaires, soulignons la *Library of Congress Classification* (LCC) (*Library of Congress Classification*, 2009), la Classification décimale universelle (CDU ou *UDC Consortium*, 2009) et la Classification décimale de Dewey (CDD ou OCLC, 2009).

Outre les schémas de classification de nature générale, certains schémas dédiés plus spécifiquement au matériel visuel ont vu le jour. Parmi ceux-ci, on retrouve le système ICONCLASS, un système international de classification pouvant être utilisé pour la description du matériel visuel (ICONCLASS, 2009). Un aspect particulier d'ICONCLASS est son système de notation qui utilise des codes de classification alphanumériques. Chaque notation fait partie d'une structure hiérarchique et offre les avantages des vocabulaires contrôlé et libre (Jørgensen, 2003). Le schéma de classification ICONCLASS présente également quelques lacunes. De plus, les classes offertes par ICONCLASS ne sont pas toujours développées également pour tous les sujets et ne conviennent pas nécessairement à tous les types d'images, puisqu'elles visent essentiellement l'art occidental et sont très peu adaptées pour la description des images ordinaires représentant des objets de la vie quotidienne.

Listes de vedettes-matière

Une vedette-matière est un mot ou un groupe de mots indiquant le sujet sous lequel tous les écrits

ou autres documents traitant du même thème sont regroupés. Un répertoire (ou liste) de vedettes-matière est constitué de vedettes-matière développées et montre les relations d'équivalence, de hiérarchie et d'associations qui existent entre elles (Dégez et Ménillet, 2001). Ce type de vocabulaire contrôlé a été conçu pour la recherche en mode précoordonné, mais il peut également être utilisé dans un système permettant la postcoordination (Chu, 2003). Parmi les répertoires de vedettes-matière de nature générale, mentionnons les *Library of Congress Subject Headings* (*Library of Congress Subject Headings* [LCSH], 2009), le *Répertoire d'autorité-matière encyclopédique et alphabétique unifié* (RAMEAU, Bibliothèque nationale de France, 2009), et le *Répertoire de vedettes-matière de l'Université Laval* (RVM, Université Laval, 2009). Toutefois, on constate que les répertoires de vedettes-matière demeurent relativement peu utilisés pour l'indexation de l'image (Jørgensen, 2003).

Thésaurus

Les thésaurus sont des vocabulaires contrôlés tirés de la langue naturelle, spécifiquement conçus pour la recherche en environnement postcoordonné. Ils contribuent à réduire les difficultés de repérage causés par la polysémie et la synonymie inhérentes à la langue naturelle (Hudon, 2006). Ils peuvent être monolingues ou multilingues. Plusieurs thésaurus sont communément utilisés pour la description de l'image numérique. Le *Art and Architecture Thesaurus* (AAT) est un vocabulaire contrôlé qui propose un ensemble structuré de termes pour la description et l'indexation d'images en art et en architecture, de matériaux et d'objets de la culture matérielle, de l'Antiquité à nos jours (Getty Foundation, 2009a). Le AAT sert à des fins de catalogage, de repérage et comme outil de recherche. Le thésaurus AAT offre les relations thésaurales traditionnelles (équivalence, hiérarchie, association), mais également des liens sémantiques basés sur les relations logiques qui existent entre les concepts, les activités et les objets. Toutefois, il faut souligner que le thésaurus AAT présente quelques lacunes dans sa couverture, surtout en ce qui concerne les personnes, les événements et les activités, des points d'accès importants dans les collections générales d'images.

Un autre exemple de thésaurus utilisé couramment pour l'indexation de matériel visuel est le *Thesaurus for Graphic Materials* (TGM) développé par la *Library of Congress* pour le catalogage et les besoins spécifiques de repérage de tout matériel visuel (estampes, photographies, dessins, images en mouvement, documents éphémères), qu'il fasse partie d'un livre, d'un manuscrit ou d'une collection visuelle. La principale utilité de ce vocabulaire contrôlé est d'aider la recherche de matériel visuel, mais également de faciliter la gestion des collections visuelles (*Library of Congress*, 2009).

Listes d'autorité et dictionnaires visuels

D'autres langages documentaires peuvent être employés pour l'indexation des images, notamment les listes d'autorité et les dictionnaires visuels (Jørgensen, 2003). Parmi les principales listes d'autorité pouvant être utilisées pour l'indexation des images, on retrouve le *Union List of Artist Names* (ULAN) (Getty Foundation, 2009b), le fichier d'autorité de la *Library of Congress* (*Library of Congress Authorities*, 2009), et le *Multilingual Glossary for Art Librarians* de l'*International Federation of Library Associations* (IFLA) (*International Federation of Library Associations and Institutions*, 2009).

En ce qui concerne les dictionnaires visuels, même s'ils ne peuvent être formellement considérés comme des vocabulaires contrôlés en raison de leur manque de structure hiérarchique ou d'équivalence, ils peuvent tout de même être utilisés pour la description de l'image ordinaire. Étant donné la nature diversifiée des collections d'images que l'on retrouve dans les bibliothèques virtuelles et les musées, certains dictionnaires visuels ont été élaborés afin de répondre aux besoins spécifiques de ces collections d'objets. Par exemple, le *Dictionnaire descriptif et visuel d'objets* de Parcs Canada (Bernard, 1997) est un outil hybride alliant les caractéristiques d'une liste d'autorité et d'un schéma de classification. Ce langage repose spécifiquement sur la fonction originelle de l'objet à décrire. Par ailleurs, des dictionnaires visuels comme le *Nouveau dictionnaire visuel* (Corbeil et Archambault, 2003) ont pour objectif de définir l'environnement matériel d'une personne évoluant dans le monde industrialisé contemporain, et qui doit non seulement connaître mais utiliser un grand nombre de termes spécialisés dans de nombreux domaines. Ce type de dictionnaire offre deux avantages importants pour l'indexation de l'image. D'une part, les illustrations qu'il contient permettent de trouver un mot à partir d'une idée ou d'une image qu'on a en tête, contrairement aux dictionnaires traditionnels qui exigent que l'on connaisse d'abord le mot auquel il réfère. D'autre part, ce type de dictionnaire est souvent offert en version bilingue ou multilingue, ce qui est rarement le cas de la plupart des langages documentaires.

L'utilisation des vocabulaires contrôlés pour l'indexation de l'image offre de nombreux avantages pour le repérage, le furetage et l'interopérabilité. Cependant, ces vocabulaires ne sont pas toujours assez précis et ne fournissent pas nécessairement les points d'accès de l'image correspondant à tous les utilisateurs, car le vocabulaire est parfois trop spécialisé et sans rapport avec le vocabulaire du chercheur d'images profane (Greisdorf et O'Connor, 2008). En outre, un autre désavantage de l'indexation à l'aide de vocabulaires contrôlés est que ceux-ci deviennent rapidement incomplets, voire désuets. En effet, les néologismes que l'on retrouve dans l'indexation en vocabulaire libre mettront souvent beaucoup de temps à apparaître dans les différents vocabulaires contrôlés utilisés pour l'indexation. Enfin, l'élaboration

et la gestion des vocabulaires contrôlés constituent deux opérations entraînant des coûts importants. Parce que le vocabulaire contrôlé ne convient pas nécessairement à tous les types d'images, il devient intéressant d'indexer l'image en vocabulaire libre.

Vocabulaire libre

La deuxième approche d'indexation de l'image constitue une solution alternative à l'indexation traditionnelle effectuée avec le vocabulaire contrôlé. Ainsi, le vocabulaire non contrôlé ou libre est constitué de termes librement choisis par le chercheur, ou encore par l'indexeur, et qui ne sont pas validés par une liste de termes faisant autorité (Arsenault, 2006). Ces termes, appelés aussi mots-clés, provenant du titre ou de toute partie significative d'un document (par exemple, la légende lorsqu'elle est disponible), sont utilisés librement pour l'indexation de l'image. L'indexeur peut aussi choisir ses propres termes pour la description documentaire. À l'instar de l'indexation en vocabulaire contrôlé, l'indexation en vocabulaire libre présente plusieurs lacunes.

Par exemple, certaines ambiguïtés apparaissent lorsqu'un même mot-clé est employé par plusieurs individus, mais dans divers contextes. De la même manière, le manque de contrôle de la synonymie entraîne l'utilisation de nombreux mots-clés différents pour décrire un même concept.

L'utilisation du vocabulaire contrôlé ou du vocabulaire libre pour l'indexation documentaire est une question longuement débattue dans la littérature. Au fil des années, plusieurs chercheurs se sont penchés sur la question du vocabulaire devant être utilisé pour le processus d'indexation. Les résultats découlant de certaines de ces études indiquent que le vocabulaire libre offre un rendement supérieur par rapport au vocabulaire contrôlé au moment du repérage (Savoy, 2005). Ce point de vue avait déjà été avancé par Manikya Rao Muddamalle (1998), alors que d'autres études (Tenopir, 1985 ; Kamps, 2004) considéraient que les vocabulaires contrôlés avaient pour effet d'améliorer la précision des résultats du repérage. Cependant, certains auteurs reconnaissaient de plus en plus l'utilité de combiner les deux vocabulaires au moment de l'indexation de l'image (Macgregor et McCulloch, 2006 ; Enser *et al.*, 2007).

Méthodologie

Une analyse des termes utilisés pour l'indexation des images a été réalisée afin de recenser les caractéristiques des deux approches d'indexation à l'étude pouvant avoir une influence sur le repérage de l'image et, qui plus est, en contexte multilingue. Trois grandes étapes ont été nécessaires à la réalisation de cet examen : l'élaboration d'une base d'images, l'indexation des images et l'analyse des termes d'indexation.

Élaboration de la base d'images

Une base d'images ordinaires a été constituée à partir des images tirées d'un catalogue en ligne d'une société commerciale canadienne ayant accepté de mettre son catalogue à notre disposition pour cette recherche. Pour les besoins de la recherche, 3 950 images puisées parmi les huit grandes sections du catalogue en ligne ont été recueillies. La sélection des images visait trois objectifs : privilégier la qualité visuelle de l'image illustrant un objet spécifique, faciliter le processus d'identification de l'image par une majorité d'individus et favoriser la diversité maximale des images incluses dans la base d'images.

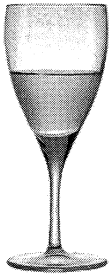
Indexation des images

Chacune des images contenue dans la base a ensuite été indexée de deux manières différentes : en vocabulaire contrôlé et en vocabulaire libre. En outre, l'indexation des images a été faite en deux langues : en français et en anglais. Ces langues ont été retenues pour deux raisons principales. D'abord, cette recherche s'est déroulée dans un pays où le français et l'anglais sont les deux langues officielles et, plus spécifiquement, dans une province où la majorité des habitants ont le français comme langue maternelle. La langue française semblait donc un choix logique. Par ailleurs, la plupart des vocabulaires contrôlés destinés à l'indexation documentaire n'existent qu'en anglais, ce qui rendait presque obligatoire la sélection de l'anglais en tant que deuxième langue d'indexation. L'indexation en vocabulaire libre français a été effectuée par un indexeur dont la langue maternelle est le français, alors que l'indexation en vocabulaire libre anglais a été effectuée par un indexeur de langue maternelle anglaise. L'indexation en vocabulaire contrôlé français et anglais a été réalisée par un indexeur maîtrisant les deux langues. Pour des raisons de disponibilité et d'économie, les indexeurs ayant participé au projet ne possédaient aucune expérience professionnelle en indexation.

Indexation en vocabulaire contrôlé

Pour l'indexation en vocabulaire contrôlé, l'indexeur a reçu comme directive d'utiliser le *Nouveau dictionnaire visuel multilingue* (Corbeil et Archambault, 2003) pour indexer chacune des 3 950 images contenues dans la base. Ce dictionnaire a été choisi pour trois raisons. Premièrement, il contient les termes pouvant convenir au type d'images incluses dans la base, c'est-à-dire des images d'objets de la vie quotidienne. Deuxièmement, le *Visuel* offre une forme de normalisation des termes qui permet une identification claire et précise des objets, tout en exerçant un contrôle maximal sur les régionalismes, les archaïsmes, les anglicismes, etc. Finalement, le *Visuel* comporte à la fois une terminologie française et anglaise (de même qu'espagnole, allemande et italienne), ce qui a facilité le processus d'indexation, tout en assurant une certaine cohérence interlinguistique. Le *Visuel* a

Figure 1
Exemple d'image et de termes d'indexation

Image extraite de la base d'images	Image correspondante tirée du <i>Dictionnaire visuel</i>	Termes d'indexation extraits du <i>Dictionnaire visuel</i>
		Français : verres à vin blanc Anglais : white wine glass

donc été employé à la fois pour l'indexation en vocabulaire contrôlé français et en vocabulaire contrôlé anglais.

Le processus d'indexation en vocabulaire contrôlé s'est déroulé comme suit : l'indexeur visualisait d'abord l'image à indexer, puis repérait dans le dictionnaire l'image correspondant le mieux à celle-ci. Les termes associés à l'image du dictionnaire étaient ensuite utilisés comme termes d'indexation, en français et en anglais. Par exemple, pour une image représentant un verre à vin blanc, les termes « verre à vin blanc » et « *white wine glass* » associés à une image similaire du *Visuel* ont été attribués à l'image à indexer, tel qu'illustré à la Figure 1. De un à cinq termes d'indexation contrôlés pouvaient être attribués aux images.

Indexation en vocabulaire libre

Le processus d'indexation en vocabulaire libre présentait des caractéristiques opposées à celles de l'indexation en vocabulaire contrôlé. Une politique spécifique pour l'indexation en vocabulaire libre a également été élaborée pour notre recherche. Cette politique prévoyait que l'indexeur visualisait d'abord l'image à indexer et utilisait ensuite ses propres termes pour la décrire, sans imposer de règle quant au choix des termes, à leur forme et à leur nombre.

Analyse des termes d'indexation

Suite au processus d'indexation, une analyse des termes d'indexation utilisés a été effectuée pour les quatre formes d'indexation : vocabulaire contrôlé français (VCF), vocabulaire libre français (VLF), vocabulaire contrôlé anglais (VCA) et vocabulaire libre anglais (VLA). Dans notre étude, une forme d'indexation est définie comme une approche d'indexation (en vocabulaire contrôlé ou libre) associée à une langue d'indexation (français ou anglais). Une grille d'analyse des termes d'indexation, développée à partir d'attributs identifiés dans la littérature sur l'indexation des images, comportait trois catégories : les attributs terminologiques se rapportant aux aspects lexicographiques des termes

d'indexation ; les attributs perceptuels liés aux aspects physiques, fonctionnels et identificateurs des images ; les attributs structurels identifiant plus précisément les relations qui existent entre les mots. Pour l'analyse terminologique, cinq types de termes d'indexation ont été examinés : les termes simples (1 seul mot / 1 seul concept, par exemple, *raquette*) ; les termes composés (plusieurs mots / 1 seul concept, par exemple *raquette de tennis*), les termes complexes (plusieurs mots / plusieurs concepts, par exemple, *casque d'écoute avec radio et antenne*), les abréviations (par exemple, *GPS*) et les néologismes (par exemple, *cinéma maison*). Deux niveaux d'indexation ont été considérés : les termes de nature générique (par exemple, *pince*) et les termes de nature spécifique (par exemple, *pince à découper*).

Sur le plan perceptuel, quatre groupes d'attributs ont été analysés. Le premier groupe était constitué des attributs physiques de l'objet représenté sur l'image, c'est-à-dire les termes faisant référence à la forme (par exemple, *assiette rectangulaire*), aux dimensions (par exemple, *grande tasse*), à la couleur (par exemple, *bicyclette bleue*), à la texture (par exemple, *simili fourrure*), au matériel (par exemple, *poêlon en téflon*) et aux parties du corps (par exemple, *bain de pied*). Le deuxième groupe se rapportait aux attributs fonctionnels des objets, c'est-à-dire la fonction (par exemple, *planche à découper*), le résultat-produit (par exemple, *bouée de sauvetage*), l'activité-action (par exemple, *clignotant*) et le lieu (par exemple, *table de jardin*). Le troisième groupe comportait les attributs identificateurs des objets représentés, c'est-à-dire la marque commerciale (par exemple, *MotoMaster*), le nom de produit (par exemple, *Frigidaire*), le commanditaire (par exemple, *t-shirt Nascar*) et le lieu géographique (par exemple, *marbre d'Italie*). Un dernier groupe d'attributs perceptuels recensait les termes d'indexation de nature temporelle ou saisonnière (par exemple, *tasse de Noël*), de nature numérique (par exemple, *cartouche d'encre HP 78*), les autres types de qualificatifs non couverts par les catégories déjà mentionnées (par exemple, *à ski, ouverte*) et les attributs liés aux personnes (par exemple, *siège d'enfant, chemise d'homme*).

Tableau I
Synthèse des caractéristiques des termes d'indexation

ATTRIBUTS	CATÉGORIES	SOUS-CATÉGORIES	VCF (N = 4 482)		VLF (N = 4 062)		VCA (N = 4 409)		VLA (N = 3 967)	
			N	%	N	%	N	%	N	%
Terminologiques	Types de termes	Termes simples	2 175	48,5	684	16,8	1956	44,4	463	11,7
		Termes composés	2 307	51,5	3 378	83,2	2 453	55,6	3 504	88,3
		Total	4 482	100,0	4 062	100,0	4 409	100,0	3 967	100,0
	Autres types de termes	Termes complexes	42	0,9	363	8,9	28	0,6	175	4,4
		Abréviations	41	0,9	333	8,2	52	1,2	396	10,0
		Néologismes	15	0,3	145	3,6	70	1,6	157	4,0
		Total	98	2,1	841	20,7	150	3,4	728	18,4
	Niveaux d'indexation	Termes génériques	2 340	52,2	867	21,3	2 056	46,6	527	13,3
		Termes spécifiques	2 142	47,8	3 195	78,7	2 353	53,4	3 440	86,7
		Total	4 482	100,0	4 062	100,0	4 409	100,0	3 967	100,0
Perceptuels	Physiques	Forme	71	1,6	135	3,3	95	2,2	263	6,6
		Dimension	69	1,5	57	1,4	22	0,5	89	2,2
		Couleur	5	0,1	348	8,6	3	0,1	123	3,1
		Texture	17	0,4	7	0,2	8	0,2	18	0,5
		Matériel	210	4,7	664	16,3	131	3,0	382	9,6
		Partie du corps	124	2,8	45	1,1	101	2,3	82	2,1
		Total	496	11,1	1256	30,9	360	8,2	957	24,1
	Fonctionnels	Fonction	697	15,6	838	20,6	658	14,9	849	21,4
		Résultat/produit	77	1,7	39	1,0	67	1,5	51	1,3
		Activité/action	339	7,6	532	13,1	334	7,6	665	16,8
		Lieu	180	4,0	307	7,6	293	6,6	474	11,9
		Total	1 293	28,8	1 716	42,2	1 352	30,7	2 039	51,4
	Identificateurs	Marque	0	0,0	1 019	25,1	0	0,0	243	6,1
		Nom de produit	0	0,0	66	1,6	0	0,0	29	0,7
		Commanditaire	0	0,0	67	1,6	0	0,0	56	1,4
		Lieu géographique	12	0,3	20	0,5	17	0,4	10	0,3
		Total	12	0,3	1 172	28,9	17	0,4	338	8,5
	Autres attributs	Temporel	7	0,2	36	0,9	3	0,1	22	0,6
		Numérique	8	0,2	249	6,1	29	0,7	77	1,9
		Qualificatif	514	11,5	314	7,7	473	10,7	238	6,0
		Personne	170	3,8	94	2,3	138	3,1	295	7,4
		Total	699	15,6	693	17,1	643	14,6	632	15,9
Structurels	Relations	Génériques	2 091	46,7	2 466	60,7	2 417	54,8	3 288	82,9
		Partitives	28	0,6	518	12,8	22	0,5	172	4,3
		D'instances	11	0,2	1148	28,3	12	0,3	317	8,0
		Associatives	53	1,2	350	8,6	327	7,4	128	3,2
		Total	2 183	48,7	4 482	110,3	2 778	63,0	3 905	98,4

Finalement, l'analyse des relations structurelles a pris en compte quatre types de relations : les relations de genre à espèce (par exemple, *lampe sur pied*), les relations de partie à tout (par exemple, *jante de roue*), les relations d'instance (par exemple, *couteau Kitchen Aid*) et les relations associatives (par exemple, *aspirateur d'auto*). La grille d'analyse a permis de mettre en relief les similitudes et les différences qui existent entre les formes d'indexation à l'étude.

Résultats

Le tableau I résume les observations faites à la suite de l'analyse effectuée sur l'ensemble des termes d'indexation attribués aux images de la base.

Dans les paragraphes suivants, nous présentons la synthèse des observations découlant de l'analyse des termes d'indexation, par forme, par approche et par langue d'indexation.

Forme d'indexation

L'indexation en vocabulaire contrôlé français (VCF) se démarque des autres formes d'indexation sur

Tableau II

Facteurs d'influence sur les caractéristiques des termes d'indexation

	CARACTÉRISTIQUES TERMINOLOGIQUES	CARACTÉRISTIQUES PERCEPTUELLES	CARACTÉRISTIQUES STRUCTURELLES
Indexeur		✓	
Langue d'indexation			
Politique d'indexation	✓		
Type d'images		✓	✓
Vocabulaire d'indexation	✓	✓	✓

trois points précis. Premièrement, sur le plan terminologique, cette forme d'indexation utilise davantage de termes d'indexation pour chaque image et ces derniers sont généralement de nature générique. Deuxièmement, sur le plan perceptuel, l'indexation VCF n'offre pas une grande variété dans le choix des attributs assignés aux images. Enfin, la forme VCF utilise peu de relations structurelles, en général.

Sur le plan terminologique, l'indexation en vocabulaire libre français (VLF) se distingue principalement des autres formes d'indexation par une attribution marquée de termes complexes. Sur le point de vue perceptuel, cette forme d'indexation est particularisée par un emploi important d'attributs identificateurs (marque commerciale, nom de produit, fabricant et lieu géographique) et d'attributs physiques (couleur et matériaux). La forme VLF tend à offrir plus de relations structurelles que les autres formes d'indexation.

L'indexation en vocabulaire contrôlé anglais (VCA) ne semble pas se différencier des autres formes d'indexation. Sur le plan terminologique, cette forme d'indexation offre des proportions de termes simples-composés et spécifiques-génériques assez semblables. Sur les plans perceptuel et structurel, la forme VCA ne se distingue pas non plus quant à l'utilisation de certains attributs et relations.

Enfin, l'indexation en vocabulaire libre anglais (VLA) se distingue principalement sur le plan terminologique par le nombre le moins élevé de termes d'indexation attribués par image et par un taux élevé d'utilisation de termes composés, de termes spécifiques, de néologismes et d'abréviations. En ce qui concerne le point de vue perceptuel, l'emploi de certains attributs de type fonctionnel (activité-action et fonction) et de type physique (texture et dimension) particularise la forme VLA. L'indexation VLA fait davantage usage des relations génériques que les autres formes d'indexation.

Approches d'indexation

L'analyse des termes d'indexation utilisés pour la description des images représentant des objets de la vie quotidienne a révélé que les approches d'indexation à l'étude se distinguent l'une de l'autre aussi bien sur les plans terminologique que perceptuel et structurel. Ainsi, l'indexation en vocabulaire contrôlé diffère de l'indexa-

tion en vocabulaire libre par une attribution d'un plus grand nombre de termes d'indexation à une même image. En outre, l'indexation en vocabulaire contrôlé offre des proportions de termes simples-composés et spécifiques-génériques assez semblables. L'approche contrôlée ne montre pas une très grande variété dans le choix des attributs perceptuels attribués aux images. En fin de compte, elle utilise peu de relations structurelles.

Par ailleurs, l'approche d'indexation en vocabulaire libre se distingue de l'indexation en vocabulaire contrôlé principalement sur le plan terminologique par un nombre moins élevé de termes d'indexation utilisés pour décrire les images de la base. L'approche libre domine l'approche contrôlée pour l'attribution de nombreux attributs physiques, identificateurs, fonctionnels et autres types d'attributs perceptuels. Finalement, l'approche libre offre généralement plus de relations structurelles que l'approche contrôlée.

Langues d'indexation

Sur le plan linguistique, l'analyse des termes d'indexation a permis d'observer que les caractéristiques des termes sont assez semblables au sein d'une même approche d'indexation, peu importe la langue. Les différences, lorsqu'elles existent, sont toutefois plus marquées avec les formes d'indexation utilisant le vocabulaire libre. Sur le plan terminologique, on constate que l'indexation en français recourt davantage à des termes complexes que l'indexation en anglais, toutes approches d'indexation confondues. En ce qui concerne les attributs perceptuels, la principale différence observée entre les langues d'indexation se situe dans la catégorie des attributs perceptuels identificateurs (marque, nom de produit, commanditaire et lieu géographique), par lesquels, en général, on remarque que l'indexation en français comporte davantage ce type d'attributs que par l'indexation en anglais. Enfin, sur le plan structurel, les relations génériques sont plus nombreuses avec l'approche libre en anglais qu'avec l'approche libre en français.

Discussion

Cinq facteurs ayant eu une incidence sur les caractéristiques des termes d'indexation attribués aux images ressortent de l'analyse précédente. Le tableau II résume

comment cette influence s'est principalement manifestée sur les plans terminologique, perceptuel et structurel.

Vocabulaire d'indexation

Le vocabulaire est sans doute le facteur ayant eu la plus grande incidence sur les termes d'indexation attribués aux images. La comparaison des termes d'indexation révèle que le vocabulaire employé pour l'indexation a une influence importante sur le choix des attributs terminologiques, perceptuels et structurels des termes d'indexation et explique une grande partie des différences observées entre les caractéristiques des termes attribués pour chaque forme d'indexation. La principale constatation qui ressort de cette analyse est que l'approche en vocabulaire libre offre une grande variabilité dans le choix des attributs perceptuels et des relations structurelles, alors que l'approche en vocabulaire contrôlé, tout en restreignant ce choix, présente une grande cohérence terminologique. Nous pouvons donc penser que la combinaison des vocabulaires contrôlé et libre, dans le cas des images utilisées pour cette recherche, constitue la meilleure solution d'indexation, puisque celle-ci allie à la fois les caractéristiques des deux vocabulaires utilisés pour l'indexation. Cette constatation rejoint d'ailleurs les conclusions que l'on retrouve dans la littérature encourageant la cohabitation des deux vocabulaires pour les documents textuels (Rao Muddamalle, 1998 ; Arsenault, 2006) et pour les images (Macgregor et McCulloch, 2006 ; Enser *et al.*, 2007). Cependant, comme l'utilisation conjointe des deux vocabulaires a aussi pour conséquence de complexifier le processus d'indexation, il faut se demander à juste titre s'il s'agit de la meilleure approche d'indexation à adopter.

Langue d'indexation

L'analyse des termes d'indexation a permis de déterminer qu'il existe des constantes dans la manière d'indexer en vocabulaire libre ou contrôlé, peu importe la langue, puisque peu de différences sur le plan terminologique, perceptuel et structurel ont été observées entre les deux langues, à l'intérieur d'une même approche d'indexation. Toutefois, lorsque des différences linguistiques sont relevées, celles-ci sont un peu plus marquées en vocabulaire libre qu'en vocabulaire contrôlé. Le peu de différence observée sur le plan linguistique en approche contrôlée peut évidemment s'expliquer par le fait que le choix du vocabulaire contrôlé utilisé pour cette recherche s'est porté sur une ressource multilingue, avec correspondance directe entre les termes français et anglais. Cela a eu pour effet de confiner l'indexation en vocabulaire contrôlé aux mêmes caractéristiques terminologiques, perceptuelles et structurelles, aussi bien en français qu'en anglais. De plus, il convient de mentionner que l'indexation en vocabulaire contrôlé français et anglais a été réalisée par un même indexeur.

Politique d'indexation

Le processus d'indexation des 3 950 images de la base reposait sur deux politiques d'indexation distinctes. Ces politiques ont été élaborées afin d'établir un ensemble de règles à respecter dans la tâche d'indexation. De manière générale, les deux politiques d'indexation conçues pour cette recherche étaient similaires. Cependant, trois différences entre ces politiques doivent être mentionnées : le vocabulaire d'indexation, la langue d'indexation dont nous avons parlé précédemment et, enfin, les règles spécifiant le nombre de termes d'indexation à attribuer aux images et le niveau de spécificité à privilégier. Sur le plan du nombre de termes attribués aux images, on remarque une différence entre les proportions enregistrées en approche contrôlée et en approche libre, malgré une directive de la politique d'indexation qui aurait pu résulter en tendance contraire. En outre, une différence sur le plan de la spécificité de l'indexation a été observée. Pour cette recherche, les consignes des deux politiques d'indexation avaient été données aux indexeurs en ce qui concernait le degré de spécificité à respecter, à une exception près : en effet, une directive supplémentaire était donnée à l'indexeur en vocabulaire contrôlé, c'est-à-dire qu'au moment de l'indexation, si aucune image du *Visuel* ne correspondait à l'image à indexer, une image plus générique de ce dictionnaire devait être cherchée et un terme plus général était alors retenu comme terme d'indexation. Cette décision méthodologique a été prise afin de s'assurer que toutes les images de la base soient indexées avec au moins un terme d'indexation tiré du vocabulaire contrôlé retenu pour cette étude. En raison de cette directive supplémentaire, il était légitime de s'attendre à ce que l'écart entre les proportions de termes génériques et spécifiques des vocabulaires contrôlé et libre présentent des différences. En effet, l'analyse du niveau d'indexation nous a permis de constater que l'indexation en contrôlé, peu importe la langue d'indexation, présente des proportions de termes génériques et spécifiques similaires, alors que l'écart entre ces proportions est beaucoup plus prononcé en libre où l'on observe une majorité de termes spécifiques. Par exemple, des termes génériques comme « clôture » ou « robinet » (*fence* et *faucet*) ont été utilisés en contrôlé, alors qu'en libre on retrouve plutôt les termes « clôture à neige » ou « robinet de lavabo de salle de bain » (*snow fence* ou *bathroom faucet*) pour la description des mêmes images. Par conséquent, nous pouvons penser que la directive d'attribuer un terme plus générique, si aucun terme spécifique n'était disponible, peut avoir gonflé de manière artificielle la proportion de termes génériques du contrôlé, au détriment de la proportion de termes spécifiques. Ce manque de spécificité de l'indexation en contrôlé est directement lié au manque de spécificité du langage documentaire choisi pour cette recherche et ne peut sans doute pas être imputé à la politique d'indexation et s'explique plutôt par le vocabulaire employé pour l'indexation.

Type d'images

Le type d'images retenu pour cette recherche a également une incidence sur le choix des termes d'indexation utilisés pour leur représentation. Rappelons d'abord que, pour cette étude, les termes d'indexation, tant en indexation libre qu'en indexation contrôlée, étaient attribués selon la technique d'indexation centrée sur l'objet, c'est-à-dire que l'accent devait être mis sur le sujet de l'image, plutôt que sur l'interprétation de celle-ci. Les termes d'indexation attribués aux images correspondent donc à des caractéristiques stables qui ne sont pas susceptibles de changer avec le temps (Jørgensen, 2003). De plus, les termes d'indexation tiennent compte du contenu visuel de l'image, c'est-à-dire le niveau pré-iconographique lié aux objets représentés (Panofsky, 1955 ; Shatford, 1986 ; Enser, 1995). Les termes d'indexation se rapportent donc aux objets ou aux situations identifiables que l'on retrouve sur l'image (Turner, 1998), ce que Shatford (1986) définit comme le niveau *of*, c'est-à-dire ce qui peut être identifié sur l'image. Le niveau *about* (interprétation des images) et le niveau de signification basé sur le savoir et l'érudition (Turner, 1998) sont exclus de l'indexation, puisque ceux-ci ne s'appliquent pas vraiment aux images utilisées pour cette recherche. En d'autres termes, les images de notre étude demandaient plutôt un mode d'indexation « dure », tel que décrit par Krause (1988), mettant l'accent sur la description de ce que l'indexeur voit sur l'image, plutôt qu'une indexation « douce », qui réfère plutôt à la signification de ce qu'il voit.

Sélection des indexeurs

Enfin, la sélection des indexeurs est le dernier facteur examiné ayant eu une influence sur les caractéristiques des termes d'indexation. Pour des considérations de disponibilité et d'économie, les indexeurs ayant participé à cette étude ne possédaient aucune expérience professionnelle en indexation. La sélection des indexeurs était plutôt basée sur leurs compétences linguistiques. Selon Lancaster (2003), l'indexation est une activité hautement subjective. De ce fait, les résultats de l'indexation varient en fonction de multiples facteurs, allant des connaissances linguistiques à la culture générale de l'indexeur. Par ailleurs, il convient de souligner que dans l'approche d'indexation contrôlée, le support visuel offert par les images du dictionnaire a sans doute contribué à faciliter la tâche de l'indexeur dans les cas où l'objet représenté ne pouvait être facilement identifié. La combinaison image-texte était alors d'un grand secours pour l'indexeur qui devait uniquement établir une correspondance entre les caractéristiques visuelles de l'image à indexer et celles d'une image correspondante du dictionnaire, et ensuite utiliser les termes d'indexation associés. Ainsi, la tâche de l'indexeur qui disposait d'un langage d'indexation contrôlé, même si elle demandait une certaine dextérité au point de vue de l'observation, n'était pas aussi exigeante sur le plan des connais-

sances linguistiques que celle de l'indexeur travaillant en vocabulaire libre.

Pour leur part, les indexeurs travaillant en vocabulaire libre ont rencontré deux catégories de difficultés au moment de l'indexation. D'une part, il est possible que les indexeurs aient pu ne pas reconnaître l'objet illustré sur l'image. Cependant, ces cas sont plutôt rares, puisque la sélection des images de cette étude était basée sur plusieurs critères spécifiques ayant pour but de faciliter l'identification des images par une majorité d'individus. Le risque de ne pas identifier l'objet représenté sur l'image et d'attribuer un mauvais terme d'indexation était par conséquent peu élevé. D'autre part, même si les indexeurs pouvaient reconnaître l'objet contenu dans l'image, il est également possible qu'ils n'aient pu nécessairement nommer celui-ci, ce qui a pu influencer le niveau de spécificité des termes d'indexation choisis.

En résumé, de tous les éléments décrits précédemment, le vocabulaire d'indexation est sûrement le facteur ayant eu la plus grande incidence sur l'attribution des termes d'indexation, alors que la langue d'indexation a joué un rôle moins important. En outre, c'est sur le plan des attributs perceptuels que l'influence de ces facteurs a surtout été remarquée. Il faut toutefois mentionner que les facteurs ne peuvent pas vraiment être considérés individuellement et que les caractéristiques des termes d'indexation attribués aux images sont plutôt le résultat de l'effet de plusieurs facteurs combinés.

Conclusion

Cette recherche a identifié les caractéristiques des approches d'indexation de l'image en vocabulaire contrôlé et en vocabulaire libre. Ces caractéristiques indiquent d'une part que ces deux approches diffèrent fondamentalement l'une de l'autre, sur les plans terminologique, perceptuel et structurel. L'examen de ces caractéristiques révèle également que si la langue d'indexation est modifiée, les caractéristiques varient très peu au sein d'une même approche d'indexation. Cette variation est légèrement plus marquée avec l'indexation en vocabulaire libre, principalement à cause du manque de contrôle inhérent à cette approche.

À la lumière de cette analyse, on peut penser que la description d'images ordinaires représentant des objets de la vie quotidienne devrait idéalement être fondée sur le plan pré-iconographique en privilégiant certaines caractéristiques perceptuelles, comme les attributs physiques et identificateurs se rattachant plus précisément à la description de l'objet illustré (couleur, marque, etc.). En outre, l'indexation de ce type d'images doit favoriser l'inclusion de la fonction de l'objet représenté, afin d'augmenter la probabilité de repérage. Toutefois, les vocabulaires contrôlés mettant essentiellement l'accent sur la fonction de l'objet sont assez rares.

L'indexation en vocabulaire contrôlé, effectuée avec les dictionnaires visuels notamment, présente certaines limites, puisque les nouveaux objets n'y sont pas toujours

représentés, ce qui a souvent comme résultat d'offrir une indexation plus générique risquant d'avoir un effet sur la précision du repérage et, par le fait même, des conséquences sur l'efficacité, l'efficience et la satisfaction du chercheur d'images.

Cette analyse visait à répondre à certaines questions ayant trait à l'indexation d'images ordinaires représentant des objets de la vie quotidienne, lorsque celles-ci sont indexées selon deux approches spécifiques. D'autres recherches semblent nécessaires pour approfondir les résultats de la présente étude. L'examen d'autres vocabulaires contrôlés pour l'indexation pourrait servir à mieux cerner les similitudes et les différences de ces deux approches d'indexation. Par exemple, il serait pertinent de poursuivre la réflexion entreprise avec cette recherche et d'examiner les caractéristiques d'une approche d'indexation utilisant un vocabulaire contrôlé uniquement textuel et non pas basé sur la combinaison image-texte. En outre, d'autres catégories d'images pourraient être examinées afin de pouvoir émettre les suggestions nécessaires pour en améliorer l'organisation et, éventuellement, le repérage. ☉

Sources consultées

Tous les URLs étaient valides le 12 février 2009.

- Arsenault, Clément. 2006. L'utilisation des langages documentaires pour la recherche d'information, *Documentation et Bibliothèques*, 52 (2) : 139-148.
- Association française de normalisation. 1996. Information et documentation – Principes généraux pour l'indexation des documents (NF Z 47-102, octobre 1993). Dans *Recueil de normes françaises – Documentation*, Tome : Présentation des publications, traitement documentaire et gestion de bibliothèques, 6^e éd. Paris, AFNOR.
- Bernard, Louise, dir. 1997. *Dictionnaire descriptif et visuel d'objets : présenté selon le Système de classification des collections historiques de Parcs Canada*, Ottawa, Patrimoine canadien, Parcs Canada.
- Bibliothèque nationale de France. RAMEAU, <<http://rameau.bnf.fr/index.htm>>.
- Boudry, Christophe et Clémence Agostini. 2004. Étude comparative des fonctionnalités des moteurs de recherche d'images sur Internet, *Documentaliste – Sciences de l'information*, 41 (2) : 96-105.
- Chu, Heting. 2005. *Information representation and retrieval in the digital age*. Medford, NJ, Information Today.
- Corbeil, Jean-Claude et Ariane Archambault. 2003. *Le nouveau dictionnaire visuel multilingue*, Montréal, Québec Amérique.
- Dégez, Danièle et Dominique Ménillet. 2001. *Thésauroglossaire des langages documentaire : un outil de contrôle sémantique*, Paris, ADBS.
- Enser, Peter G. B. 1995. Progress in documentation : Pictorial information retrieval. *Journal of Documentation*, 51 (2) : 126-170.
- Enser, Peter G. B. et al. 2007. Facing the reality of semantic image retrieval, *Journal of Documentation*, 63 (4) : 465-481.
- Furnas, G. W. et al. 1987. The vocabulary problem in human-system communication. *Communications of the ACM*, 30 (11) : 964-971.
- Getty Foundation. 2009a. *Art & Architecture Thesaurus Online*, <http://www.getty.edu/research/conducting_research/vocabularies/aat/>.
- . 2009b. *Union List of Artist Names*. <http://www.getty.edu/research/conducting_research/vocabularies/ulan/>.
- Goodrum, Abby A. 2000. Image information retrieval : an overview of current research. *Informing Science*, 3 (2) : 63-67.
- Greisdorf, Howard F. et Brian C. O'Connor. 2008. *Structures of images collections : from Chauvet-Pont d'Arc to Flickr*, Westport, Conn., Libraries Unlimited.
- Hudon, Michèle. 2006. Le thésaurus : au carrefour des sciences de l'information et de la terminologie. Dans *Terminologie et accès à l'information*, éd. par Widad Mustafa El Hadi, 71-98. Paris : Hermès Science.
- Jansen, Bernard J. 2008. Searching for digital images on the web. *Journal of Documentation*, 64 (1) : 81-101.
- Jørgensen, Corinne. 2003. *Image retrieval : Theory and research*, Lanham, MD, Scarecrow Press.
- Kamps, Jaap. 2004. Improving retrieval effectiveness by reranking documents based on controlled vocabulary. Dans *Advances in Information Retrieval : 26th European Conference on IR research, ECIR 2004*, sous la direction de Sharon McDonald et de John Tait, 283-295, Berlin, Springer.
- Krause, Michael G. 1988. Intellectual problems of indexing picture collections. *Audiovisual Librarian*, 14 (4) : 73-81.
- ICONCLASS. <<http://www.iconclass.nl/>>.
- International Federation of Library Associations and Institutions. *Multilingual Glossary for Art Librarians*, <<http://www.ifla.org/VII/s30/pub/mg1.htm>>.
- Lancaster, F. W. 2003. *Indexing and abstracting in theory and practice*. Champaign, IL, Graduate School of Library and Information Science, University of Illinois.
- Library of Congress Authorities. <<http://authorities.loc.gov>>.
- Library of Congress Classification. <<http://www.loc.gov/catdir/cpsol/lcc.html>>.
- Library of Congress Subject Headings. <<http://www.loc.gov/aba/cataloging/subject/>>.
- Library of Congress. 2009. *Thesaurus for Graphic Materials I : subject terms (TGM I)*, <<http://www.loc.gov/rr/print/tgm1/>>.
- Macgregor, George et Emma McCulloch. 2006. Collaborative tagging as a knowledge organisation and resource discovery tool, *Library Review*, 5 (5) : 291-300.
- Maniez, Jacques. 1987. *Les langages documentaires et classificatoires : conception, construction et utilisation dans les systèmes documentaires*, Paris, Éditions d'Organisation.
- Matusiak, Krystyna K. 2006. Towards user-centered indexing in digital image collections, *OCLC Systems & Services*, 22 (4) : 238-98.
- OCLC. *Dewey Services*, <<http://www.oclc.org/dewey/>>.
- Panofsky, Erwin. 1955. *Meaning in the visual arts : papers in and on art history*, Garden City, N.Y., Doubleday.
- Rao Muddamalle, Manikya. 1998. Natural language versus controlled vocabulary in information retrieval : A case study in soil mechanics, *Journal of the American Society for Information Science*, 49 (10) : 881-887.
- Savoy, Jacques. 2005. Bibliographic database access using free-text and controlled vocabulary : an evaluation, *Information Processing and Management*, 41 (4) : 873-890.
- Shatford, Sara. 1986. Analyzing the subject of a picture : a theoretical approach, *Cataloging & Classification Quarterly*, 6 (3) : 39-61.
- Tenopir, Carol. 1985. Full text database retrieval performance, *Online Review*, 9 (2) : 149-164.
- Tsai, Chih-Fong. 2003. Stacked generalisation : a novel solution to bridge the semantic gap for content-based image retrieval, *Online Information Review*, 27 (6) : 442-445.
- Turner, James. 1994. *Determining the subject content of still and moving image documents for storage and retrieval : an experimental investigation*, thèse de doctorat, Toronto, Université de Toronto.
- . 1998. *Images en mouvement : stockage, repérage, indexation*, Sainte-Foy, Presses de l'Université du Québec.
- UDC Consortium, <<http://www.udc.org/>>.
- Université Laval. *Répertoire de vedettes-matière*. <http://www.bibl.ulaval.ca/mieux/chercher/ch_vedettes_matiere/cache/bypass?print=1>.