

Département d'histoire
Faculté des lettres et sciences humaines
Université de Sherbrooke

*L'apport des nouvelles technologies de reconnaissances d'images et d'indexation à la
recherche historique et à la diffusion des archives photographiques*

Par
Gabrielle Thériault
Mémoire pour obtenir
La Maîtrise *ès arts* (histoire)

Université de Sherbrooke
Décembre 2020

Résumé

L'indexation et la diffusion des images sont en constante évolution dans le domaine de l'archivistique et de la recherche en histoire. En effet, la numérisation des archives photographiques et l'explosion de la photographie numérique nous obligent à réfléchir à l'intégration de nouvelles méthodes pour l'indexation de ces images. Ce travail de recherche se penche donc sur la perspective d'améliorer l'accessibilité des archives photographiques et leur potentiel pour la recherche historique grâce à l'apport des nouvelles technologies de reconnaissances d'images et d'indexation. Afin d'éviter que cette richesse archivistique demeure sous-exploitée, les centres auront besoin de nouveaux outils pour favoriser la diffusion de ces images.

Ce mémoire propose donc une réflexion sur les concepts et les méthodes d'indexation et sur l'état des techniques d'intelligence artificielle permettant la reconnaissance d'images. Une telle approche permettra d'ajouter une liste de mots automatique dans les échelons supérieurs de la taxonomie, offrant ainsi une plus grande cohérence et une meilleure uniformité entre les collections. L'indexation humaine des sujets plus pointus faciliteront ensuite la recherche ciblée dans les collections et ainsi rendra les documents plus accessibles aux chercheurs. Ceux-ci pourront ainsi exécuter des recherches plus prévisibles dans les grandes catégories d'information et disposer d'informations précises pour les recherches ciblées. Bref, l'ajout des technologies de reconnaissance de formes et d'indexation automatisée va favoriser la diffusion des images, mais aussi, la démocratisation du savoir. Elles établissent une complémentarité entre le travail des archivistes, la mise en ligne des sources et les besoins des chercheurs.

Mots clés : indexation, reconnaissance d'images, archives, automatisation, accessibilité, classement, gestion des données, chaîne de traitement.

Remerciements

La réalisation de ce mémoire a été possible grâce à l'appui de mon directeur de recherche, le professeur Léon Robichaud. Merci pour votre présence ainsi que pour les nombreux conseils et pistes de réflexions que vous m'avez apportés tout au long de la rédaction.

Je tiens aussi à témoigner toute ma gratitude à mes lecteurs, Harold Bérubé et Tristan Landry.

Je remercie aussi la Société d'histoire de Sherbrooke (Musée d'histoire de Sherbrooke) pour m'avoir laissée plonger dans le cœur de leurs archives afin de comprendre le fonctionnement de leur base de données en ligne.

Finalement, je ne peux finir mes remerciements sans mentionner l'appui et le dévouement inestimable de Mme Francine Waelles, conjointe de feu l'historien et professeur Jean-Pierre Kesteman. Merci pour ton soutien, ton écoute et aussi, pour avoir pris le temps de suivre la progression de mon travail en m'apportant des bases solides en rédaction du français.

Table des matières

Résumé	I
Remerciements	II
Liste de tableaux et figures.....	V
Introduction	1
Mise en situation et problématique.....	1
État de la question	3
<i>L'importance de l'image dans le processus de recherche.....</i>	<i>3</i>
<i>L'accessibilité des sources.....</i>	<i>5</i>
Hypothèse	7
Démarche.....	9
Chapitre 1 : Le classement et l'accès aux banques de photographies historiques ...	13
1.1 Les défis d'un centre d'archives régional	14
1.2 Les archives photographiques dans les institutions métropolitaines	26
1.3 Les recommandations sur la diffusion des archives en ligne	32
1.4 Conclusion.....	36
Chapitre 2 : Les fondements de l'indexation.....	39
2.1 Les taxonomies internationales pour l'indexation des images.....	41
2.2 L'intégration de la folksonomie	53
2.3 Conclusion.....	61
Chapitre 3 : Le potentiel de la reconnaissance d'image	63
3.1 Logiciels de reconnaissance d'objets	64
3.2 Les défis de la reconnaissance des images sur le terrain	74
3.3 Logiciels existants pour la reconnaissance d'objets.....	78
3.4 Conclusion.....	82
Conclusion : de la reconnaissance à l'indexation	84
Vers une nouvelle chaîne de traitement.....	87
L'avenir de la recherche numérique	91
BIBLIOGRAPHIE.....	93
ANNEXE A.....	103
ANNEXE B.....	105
ANNEXE C.....	106

ANNEXE D	107
ANNEXE E	108
ANNEXE F	110
ANNEXE G	111
ANNEXE H	112
ANNEXE I	113
ANNEXE J	114

Liste de tableaux et figures

Tableau 1.1 : Tri de documents, p. 17

Tableau 1.2 : Zone de descriptions, p. 18

Figure 1.3 : Résultats d'une recherche, p. 21

Figure 1.4 : Fiche #13143, p. 22

Figure 1.5 : Fiche #13143, détails, p. 23

Figure 2.1 : Rappel de l'image, fiche #13143, p. 45

Tableau 2.2 : Liste d'autorités de l'institut Getty, p. 49

Introduction

Mise en situation et problématique

La pratique de l'histoire a évolué progressivement au fil du temps. En effet, les techniques de recherche ont changé entre autres avec la démocratisation de l'ordinateur et l'apport des nouvelles technologies dans les centres d'archives¹. Toutefois, bien que ces avancées aient permis une meilleure accessibilité des sources, il reste encore beaucoup de travail à faire en ce qui a trait aux plateformes de recherche en ligne. En effet, ces dernières sont souvent faites de manière très générique avec une indexation trop large ou dans un cas contraire, très spécifique, adapté aux besoins du centre d'archives. Avec cette disparité, il devient donc plus ardu pour un chercheur d'effectuer des recherches cohérentes mais surtout, uniformes d'une base de données à l'autre. Le chercheur est soit confronté à une surabondance de résultats qu'il doit ensuite vérifier, ou passe à côté d'une image pertinente, car elle a été indexée en utilisant des mots spécifiques à une personne. Il est donc nécessaire de se pencher sur cette problématique en réfléchissant aux diverses options qui s'offrent à nous.

Il est important de savoir que l'indexation et la diffusion des images sont en constante évolution dans le domaine de l'archivistique et de la recherche en histoire. En effet, la numérisation des archives photographiques et l'explosion de la photographie numérique nous obligent à réfléchir à l'intégration de nouvelles méthodes pour l'indexation de ces images.

¹ La revue *Histoire et mesure* mentionnait dans les années 80 que : « l'ordinateur aurait d'ores et déjà profondément transformé la pratique de nombreux historiens » (Éditorial, *Histoire & mesure*, vol. I, n°1, 1986, p. 5-6), cité par Franziska Heimbürger, et Émilien Ruiz, « Faire de l'histoire à l'ère numérique : retours d'expériences », *Revue d'histoire moderne et contemporaine*, vol. 5, n° 58, 2011, p. 70.

Notre travail de recherche vise donc à réfléchir sur la perspective d'améliorer l'accessibilité des archives photographiques grâce à l'apport des nouvelles technologies de reconnaissances d'images et d'indexation². Plusieurs études mentionnent les différents défis et besoins qu'apporte l'indexation des images autant pour les chercheurs que pour les archivistes³. Que ce soit pour le classement et l'accessibilité des sources dans les centres d'archives⁴ ou pour la recherche dans le cadre du travail des historiens⁵, il devient évident que le besoin est de plus en plus criant en matière d'accessibilité et de la place de l'image en histoire⁶. Afin de tirer profit des avantages que peuvent procurer la numérisation des archives photographiques et l'ubiquité d'images créées à partir d'appareils numériques, les archivistes auront besoin de nouveaux outils afin d'éviter que

² Index: « Liste alphabétique d'auteurs, de matières, de mots clés, etc., apparaissant dans un ouvrage, avec des références permettant de les retrouver », voir « Index », *Dictionnaire Larousse* [En ligne], 2019, sur le site www.larousse.fr, consulté le 20 janvier 2019, < <https://www.larousse.fr/dictionnaires/francais/index/42560#definition> >.

³ Olivier Le Deuff, *Du tag au like, La pratique des folksonomies pour améliorer ses méthodes d'organisation de l'information*, France, Éditions FYP, 2012, 159 p. ; Rodrigo Almeida, *et al.*, « Interactions et métadonnées riches pour les bibliothèques numérisées », *Document numérique*, vol. 9, n° 2, 2006, p. 83-109 ; James M. Turner, *Images en mouvement, stockage. Repérage. Indexation*, Sainte-Foy, Presses de l'Université du Québec, 1998, 102 p. ; Normand Charbonneau et Mario Robert dirs., *La gestion des archives photographiques*, Sainte-Foy, Presses de l'Université du Québec, 2011, 306 p.

⁴ Plusieurs chercheurs mentionnent la montée toujours grandissante des nouvelles technologies dans le domaine de l'archivistique et de leur importance dans la préservation, l'accessibilité et la diffusion du savoir historique. Sophie Roy, « La diffusion d'archives à l'aide des technologies du Web 2.0 : le cas du Musée McCord », Mémoire de maîtrise (Sciences de l'information), Montréal, Université de Montréal, 2012, 158 p. ; Yvon Lemay et Anne Klein, « La diffusion des archives ou les 12 travaux des archivistes à l'ère du numérique », *Les Cahiers du numérique*, vol. 8, n° 3, 2012, p. 15-48. ; Carol Couture, « Le concept de document d'archives à l'aube du troisième millénaire », *Archives*, vol. 27, n° 4, 1996, p. 3-19 ; Carol Couture et Marcel Lajeunesse, *L'archiviste à l'ère du numérique. Les éléments fondamentaux de la discipline*, Québec, Presses de l'Université du Québec, 2014, 278 p.

⁵ Carol Couture et Jean-Yves Rousseau, *Les archives au XX^e siècle. Une réponse aux besoins de l'administration et de la recherche*, Montréal, Université de Montréal, Secrétariat général et Service des archives, 1982, VI-491 p. ; Yvon Lemay et Anne Klein, « La diffusion des archives... », p. 15-48 ; Franziska Heimburger, et Émilien Ruiz, « Faire de l'histoire à l'ère numérique : retours d'expériences », p. 70-89 ; Pascal Scallon-Chouinard, « Développer les sciences humaines numériques au Québec », *Histoire Engagée* [En ligne], 5 juin 2013, sur le site *Histoireengagee.ca*, consulté le 20 janvier 2017, < <http://histoireengagee.ca/?p=2899> >.

⁶ Comme le mentionne Christian Delporte : « [L'image] révèle les modes de pensée d'une société », dans Christian Delporte, Laurent Gervereau et Denis Maréchal, dirs., *Quelle est la place des images en histoire?*, Paris, Éditions Nouveau monde, 2008, p. 12.

cette nouvelle richesse archivistique demeure sous-exploitée. Puisque le défi est de plus en plus grand en matière d'accessibilité, je me pose donc la question à savoir : comment peut-on à la fois améliorer les systèmes d'indexation ainsi que la mise en ligne des images dans les centres d'archives pour mieux répondre aux besoins des chercheurs? En d'autres termes, comment un programme automatisé de reconnaissance d'images et d'indexation peut-il être intégré à la chaîne de numérisation, de classement et de diffusion des archives afin de faciliter le travail des historiens et des archivistes?

État de la question

L'importance de l'image dans le processus de recherche

Le concept de l'image peut être vu dans une perspective historique, mais aussi sociale, économique et politique. Selon plusieurs chercheurs, l'image a une importance tout aussi notable que le texte écrit⁷. En effet, l'image est un témoin de l'histoire comme le mentionne Marie Chominot. Elle est une source qui transmet les éléments clés d'un événement de sa production à sa diffusion⁸. Claire Sécaïl-Traques démontre qu'elle est créatrice de la mémoire collective d'une société⁹. L'historien peut alors se poser la

⁷ Susan Close, « Reading Photographs: Playing the Visual Literacy Game », *CV Photo* 52, n° 52, 2000, p. 5; Christian Delporte *et al.*, *Quelle est la place des images en histoire?*, 480 p. ; Laurent Gervereau, *Les images qui mentent, histoire du visuel au XX^e siècle*, Paris, Éditions du Seuil, 2000, 458 p. ; Francis Haskell, *L'historien et les images*, Trad. de l'anglais par Alain Tachet et Louis Évrard, Paris, Gallimard, 1995, 781 p. ; Ilse About et Clément Chéroux « L'histoire par la photographie », *Études photographiques* [En ligne], n° 10, novembre 2002, sur le site *études photographiques*, consulté le 8 mars 2017, < <http://etudesphotographiques.revues.org/261> >.

⁸ Marie Chominot, « L'image photographique, une source pour écrire l'histoire de la guerre d'Algérie/ guerre de libération ? », dans Hassan Remaoun et Mahamed Bensalad dirs., *Image, Histoire, Mémoire. Les représentations iconographiques en Algérie et au Maghreb. Actes de colloque international organisé par la CRASC*, Algérie, CRASC, 2007, p. 77-78.

⁹ Claire Sécaïl-Traques, « Image, représentations, mémoire de l'événement », *Vingtième Siècle, revue d'histoire*, n°72, 2001, p. 136-137.

question à savoir : « [...] que vaut l'image pour celui qui la produit, la vit, la lit ou la transmet [?] [...] »¹⁰.

Au fil des recherches historiques, l'image a souvent été sous-estimée par la communauté historique. En effet, plusieurs études mentionnent comment elle est souvent prise en second plan comme source¹¹. Les historiens Ilse About et Clément Chéroux affirment que : « Si l'histoire *de* la photographie est vivace, l'histoire *par* la photographie demeure infertile. Trop peu d'historiens se consacrent à l'analyse d'archives photographiques. Rares sont ceux qui utilisent la photographie au-delà de sa valeur illustrative¹² ». Bien sûr, elle n'est pas totalement absente dans l'historiographie. En effet, quelques historiens ont consacré leurs recherches à l'importance de l'image dans la discipline de l'histoire¹³. C'est le cas de Francis Haskell, lequel intègre l'œuvre d'art à la discipline historique en tant que source grâce à sa production physique, mais aussi par la représentation qu'elle fait des connaissances du passé¹⁴. L'image n'est donc plus un simple auxiliaire à la recherche, elle devient un témoin de l'histoire¹⁵. Christian Delporte mentionne qu'elle « [...] révèle les modes de pensée d'une société¹⁶ ». De sa production à sa diffusion, elle transmet les éléments clés d'un événement¹⁷. Marc Ferro affirme que : « [...] ce sont souvent les images plus que les écrits qui marquent les mémoires¹⁸ ».

¹⁰ *Ibid.*, p. 137.

¹¹ Ilse About et Clément Chéroux, Marie Chominot, Susan Close, Claire Sécail-Traques, etc.

¹² Ilse About et Clément Chéroux « L'histoire par la photographie », p. 1 et 5. Voir aussi : Élisabeth Parinet, « Pour une diplomatique de l'image », dans Christian Delporte et al., *Quelle est la place des images en histoire?*, p. 97.

¹³ Nous pensons, entre autres à l'ouvrage collectif : Christian Delporte et al., *Quelle est la place des images en histoire?*, 480 p.

¹⁴ Francis Haskell, *L'historien et les images*, ..., p. 11-19.

¹⁵ Marie Chominot, « L'image photographique, une source pour écrire l'histoire de la guerre d'Algérie/guerre de libération ? », p. 77-78.

¹⁶ Christian Delporte et al., *Quelle est la place des images en histoire?*, p. 12.

¹⁷ Marie Chominot, « L'image photographique, une source pour écrire l'histoire de la guerre d'Algérie/guerre de libération ? », p. 75 et 77-78.

¹⁸ Marc Ferro, *Cinéma, une vision de l'histoire*, France, Éditions Du Chêne, 2003, p. 7.

Malgré tout, l'image a quand même ses limites en tant que source historique. Elle est un moment fixe dans l'histoire, en représentant une partie d'un événement avec un cadrage, nous empêchant de voir ce qui s'est passé en dehors des limites de cette dernière¹⁹. Que s'est-il passé avant? Qu'est-il arrivé après? La photographie peut devenir esthétique, spécifique, conçue pour charmer ou même provoquer²⁰. Mais avant de pouvoir l'utiliser en histoire, la photographie doit être repérable. Comme nous le verrons, la nature des index et des moteurs de recherche ne permet pas de retrouver des images concernant le même lieu ou les mêmes objets à partir des mêmes mots-clés. Ma recherche s'inscrit donc dans ce besoin de mise en valeur de l'image par l'amélioration des systèmes d'indexation dans les centres d'archives en vue de faciliter le repérage de manière plus uniforme et systématique.

L'accessibilité des sources

La diffusion des archives à l'ère numérique est, selon les historiens Yvon Lemay et Anne Klein, l'opportunité d'un renouvellement²¹. Elle apporte plusieurs transformations, particulièrement en ce qui a trait à l'accessibilité de la documentation, autant photographique, audio, vidéo que manuscrite²². Avec la création d'une archive en ligne, il devient possible, comme le mentionne Rémy Besson, de « mettre de l'avant la complémentarité entre l'innovation technologique et une série de gestes créateurs posés par les individus impliqués dans son élaboration²³ ». Cette complémentarité entre la

¹⁹ Laurent Gervereau, *Les images qui mentent, histoire du visuel au XX^e siècle*, p. 319.

²⁰ Élisabeth Parinet, « Pour une diplomatique de l'image », p. 101-102. Voir aussi : Laurent Gervereau, *Les images qui mentent*, p. 319.

²¹ Yvon Lemay et Anne Klein, « La diffusion des archives ... », p. 16.

²² Rodrigo Almeida, *et al.*, « Interactions et métadonnées riches pour les bibliothèques numérisées », p. 99; James M. Turner, *Images en mouvement*, 102 p.

²³ Rémy Besson, « Regard intermédiaire sur le devenir numérique des archives », *Intermédialités : histoire et théorie des arts, des lettres et des techniques*, n° 23, 2014, para. 2.

gestion de la documentation et les nouvelles technologies se marie facilement avec la technologie de la reconnaissance d'images. Les recherches menées depuis plusieurs années dans le domaine de la reconnaissance de formes offrent de nombreuses solutions dont la fragmentation d'images, la segmentation et même l'intelligence artificielle. De plus, les nouvelles technologies favorisent la diffusion et la démocratisation du savoir²⁴. La mise en ligne des ressources permet la réédition de plusieurs archives, ainsi qu'une meilleure protection des originaux²⁵. La numérisation a apporté des changements importants dans la manière de classer, traiter et transmettre des documents²⁶. Ces évolutions techniques dans le domaine donnent des possibilités nouvelles, mais il faut toutefois utiliser des logiciels spécifiques, connaître leurs caractéristiques et aussi les spécialistes qui savent les utiliser. Que ce soit l'organisation, le classement et les procédures à suivre²⁷, il ne faut pas non plus négliger les métadonnées qui « décrivent différents attributs d'informations et leur donnent signification, contexte et organisation²⁸ ». À l'instar de l'ancienne fiche de bibliothèque, elles permettent de contrôler et de décrire les collections²⁹.

²⁴ Jean-Marc Alain, « La convergence : enjeux et défis », *Documentation et bibliothèques*, vol. 47, n° 2, 2001, p.72; voir aussi : Rolande Blair, « La gestion des documents électroniques : le point de vue de la Bibliothèque nationale du Canada », *Documentation et bibliothèques*, vol. 42, n° 3, 1996, p. 127.

²⁵ Michel Cotte, « Les outils numériques au service de l'histoire des techniques », *Revue d'histoire des techniques*, vol. 1, n° 2, 2012, p. 13.

²⁶ Marcelo Dascal, « Culture numérique. Enjeux pragmatiques et philosophiques », *Diogène*, vol. 3, n° 211, 2005, p. 31. Par transmission, nous pensons entre autres à l'étude de Boyer, Cheetham et Johnson mentionnant la mise en ligne, grâce à un serveur web collaborant avec d'autres organisations comme Google Earth, de plusieurs photographies de divers bâtiments et activités qui eurent lieu dans la ville de Philadelphie depuis 1850. Grâce à cette plateforme, le public peut s'intéresser, d'une manière interactive, à l'histoire de la ville. Deborah Boyer, Robert Cheetham et Mary Johnson, « Using GIS to Manage Philadelphia's Archival Photographs », *The American Archivist*, vol. 74, n° 2, 2011, p. 652-663.

²⁷ Michel Cotte, « Les outils numériques au service de l'histoire des techniques », p. 15.

²⁸ Bibliothèques de l'Université Cornell, *De la théorie à la pratique. Didacticiel d'imagerie numérique* [site Web], 2000-2003, consulté le 1 octobre 2018, < <http://preservationtutorial.library.cornell.edu/tutorial-french/tutorial-french.pdf> >.

²⁹ Gill Tony, « Metadata and the Web, An Introduction to Metadata: Pathways to Digital Information, edited by Murtha Baca », *Online Edition (Version 3.0)* [En ligne], 2008, sur le site *Getty éducation*, consulté le 20

L'abondance de données numériques que nous accumulons depuis les années 90 nous amène à réfléchir sur la manière que l'on doit traiter les documents. Particulièrement dans une société où la plateforme informatique devient un outil essentiel pour la recherche et la gestion de données. Des historiens comme Ian Milligan expliquent que les sources historiques qui sont basées sur le fameux *World Wide Web* représentent un défi important pour les archivistes et les historiens en matière de techniques de classement, d'accessibilité, de stockage, mais aussi, de recherches³⁰. En effet, comment un document hébergé sur le web peut-il devenir une source historique? La restructuration de la pratique historique grâce au classement numérique a permis une augmentation significative à l'accès à l'information. Toutefois, afin de faire des analyses de données précises sur certains sujets, nous devons avoir des outils de classement adéquat et adaptés aux nouveaux besoins des chercheurs. Dans cette mesure, les archivistes doivent s'adapter en arrière-plan afin de rendre le document facilement accessible. Pour reprendre les mots de Milligan: « *This is a useful segue into the world of Big Data and what historians and others need to know to leverage it*³¹ ».

Hypothèse

Partant des différentes lectures concernant les approches scientifiques et informatiques sur la question, nous pensons qu'il est possible d'intégrer un logiciel de reconnaissance d'images dans la chaîne de traitement afin d'améliorer le repérage d'images pertinentes. En effet, les débats entourant les avantages respectifs de la

janvier 2017, <
http://www.getty.edu/research/publications/electronic_publications/intrometadata/metadata.html >.

³⁰ Ian Milligan, *History in the age of abundance? How the web is transforming historical research*, Montreal & Kingston, McGill-Queen's University Press, 2019, p. 213 et 214.

³¹ *Ibid.*

taxonomie et de la folksonomie ont démontré que la masse d'images à indexer requiert de nouveaux outils³². Bien que le système d'indexation sociale soit récent et révolutionnaire, plusieurs chercheurs maintiennent que pour qu'un système indexé soit viable à long terme, le classement se doit d'être contrôlé de manière précise³³. Nous proposons donc d'intégrer un programme d'indexation automatisé au processus humain de classification³⁴ en vue de créer un meilleur apport référentiel dans le processus d'identification des champs indexés et un meilleur accès aux images³⁵. De plus, l'intégration d'une liste de mots génériques pourrait permettre l'incorporation d'un logiciel de reconnaissance de formes afin d'effectuer un premier classement de données. En effet, la reconnaissance d'image donne une vue d'ensemble des éléments génériques tels les bâtiments, les rues, la végétation, l'architecture, etc. Nous estimons donc qu'une première analyse avec le logiciel permettrait à l'être humain de compléter par la suite avec des éléments beaucoup plus spécifiques à l'image. Cette double indexation -- utilisant les forces de l'informatique pour le traitement général et uniforme de grands corpus ainsi que celles des archivistes pour l'ajout d'informations nécessitant une connaissance plus fine du contexte -- offrira des résultats plus prévisibles pour les chercheurs et facilitera l'utilisation de l'image tant à des

³² La taxonomie représente une indexation pré-déterminée par un système. Nous pouvons prendre comme exemple connu l'index des mots à la fin d'un ouvrage. La folksonomie représente pour sa part l'indexation par l'utilisateur, selon une pratique courante sur les sites de partages de photos tels que Flickr, Instagram, etc. Voir aussi : Roxane Borgès Da Silva, « Taxonomie et typologie : comment rapprocher les cadres conceptuels des différentes écoles de pensée ? », *Santé Publique*, vol. 26, n°1, 2014, p. 87-88.

³³ Olivier Le Deuff, « Folksonomies. Les usagers indexent le Web », *Bulletin des bibliothèques de France (BBF)* [En ligne], n° 4, 2006, sur le site *BBF*, consulté le 28 février 2018, < <http://bbf.enssib.fr/consulter/bbf-2006-04-0066-002> >; Olivier Le Deuff, *Du tag au like*, 159 p.

³⁴ Classification : « La classification a pour but d'identifier et de mettre en ordre intellectuellement les documents d'archives à l'intérieur d'un fonds ou d'une collection. Elle cherche à répondre à des besoins administratifs, de repérage ou de recherche et d'aide au traitement des documents ». Dans Normand Charbonneau et Mario Robert dirs., *La gestion des archives photographiques*, p. 83.

³⁵ Ces logiciels permettent l'intégration d'une liste de mots génériques très longue, voire même trop longue pour que l'indexation humaine soit en mesure de le faire sans aide de logiciels.

fins plus traditionnelles d'illustration, mais surtout à des fins d'analyse des contenus iconographiques.

Démarche

Afin de mener à bien la réflexion, ce mémoire sera donc divisé en trois chapitres. Dans le premier chapitre, il sera question de l'analyse des bases de données d'indexation afin de comprendre comment ces dernières fonctionnent et comment nous pourrions associer ce type de procédé à un logiciel de reconnaissance de formes. Dans la première sous-section, nous porterons un regard sur le système d'indexation employé par le centre d'archives régional choisi dans le cadre de la recherche soit, la Société d'histoire de Sherbrooke (Musée d'histoire de Sherbrooke). Cette partie nous permettra d'établir les bases du principe d'indexation d'un centre d'archives régional et de comprendre comment fonctionne leur système de classement. Par le fait même, nous ferons une recherche sur leur base de données en ligne afin de voir quelles sont les informations disponibles au public et nous analyserons la forme et le contenu d'un fonds spécifique pour comprendre leur méthode de classement et leur adéquation avec les besoins des chercheurs. Finalement, nous porterons un regard sur la pertinence d'intégrer un logiciel de reconnaissance de formes ainsi qu'un système automatisé d'indexation au procédé déjà en place. Dans la deuxième sous-section, nous analyserons des bases de données d'institutions plus grandes, soit le musée McCord et les Archives de la Ville de Montréal, lesquelles sont confrontées à la gestion d'un très grand nombre de documents iconographiques. Il sera question tout d'abord d'appliquer une vue d'ensemble afin de comprendre leur fonctionnement et par la suite, nous les comparerons afin de comprendre pourquoi l'indexation et les bases de données en ligne ont une importance considérable

dans le repérage des images pour la recherche. À une époque où les centres d'archives régionaux acquièrent de plus en plus d'images, quelles leçons peut-on tirer de l'expérience des centres de plus grande taille? À travers toutes ces réflexions, nous nous poserons la question à savoir : est-ce que ces systèmes de classement répondent aux besoins des historiens dans leurs recherches et comment faire pour les intégrer de manière efficace?

Dans le deuxième chapitre, nous mettrons en perspective les types d'indexations possibles dans un système de classement d'images soit, la taxonomie et la folksonomie. Dans la première sous-section, nous mettrons de l'avant deux systèmes potentiels d'indexation automatisés³⁶, soit Iconclass³⁷ et le Art and Architecture Thesaurus (AAT) de l'Institut Getty³⁸. Cette analyse nous permettra de cerner l'importance d'une bonne indexation dans l'identification d'images en ligne et l'intérêt des taxonomies internationales pour rendre les recherches plus prévisibles pour les chercheurs. Dans la deuxième sous-section, nous réfléchirons à la pertinence de l'indexation par l'utilisateur que l'on nomme la folksonomie afin de comprendre la différence entre une indexation automatisée taxonomique et une indexation définie par le public. Il sera question tout d'abord de connaître ce que représente ce procédé et comment il est possible de l'appliquer dans la vie de tous les jours. Par la suite, nous étudierons ses caractéristiques en fonction des besoins des chercheurs en histoire. Cette partie nous permettra de mieux comprendre les diverses fonctions entourant les types d'indexations et leurs apports dans un contexte de recherche.

³⁶ Systèmes taxonomiques.

³⁷ Iconclass, *History of Iconclass* [site Web], 2012, consulté le 18 juillet 2018, <<http://www.iconclass.nl/about-iconclass/history-of-iconclass/history-of-iconclass>>.

³⁸ The Getty Research Institute, *About the Research Institute* [site Web], 2017, consulté le 1 octobre 2018, <<http://www.getty.edu/research/institute/>>.

Dans le troisième chapitre, nous étudierons les divers logiciels de reconnaissances d'images potentiels pour la numérisation des documents iconographiques. Dans cette partie, il sera alors possible de comprendre comment fonctionne la reconnaissance d'objet et pourquoi il serait intéressant de l'intégrer au processus de numérisation. Dans la première sous-section, nous porterons un regard sur les principes techniques³⁹ de la reconnaissance de formes afin d'en comprendre les grandes lignes et les raisons qui portent à croire en l'efficacité de la technique dans une approche archivistique. Par le fait même, il sera aussi question de l'analyse des approches hybrides déjà existantes dont nous trouvons des exemples dans l'imagerie médicale. Par la suite, dans la deuxième sous-section, nous analyserons quelques études de cas concernant des travaux effectués auparavant sur le principe de reconnaissance de formes afin de voir s'il est possible de l'utiliser dans un contexte archivistique. Nous serons alors en mesure de comprendre l'importance de cette technologie dans les recherches, mais aussi de manière plus générale dans les avancées technologiques. Dans la troisième sous-section, nous analyserons quelques outils déjà sur le marché afin de voir s'ils seraient adéquats pour les besoins des centres d'archives en région. Ce chapitre va donc nous permettre d'effectuer un lien entre les systèmes d'indexation ainsi que la reconnaissance de formes afin de prendre conscience de ce que la technologie peut nous apporter et comment elle peut favoriser à la fois la recherche et l'analyse historique.

En guise de conclusion, nous ferons une mise en relation des observations faites tout au long du mémoire. Nous évaluerons alors s'il est possible d'offrir un outil permettant l'amélioration de la recherche d'images auprès de centre d'archives régionaux.

³⁹ Nous pensons alors à l'intelligence artificielle et à la segmentation d'images.

Dans un premier temps, nous effectuerons une comparaison entre les principes de reconnaissance de formes et les systèmes d'indexation automatisés. Nous nous poserons alors la question à savoir : quels sont leurs avantages et leurs désavantages en contexte archivistique et de recherche et s'il est possible de les mettre en relation afin d'améliorer l'efficacité de la mise en ligne des ressources? Dans un deuxième temps, nous reviendrons sur les données recueillies dans l'analyse de photos sherbrookoises en lien avec les défis d'un centre d'archives régional ainsi que ceux des institutions plus grandes. Nous pourrons alors voir comment il est possible d'intégrer une nouvelle forme de traitement des images dans la chaîne déjà établie du processus de numérisation. Finalement, dans un troisième temps, il sera question des bonifications et des recommandations ressorties à la suite des analyses et des perspectives d'avenir envisagées. Quels sont les avantages pour les chercheurs et comment les mettre en application. Aucun logiciel n'étant présentement en mesure de répondre à nos besoins, ce chapitre formera en quelque sorte le devis pour la conception d'un outil qui appuiera les archivistes dans leur travail et qui facilitera la recherche par les chercheurs et par le public.

Au terme de ce travail, nous serons en mesure de déterminer s'il est possible de concevoir un outil permettant une meilleure accessibilité des images auprès des chercheurs et du public. Par cette approche, nous pourrons mettre en lumière les différentes possibilités accessibles aux centres d'archives et les avantages qu'elles pourraient apporter grâce à l'amélioration de l'accessibilité des sources iconographiques.

Chapitre 1 : Le classement et l'accès aux banques de photographies historiques

Pour un chercheur, la découverte d'une image pertinente à sa recherche repose sur la manière dont le document a été classé et décrit. Le choix des métadonnées associées à une photographie a donc un impact sur sa « découvrabilité⁴⁰ ». L'indexation, activité au cœur du travail des archivistes, est l'étape du traitement qui facilitera ou non le repérage des documents. Pour l'historien qui utilise les outils de recherche proposés par les centres d'archives ou même les musées, l'indexation permet non seulement de trouver rapidement l'information, mais aussi de détecter des sources qui auraient pu être mises de côté lors d'une recherche effectuée de manière plus générale. Avec la technologie d'aujourd'hui, les outils informatiques permettent aux chercheurs de repérer rapidement l'information dont ils ont besoin dans le cadre de leurs travaux⁴¹, mais ces résultats ne sont pas encore optimaux, comme nous le verrons. La recherche informatique peut même donner une fausse impression de pouvoir trouver tout ce qu'on cherche alors que les résultats dépendent de la qualité de l'indexation. Il en devient donc essentiel que l'indexation soit précise dans les termes employés, mais aussi, qu'elle ait une certaine uniformité dans son approche. Nous proposons de ce fait d'examiner quelques banques de données pour comprendre les mécanismes, les forces et les lacunes des systèmes en place dans quelques institutions. En identifiant les limites actuelles des systèmes nous pourrions pousser la

⁴⁰ Traduction directe du concept anglais de « *discoverability* ». « Potentiel pour un contenu, un produit ou un service de capter l'attention d'un internaute de manière à lui faire découvrir des contenus autres ». Office québécois de la langue française, *Fiche terminologique, découvrabilité* [site Web], 2012, consulté le 28 mars 2019, <http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26541675>.

⁴¹ Voir : Ilse About et Clément Chéroux, « L'histoire par la photographie », 20 p.

réflexion quant aux systèmes de classement et à l'apport de la reconnaissance des images pour améliorer le processus d'indexation.

1.1 Les défis d'un centre d'archives régional

Bien qu'il ne dispose pas d'une très grande collection, un centre d'archives régional présente une échelle appropriée pour évaluer les modes d'indexation en usage par une petite équipe. Dans le cadre de ce mémoire, nous avons analysé le cas de la Société d'histoire de Sherbrooke. Fondée le 2 mars 1927, l'institution portait alors le nom de Société d'histoire des Cantons-de-l'Est⁴². Dès ses débuts, elle avait pour vocation la diffusion de l'histoire de Sherbrooke et des Cantons-de-l'Est, et la conservation du patrimoine. L'institution a pour mission aujourd'hui de : « Conserver, étudier, mettre en valeur et diffuser le patrimoine historique documentaire et ethnologique de la région sherbrookoise; ce faisant, susciter l'intérêt et le sentiment d'appartenance des Sherbrookoises et Sherbrookoises pour leur ville et son histoire et contribuer à faire de Sherbrooke une destination touristique reconnue⁴³ ». De plus, la Société d'histoire détient aussi un service d'archives privés qui reçut son agrément des Archives nationales du Québec en 1990. Le service a pour mission de : « [...] conserver les collections d'archives, de les compléter et de les rendre accessibles au public et aux chercheurs⁴⁴ ». Rendre accessible au public et aux chercheurs est un élément clé en ce qui a trait à la base de

⁴² La Société d'histoire de Sherbrooke, *Histoire Sherbrooke, à propos de la Société d'histoire de Sherbrooke* [site Web], 2018, consulté le 15 février 2019, <http://www.histoiresherbrooke.org/a_propos_de_la_shs>. N. B. La Société d'histoire de Sherbrooke a pris le nom de Musée d'histoire de Sherbrooke (MHIST) pendant la rédaction du mémoire.

⁴³ *Ibid.*

⁴⁴ *Ibid.*

données en ligne proposée par l'institution. De ce fait, nous analyserons les diverses possibilités offertes aux chercheurs en effectuant des recherches sur cette dernière.

Intégrer un centre d'archives régionales à la recherche correspond aussi à ce besoin de mettre de l'avant les archives plus spécifiques à une région qui reçoivent un intérêt de la part de la population. Yvon Lemay et Anne Klein ont signalé les éléments à prendre en compte afin de rendre accessibles les archives au public élargi, sans négliger la place de l'utilisateur. Selon les enquêtes menées par l'Observatoire de la culture et des communications du Québec en 2004, 11,4% des Québécois affirment avoir visité un centre d'archives, autant pour des recherches en histoire ou en généalogie⁴⁵. De plus, on y mentionne que : « [...] la grande majorité des chercheurs fréquentent les centres régionaux des archives nationales ainsi que les services des sociétés d'histoire et autres associations d'histoire locale. En revanche, les services d'archives liés au réseau de la santé et de l'éducation ainsi que ceux des secteurs public et gouvernemental attirent beaucoup moins d'utilisateurs⁴⁶ ». L'amélioration de l'indexation des archives devrait donc faciliter l'accès aux données et potentiellement augmenter la fréquentation grâce à une meilleure « découvrabilité » dans un contexte où il ne faut pas négliger les centres locaux et régionaux.

⁴⁵ R. Garon, *Le public des institutions patrimoniales. État des lieux du patrimoine, des institutions muséales et des archives*, Québec, Institut de la statistique du Québec, Observatoire de la culture et des communications du Québec, Cahier 9, 2010, p. 14 cité par Yvon Lemay et Anne Klein, « La diffusion des archives ... », p. 32. « [...] Et c'est une proportion relativement élevée lorsqu'elle est comparée à des données venant d'autres pays ».

⁴⁶ L. Gagnon-Arguin, *Les archives au Québec, des ressources documentaires à découvrir, État des lieux du patrimoine, des institutions muséales et des archives*, Québec, Institut de la statistique du Québec, Observatoire de la culture et des communications du Québec, Cahier 6, 2008 cité par Yvon Lemay et Anne Klein, « La diffusion des archives ... », p. 32.

Les méthodes d'indexation des différents centres d'archives sont peu accessibles au public, ce qui justifie aussi le choix d'une institution pour laquelle nous avons accès aux informations concernant les pratiques internes. Nous avons donc rencontré l'archiviste⁴⁷ de la Société d'histoire de Sherbrooke afin de bien comprendre le protocole de l'institution analysée.

Le centre d'archives a établi un protocole clairement défini pour les stagiaires, membres du personnel et étudiants qui effectuent le classement et l'indexation des archives. En effet, lorsqu'il est question de concevoir les fiches techniques des images à traiter, un cahier montrant les étapes à suivre est remis afin d'uniformiser la démarche. Le guide décrit chaque étape à suivre lors du traitement des images : le tri des documents, comment les titrer et les coter, comment les regrouper, comment compléter la description et nous retrouvons aussi les diverses procédures plus manuelles comme la mise en enveloppe et le rangement dans la voute⁴⁸.

Les éléments indiqués dans le tri⁴⁹ ne sont pas à négliger, car ils déterminent les items que nous retrouvons ensuite dans les fonds qui contiennent des images et, par la bande, les sujets qui seront traités.

⁴⁷ Karine Savary, conservatrice : karine.savary@histoiresherbrooke.org

⁴⁸ SHS, *Guide pour le traitement des archives iconographiques*, Sherbrooke, SHS, 1987-1988 (révisé en 2011), 45 p.

⁴⁹ La personne qui effectue le tri doit posséder une culture historique adéquate afin de pouvoir juger de la valeur du document mais aussi, pour bien classer ce dernier selon Normand Charbonneau et Mario Robert dirs., *La gestion des archives photographiques*, p. 101.

Tableau 1.1 : Tri des documents

Critères des documents à conserver
- Illustrant l'avènement d'un procédé technique et son évolution
- Dont l'acteur a joué un rôle significatif dans l'histoire.
- Qui constitue une source de renseignements uniques sur des individus, des événements, des sujets ou des lieux.
- Qui fournissent des informations sur la collectivité québécoise de manière significative et révélatrice en matière d'ethnographie, d'ethnologie, d'anthropologie, de science et de technique.
- Valeur esthétique.
- Pris par un grand photographe.
Critères des documents à éliminer
- Défauts techniques majeurs.
- Faisant double emploi ou qui n'apporte pas d'informations supplémentaires.
- Document altéré par des agents de détérioration chimique ou biologique.
- Les doublons et multiples. (Conserver seulement ceux exceptionnels).
- Les négatifs en nitrate (Transférer avant).
- Les négatifs à pigmentation instable (polaroid) (Transférer avant).
- Ceux non identifiables.

Source : SHS, *Guide pour le traitement des archives iconographiques*, ..., p. 3 et 4

C'est ensuite dans le bordereau de description que l'on doit indiquer les métadonnées⁵⁰ pertinentes : le titre du fonds, la localisation, la cote du fonds, le code du support, le droit de consultation, le titre du sujet, la mention de responsabilité, la date de création, le numéro du négatif, les copies disponibles, la collation⁵¹, les notes, une description sommaire, les descripteurs, la date du traitement et le documentaliste⁵². Parmi ces éléments, nous retiendrons la description sommaire ainsi que les descripteurs. Voici un tableau résumant la portée et le contenu des catégories prises en considération lors de la rentrée des données:

⁵⁰ Les métadonnées « [...] décrivent différents attributs d'informations et leur donnent signification, contexte et organisation »; Bibliothèques de l'Université Cornell, *De la théorie à la pratique. Didacticiel d'imagerie numérique* ... Voir aussi: Office québécois de la langue française, *Fiche terminologique, métadonnées* [site Web], 2012, consulté le 2 novembre 2019, < http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8869869 >.

⁵¹ SHS, *Guide pour le traitement des archives iconographiques*, p. 10. Nombre d'unités, type de support, couleur dimension, etc.

⁵² *Ibid.*, p. 7-14.

Tableau 1.2 : Zone de descriptions

La description sommaire	La description sommaire correspond à la description globale de l'image. <i>Quels sont les éléments qui composent l'image?</i> Partant du plus large au plus précis, la personne qui effectue la description doit englober l'ensemble de l'image. Lorsqu'il s'agit d'une photo de groupe, les nommer de gauche à droite, de l'avant vers l'arrière. De plus, il est aussi question du point de vue de l'image. En effet, il permet de constater ce que le photographe voulait immortaliser sur la photographie, la position adoptée par ce dernier ainsi que son point de vue.
Les descripteurs	Les descripteurs constituent les lieux, les sujets et les groupes de personnages. Il y a les grands thèmes, les sous-thèmes et les sous-sous-thèmes. (Ex : Construction diverses / édifices religieux / églises). C'est l'indexation de l'image. Aller du plus large au plus précis.

Source : SHS, *Guide pour le traitement des archives iconographiques*, p. 7-14.

La description sommaire ainsi que les descripteurs représentent une indication sur le contenu de l'image numérisée. Pour l'archiviste, ces catégories constituent une identification supplémentaire dans le classement. Autant par les mots d'indexation qui permettent de rechercher plus facilement les catégories associées à l'image que la description sommaire qui permet de mieux situer l'image dans son contexte historique mais aussi temporel. Pour l'historien qui utilise la base de données en ligne proposée par l'institution, ces identifications permettent de mieux comprendre l'image. Plus il y a d'informations, plus il sera facile de situer la photographie dans son contexte historique. Le chercheur ou la chercheuse peut alors mieux repérer les grands thèmes qui la composent. Il faudra toutefois plus de temps pour ajouter ces métadonnées.

De façon plus didactique, **les descripteurs** s'orientent vers les termes d'indexation tandis que **la description sommaire** constitue un ajout à l'information ainsi que le point

de vue de l'image. Les descripteurs pour leurs parts pourraient être bonifiés avec un logiciel qui permettrait une pré-indexation de l'image. Pourquoi? Comme nous le verrons plus loin, avec une première indexation qui sera composée de mots génériques, il devient plus facile pour le technicien ou la technicienne qui numérise et entre les informations de savoir par où commencer lors de la rentrée de données. Que ce soit bâtiment, rue, ville, commerce, etc. Avec des mots de bases indiqués par le logiciel, le technicien ou la technicienne pourra alors se repérer beaucoup plus facilement dans ses recherches. En combinant la reconnaissance d'image qui passe d'abord par le point de vue de l'image et l'indexation prédéfinie avec les mots génériques, l'indexation deviendrait alors beaucoup plus complète et uniforme.

La liste des termes inscrits dans le guide reflète les sujets privilégiés par les archives. L'annexe A présente quelques grandes thématiques et quelques sous-points qui y sont rattachés que nous avons relevés, lesquels témoignent des priorités en matière d'indexation. La dernière partie du guide présente comment inscrire les éléments dans le système⁵³. Les procédures orientent le technicien ou la technicienne du début à la fin, autant pour la saisie de données dans la base de données que dans le dépôt physique des documents.

La base de données actuelle reprend un ancien système créé à l'aide du logiciel Edibase⁵⁴, créé par la firme québécoise Documensa⁵⁵. Conçu pour la gestion documentaire, ce logiciel avait aussi la particularité d'imposer l'utilisation de majuscules.

⁵³ SHS, *Guide pour le traitement des archives iconographiques*, p. 39.

⁵⁴ Entrevue avec Karine Savary.

⁵⁵ « Text Retrieval with New Wrinkles : Documensa's Edibase SGML », *XML.COM* [En ligne], 10 janvier 1998, sur le site *XML.com*, consulté le 13 décembre 2019, < <https://www.xml.com/pub/a/SeyboldReport/ip020516.html> >.

Cette approche ne répondant plus aux besoins du 21^e siècle, l'archiviste a développé une nouvelle base de données en 2011⁵⁶. Cette dernière, bien que n'étant pas inscrite sur la liste des logiciels recommandés pour les services d'archives privées, répond bien aux besoins de l'institution. L'approche du centre d'archives vise donc à aller au-delà de la simple description de l'image, afin de maximiser l'efficacité des recherches et permettre une meilleure diffusion des données numérisées. En allant au-delà de la simple description statique, l'archiviste offre une meilleure compréhension de l'image pour le chercheur ou la chercheuse qui utilise la base de données à des fins de recherche.

Les recherches simples sur la base de données en ligne traversent tous les champs de tous les modules : les archives, l'audiovisuel, la bibliothèque, les cartes, les objets, les périodiques, les documents sonores, la description, etc. Par exemple, nous pouvons retrouver des termes comme « Wellington » dans les documents textuels et même audiovisuels. L'option de recherche avancée permet de limiter la recherche à une collection ou à un champ spécifique (Annexe B). Nous pouvons préciser un lieu, un personnage ou même un fonds complet. Nous avons choisi, pour les fins de cet exercice, le terme Wellington. Les résultats sommaires (figure 1.3) indiquent la variété des fonds dans lesquels ce terme, associé à la principale rue commerciale du centre-ville de Sherbrooke, apparaît. L'annexe C présente une liste de résultats pour les documents iconographiques. Bien que ce terme soit précis, il est tellement utilisé dans le contexte sherbrookoïse qu'il génère une très grande quantité de résultats.

⁵⁶ Depuis 2019, l'institution travaille au renouvellement de la base de données afin qu'elle réponde de manière encore plus précise aux besoins du centre d'archives.

Figure 1.3 : Résultats d'une recherche

Archives (21)
Audiovisuels (9)
Bibliothèque (4)
Cartes (14)
Images (903)
Objets (23)
Périodiques (10)
Sonores (4)

Source : SHS, *Collection* [Banque de données], sur le site *Société d'histoire de Sherbrooke*, consulté le 13 décembre 2019, <<https://www.histoiresherbrooke.ca/search.php>>.

Afin d'approfondir notre analyse de l'indexation des images et de leur repérage à la Société d'histoire de Sherbrooke, nous avons utilisé un fonds comprenant une variété d'images : le Fonds Frederick James Sangster (IP465). On y retrouve plusieurs illustrations de paysages de la ville de Sherbrooke et des environs, réparties en 107 photos, 2 négatifs, 3 mosaïques, 2 cartes postales, 1 lithographie et 4 cm de texte. Dans la base de données en ligne, nous ne pouvons pas (au moment de la rédaction de ce mémoire) avoir accès au descriptif du fonds. D'autres centres d'archives, tels que Bibliothèque et Archives nationales du Québec, affichent la portée et le contenu des fonds sur le moteur de recherche⁵⁷. Une modification à l'interface en ligne permettrait de rendre cette information accessible aux chercheurs. Dans le cas du Fonds IP465, la notice inscrite dans la base de données apporte les détails suivants :

Notes sur l'auteur :

Né en 1907, Frederick James Sangster est le fils de James Riddle Sangster, directeur de la *Page-Sangster Printing Co. Ltd.* Éduqué aux écoles primaire et secondaire de Sherbrooke, Frederick James complète sa formation au *Carnegie Institute of Technology*, de Pittsburgh, en Pennsylvanie, où il obtient un diplôme en imprimerie. Il est associé à la firme de son père dès 1930. Il en prendra la direction. Époux de Reta Janet Michie, il est le père de 3 enfants. Frederick James Sangster est décédé le 9 juin 1998⁵⁸.

⁵⁷ Voir par exemple la notice concernant le Fonds Cour provinciale du District de Saint-François. Bibliothèque et Archives nationales du Québec, TL 27, <<https://advitam.banq.qc.ca/notice/446437>>.

⁵⁸ Tiré de la description du fonds dans la base de données de la Société d'histoire de Sherbrooke.

Le fonds a été versé au cours des années 1996 à 1998 : « Une convention de don signée par Donald Sangster, fils de Frederick James, a confirmé le don ».

Portée et contenu:

Le fonds est principalement constitué de photos anciennes. Certaines photos sont des vues des quartiers occupés par les Sangster: d'abord le secteur Fairmount, puis le quartier Liverpool où sera construite la résidence Sangster, sur la rue Belvédère Nord. D'autres photos présentent des clubs sociaux, parcs et vues diverses de Sherbrooke et des Cantons-de-l'Est. La plupart des photos sont l'oeuvre de J.E. Morkill qui collabore à une série de publications sur les attraits de Sherbrooke et à quelques calendriers régionaux, édités par la firme *Page-Sangster*. Les documents textuels sont des factures et autres documents relatifs à la construction, en 1898, de la résidence Sangster, sur la rue Belvédère Nord⁵⁹.

À partir de ce fonds, nous avons sélectionné une image comprenant divers éléments du paysage urbain, question d'établir des bases concrètes concernant l'importance d'une bonne indexation dans le processus de recherche d'images dans une base de données.

Figure 1.4 : Fiche #13143



Source : SHS, *Résultats de la recherche, fiche 13143* [Banque de données], sur le site *Société d'histoire de Sherbrooke*, consulté le 13 décembre 2019, < <https://www.histoiresherbrooke.ca/sheet.php?uid=19805> >.

⁵⁹ *Ibid.*

Figure 1.5 : Fiche #13143, détails

NO:13143
TITRE_DU_FONDS: FREDERICK JAMES SANGSTER
COTE_DU_FONDS: IP465
LOCALISATION: R
CODE_SUPPORT:PN
NO_DOCUMENT:31D 25
CONSULTATION: L
SUJET_PRINCIPAL:LA RUE WELLINGTON NORD AU COIN DE FRONTENAC
DATE_CREATION: PROBABLEMENT, 1925, AVRIL, 25
COULEUR: NON
ETAT_DOCUMENT: TRES BON, JAUNI, PALISSEMENT
DESCRIPTION: A GAUCHE, L'IMMEUBLE WHITING PUIS LE BLOC ODELL. AU NIVEAU DE LA RUE, DANS L'EDIFICE ODELL, LE COMMERCE WILSON H.C. AND SONS. SUR LA RUE, DES AUTOMOBILES ET UN TRAMWAY. SUR LES TROTTOIRS, DES INDIVIDUS. AU CENTRE, L'IMMEUBLE DU QUEBEC CENTRAL RAILWAY ET A SA DROITE, UN HOMME EST GRIMPE DANS LE POTEAU ELECTRIQUE. A NOTER, UN ORME A ETE ABATTU.
LIEUX: SHERBROOKE; RUE FRONTENAC; RUE WELLINGTON NORD
SUJETS: PAYSAGES URBAINS; RUES
NOTES: POUR PLUS D'INFORMATIONS VOIR *LA TRIBUNE* DU 25 AVRIL 1925 P.3.

Source : SHS, *Résultats de la recherche, fiche 13143.*

La fiche d'information est très détaillée et décrit ce que nous voyons sur la photographie, incluant le nom des bâtiments et même de certains commerces. La personne qui a identifié cette image a mis beaucoup d'informations complémentaires, ce qui est intéressant pour un chercheur qui veut en savoir plus sur le sujet. Au niveau de l'indexation, il y a toutefois peu de descriptions de nature générique, lesquelles sont limitées à « paysages urbains » et « rues ». Dans la section note, on nous suggère de consulter un article de *La Tribune* pour avoir plus d'informations, information très utile lorsque nous voulons pousser plus loin la recherche⁶⁰.

Cette base de données en ligne est selon-nous bien adaptée aux besoins des chercheurs de manière générale, et encore plus lorsqu'on utilise la recherche avancée où il nous est possible d'être plus précis dans les termes utilisés. Les informations transmises sont riches et bien ordonnées. Toutefois, il serait intéressant d'avoir facilement accès aux

⁶⁰ *La Tribune*, 25 avril 1925, p. 3.

descriptions des fonds et collections ainsi qu'aux descriptifs des cotes afin de mieux cerner une image ainsi que comprendre le classement des documents. Finalement, ajouter aussi un guide d'aide à la recherche dans la base de données en ligne serait un atout afin de permettre une recherche rapide et efficace pour tous.

Malgré le développement des systèmes informatisés, de nombreuses personnes soumettent des demandes directement à l'archiviste. Les demandes d'utilisation des photographies varient beaucoup en fonction de la clientèle. Elles peuvent venir de particuliers qui sont à la recherche de membres de leur famille ou bien d'informations concernant leur maison d'enfance. Elles peuvent venir aussi de commerçants, d'institutions et d'entreprises qui désirent trouver et afficher dans l'espace public des photographies d'époque de l'édifice dans lequel ils sont installés afin de mettre en contexte leur entreprise. Finalement, il y a aussi des demandes d'utilisation pour des expositions, des événements spéciaux, des anniversaires, des publications (cahiers d'exposition et circuits touristiques), des articles de journaux et autres⁶¹. Ce type de demandes précises ont probablement orienté la nature des termes utilisés dans l'indexation. Ces recherches courantes permettent donc de repérer des images reliées à des gens, à des lieux ou à des phénomènes. Le système peut répondre à ces besoins, mais n'a pas été conçu pour faciliter l'analyse du corpus dans une perspective historique.

Il est certain que pour notre recherche nous avons eu accès à divers documents nous permettant de nous repérer facilement dans la base de données en ligne. C'est pourquoi nous avons décidé d'effectuer un test chez des personnes qui connaissent les bases de données en ligne et chez d'autres personnes qui y sont moins habituées afin de

⁶¹ Entrevue avec Karine Savary.

voir si elle correspond à leurs besoins⁶². Pour ce faire, nous avons montré la photographie choisie comme test préalable pour le projet. Sans dévoiler les métadonnées qui y sont associées, les personnes devaient la retrouver par eux-mêmes. Le test a permis de réfléchir aux étapes suivies afin de retrouver l'image.

Pour la majorité, le fait de connaître le centre-ville de Sherbrooke aura été un atout majeur dans la recherche. Cette familiarité avec le secteur en a amené plusieurs à utiliser la recherche avancée. Les termes les plus employés sont : Carré Strathcona, Wellington, Wellington Nord et Frontenac. Un de ces termes qui a été écrit plusieurs fois, mais qui n'a toutefois pas donné les résultats escomptés était Carré Strathcona. Quelques personnes ont même vérifié sur Google Maps pour confirmer l'appellation du secteur, lequel porte toujours le toponyme Carré Strathcona. Première constatation, le terme est donc manquant dans la description de la fiche technique. La plupart allaient au-delà du terme seul, on y ajoutait plusieurs termes (Wellington, Frontenac, Centre-ville et arbre). Aucune aide en ligne n'étant disponible pour expliquer comment utiliser plus d'un terme, qui sont tous des termes liés à l'image, les usagers n'arrivaient pas à retrouver l'image recherchée. Toutefois, quand on y mettait seulement le terme Wellington, nous retrouvions facilement l'image. Pour arbre, nous avons remarqué que dans la description, on y indique l'espèce d'arbre (orme) mais pas le mot générique arbre. Pourtant, pour une image du même secteur, on retrouve à la fois « Carré Strathcona » et « arbre »⁶³, ce qui indique un manque d'uniformisation des termes utilisés. Notons que pour cette seconde image, les sujets sont

⁶² 7 personnes âgé entre 25 et 90 ans. Il s'agit autant de travailleurs, d'étudiants ou de retraités.

⁶³ SHS, *Résultats de la recherche, fiche 11833* [Banque de données], sur le site *Société d'histoire de Sherbrooke*, consulté le 18 décembre 2019, < <https://www.histoiresherbrooke.ca/sheet.php?uid=18490> >.

« événements », « événements économiques » et « marchés publics ». Les termes utilisés pour la première image analysée (« paysages urbains » et « rues ») n’y apparaissent pas.

Avec les résultats du test, nous constatons qu’il n’est pas aussi facile que l’on aurait pu s’y attendre de retrouver l’image demandée. Comme le mentionne le Bureau canadien des archivistes : « [...] l’indexation ne suffit pas toujours à rendre fidèlement le contenu, tout simplement parce qu’elle supprime de nombreux concepts et empêche l’expression fine des liens entre ceux-ci⁶⁴ ». Il devient donc important de prendre en considération tous les détails qui se retrouvent dans les images indexées soit, d’ajouter des suites de mots, des commentaires en annotation⁶⁵, des résumés indicatifs sur la nature du document, les termes mis de l’avant⁶⁶ ainsi que des résumés informatifs⁶⁷.

1.2 Les archives photographiques dans les institutions métropolitaines

Afin de réaliser un comparatif intéressant de la plateforme informatique de la Société d’histoire de Sherbrooke, nous avons choisi de faire un parallèle avec des institutions plus grandes, soit le musée McCord et les Archives de la Ville de Montréal. Bien que la taille des institutions soit nettement différente, l’important dans cette comparaison est la forme que prennent les plateformes informatiques accessibles au public et aux chercheurs.

Le Musée McCord offre aux usagers du web une collection de très grande envergure : 30 835 peintures, estampes et dessins numérisés sur 85 690 présents dans leur

⁶⁴ Comité de planification sur les normes de description, « L’indexation en archivistique, rapport du groupe de travail sur l’indexation par sujet », *Bureau canadien des archivistes*, n° 4, 1992, p. 83.

⁶⁵ *Ibid.*

⁶⁶ *Ibid.*

⁶⁷ *Ibid.*, p. 87-88.

collection; 980 photographies de costumes, mode et textiles sur 20 680; 5 835 concernant les cultures autochtones sur 16 500; 7 780 concernant les arts décoratifs sur 41 090; 83 340 numérisations de photographies sur 1 317 610, sans compter quelques documents d'archives textuelles⁶⁸.

Dans le cadre du mémoire, ce qui nous intéresse le plus est la partie concernant les archives photographiques numérisées. En effet, dans le descriptif général de la collection, on y mentionne : « Collection de photographies canadiennes regroupant négatifs et tirages d'époque, stéréographies, photographies peintes et photographies composites. Daguerrotypes et autres procédés anciens, albums de famille, photographies vernaculaires, séries documentaires, portfolios d'artistes, équipement photographique et appareils photo datant de 1840 à nos jours⁶⁹».

Plus spécifiquement, la collection regroupe plus de 1 317 610 photographies qui portent principalement sur l'histoire sociale de Montréal, mais qui s'étendent au Québec et au Canada. On y mentionne que l'on peut retrouver dans la collection autant des photographies datant de la période du daguerrotypes que des photographies prises par des appareils numériques en passant par la diapositive. Dans le descriptif de la collection photographique, on mentionne principalement les archives photographiques Notman qui sont le cœur de leur collection. Quelques 450 000 photographies proviennent du studio de photographie fondé par William Notman en 1856⁷⁰.

⁶⁸ Musée McCord, *À propos des collections et de l'offre numérique en ligne* [En ligne], 2018, consulté le 25 juillet 2018, <<http://collections.musee-mccord.qc.ca/fr/clefs/collections/>>.

⁶⁹ *Ibid.* Voir aussi : Musée McCord, *Photographie* [En ligne], 2018, consulté le 25 juillet 2018, <<http://www.musee-mccord.qc.ca/fr/collections/photographie/>>.

⁷⁰ Musée McCord, *Photographie* [En ligne], 2018, consulté le 25 juillet 2018, <<http://www.musee-mccord.qc.ca/fr/collections/photographie/>>.

Lorsque nous explorons la collection, nous tombons sur la page principale de recherche. Nous pouvons faire des recherches avancées ou même explorer par régions, périodes, artistes, etc⁷¹. Dans les recherches avancées, nous devons sélectionner l'onglet photographie. Par la suite, nous sommes dirigés vers la page de recherche. Plusieurs options s'offrent alors à nous : mots-clés, objets, catégorie, sujet, etc. (Annexe D). Lorsque nous avons indiqué « Sherbrooke » dans la barre de recherche, la première image de la liste fut « Vue de Sherbrooke depuis l'est, Qc, vers 1858 ». Le système semble d'abord privilégier les documents concernant un toponyme de type ville à celui de la rue Sherbrooke située dans la ville de Montréal, mais ce choix n'est pas expliqué. Le système a-t-il détecté que la requête provenait d'un ordinateur situé à Sherbrooke ou y a-t-il hiérarchisation des termes toponymiques? Cet exemple montre l'importance d'une indexation précise, hiérarchisée et sémantique lorsqu'il est question de termes identiques dans une même base de données, mais aussi le besoin de concevoir un outil de recherche qui permet d'être plus précis dans les termes utilisés.

La première photographie étant pertinente pour notre démarche, nous l'avons choisie pour notre analyse. En plus du titre complet (« Vue de Sherbrooke depuis l'est, Qc, vers 1858 »), on y retrouve quelques mots clés : photographie, scène urbaine et vue⁷², soit des termes très génériques. L'utilisation de « scène urbaine » au McCord plutôt que « paysages urbains » à la SHS met en relief la difficulté de faire des recherches à partir de différents moteurs de recherche. Les autres métadonnées plus standardisées (le

⁷¹ Musée McCord, *Collections et recherche* [Banque de données], sur le site *Musée McCord*, consulté le 25 juillet 2018, <http://collections.musee-mccord.qc.ca/scripts/search_results.php?Lang=2&keywords=departmentID:00016>.

⁷² Musée McCord, *MP-1974.61 | Vue de Sherbrooke depuis l'est, Qc, vers 1858* [Banque de données], sur le site *Musée McCord*, consulté le 25 juillet 2018, <<http://collections.musee-mccord.qc.ca/fr/collection/artefacts/MP-1974.61>>.

propriétaire, l'année, le type de support, la grandeur, la cote, le copyright, etc.) complètent le portrait des informations disponibles. Le Musée McCord offre aussi une section intitulée « Clefs pour l'histoire », laquelle présente l'histoire de la ville de Sherbrooke qui inclut bien sûr la source historique d'où proviennent les informations. Nous y retrouvons quatre sections historiques intitulées : quoi, où, quand et qui. Dans ces dernières, il est question de l'histoire de Sherbrooke de manière générale. Malheureusement, il y a peu d'informations historiques qui concernent directement la photographie consultée.

Bref, le Musée Mccord propose une collection très riche dont l'indexation sommaire comprend des termes génériques utiles pour certains types de recherche. Tout comme la Société d'histoire de Sherbrooke, le musée se concentre principalement sur la ville où elle se situe. Toutefois, plusieurs mots d'indexation pourraient être ajoutés afin d'aider à bonifier la recherche. Malgré tout, les compléments pédagogiques permettent une meilleure compréhension de la photographie ainsi qu'une meilleure géolocalisation. Une telle information serait très utile à la recherche si elle était accessible à partir du moteur. On constate ici la différence entre le caractère générique de l'indexation d'une grande collection par rapport aux données précises d'une petite collection.

Créé en 1913, le Service d'archives de la Ville de Montréal assure le traitement, la conservation et la communication de documents institutionnels et privés⁷³. Gérant une collection couvrant une période de 1796 à aujourd'hui, la banque de données décrit 4 km de documents et 1 million de photographies. Le service d'archives a pour objectif de :

[...] continuer à augmenter, traiter, préserver et mettre en valeur le patrimoine archivistique de la Ville de Montréal. Nous

⁷³ Archives Montréal, *La section des archives* [En ligne], 2018, consulté le 25 juillet 2018, <<http://archivesdemontreal.com/la-section-des-archives/>>.

entendons simultanément favoriser l'acquisition⁷⁴, la conservation et la diffusion des documents sous forme numérique, tout en renforçant nos collaborations avec des partenaires des milieux archivistiques et historiques. Nous souhaitons particulièrement donner accès au plus grand nombre de documents possible via le Web, afin que tous les citoyens soient en mesure de s'approprier et d'utiliser leur patrimoine archivistique⁷⁵.

Les Archives de la ville de Montréal proposent une section d'aide à la recherche⁷⁶. On y explique comment effectuer de la recherche par leur catalogue et leurs répertoires. Passant de la navigation générale aux symboles de recherches, on y voit comment effectuer une recherche simple et avancée, comment consulter les résultats de la recherche, comment utiliser le presse-papier et finalement, comment fonctionnent l'impression et le téléchargement de documents. Dans le cas de notre recherche sur Sherbrooke, si nous désirons rechercher spécifiquement la ville de Sherbrooke, le guide nous conseille l'utilisation de guillemets anglais afin de définir les termes employés. Toutefois, une requête utilisant **ville de Sherbrooke** sans guillemet place un résultat concernant cette municipalité en premier lieu alors que la requête '**ville de Sherbrooke**' présente d'abord des documents associés à la rue Sherbrooke à Montréal.

Lorsque nous effectuons notre recherche dans la barre de recherche simple, nous sommes dirigés vers une page générale. Pour les besoins de l'analyse, nous avons indiqué le terme « Sherbrooke » pour voir si la recherche allait être aussi vaste que pour le musée McCord. Dans les résultats obtenus, nous avons constaté ici aussi la grande variété de

⁷⁴ Acquisition : « [...] ensemble des mesures, méthodes et procédures employés par un service d'archives dans le but d'accroître le nombre de ses fonds et collections pour en permettre l'exploitation ». Selon les auteurs, plutôt que d'accroître les collections, il s'agirait plutôt « [...] d'enrichir le patrimoine documentaire de l'institution [...] ». Dans : Normand Charbonneau et Mario Robert dirs., *La gestion des archives photographiques*, ..., p. 47.

⁷⁵ Archives Montréal, *La section des archives* ...

⁷⁶ Archives Montréal, *Aide à la recherche* [En ligne], 2018, consulté le 25 juillet 2018, <<https://archivesdemontreal.ica-atom.org/help>>.

résultats proposés, soit 1419 documents. Un service d'archives municipales étant principalement centré sur son territoire, le terme Sherbrooke est principalement lié à la rue Sherbrooke de Montréal⁷⁷. Dans la même veine, le terme Montréal dans une base de données sherbrookoise mènerait à des documents concernant la rue du même nom ou des cartes postales publiées dans la ville de Montréal. Tout comme avec la recherche générale au Musée d'histoire de Sherbrooke, les résultats concernent autant les documents textuels que les photographies, une étape additionnelle étant nécessaire pour filtrer le tout en fonction du type d'archives désirées. Nous avons par la suite effectué la recherche dans la section de recherche avancée. Plusieurs options s'offraient alors à nous : la limite des résultats, la date, description, etc. Le formulaire de recherche avancée comprend une zone de recherche par mots-clés, mais elle n'est pas utilisée, les gestionnaires lui ayant préféré une indexation plein texte⁷⁸ des descriptions. Le document analysé, une carte de la ville de Sherbrooke intitulée « Sherbrooke, P. Q. »⁷⁹, répond aux standards de base en archivistique : le titre, le niveau de description, la cote, la description physique du document (l'aspect plus matériel), le nom du producteur, la portée et le contenu du document (l'aspect plus historique du document), la date d'acquisition, etc⁸⁰. Le choix d'utiliser la recherche plein-texte dans tous les descripteurs plutôt qu'une recherche par mots-clés est à noter.

⁷⁷ Archives Montréal, *Résultats* [Banque de données], sur le site *Archives de Montréal*, consulté le 25 juillet 2018, <<https://archivesdemontreal.ica-atom.org/informationobject/browse?topLod=0&query=Sherbrooke>>.

⁷⁸ Le fait de rechercher à travers toutes les données enregistrées. Voir: Michel Lévesque, « L'indexation : Luxe ou nécessité? », *Archives*, vol. 33, n° 1, 2001-2002, p. 17 à 45.

⁷⁹ Archives Montréal, *Pièce P007 - Sherbrooke, P.Q. . - [190-?]* [En ligne], 2018, consulté le 25 juillet 2018, <<https://archivesdemontreal.ica-atom.org/sherbrooke-p-q-190>>.

⁸⁰ *Ibid.*

Bref, les Archives de la Ville de Montréal proposent une collection très intéressante dont le moteur de recherche est centré sur l'indexation des titres et des descriptions plutôt que sur des mots clés standardisés. L'ajout de termes d'indexation permettrait de peaufiner la recherche et spécifier la portée et le contenu de l'image présentée. De plus, tout comme la Société d'histoire de Sherbrooke, cette institution se concentre principalement sur la conservation du patrimoine archivistique de la ville où elle se situe, restant ainsi dans la ligne directrice de sa mission. Étant donné que les documents concernant une ville plus éloignée ne sont généralement pas conservés par une institution à vocation municipale, voire régionale, les désignations toponymiques plus claires et mieux hiérarchisées faciliteraient le repérage de tels documents, notamment dans une perspective de mise en relations de collections.

1.3 Les recommandations sur la diffusion des archives en ligne

Les archives se penchent depuis longtemps sur les bonnes pratiques visant à mettre de l'avant les archives numérisées au public et aux chercheurs. Ce rôle s'inscrit dans leurs diverses fonctions, les archivistes étant :

[...] des spécialistes qui traitent de l'information organique et consignée. [...] Ces fonctions concernent les archives qui se définissent comme : « [...] l'ensemble des documents, quelle que soit leur date ou leur nature [papier, numérique, audiovisuel, photographique, etc.] produits ou reçus par une personne ou un organisme pour ses besoins ou l'exercice de ses activités et conservés pour leur valeur d'information générale » (Loi sur les archives, RLRQ, chapitre A-21.1, article 2)⁸¹.

Que ce soit par l'analyse, la conception, la planification, la gestion, l'organisation, la supervision, le contrôle ou même l'évaluation⁸², l'archiviste doit établir des directives

⁸¹ Association des Archivistes du Québec, *Archivistes* [site Web], 2016, consulté le 20 février 2019, <<https://archivistes.qc.ca/devenir-archiviste/>>.

⁸² *Ibid.*

claires et précises entourant l'acquisition, le traitement, la préservation et la diffusion des archives⁸³. Selon Carol Couture et Marcel Lajeunesse, la fonction primordiale en archivistique est la conservation des archives⁸⁴. Lorsque l'archiviste reçoit des documents, il se doit de respecter l'intégrité et la structure interne de l'ensemble de ces derniers. En d'autres termes, il ne doit pas le démembrer et doit le laisser dans son ordre originel⁸⁵. En plus de respecter l'intégrité du fonds, il doit aussi respecter le principe de territorialité. Les archives doivent rester dans la région d'où elles proviennent⁸⁶. Si nous reprenons nos exemples plus haut, la Société d'histoire de Sherbrooke se concentre sur la mise en valeur du patrimoine historique et ethnologique de la région de Sherbrooke alors que les Archives de la ville de Montréal se concentrent principalement sur le patrimoine de la ville de Montréal. Ces deux institutions sont toutes deux situées dans la ville concernée, bien que des éléments de leurs collections puissent concerner un autre lieu. Quant au Musée McCord, tout en étant le « Musée d'histoire sociale de Montréal », il a accumulé au fil des ans une collection qui dépasse largement le seul cadre montréalais, à une époque où il était désigné sous le nom de « Musée McCord d'histoire canadienne ».

Pour les institutions, « [...] plus que jamais, des outils de gestion doivent être développés et mis en place par les archivistes. Politique de diffusion, programme de numérisation, plan de communication, stratégie de développement, etc., deviennent tout à fait indispensables afin de pouvoir agir de manière cohérente et efficace⁸⁷». Les

⁸³ Association des Archivistes du Québec, *Archivistes ...*. Voir aussi : Carol Couture et Marcel Lajeunesse, *L'archivistique à l'ère du numérique*, p. 120 : Grandes fonctions de l'archivistique : création, acquisition, évaluation, classification, description, diffusion et préservation de l'information.

⁸⁴ Carol Couture et Marcel Lajeunesse, *L'archivistique à l'ère du numérique ...*, p. 60.

⁸⁵ *Ibid.*, p. 64.

⁸⁶ *Ibid.*, p. 117-118.

⁸⁷ Yvon Lemay et Anne Klein, « La diffusion des archives ... », p. 21-22.

archivistes doivent élaborer et développer des outils propres aux archives afin de mieux les repérer, les classer, etc⁸⁸. Pour donner un meilleur accès aux usagers, l'archiviste doit organiser matériellement mais aussi intellectuellement les fonds à l'aide d'index, de guides, des catalogues, d'inventaires, etc⁸⁹. Comme le mentionnent Carol Couture et Marcel Lejeunesse, internet est maintenant LE mode de diffusion par excellence pour informer et diffuser le contenu archivistique au public⁹⁰.

Le contexte numérique vient jouer un rôle important à l'heure actuelle. Pour reprendre les mots des chercheurs Yvon Lemay et Anne Klein : « Non seulement il leur faut désormais aller au-devant de l'utilisateur et sortir le matériel hors de leurs murs en le rendant accessible sur des plateformes de partage de ressources mais, de plus, ils doivent accorder les autorisations pour que ce matériel puisse être réutilisé⁹¹ ». Pour ce faire, l'archiviste doit connaître les nouvelles pratiques notamment avec l'arrivée des outils informatiques, mais aussi, connaître les usagers des centres d'archives. Louise Gagnon-Arguin mentionne que: « [...] l'identification des besoins des usagers fait partie intégrante de la gestion des archives [...]»⁹². « Qui consulte les archives? Que recherchent-ils? Pourquoi? Comment s'y prennent-ils pour obtenir l'information recherchée?»⁹³. Voilà les questions posées par le Bureau Canadien des archivistes au début des années 90. Selon leurs études, les usagers des archives viennent autant du milieu de l'enseignement que de la recherche.

⁸⁸ Association des Archivistes du Québec, *Archivistes ...*

⁸⁹ Carol Couture et Marcel Lajeunesse. *L'archivistique à l'ère du numérique...*, p. 64.

⁹⁰ *Ibid.*, p. 161.

⁹¹ Yvon Lemay et Anne Klein, « La diffusion des archives ... », p. 23.

⁹² Louise Gagnon-Arguin, « Études d'usagers », dans Normand Charbonneau et Mario Robert dirs., *La gestion des archives photographiques*, p.221.

⁹³ Comité de planification sur les normes de description, « L'indexation en archivistique ... », p. 5.

Dès 1992, il était question d'une clientèle de plus en plus hétérogène. En effet, il devenait évident que l'archiviste comme intermédiaire entre l'utilisateur et le fonds devait progressivement laisser la place à un système permettant un repérage efficace et simple pour les usagers des centres⁹⁴. Malgré l'arrivée grandissante des technologies de traitement et de gestion des données, les archivistes ont dû conserver les principes de préservation et les différentes tâches qui y sont associées. La différence majeure est le rôle d'intermédiaire qu'a l'archiviste entre l'utilisateur et le document⁹⁵. Autrefois, le chercheur consultait aussi les archivistes pour les guides dans les instruments de recherche papier et pour les aider à découvrir la perle rare. Aujourd'hui, nous avons la possibilité de faire des recherches directement à partir de notre ordinateur à la maison⁹⁶ et une interface informatique devient l'intermédiaire principal entre le chercheur et les archives. C'est pourquoi, pour reprendre les termes de la déclaration universelle sur les archives mentionnées dans l'article de la chercheuse Laure Amélie Guitard, « [l']accès le plus large aux archives doit être maintenu et encouragé ⁹⁷».

Selon Richard Lytle, dans son étude menée au début du XX^e siècle, les généalogistes seraient en premier rang pour la fréquentation dans les centres d'archives, suivent des administrateurs puis des étudiants et des biographes⁹⁸. Pour Normand Charbonneau, les utilisateurs d'archives photographiques du XXI^e siècle seraient : « [...] des historiens de l'art et de l'architecture, des directeurs artistiques, costumiers et

⁹⁴ *Ibid.*, p. 26-27.

⁹⁵ Michèle Hudon, « Indexation et langages documentaires dans les milieux archivistiques à l'ère des nouvelles technologies de l'information », *Archives*, vol. 29, n°1, 1997-1998, p. 77.

⁹⁶ Laure Amélie Guitard, « Indexation par sujet en archivistique ... », p. 210-211.

⁹⁷ *Ibid.*, p. 211. Voir aussi : Conseil international des archives, *Déclaration universelle sur les archives*, Paris, 36^e sessions de la conférence générale de l'UNESCO, 10 novembre 2011, 1 p.

⁹⁸ Richard Lytle, « Intellectual Access to Archives: I. Provenance and Content Indexing Methods of Subject Indexing », *American Archivist* 43, 1 (Winter, 1980): 66. Dans *Ibid.*, p. 5-6.

décorateurs de théâtre, de cinéma et de la télévision, des architectes, des artistes (sculpteurs ou peintres), des politologues, des journalistes de la presse écrite ou parlée et des généalogistes⁹⁹». Les outils de repérages doivent être adaptés aux usagers des centres. En effet, ils se doivent d'être conviviaux et simples afin de permettre une utilisation efficace et ordonnée¹⁰⁰. Les chercheurs sont toutefois une minorité des usagers et les outils développés ne répondent pas toujours à leurs besoins parfois très large et parfois très ciblés.

Mettre les bases technologiques et connaître les usagers des centres ne sont pas les seules choses à savoir pour avoir une bonne pratique de diffusion archivistique : « L'ère du numérique n'est-elle pas d'abord et surtout une ère de l'archivage, autrement dit un contexte qui appelle à une redéfinition, une reconfiguration du domaine des archives ?¹⁰¹ ». Cette reconfiguration s'applique évidemment aux lieux où les archives sont placées, mais il y a aussi une reconfiguration de la pratique¹⁰². L'utilisation optimale des outils informatiques pour promouvoir l'accessibilité des archives permettra ainsi de s'ouvrir à une plus grande clientèle et de faciliter l'accès pour des types variés de recherche.

1.4 Conclusion

En guise de conclusion de ce premier chapitre, nous pouvons dire que les systèmes analysés sont conçus en fonction des contenus disponibles, des ressources institutionnelles

⁹⁹ Normand Charbonneau, « La gestion des documents photographiques. Évaluation, sélection et préservation », *Archives*, vol. 25, n°4, 1994, p. 5. Tiré du texte de Louise Gagnon-Arguin, « Études d'usagers », dans Normand Charbonneau et Mario Robert dirs., *La gestion des archives photographiques*, p. 224.

¹⁰⁰ James M. Turner, *Images en mouvement*, p. IX.

¹⁰¹ Yvon Lemay et Anne Klein, « La diffusion des archives ... », p. 26.

¹⁰² *Ibid.*, p. 26. Voir aussi : Carol Couture et Marcel Lajeunesse. *L'archivistique à l'ère du numérique ...*, p. 175.

et de la procédure adaptée en fonction des besoins de l'institution. À l'ère du numérique, les besoins des archivistes et des historiens ont évolué en fonction de la demande toujours grandissante du besoin d'accessibilité à l'information. En effet, que ce soit, pour les archivistes, la nécessité d'indexer et de mettre en ligne les documents ou, pour les historiens, le besoin d'obtenir rapidement et efficacement de l'information. Malgré tout, les bases de données que nous avons vues ont toutes un problème commun dans leur conception. Elles souffrent du fait qu'elles ne sont pas sémantiques. Lorsque nous les utilisons, nous recherchons des chaînes de caractères et non des expressions avec un sens quelconque. Les systèmes proposent des termes en fonction des besoins de l'institution, que ce soit les Archives de la ville de Montréal où la majorité des termes se raccrochent à la ville de Montréal ou la Société d'histoire de Sherbrooke où les thèmes sont plus orientés vers les Cantons-de-l'Est et plus particulièrement la ville de Sherbrooke. Les institutions de grande taille ont tendance à privilégier les descriptions plus générales, probablement à cause du grand volume de leurs collections, alors qu'une institution de petite taille plonge rapidement dans les détails. Par conséquent, la hiérarchisation de la terminologie, laquelle permettrait aux usagers de se repérer de manière plus constante dans différents systèmes, n'est pas appliquée.

Des améliorations sont donc possibles pour rendre le contenu documentaire plus accessible. Qu'elles soient liées à l'évolution des besoins des usagers ou bien à la montée toujours grandissante des technologies informatiques, il devient évident que les institutions doivent s'approprier les nouveaux outils afin de favoriser le partage et établir un lien avec une plus grande clientèle. Malgré tout, il ne faut pas oublier qu'il est tout de même important d'établir un lien direct avec les usagers. En effet, le numérique permet

une recherche plus rapide et efficace pour les chercheurs, mais pour d'autres, le contact reste un moment précieux entre le passé et les générations d'aujourd'hui. Nous pensons aux groupes scolaires en visite au centre découvrant un daguerréotype ou bien une vidéo en pellicule¹⁰³.

Bref, en tant que chercheurs en histoire, nous voyons, à travers nos études, qu'il devient de plus en plus criant d'améliorer l'accessibilité des sources mais aussi, les méthodes d'accès. Les bases de données, toutes institutions confondues, sont soit trop spécifiques, ce qui demande un vocabulaire très avancé, soit elles ne le sont pas assez, ce qui peut empêcher de trouver efficacement une source qui pourrait être pertinente. La problématique de l'indexation date depuis plusieurs années et les archivistes en sont conscients. Il y a 20 ans, les technologies ne permettaient pas de résoudre les problèmes liés à l'indexation et à la numérisation. Aujourd'hui, même avec les avancées informatiques, il est difficile d'améliorer l'indexation car il y a un manque considérable dans la hiérarchie des mots ainsi que dans la sémantique¹⁰⁴.

¹⁰³ *Ibid.*, p. 31.

¹⁰⁴ Le chercheur Abdelkader Hamadi parle pour sa part de fossé sémantique, soit : « [...] le manque de concordance entre les informations que la machine peut extraire d'un document numérique et des interprétations humaines ». Abdelkader Hamadi, « Utilisation du contexte pour l'indexation sémantique des images et vidéos », Thèse de doctorat (Informatique), Grenoble, Université Joseph Fourier Grenoble -1-, 2014, p. 3.

Chapitre 2 : Les fondements de l'indexation

Après avoir examiné les bases de données des institutions ainsi que les bonnes pratiques archivistiques, nous pouvons analyser comment l'indexation des images sert à la décrire dans une base de données. Les archivistes doivent élaborer et développer des outils propres aux archives afin de mieux les repérer, les classer, etc. Pour donner un meilleur accès aux usagers, il faut organiser matériellement mais aussi intellectuellement les fonds à l'aide d'index, de guides, des catalogues, d'inventaires, etc.

Déjà, au début des années 90, on mentionne l'importance d'une bonne indexation afin de mieux repérer les documents¹⁰⁵. Les études et divers outils ont pour objectifs de : « [...] donner aux archivistes des connaissances les rendant aptes à appliquer les principes de l'analyse documentaire aux descriptions de documents d'archives, à différents niveaux et pour diverses catégories de documents, et à développer chez eux des savoirs techniques les rendant aptes à produire des index normalisés¹⁰⁶». En effet, James M. Turner mentionne que si nous n'agissons pas maintenant sur les problèmes de stockage et de repérages, ces derniers vont prendre de plus en plus de place dans les enjeux actuels¹⁰⁷. Avec les changements dans les procédés photographiques, il devient important de mettre en place des nouvelles méthodes afin de mieux contrôler et préserver les sources¹⁰⁸. Les changements ne sont pas que dans la technique photographique, ils sont aussi dans la pratique de l'indexation¹⁰⁹. En effet, comme le mentionne James M. Turner : « Ce sont

¹⁰⁵ Comité de planification sur les normes de description, « L'indexation en archivistique ... » ; James M. Turner, *Images en mouvement*

¹⁰⁶ Michèle Hudon, « Indexation et langages documentaires ... », p. 76.

¹⁰⁷ James M. Turner, *Images en mouvement*, p. x.

¹⁰⁸ *Ibid.*, p. 1.

¹⁰⁹ *Ibid.*, p. 76.

les progrès technologiques qui ont toujours été à l'origine des changements théoriques et pratiques dans notre domaine [la théorie et la pratique de l'indexation]¹¹⁰». Dans l'étude de Michèle Hudon, on définit l'indexation comme : « [...] une activité de nature cognitive qui consiste à décrire le contenu d'un document à l'aide d'un langage d'indexation pour faciliter la mémorisation dans un fichier, en vue d'une recherche ultérieure, [...]»¹¹¹. Pour Normand Charbonneau et Mario Robert il s'agit de : « [...] la représentation et la mise en vedette des concepts ou catégories d'accès tirés de l'analyse des éléments de la notice de description, ou tirés des documents eux-mêmes, afin de faciliter l'accès grâce aux éléments d'un langage documentaire¹¹²». Laure Amélie Guitard va plus loin en mentionnant que l'indexation est un processus linguistique qui s'effectue en deux étapes : la sélection de concepts avec l'analyse du document¹¹³ (constituant l'index) et la traduction des concepts en langage d'indexation¹¹⁴. Les concepts correspondent aux besoins des usagers car ils représentent la valeur première dans la recherche. Cette première lecture permet de prendre connaissance des éléments qui composent le document¹¹⁵. Une indexation complète du contenu d'une image peut donc nécessiter un grand nombre de concepts afin de répondre aux besoins de différents types d'historiens.

¹¹⁰ *Ibid.*, p. 76.

¹¹¹ Michèle Hudon, « Indexation et langages documentaires ... », p. 77. Définition adaptée de Jacques Chaumier, *Travail et méthodes du documentaliste, connaissance du problème, applications pratiques*, Paris, ESF, 1996. 5e éd., p.101. Voir aussi : Laure Amélie Guitard, « Indexation par sujet en archivistique ... », p. 202.

¹¹² Normand Charbonneau et Mario Robert, « Indexation », dans Normand Charbonneau et Mario Robert dir., *La gestion des archives photographiques*, p. 155.

¹¹³ « L'indexation a pour produit l'index. Un index est un relevé systématique des concepts pertinents pour la recherche en fonction des usagers du milieu où se pratique l'indexation ». Laure Amélie Guitard, « Indexation par sujet en archivistique ... », p. 202-203. Voir aussi : Normand Charbonneau et Mario Robert, « Indexation », p. 154. Voir le tableau concernant l'analyse de contenu en archivistique.

¹¹⁴ Laure Amélie Guitard, « Indexation par sujet en archivistique ... », p. 203 et 207. Voir aussi : Michèle Hudon, « Indexation et langages documentaires ... », p. 77.

¹¹⁵ Laure Amélie Guitard, « Indexation par sujet en archivistique ... », p. 204. Voir aussi : Comité de planification sur les normes de description, « L'indexation en archivistique ... », p. 92-93. (Norme ISO 5963-1985).

Le Bureau canadien des archivistes insiste sur les risques qu'un concept puisse être représenté par plusieurs synonymes¹¹⁶. Pour contrer ce problème, l'indexeur doit établir une liste de termes que l'on nomme aussi des thésaurus¹¹⁷ qui va permettre une meilleure définition du concept. L'indexation doit être exhaustive, spécifique, exacte et uniforme en étant constante dans les mots employés pour indexer les photographies (vocabulaire contrôlé)¹¹⁸. De plus, on identifie les termes d'indexation en fonction entre autres des usagers et de l'institution¹¹⁹. Plusieurs systèmes d'indexations informatisés fonctionnent selon l'emploi de termes justes souvent choisis par l'indexeur. Si l'utilisateur n'utilise pas les termes exacts, les résultats de sa recherche seront limités voire même nuls¹²⁰. Le classement des images doit aussi tenir compte de l'emplacement géographique, de l'époque (qu'ils soient représentés de manière figurée dans une œuvre artistique et dans la réalité avec une photographie), de l'angle de la prise de vue, etc¹²¹.

2.1 Les taxonomies internationales pour l'indexation des images

Une indexation cohérente repose sur la taxonomie, soit « La science des lois de la classification » visant à établir une classification formée de « suite d'éléments formant des listes qui concernent un domaine, une science¹²² ». Roxane Borgès Da Silva mentionne qu'elle est bâtie selon une base de données formelle, hiérarchisée et utilise un système

¹¹⁶ Comité de planification sur les normes de description, « L'indexation en archivistique, ... », p. 74 et p. 106 à 110.

¹¹⁷ Le thésaurus : « [...] une liste de mots simples ou d'expressions ayant valeur de termes dans un domaine limité de la connaissance, le terme étant défini ici comme la représentation linguistique (symbolique) d'un concept unique ». Michèle Hudon, *Le Thésaurus : Conception, élaboration, gestion*, Montréal, Les Éditions ASTED, 1994, p. 17 citée par Normand Charbonneau et Mario Robert, « Indexation », p. 166. En anglais, un *thesaurus* est aussi un dictionnaire des synonymes.

¹¹⁸ Normand Charbonneau et Mario Robert, « Indexation », p. 157 à 160.

¹¹⁹ Laure Amélie Guitard, « Indexation par sujet en archivistique ... », p. 203.

¹²⁰ James M. Turner, *Images en mouvement*, p. 37 et 39.

¹²¹ *Ibid.*, p. 6.

¹²² Larousse, « taxonomie », *dictionnaire Larousse* [En ligne], 2019, sur le site www.larousse.fr, consulté le 20 janvier 2019, < <https://www.larousse.fr/dictionnaires/francais/taxinomie/76893> >.

d'analyse quantitative¹²³. La taxonomie peut alors traiter un plus grand nombre de données et créer des groupes homogènes¹²⁴. Selon Olivier Le Deuff, la première utilisation de la taxonomie aurait eu lieu en 1813 dans l'ouvrage d'Augustin Pyrame de Candolle, un botaniste suisse¹²⁵. En effet, la taxonomie portait, au départ, plus spécifiquement sur la classification des êtres vivants, à la classe plus biologique. De nos jours, cette forme classificatoire est aussi utilisée dans un cadre plus formel numérique en organisant et en classifiant des données relatives d'entreprises ou d'organisations¹²⁶. Appliquer une taxonomie peut être avantageux pour des images comme des œuvres d'art qui ont toutes des métadonnées spécifiques. Dans le cas de la photographie, chacune a sa spécificité propre ce qui rend difficile l'application d'un vocabulaire contrôlé selon les pratiques en histoire de l'art¹²⁷. Elles requièrent donc un langage universel comme nous avons vu précédemment¹²⁸. Toutefois, l'élaboration d'un outil de travail devient indispensable afin de rester homogène dans la démarche d'indexation¹²⁹. Afin de répondre à ces contraintes paradoxales, une taxonomie hiérarchique peut répondre aux besoins des chercheurs en offrant à la fois des classifications générales et spécifiques. Parmi les systèmes de classification existants, deux ont attiré notre attention pour ce mémoire : Iconclass et le Art and Architecture Thesaurus (AAT) du Getty Research Institute. Créé aux États-Unis, ce dernier est recommandé par le Réseau canadien d'information sur le patrimoine alors

¹²³ Roxane Borgès Da Silva, « Taxonomie et typologie : est-ce vraiment des synonymes ? », *Santé Publique*, vol. 25, n°5, 2013, p. 635. Voir aussi : Roxane Borgès Da Silva, « Taxonomie et typologie ... », p. 87 et Olivier Le Deuff, *Du tag au like*, p. 30-31.

¹²⁴ Roxane Borgès Da Silva, « Taxonomie et typologie ... », p. 633 et 636.

¹²⁵ Olivier Le Deuff, *Du tag au like*, p. 30.

¹²⁶ *Ibid.*, p. 31.

¹²⁷ James M. Turner, *Images en mouvement*, p. 42.

¹²⁸ Normand Charbonneau et Mario Robert, « Indexation », p. 153.

¹²⁹ *Ibid.*, p. 165.

que le premier, issu des Pays-Bas, nous semble plus accessible pour les petites institutions et découle d'une approche européenne plus multilingue à l'origine.

Le système Iconclass a été développé au début des années 1950 par le professeur d'histoire de l'art Henri Van de Waal de l'Université de Leyde aux Pays-Bas. Le but était d'avoir un aperçu rapide des sujets et des thèmes d'une image. Le système qui portera le nom d'Iconclass verra finalement le jour après 1972 et sera publié par l'Académie royale néerlandaise des Arts et des Sciences entre 1973 et 1985. L'Université d'Utrecht publiera par la suite entre 1990 et 2001 plusieurs éditions informatisées du système. Le système Iconclass est utilisé par diverses institutions partout dans le monde afin de classer leurs collections. Il permet de chercher et d'identifier des images, mais aussi d'attribuer des éléments de classement visant un sens très large et des taxons plus spécifiques. De plus, le navigateur multilingue *Iconclass Libertas* lancé par l'organisme KNAW (depuis oct. 2001) permet une plus grande accessibilité depuis sa diffusion en ligne à partir de novembre 2004 en quatre langues (anglais, allemand, français et italien). Relevant du Rijksbureau voor Kunsthistorische Documentatie depuis septembre 2006, cette taxonomie est disponible dans le nouveau *Iconclass Browser* depuis le 10 novembre 2009¹³⁰.

Le système d'indexation d'Iconclass a plusieurs caractéristiques à prendre en considération. En effet, la base de données est un système de classification spécifique élaboré pour classer et indexer les œuvres d'art et les images. Accessible via leur site web, elle existe aussi sous format de données ouvertes et liées (*Linked Open Data* ou LOD). De nombreux spécialistes utilisent cet outil : « *Art historians, researchers and*

¹³⁰ Iconclass, *Home Iconclass* [site Web], 2012, consulté le 18 juillet 2018, <<http://www.iconclass.nl/home>>.

*curators use it to describe, classify and examine the subject of images represented in various media such as paintings, drawings and photographs*¹³¹». Le programme sert à marquer une image à l'aide de **concepts**. La mise en place de concepts correspond à la première étape lors d'une indexation comme nous l'avons vu précédemment. Par les concepts, il nous est possible par la suite d'ajouter le langage d'indexation à l'aide d'une syntaxe et d'un lexique. Depuis ses débuts, le système ne cesse de s'accroître. En effet, nous y retrouvons aujourd'hui 28 000 définitions, 10 divisions principales, 14 000 mots-clés ainsi que 40 000 références portant sur l'iconographie. «*The definitions are used to index, catalogue and describe the subjects of images represented in works of art, reproductions, photographs and other sources*¹³²».

Si on applique la taxonomie Iconclass à l'image tirée du Fonds Frederick James Sangster, les éléments en gras correspondent à ce que nous avons relevé dans la classification Iconclass. Nous avons complété cette liste avec des concepts plus spécifiques liés au contexte local tirés des informations supplémentaires à travers les différents tomes sur l'histoire de Sherbrooke écrits par l'historien Jean-Pierre Kesteman¹³³ ainsi que dans le *Guide historique du vieux Sherbrooke* écrit par le même auteur¹³⁴.

¹³¹ *Ibid.*

¹³² *Ibid.*

¹³³ Jean-Pierre Kesteman, 2000-2002 - *Histoire de Sherbrooke*, Sherbrooke, GGC Éditions, 4 vol. (coll. Patrimoine) t. 1 : *De l'âge de l'eau à l'ère de la vapeur, 1802-1866*, 2000, 353 p. / t. 2 : *De l'âge de la vapeur à l'ère de l'électricité, 1867-1896*, 2001, 280 p. / t. 3 : *La ville de l'électricité et du tramway, 1897-1929*, 2002, 292 p. / t. 4 : *De la ville ouvrière à la métropole universitaire, 1930-2002*, 2002, 486 p.

¹³⁴ Jean-Pierre Kesteman, *Guide historique du vieux Sherbrooke*, Sherbrooke, Société d'histoire de Sherbrooke, 2001, 2^e éd. (1985), 269 p.

Figure 2.1 : Rappel de l'image, fiche #13143



Source : SHS, *Résultats de la recherche, fiche 13143*.

Dans l'aspect histoire : 61ADFHIK

- **Histoire, scène, événement, événement historique, lieu, région, ville**, Sherbrooke, Estrie, Québec, Canada, Cantons-de-l'Est, **rue**, Wellington, Wellington Nord, Frontenac, Factory Street, rue de la Facterie, King's Highway, Chemin du Roi, **bâtiments, bâtiments historiques**, Édifice Odell, Édifice Whiting, Quebec Central Railway Building, Édifice Bell Telephone, **Compagnies, organisations**, quincaillerie Luke & Mitchell, J. S. Mitchell, H. C. Wilson, Magasin général, Hôtel de ville, Palace of illusion, magasin de chaussures Wiggett, Banque de Montréal, Quebec central, Bell Telephone, Faculté de Droit UdeS, Université de Sherbrooke, Musée des beaux-arts, Petit Théâtre de Sherbrooke, Red School House, Chapelle méthodiste, station de pompiers, Bellevue Hotel, carré Strathcona, tramway, Sherbrooke Street Railway, Sherbrooke Railway & Power, rails, ligne, ligne Belt, urbaine, **situation, événement historique**, arbre, arbre couché, jour, abattage, entretien.

Société : 46B1132

- **Acheter, affaires, bâtiment**, Édifice Odell, Édifice Whiting, Quebec Central Railway Building, Édifice Bell Telephone, **civilisation**, Sherbrookoise, québécoise, **commerce**, quincaillerie Luke & Mitchell, J. S. Mitchell, H. C. Wilson, Magasin général, magasin de chaussures Wiggett, **culture**, centre-ville, pôle commercial, **finance**, Banque de Montréal, **marché**, carré Strathcona, **occupations humaines**, marche, **société**, sherbrookoise, **économie, édifice, vendre, vitrine, magasin, montrer**,

Humain : 31D 32B 36A1 36B

- **Biologie, vie, âge, être humain, race humaine, nationalité**, canadienne, québécoise, **environnement**, ville, centre-ville, **conservation, nettoyage environnemental, nettoyer, protection, destruction**, abattage

Nature : 21B 22E 23T

- **Nature, terre, élément, phénomène, phénomène naturel, température, histoire, nature, temps, époque**, 1925, XXième siècle,

Cet exercice nous amène à constater que cette approche a permis d'identifier un plus grand nombre de concepts que lors d'une description guidée par un manuel défini pour un usage local. En complétant l'analyse générale avec des ouvrages sur l'histoire locale, nous avons pu par la suite préciser chaque concept avec des termes d'indexation plus spécifiques au contexte sherbrookoïse, la démarche hiérarchique aidant aussi à nous guider dans le processus.

Il est certain que le système d'indexation Iconclass a un aspect très globalisant dans les termes utilisés. En effet, les éléments plus spécifiques doivent être ajoutés par la personne qui indexe la source iconographique. Par exemple, nous pouvons avoir le terme général de ville ou région. De ce fait, nous allons ajouter Sherbrooke, Estrie, Québec, Cantons-de-l'Est, etc. Compte tenu de la nature de la collection, ces termes pourraient être présents par défaut, l'indexeur n'ayant à les changer que lorsque c'est nécessaire. En présentant les concepts plus généraux, Iconclass devient un aide-mémoire pour guider l'inscription de termes plus précis pour l'indexation.

Le caractère uniforme et hiérarchique du système permet de contrôler plus facilement les termes choisis lors de la première indexation. Pour un chercheur, cet outil, qui est utilisé dans plusieurs institutions à travers le monde, permet de présenter l'information à partir d'une terminologie constante quelle que soit l'institution ou l'indexeur. Une meilleure indexation mène à une visibilité plus uniforme et facilitera la mise en relation des informations à travers les données ouvertes et liées¹³⁵. Bref, avec un système standardisé, il devient possible d'établir de nouveaux types d'analyses et d'élargir les

¹³⁵ Philippe Michon, « Vers une nouvelle architecture de l'information historique : L'impact du Web sémantique sur l'organisation du Répertoire du patrimoine culturel du Québec », Mémoire de maîtrise (histoire), Sherbrooke, Université de Sherbrooke, 2016, 134 p.

champs de recherche. Avec des collections cohérentes comme celles utilisées par les chercheurs Ginosar, Rakelly, Sachs, Yin et A. Efros¹³⁶, il devient facile et rapide d'effectuer une pré-indexation des éléments constituant les photographies. Dans leur expérience, le corpus est constitué de portraits de finissants et ils sont analysés grâce à un système de codage. Avec cette analyse, il est possible de voir comment les étudiants se placent pour la photographie au fil du temps et dans quelle période un style a été plus populaire versus un autre. Un autre exemple intéressant est la plateforme web *Civil War Photo Sleuth*¹³⁷. Cette dernière a pour objectif d'identifier les portraits des hommes qui ont participé à la Guerre de Sécession. Que ce soit directement sur le visage ou les éléments qui composent l'uniforme, le programme fait une corrélation entre la photo et les autres images de la base de données déjà indexées.

Le deuxième référent d'indexation choisi dans le cadre du projet est celui du Getty Research Institute (Getty) qui définit ainsi sa mission:

The Getty Research Institute is dedicated to furthering knowledge and advancing understanding of the visual arts and their various histories through its expertise, active collecting program, public programs, institutional collaborations, exhibitions, publications, digital services, and residential scholars programs. Its Research Library and Special Collections of rare materials and digital resources serve an international community of scholars and the interested public. The Institute's activities and scholarly resources guide and sustain each other and together provide a unique environment for research, critical inquiry, and scholarly exchange¹³⁸.

¹³⁶ Shiry Ginosar *et al.*, « A Century of Portraits: A Visual Historical Record of American High School Yearbooks », *International Conference on Computer Vision*, 7 p.

¹³⁷ Erica X Eisen, « Who's Behind That Beard? Historians are using facial recognition software to identify people in Civil War photographs », *Slate* [En ligne], 15 novembre 2018, sur le site *slate.com*, consulté le 4 novembre 2020, < <https://slate.com/technology/2018/11/civil-war-photo-sleuth-facial-recognition.html> >.

¹³⁸ The Getty Research Institute, *About the Research Institute* [site Web], 2017, consulté le 1 octobre 2018, < <http://www.getty.edu/research/institute/> >.

Mettant de l'avant un environnement collaboratif pour la recherche et la publication, Getty se concentre, tout comme Iconclass, dans le domaine de l'histoire de l'art. Visant à améliorer constamment les connaissances et la compréhension de l'art visuel, le système est conçu afin de donner une terminologie structurée pour des sujets comme l'architecture, les documents d'archives, l'art, la conservation, les arts décoratifs, les substituts visuels et les documents bibliographiques¹³⁹. Le système aide les chercheurs grâce au respect de normes internationales tant dans sa structure originale que dans le développement du vocabulaire. Leur public cible est composé de chercheurs en histoire de l'art, de spécialistes œuvrant dans les musées, les bibliothèques, des conservateurs, mais aussi des étudiants et des gens du grand public¹⁴⁰.

Les catégories proposées peuvent être utilisées lors du catalogage comme norme de valeur¹⁴¹. Les vocabulaires peuvent aussi être utilisés dans le cadre d'une récupération de bases de données et même en environnement lié¹⁴². De plus, le système de classification est accessible via leur site web, mais aussi en format *Linked Open Data* (LOD) tout comme Iconclass. Leur objectif est de rendre librement accessibles les connaissances¹⁴³. Pour chacune des catégories proposées, on mentionne comment l'utiliser et quel est son objectif. Chaque étape de la recherche y est mentionnée, les termes à privilégier, ceux à éviter, etc. En note de bas de page se trouve le lien de la catégorie utilisée¹⁴⁴. Sur le site,

¹³⁹ The Getty Research Institute, *Search Tools and Databases* [site Web], 2017, consulté le 1 octobre 2018, <<http://www.getty.edu/research/tools/>>.

¹⁴⁰ *Ibid.*

¹⁴¹ The Getty Research Institute, *About the AAT* [site Web], 2017, consulté le 1 octobre 2018, <<http://www.getty.edu/research/tools/vocabularies/aat/about.html>>.

¹⁴² *Ibid.*

¹⁴³ *Ibid.*

¹⁴⁴ The Getty Research Institute, *How to Use the AAT Online* [site Web], 2017, consulté le 1 octobre 2018, <<http://www.getty.edu/research/tools/vocabularies/aat/help.html>>.

plusieurs catégories sont offertes, l'AAT étant une des listes d'autorités gérées par l'institut Getty.

Tableau 2.2 : Liste d'autorités de l'institut Getty

The Getty Thesaurus of Geographic Names (TGN)¹⁴⁵. Géographique.	Elle peut être intégrée à un système d'information géographique (SIG).
The Cultural Objects Name Authority (CONA)¹⁴⁶. Noms des objets culturels.	Elle a un lien avec les autres sources de vocabulaires Getty. Elle touche entre-autres les œuvres d'art et l'architecture.
The Union List of Artist Names (ULAN)¹⁴⁷.	On y retrouve une liste d'autorité des noms d'artistes.
The Getty Iconography Authority (IA)¹⁴⁸. Mythique.	Dans cette catégorie on retrouve des noms propres ainsi que des termes liés à la religion, aux récits mythiques, etc.
The Art & Architecture Thesaurus (AAT)¹⁴⁹.	Enfin, cette catégorie est axée principalement sur l'art, l'architecture, la conservation, l'archéologie, etc. ¹⁵⁰ .

L'AAT est initialement fondée par des experts en architecture et directeurs de bibliothèques soit, Toni Petersen, Dora Crouch et Pat Molholt¹⁵¹. Présentée en 1990 puis imprimée en 1994¹⁵², l'AAT passe sous forme numérique en 1997 à cause d'une croissance constante des concepts¹⁵³. En effet, à elle seule, l'édition de 1994 contient au-

¹⁴⁵ The Getty Research Institute, *Getty Thesaurus of Geographic Names® Online* [site Web], 2017, consulté le 1 octobre 2018, <<http://www.getty.edu/research/tools/vocabularies/tgn/index.html>>.

¹⁴⁶The Getty Research Institute, *Cultural Objects Name Authority® Online* [site Web], 2017, consulté le 1 octobre 2018, <<http://www.getty.edu/research/tools/vocabularies/cona/index.html>>.

¹⁴⁷ The Getty Research Institute, *Union List of Artist Names® Online* [site Web], 2017, consulté le 1 octobre 2018, <<http://www.getty.edu/research/tools/vocabularies/ulan/index.html>>.

¹⁴⁸ The Getty Research Institute, *Cultural Objects Name Authority® Online* [site Web].

¹⁴⁹ The Getty Research Institute, *Art & Architecture Thesaurus® Online* [site Web], 2017, consulté le 1 octobre 2018, <<http://www.getty.edu/research/tools/vocabularies/aat/index.html>>.

¹⁵⁰ The Getty Research Institute, *About the AAT* [site Web], 2017, consulté le 1 octobre 2018, <<http://www.getty.edu/research/tools/vocabularies/aat/about.html>>.

¹⁵¹ The Getty Research Institute, *Art & Architecture Thesaurus*

¹⁵² Toni Petersen, dir., *Art & Architecture Thesaurus*, New York, Oxford University Press, 1994, 2e éd.(1990), 3182 p.

¹⁵³ The Getty Research Institute, *About the AAT*

delà de 90 000 mots et expressions¹⁵⁴. Les travaux ont commencé en réponse au besoin croissant d'automatisation des procédures de catalogage et d'indexation dans les années 70¹⁵⁵. Afin d'avoir une cohérence dans le vocabulaire utilisé ainsi qu'une saisie de données efficace, les catalogueurs ont rapidement demandé de l'aide dans leur démarche. Pour ce faire, il a été initialement question des termes déjà employés en histoire de l'art et en architecture. Afin de cerner les bons termes, une équipe comportant des bibliothécaires, des historiens de l'art, des historiens de l'architecture, des archivistes, des architectes, des conservateurs, et des spécialistes des thésaurus ont donné leurs avis afin de créer une ressource compatible pour tous¹⁵⁶. Au fil du temps et des collaborations, le vocabulaire continue à grandir et à s'actualiser. Afin de répondre aux besoins des utilisateurs, les thésaurus de l'AAT sont construits autour de normes internationales et nationales.

« *The AAT is a thesaurus in compliance with ISO and NISO standards*¹⁵⁷ ». Chaque élément composant le vocabulaire de l'AAT est un concept avec un identifiant unique tout comme Iconclass. Un des objectifs de l'institution est de rendre l'AAT multiligne, car sa structure, semblable à Iconclass, peut le permettre¹⁵⁸. En juillet 2017, l'AAT comptait 375 000 termes et 60 000 enregistrements¹⁵⁹.

De plus, le Réseau canadien d'information sur le patrimoine indique que l'AAT est largement utilisé dans les musées canadiens et qu'il a été élaboré avec l'aide de

¹⁵⁴ Toni Petersen, dir., *Art and Architecture Thesaurus*.

¹⁵⁵ The Getty Research Institute, *About the AAT* [site Web],

¹⁵⁶ *Ibid.*

¹⁵⁷ The Getty Research Institute, *About the AAT ...*

¹⁵⁸ *Ibid.*

¹⁵⁹ The Getty Research Institute, *AAT: Frequently Asked Questions* [site Web], 2017, consulté le 1 octobre 2018, <http://www.getty.edu/research/tools/vocabularies/aat/aat_faq.html>.

nombreuses terminologies qui étaient déjà existantes ainsi qu'à partir du vocabulaire d'un ouvrage nommé *Revised Nomenclature*. Nous pouvons y lire que :

[l]e Réseau canadien d'information sur le patrimoine (RCIP) a fourni environ 2 600 termes en français qui y sont présentés comme équivalents français des termes les plus courants. Cette version bilingue de l'AAT est utilisée pour faire des recherches dans la Base de données des sciences humaines d'Artefacts Canada. Le RCIP recommande aux musées qui ont de vastes collections en sciences humaines d'utiliser le AAT. La terminologie qu'on y trouve peut être utilisée dans de nombreuses zones de la base de données des sciences humaines¹⁶⁰.

À des fins de comparaison, nous avons aussi appliqué la terminologie de l'AAT à l'image choisie pour notre test (pour référence à l'image voir figure 4)¹⁶¹, en indiquant à nouveau en gras les termes de cette taxonomie. Nous avons ajouté, par la suite, tout comme pour le test effectué avec Iconclass, de l'information plus spécifique à la photographie analysée tirées d'ouvrages consacrés à l'histoire locale¹⁶². De plus, afin d'effectuer une comparaison avec Iconclass, nous avons repris les termes proposés par ce dernier afin d'effectuer notre recherche de termes sur Getty. À titre de rappel, voici ce qui était inscrit dans la **catégorie Histoire 61ADFHIK** de Iconclass: **Histoire, scène, événement, événement historique, lieu, région, ville, rue, bâtiments, bâtiments historiques, Compagnies, organisations, situation, événement historique**. Pour les besoins, nous avons pris les termes en gras et les avons par la suite indiqués dans la barre de recherche de l'AAT (Voir annexe E).

¹⁶⁰ Gouvernement du Canada, *Normes principales pour les musées canadiens* [site Web], 2020, consulté le 1 octobre 2018, <<https://www.canada.ca/fr/reseau-information-patrimoine/services/normes-documentation-collections/guide-rcip-normes-musees/normes-principales-musees-canadiens.html>>.

¹⁶¹ IP465 PN_31D_25_13143

¹⁶² Jean-Pierre Kesteman, 2000-2002 - *Histoire de Sherbrooke, Sherbrooke*, GGC Éditions, 4 vol. (coll. Patrimoine) ; Jean-Pierre Kesteman, *Guide historique du vieux Sherbrooke*, 269 p.

Tout comme Iconclass, l'AAT est un système apportant des termes généraux. Les éléments spécifiques doivent être ajoutés par la suite à la main par l'indexeur. De plus, les termes génériques sont des aide-mémoires qui guident l'indexeur vers des termes plus précis qui correspondent au document iconographique indexé. Finalement, chaque terme est identifié par un identifiant unique (ID)¹⁶³.

Pour reprendre nos observations faites pour la base de données Iconclass : « [...] elle permet de contrôler plus facilement les termes choisis lors de la première indexation ». Finalement, il est plus difficile de se repérer dans le système de classement Getty que celui d'Iconclass. En effet, le système est beaucoup plus complexe, ce qui oblige un temps d'adaptation afin de savoir comment l'utiliser correctement. Malgré tout, le système propose une panoplie de termes en adéquation avec ce que nous recherchons. Pour chacun d'eux, il y a aussi la source ou la référence, ce qui permet de pousser plus loin la réflexion.

Les centres d'archives ont adopté différents systèmes de classification pour des raisons liées à l'expertise interne et aux ressources disponibles. Il serait à peu près impossible pour une institution de changer de système de classement en utilisant les maigres ressources humaines dont elles disposent. La reconnaissance des images permettrait toutefois d'ajouter un classement plus conforme aux normes internationales en termes de grandes catégories et faciliterait ainsi l'échange d'informations et permettrait aux chercheurs d'utiliser leur expérience de recherche d'un système pour l'appliquer dans un nouveau contexte. Le volet sémantique désormais intégré aux deux systèmes va de plus clarifier le sens des concepts et faciliter l'interopérabilité des systèmes, décuplant le potentiel des historiens pour des recherches sur des thématiques très larges.

¹⁶³ Voici quelques exemples de termes et de leur identifiant : Street scenes 300386103 / Market Place 300000250 / Streetcar Systems 300155943 / Streets 300008247

2.2 L'intégration de la folksonomie

À la suite de l'indexation taxonomique faite grâce aux systèmes de classement Iconclass et AAT, nous avons vérifié l'apport de l'indexation folksonomique sur les réseaux sociaux. Pour les besoins du test, nous avons utilisé une image provenant du centre d'archives de la Société d'histoire de Sherbrooke. Elle correspond à la même image utilisée lors de l'indexation taxonomique précédente afin de constater les similitudes et les différences entre les deux méthodes.

Tout d'abord, rappelons ce qu'est la folksonomie. En effet, si l'indexation taxonomique existe depuis des siècles, la folksonomie fait son apparition en 2004, avec l'arrivée du Web 2.0¹⁶⁴. Donnant des possibilités pratiquement infinies aux internautes sans être régie par un système cadré comme la taxonomie¹⁶⁵, la folksonomie met en commun deux termes : folk qui signifie peuple et taxonomie qui signifie la règle de classification¹⁶⁶. Ce type d'indexation permet à l'utilisateur de retrouver des documents numériques facilement avec l'aide de mots-clés qu'il aura préalablement attribués¹⁶⁷. L'office de la langue française définit la folksonomie comme un : « [s]ystème de classification collaborative et spontanée de contenus Internet, basé sur l'attribution de mots-clés librement choisis par des utilisateurs non spécialistes, qui favorise le partage de ressources et permet d'améliorer la recherche d'information¹⁶⁸ ». Des chercheurs comme Olivier Le Deuff tentent de répondre à de nombreuses questions concernant la pertinence

¹⁶⁴ Olivier Le Deuff, *Du tag au like*, p. 30.

¹⁶⁵ *Ibid.*, p. 33.

¹⁶⁶ *Ibid.*, p. 30.

¹⁶⁷ *Ibid.*, p.32.

¹⁶⁸ Gouvernement du Québec, *Office québécoise de la langue française* [site Web], 2013, consulté le 30 mars 2019, <https://www.oqlf.gouv.qc.ca/ressources/bibliotheque/dictionnaires/terminologie_blogue/folksonomie.html>.

de la folksonomie. En effet, ce dernier se demande, si « [l]e phénomène [est] durable ou n'est-ce qu'un effet de mode? Les systèmes de classification et les thésaurus sont-ils menacés par ces systèmes ?¹⁶⁹». Qualifiant la folksonomie comme étant la taxonomie du peuple¹⁷⁰, l'auteur souligne son avantage pour un accès rapide et efficace à l'information¹⁷¹. Toutefois, l'indexation standardisée resterait beaucoup plus durable grâce aux normes qui l'entourent¹⁷². D'autres, comme Fabrice Pirolli, nous rappellent que les folksonomies « [...] contribuent à établir de nouvelles pratiques de description, de partage et d'organisation des ressources informationnelles¹⁷³». L'auteur apporte l'idée que l'indexation sociale « [...] correspond à un déplacement de pratiques documentaires traditionnellement dévolues aux professionnels de l'information vers l'utilisateur final, souvent novice en la matière¹⁷⁴». La folksonomie est un langage naturel et flexible et demande moins de préparation pour l'indexeur. Toutefois, un inconvénient majeur est sans nul doute ce que l'on nomme la synonymie. Cette dernière correspond à « [...] un concept peut être représenté par plusieurs termes [...] », et la polysémie, soit le fait qu'« [...] un terme peut représenter plusieurs concepts »¹⁷⁵.

Afin de donner un exemple concret de la notion de folksonomie, prenons l'application Instagram. Fondée en 2010, elle a comme principal objectif le partage de photos et de vidéos par les utilisateurs de l'application¹⁷⁶. Appartenant depuis 2012 au

¹⁶⁹ Olivier Le Deuff, « Folksonomies. Les usagers indexent le web », p. 66.

¹⁷⁰ Olivier Le Deuff, *Du tag au like*, p. 29.

¹⁷¹ Olivier Le Deuff, « Folksonomies. Les usagers indexent le web », p. 70.

¹⁷² *Ibid.*, p. 66-70.

¹⁷³ Fabrice Pirolli, « Pratiques d'indexation sociale et démarches de veille informationnelle », *Études de communication*, n° 36, 2011, para. 1 et 9.

¹⁷⁴ *Ibid.*, para. 6.

¹⁷⁵ Normand Charbonneau et Mario Robert, « Indexation », p. 165.

¹⁷⁶ Julien Bellver, « Facebook se paye Instagram pour 1 milliard de dollars », *Pure média* [En ligne], 10 Avril 2012, sur le site *oxap.com*, consulté le 20 avril 2019, < <https://www.oxap.com/actu/facebook-se-payee-instagram-pour-1-milliard-d-euros/440310> >.

géant Facebook, l'application est principalement accessible via les plateformes cellulaires, mais aussi sur ordinateur de façon plus restreinte¹⁷⁷. Les utilisateurs peuvent laisser des commentaires, identifier les photos et vidéos et aimer les publications. Instagram favorise ce que l'on nomme la phonéographie, le partage de photos directement par l'appareil mobile¹⁷⁸. Le nom de l'application INSTAgram, comprend d'ailleurs une référence au mot instantané. L'application est conçue pour encourager l'ajout de tags qui permettent de retrouver l'image. Plus l'utilisateur va mettre des tags dans sa publication, plus la photo sera diffusée sur le réseau. Par exemple, si nous indiquons #sherbylove, tous les utilisateurs abonnés à cette métadonnée pourront voir dans leur fil d'actualité la photographie. Le tag permet à la fois de retrouver l'image, mais aussi, de la diffuser sur la plateforme¹⁷⁹.

Une autre plateforme bien connue est Flickr¹⁸⁰. Probablement un des meilleurs exemples d'utilisation de la folksonomie, la plateforme fut développée en 2004 par une entreprise canadienne nommée Ludicorp¹⁸¹. Yahoo va racheter la plateforme en 2005, la soumettant ainsi aux lois américaines¹⁸². En 2018, dans le cadre du processus de démantèlement de Yahoo, Flickr rejoint le groupe SmugMug, lequel vise aussi à mettre de l'avant le travail des photographes à travers le monde par le développement, la

¹⁷⁷ Facebook, « Facebook to Acquire Instagram », *Facebook Newsroom* [En ligne], 9 avril 2012, sur le site *newsroom.fb.com*, consulté le 20 avril 2019, <<https://newsroom.fb.com/news/2012/04/facebook-to-acquire-instagram/>>.

¹⁷⁸ Marlène Duret, « Phonéographie. Qui "instagramme" ses photos et quels sont les sujets fédérateurs de cette application ? », *Le monde* [En ligne], 01 avril 2013, sur le site *Lemonde.fr*, consulté le 20 avril 2019, <https://www.lemonde.fr/vous/article/2013/04/01/phoneographie_3151452_3238.html>.

¹⁷⁹ Olivier Le Deuff, *Du tag au like*, p. 131-132. Aussi : Observations personnelles.

¹⁸⁰ Flickr, *Flickr, Site officiel* [site Web], 2019, consulté le 10 avril 2019, <<https://www.flickr.com/>>.

¹⁸¹ Olivier Le Deuff, *Du tag au like*, p. 117.

¹⁸² Tomio Geron, « A Look Back At Yahoo's Flickr Acquisition For Lessons Today », *Techcrunch* [En ligne], 2014, sur le site *techcrunch.com*, consulté le 10 avril 2019, <<http://tcrn.ch/1qC81Pe>>. Voir aussi: Olivier Le Deuff, *Du tag au like*, p. 117.

communication et le partage¹⁸³. L'objectif de Flickr est de : « [...] permettre aux gens de partager leurs photos avec les personnes de leur choix et notre but est de proposer de nouvelles méthodes d'organisation pour vos photos et vos vidéos¹⁸⁴ ». L'organisation des photos et des vidéos repose sur la collaboration entre les utilisateurs. L'intégration de tags ou de commentaires est libre tant que l'utilisateur donne l'autorisation aux usagers collaborateurs de l'album¹⁸⁵. Flickr soutient, depuis 2004, les licences Creative Common (CC) qui permettent aux utilisateurs de choisir de partager ou non leurs contenus selon des modalités plus flexibles que le copyright traditionnel¹⁸⁶.

Bref, Flickr représente bien le principe de partage photographique à travers une communauté d'internautes passionnés. L'utilisateur a le pouvoir et le contrôle sur les images qu'il publie et la plateforme permet un partage mondial d'informations et de données. Grâce à Flickr, l'accessibilité des images n'est plus un enjeu mais une habitude bien ancrée. Pour un chercheur, cette plateforme peut fournir des éléments intéressants, particulièrement avec l'apport important d'institutions muséales et archivistiques telles que les Archives Municipales d'Halifax et le musée McCord de Montréal.

¹⁸³Matthew Roth, « Together is where photographers belong! », *Flickr Blog* [En ligne], 20 avril 2018, sur le site *blog.flickr*, consulté le 15 avril 2019, <<http://blog.flickr.net/en/2018/04/20/together-smugmug-flickr/>>. Voir aussi : Smugsmug, *Smugsmug, Site officiel* [site Web], 2019, consulté le 15 avril 2019, <<https://www.smugmug.com/>>.

¹⁸⁴Flickr, *About* [site Web], 2019, consulté le 10 avril 2019, <<https://www.flickr.com/about>>.

¹⁸⁵*Ibid.*, voir aussi : Olivier Le Deuff, *Du tag au like, ...*, p. 116. « Sur Flickr, les membres importent des photos, les partagent en toute sécurité, y ajoutent des métadonnées, comme des informations de licence, une géolocalisation, des personnes, des tags, etc., puis interagissent avec leur famille, leurs amis, leurs contacts ou toutes les personnes de la communauté ». Dans Flickr, *Le guide du développeur Flickr* [site Web], 2019, consulté le 15 avril 2019, <<https://www.flickr.com/services/developer>>.

¹⁸⁶Leticia Roncero, « Photography and Licensing: Q&A with Creative Commons CEO Ryan Merkley », *Flickr blog* [En ligne], 14 décembre 2018, sur le site *blog.flickr*, consulté le 15 avril 2019, <<https://blog.flickr.net/en/2018/12/14/photography-and-licensing-qa-with-creative-commons-ceo-ryan-merkley/>>.

Si Flickr vise le partage, le logiciel Picasa, développé à l'origine par Idealab avait pour objectif principal la gestion d'images¹⁸⁷. Racheté par Google en juin 2004¹⁸⁸, l'application devient Google Picasa Web, en 2006, visant à modifier, partager et rechercher des images¹⁸⁹. Toutefois, en mars 2016, Google prend la décision d'abandonner Picasa pour se concentrer entièrement sur Google Photo¹⁹⁰. Les concepteurs ont confirmé que le principe de téléchargement et d'organisation allait rester le même que Picasa, mais qu'il y aurait de nouvelles méthodes de recherche et de partages d'images beaucoup plus avancées¹⁹¹. Le passage de Picasa à Google Photo n'a pas que des avantages. Les outils de retouches photos proposés sont beaucoup moins performants. Certaines alternatives sont alors proposées afin de mixer Google Photo avec le programme Gimp (programme de manipulation d'images [GNU]) de retouche ainsi que Flickr pour un partage web¹⁹². Google Photos propose une version améliorée du système de reconnaissance intégré à Picasa 3.5¹⁹³, mais permet aussi l'ajout automatique de mots-clés¹⁹⁴. En plus d'organiser automatiquement les photographies avec la reconnaissance de

¹⁸⁷ Idealab, *Idealab, site officiel* [site Web], 2019, consulté le 15 avril 2019, <<https://www.idealab.com/>>.

¹⁸⁸ Alexandre Boni et Nicolas Stemart, *Google, Le guide complet*, Paris, Éditions Micro Application, 2008, p. 142.

¹⁸⁹ *Ibid.*

¹⁹⁰ Picasa, *Picasa, site officiel* [site Web], 2016, consulté le 20 avril 2019, <<https://picasa.google.fr/>>. Voir aussi : Google, « Qu'advient-il de Picasa, de Picasa Albums Web et de l'API Picasa Albums Web? », *support google* [En ligne], 2016, sur le site *support.google.com*, consulté le 20 avril 2019, <<https://support.google.com/picasa/answer/6383491?hl=fr>>.

¹⁹¹ Google, « Qu'advient-il de Picasa, ... ».

¹⁹² J. D. Biersdorfer, « Moving On From Picasa », *The New York Times* [En ligne], 21 juin 2018, sur le site *nytimes.com*, consulté le 20 avril 2019, <<https://www.nytimes.com/2018/06/21/technology/personaltech/picasa-replace-photos.html>>.

¹⁹³ Guillaume Champeau, « Picasa reconnaît les visages et tag automatiquement vos photos », *Numerama* [En ligne], 23 septembre 2009, sur le site *numerama.com*, consulté le 20 avril 2019, <<https://www.numerama.com/magazine/14036-picasa-reconnait-les-visages-et-tag-automatiquement-vos-photos.html>>.

¹⁹⁴ Julien Lausson, « Google tire un trait sur Picasa, son service d'albums photos », *Numerama* [En ligne], 15 février 2016, sur le site *numerama.com*, consulté le 20 avril 2019, <<https://www.numerama.com/tech/145584-google-tire-un-trait-sur-picasa-son-service-dalbums-photos.html>>.

formes, les internautes peuvent faire des recherches selon les lieux ou même les sujets, le tout en utilisant la reconnaissance d'objets¹⁹⁵. Cette technologie démontre le potentiel de la reconnaissance d'objets pour le classement des archives photographiques, non seulement personnelles, mais aussi institutionnelles. Si Google Photo est destiné à un usage individuel, le système de reconnaissance des objets est toutefois lié à leur système de publicité, d'où des enjeux de vie privée¹⁹⁶.

L'analyse plutôt brève de ces plateformes permet de poser un regard sur l'importance grandissante de la place de l'utilisateur dans le processus de diffusion et d'accessibilité des sources. Ce ne sont que quelques plateformes libres qui permettent le partage et/ou la réutilisation de photographies. Certaines sont plus personnelles et d'autres plus axées sur la réutilisation des sources. Elles ont toutes le même objectif de partage et d'ajout d'informations. Le musée McCord ainsi que d'autres centres comme Archives Montréal utilisent aussi des plateformes populaires comme Flickr pour promouvoir le partage et la diffusion de leurs images¹⁹⁷. Comme nous avons vu précédemment, la plateforme Flickr est sans contredit une source inestimable d'images et de partage touchant des institutions partout à travers le monde. On constate donc que la folksonomie prend une place importante dans le web 2.0 ainsi que dans le processus de partage d'informations.

¹⁹⁵ Google, *Google, Site officiel* [site Web], 2019, consulté le 21 avril 2019, <<https://www.google.com/photos/about/?hl=fr>>.

¹⁹⁶ Simon Hill, « Google Photos : Should You Be Worried about Privacy? », *Android Authority* [En ligne], 16 juin 2015, sur le site *androidauthority*, consulté le 28 mai 2019, <<https://www.androidauthority.com/google-photos-worried-privacy-616339/>>.

¹⁹⁷Flickr, *Musée McCord* [site Web], 2019, consulté le 21 avril 2019, <<https://www.flickr.com/photos/museemccordmuseum/>>. Voir aussi : Flickr, *Archives Montréal* [site Web], 2019, consulté le 21 avril 2019, <<https://www.flickr.com/photos/archivesmontreal/with/47616334601/>>.

Afin de valider l'apport de la folksonomie par rapport aux méthodes plus classiques testées dans ce chapitre, nous avons soumis la même image tests¹⁹⁸ auprès de quelques personnes. Nous avons choisi un média social que nous n'avons pas beaucoup mentionné précédemment : Facebook. Il permettait de rejoindre rapidement un bassin de personnes dont le niveau de connaissances sur l'histoire et sur Sherbrooke était très varié.

Durant une semaine, nous avons mis l'image sur notre mur Facebook afin de voir comment les abonnés allaient décrire la photographie. Nous avons toutefois mis quelques explications préalables afin d'orienter les gens dans la démarche d'indexation :

8 juillet 2018 : Bonjour à tous! Dans le cadre de ma recherche pour mon mémoire de maîtrise, je dois effectuer une analyse sur l'indexation d'image. Dans un premier temps, je dois analyser l'indexation à l'aide d'une base de données prédéterminée. Dans un deuxième temps, il est principalement question d'indexation par l'utilisateur. De ce fait, j'aimerais votre collaboration afin d'accomplir cette dernière étape. Je mets donc une image provenant des archives de la Société d'histoire de Sherbrooke. Je vous demande donc d'indiquer en commentaire les différents termes qui vous viennent à l'esprit lorsque vous regardez l'image. Que ce soit, des éléments naturels, historiques, événementiels, etc. Ça peut être une date, un lieu, un nom de bâtiment, etc. Par indexation, pensez aux indexes à la fin de certains ouvrages ou même aux tags sur Instagram, cela vous aidera à mieux vous orienter. :) Je vous remercie à l'avance pour votre collaboration! Cela m'est d'une grande aide!

La première constatation est que les usagers sont soit très précis dans leurs termes d'indexation, soit ils sont beaucoup plus vagues ou emploient des mots plus génériques. Dans le test, nous avons constaté que certains allaient jusqu'au nom des édifices et aux lieux comme le carré Strathcona ou l'édifice Odell et d'autres portaient plus leur regard sur l'aspect géographique de la scène comme « centre-ville » et « rue Wellington ». De plus, nous avons aussi constaté que peu de gens ont tenté l'expérience de l'indexation.

¹⁹⁸ SHS, *Résultats de la recherche, fiche 13143*.

Peut-être par peur de se tromper ou simplement car l'intérêt n'était pas là. Somme toute, les participants qui ont volontairement effectué le test ont clairement démontré la différence dans la perception que chacun fait de l'image (Annexe F).

Cette expérience rejoint les constats des chercheurs quant aux résultats très variables en termes de participation, d'engagement et de termes d'indexation. Elle confirme aussi que la folksonomie peut aider à compléter l'information à partir de connaissances personnelles très précises de quelques individus. En effet, lorsqu'un historien effectue ses recherches, il doit nécessairement faire sa propre indexation afin de retrouver adéquatement les éléments dans ses dossiers. Une question se pose alors: ne devrait-il pas partager ses ressources de mots clés avec les institutions archivistiques? Cet échange pourrait mener à un plus grand dialogue entre archivistes et historiens et ainsi enrichir le classement des ressources.

À la lumière de l'expérimentation, nous pouvons dire que la folksonomie à elle seule ne peut assurer adéquatement une indexation dans une institution muséale. En effet, des auteurs comme Olivier Le Deuff mentionnent que les folksonomies n'ont pas influencé la recherche dans les moteurs de recherche. Elles ne remplaceront pas non plus une indexation définie et spécifique. Elles occupent une réalité bien différente selon la manière dont elles sont utilisées. Que ce soit dans un blog, un intranet, une plateforme en ligne, etc¹⁹⁹. Indexer une image sur les réseaux sociaux tels qu'Instagram ne peut être comparé à une indexation dans une base de données d'une institution comme la Société d'histoire. En effet, l'indexeur travaillant pour une institution n'aura sans doute pas les

¹⁹⁹ Olivier Le Deuff, *Du tag au like*, p. 55.

mêmes réflexions concernant les termes à employer que celui qui publie de manière instantanée par ses réseaux sociaux.

Cependant, un des avantages importants de la folksonomie est son aspect collaboratif. L'utilisateur permet d'enrichir les connaissances par l'ajout de données aux images publiées²⁰⁰. Le Deuff mentionne un autre élément important concernant le parallèle entre la taxonomie et la folksonomie. En effet, « [l']indexation professionnelle et les folksonomies sont davantage complémentaires qu'antinomiques²⁰¹ ». La folksonomie permet une recherche rapide de l'information tandis que la taxonomie apporte une durabilité dans le classement des données²⁰². L'indexation par l'utilisateur peut donc apporter des termes plus en phase avec le vocabulaire d'une époque donnée que celui d'une taxonomie standardisée²⁰³. En effet, nous prenons de plus en plus conscience de l'importance qu'a le choix des mots employés autant pour le classement d'une image que pour la diffusion de cette dernière.

2.3 Conclusion

Bref, de nos jours, les besoins d'uniformisations et de formations en matière d'indexation dans les centres d'archives sont de plus en plus criants. Comme le mentionne James M. Turner, l'uniformisation n'était pas un problème de premier ordre dans le passé car l'archiviste était l'intermédiaire entre la personne qui consulte et l'archive elle-même²⁰⁴. Grâce aux systèmes informatisés et numérisés, les chercheurs peuvent se perdre facilement à travers les réseaux propres à chaque centre car ils ont alors un lien direct avec

²⁰⁰ *Ibid.*, p. 59 et 113.

²⁰¹ *Ibid.*, p. 79.

²⁰² Olivier Le Deuff, « Folksonomies. Les usagers indexent le web », p. 66-70.

²⁰³ Olivier Le Deuff, *Du tag au like*, p. 90.

²⁰⁴ James M. Turner, *Images en mouvement*, p. 76-77.

les ressources²⁰⁵. De plus, pour ce qui est des programmes d'indexations, nous constatons qu'il serait intéressant d'intégrer un système général comme Iconclass ou l'AAT au processus humain d'indexation. En effet, ces derniers deviendraient alors beaucoup plus complets et il sera alors plus facile de repérer les images dans la base de données. Les logiciels professionnels permettent l'intégration d'une liste de mots génériques très longue, voire même trop longue pour que l'indexation humaine puisse y arriver sans l'aide de logiciels. La liste de mots permettrait de prendre conscience de certains éléments que nous pourrions oublier dans une indexation seulement folksonomique. Finalement, à la lumière des expérimentations effectuées, les mots proposés sont assez standardisés pour pouvoir être inscrits à l'aide de la reconnaissance d'image. En effet, comme nous allons le voir dans le prochain chapitre, la reconnaissance d'image permet une vue d'ensemble des éléments génériques tels les bâtiments, les rues, la végétation, l'architecture, etc. Une première analyse avec le logiciel permettrait à l'être humain de compléter par la suite avec des éléments beaucoup plus spécifiques à l'image. Une telle indexation sera alors d'autant plus appropriée pour les recherches plus générales à travers différents corpus conservés dans différentes institutions.

²⁰⁵ *Ibid.*

Chapitre 3 : Le potentiel de la reconnaissance d'image

L'intelligence artificielle est devenue la plus récente tendance informatique depuis quelques années et ses techniques d'apprentissage machine, combinée à la segmentation de l'image, ont apporté des progrès importants à la reconnaissance d'images. L'intelligence artificielle a pour objectif de simuler des facultés cognitives de l'humain²⁰⁶. Une présentation des principes de ces approches permettra de cerner leurs potentiels dans le processus de numérisation et d'indexation des images. Pour mettre à profit les avantages que peuvent procurer la numérisation des archives photographiques et l'ubiquité d'images créées à partir d'appareils numériques, les archivistes auront besoin de nouveaux outils afin d'éviter que cette nouvelle richesse archivistique demeure sous-exploitée. Bien que la combinaison de l'œil et du cerveau humain soit plus efficace que l'ordinateur pour identifier et reconnaître facilement des éléments sur une photographie, la machine peut plus facilement traiter de grands corpus de manière uniforme.

Ce chapitre s'écarte de la norme en histoire par son recours aux blogues comme ressource. L'objectif étant de présenter l'IA de manière vulgarisée pour le monde de l'histoire, ces ressources semblaient mieux adaptées que les thèses en informatique ou les documents techniques qui nous plongent dans un niveau de détail technique qui n'est pas nécessaire à la compréhension des principes généraux de la technologie. Précisions de plus que notre utilisation du terme intelligence artificielle renvoie à l'apprentissage machine ou apprentissage profond, bien que l'intelligence artificielle puisse aussi décrire d'autres approches. Ce choix découle du fait que l'apprentissage profond est l'approche

²⁰⁶ Jean-Gabriel Ganascia, *IA vers une domination programmée?*, Paris, Éditions Le cavalier Bleu, 2017, 2e éd. (2007), p. 9.

en intelligence artificielle qui est la plus pertinente pour la reconnaissance des collections de photos d'archives. Les acronymes IA pour intelligence artificielle et AP pour apprentissage profond seront aussi utilisés afin d'alléger le texte.

3.1 Logiciels de reconnaissance d'objets

Les recherches dans le domaine de la reconnaissance de formes sont très variées. En effet, la vision par ordinateur (*computer vision*) est un champ de recherche en soi qui mobilise de nombreux spécialistes²⁰⁷. En résumé, la reconnaissance d'images / de formes, que l'on peut aussi appeler la détection d'objets, consiste à reconnaître et classifier les formes transmises à l'aide de données brutes, par exemple des pixels, et de la catégoriser selon les données apprises par l'ordinateur²⁰⁸. Dans le cadre de ce projet, nous avons retenu trois méthodes qui permettent d'effectuer une reconnaissance d'images dans un contexte d'archives de photographies anciennes : l'intelligence artificielle, la segmentation et les approches hybrides.

L'idée qu'une machine puisse réaliser des gestes humains ne date pas d'hier. Dès l'Antiquité, des philosophes et des poètes tels qu'Homère, proposent déjà des êtres artificiels ressemblant à l'humain. Dans l'*Illiade*, nous retrouvons des servantes conçues et

²⁰⁷ Thomas S. Huang, « Computer Vision: Evolution and Promise », dans Carlo E. Vandoni, éd., *19th CERN School of Computing (conference), Organisation européenne pour la recherche nucléaire, CERN european organization for nuclear research*, Egmond aan Zee, The Netherlands, Éditions Geneva, n°8, 8-21 septembre 1996, p. 21-25.

²⁰⁸ Bharath Ramsundar et Reza Bosagh Zadeh, *TensorFlow pour le deep learning, La regression linéaire à l'apprentissage par renforcement*, Paris, Éditions First interactive, 2018, p. 132. Voir aussi : Stephane Nachez, « Classifier des images avec TensorFlow », *ActuIA* [En ligne], 2 juillet 2017, sur le site *actuia.com*, consulté le 1 octobre 2019, <<https://www.actuia.com/tutoriel/classifier-images-tensorflow/>>, voir aussi : Evelyn Grevelink, « A Closer Look at Object Detection, Recognition and Tracking », *Intel Software Developer zone* [En ligne], 18 décembre 2017, sur le site *software.intel.com*, consulté le 1 octobre 2019, <<https://software.intel.com/en-us/articles/a-closer-look-at-object-detection-recognition-and-tracking>>.

fabriquées par Héphaïstos²⁰⁹. Au milieu du XVIII^e siècle, l'horloger suisse Pierre Jaquet-Droz élabore trois androïdes : un écrivain, une musicienne et un dessinateur²¹⁰. Avec l'aide de ses fils, Jaquet-Droz élabora des automates dont l'écrivain qui était composé de 6000 pièces²¹¹. Plus proche des principes mêmes de l'IA, l'auteur Arthur C. Clarke inclut dans son roman « 2001 : l'Odyssée de l'espace » (1968) HAL 9000, un ordinateur de vaisseau spatial ayant des capacités humaines : parler, reconnaître les visages, raisonner, etc.²¹². De nos jours, les IA les mieux connues du grand public sont celles de Deep Blue (IBM) et AlphaGo (DeepMind). Le premier a démontré le potentiel de l'IA en l'emportant contre le champion du monde d'échecs, Garry Kasparov en 1997²¹³. Récemment, AlphaGo, créé par David Silver, a réussi à battre le champion du monde du jeu de Go²¹⁴ en 2015, 2016 et en 2017²¹⁵.

La conception moderne de la discipline de l'IA apparaît en 1956 lors d'une école d'été organisée par John McCarthy, Marvin Minsky, Nicolas Rochester et Claude Shannon au Dartmouth College au New Hampshire aux États-Unis²¹⁶. L'objectif de

²⁰⁹ Jean-Gabriel Ganascia, *IA vers une domination programmée?*, Paris, Éditions Le cavalier Bleu, 2017, 2e éd. (2007), p. 37. Voir aussi : Ian Goodfellow, Yoshua Bengion et Aaron Courville, *L'apprentissage profond*, France, Éditions Massot et Quantmetry, 2018, p. 27.

²¹⁰ Luis de Miranda dir., *Intelligence artificielle et robotique en 30 secondes, 50 avancées majeures, expliquées en moins d'une minute*, Montréal, Édition Hurtubise, 2019, p. 16.

²¹¹ *Ibid.*, p. 16. « Il tient une plume d'oie qu'il plonge dans un encrier et peut écrire jusqu'à 40 caractères [...]. Le Dessinateur utilisait un crayon pour dessiner un portrait du roi de France, Louis XVI, et de la reine Marie-Antoinette, [...]. La Musicienne était une femme élégante dont les doigts mobiles pressaient les touches d'un orgue véritable ».

²¹² *Ibid.*, p. 24.

²¹³ *Ibid.*, p. 36. Voir aussi : Philadelphie (AFP), « Garry Kasparov remporte la 5^e partie contre « Deep Blue » », *Le Devoir* (Montréal), 17 février 1996, vol. LXXXVI, n°40, p. A5 / François Brousseau, « La victoire de la ferraille Deep Blue remporte le match contre Kasparov », *Le Devoir* (Montréal), 12 mai 1997, vol. LXXXVIII, n°104, p. A1 et A8. Voir aussi: Ian Goodfellow, ..., *L'apprentissage profond*, p. 28.

²¹⁴ *Ibid.* Voir aussi: DeepMind, *AlphaGo* [site Web], 2020, consulté le 26 avril 2020, < <https://deepmind.com/research/case-studies/alphago-the-story-so-far> >.

²¹⁵ *Ibid.*, « [...] en 2017, [l'ordinateur] joua en ligne contre plusieurs joueurs internationaux d'envergure, remportant 60 victoires consécutives ».

²¹⁶ Jack Challoner, *L'intelligence artificielle, un guide d'initiation au futur de l'informatique et de la robotique*, France, Éditions Pearson, 2003, p. 7. Voir aussi : Jean-Gabriel Ganascia, *IA vers une domination programmée?*, p. 9 et 41.

l'école était de « [...] créer une nouvelle science destinée à modéliser sur ordinateur la prise de décisions [...]»²¹⁷. Par prise de décisions, il s'agit bien sûr de prise de décisions cognitives. D'une manière plus précise, le chercheur Jean-Gabriel Ganascia apporte l'idée que : « [...] l'IA reposait sur la conjecture selon laquelle toutes les facultés cognitives, en particulier le raisonnement, le calcul, la perception, la mémorisation, voire même la découverte scientifique ou la créativité artistique, pourraient être décrites avec une précision telle qu'il devrait être possible de les reproduire à l'aide d'un ordinateur²¹⁸». Le concept de l'IA est passé par de nombreuses étapes au cours de son histoire²¹⁹. En effet, avec toutes ces idées déjà bien ancrées depuis l'Antiquité, nous pouvons nous poser la question à savoir si les ordinateurs, dans un futur proche, auront une conscience. Jack Challoner affirme que les ordinateurs traditionnels « répondent aux informations qu'ils reçoivent [...] grâce à un ensemble d'instructions, appelées programmes²²⁰» alors que les ordinateurs utilisant l'IA doivent s'adapter pour répondre à des circonstances imprévues par les programmeurs, bien que le code de base ait été conçu par des humains.

De l'IA découlent certains types d'apprentissages (Voir annexe G), dont l'apprentissage automatique (*machine learning*) qui est issu de données²²¹ et l'apprentissage profond (AP) (Deep learning)²²². Ce dernier s'inspire du fonctionnement cérébral du cerveau et il est structuré par couches selon l'abstraction des connaissances,

²¹⁷ *Ibid.*, p. 41.

²¹⁸ *Ibid.*, p.9.

²¹⁹ *Ibid.*, p. 41-42.

²²⁰ Jack Challoner, *L'intelligence artificielle*, p. 6. « Les ordinateurs modernes sont programmés intelligemment pour agir intelligemment ». Dans *Ibid.*, p. 7.

²²¹ Laurent Charlin, « Intelligence artificielle : une mine d'or pour les entreprises », *HEC MTL Gestion*, vol. 42, n° 1, 2017, p. 76-79. Voir aussi : Ian Goodfellow, ..., *L'apprentissage profond*, p. XVII. Voir aussi : Evelyn Grevelink, « A Closer Look at Object Detection, Recognition and Tracking ».

²²² Marie-Anne Valiquette, « L'intelligence artificielle et l'imagerie médicale », *Article de recherche, technologies de la santé*, sur le site *Substance* [En ligne], 6 février 2018, sur le site *Actualité scientifique et innovation de l'ÉTS*, consulté le 2 mars 2018, <<https://substance.etsmtl.ca/lintelligence-artificielle-imagerie-medicale#respond>>.

du plus général au plus précis²²³. L'AP travaille avec ce que l'on appelle la hiérarchie des concepts. Grâce aux concepts plus simples, l'ordinateur est en mesure d'en élaborer de plus complexes²²⁴. Par exemple, lorsque l'ordinateur doit décomposer une image, il va analyser les couches visibles, principalement les pixels. Par la suite, la machine va élaborer une série de couches cachées en fonction de l'image présentée. Les couches cachées peuvent être les contours, les coins, des parties d'objets, etc²²⁵. Yann Lecun rappelle une question fondamentale quant au potentiel des ordinateurs pour reconnaître des images : « Comment une machine peut-elle identifier un chien ou une chaise dans le tableau de nombre d'une image quand l'apparence d'un chien ou d'une chaise et des objets qui les entourent peuvent varier infiniment ?²²⁶ » Selon ce dernier, il est particulièrement difficile pour un programme de s'adapter à plusieurs situations.

L'apprentissage automatique pour sa part est un système entraînable. Il peut donc apprendre ce qu'est un chien, un chat, un arbre, un escalier, etc²²⁷. Ce type d'apprentissage repose sur des connaissances initiales codées en dur, donc non modifiables et supervisées. L'ordinateur crée ses propres modèles en fonction des données envoyées par le programmeur²²⁸. Nous devons donc lui montrer les éléments qu'il devra assimiler, afin qu'il s'ajuste au fur et à mesure de son apprentissage²²⁹. À force d'apprendre les éléments,

²²³ Laurent Charlin, « Intelligence artificielle... », p. 78. Voir aussi : Collège de France, « Yann LeCun. Recherche sur l'intelligence artificielle », *Collège de France* [site Web], 2018, sur le site *Collège de France*, consulté le 22 janvier 2018, < <https://www.college-de-france.fr/site/yann-lecun/Recherches-sur-l-intelligence-artificielle.htm> > p.3. Voir aussi : Yann LeCun, Yoshua Bengio Et Geoffrey Hinton, « L'apprentissage en profondeur », *Nature*, vol. 521, 28 mai 2015, p. 436 – 444.

²²⁴ Ian Goodfellow, ..., *L'apprentissage profond*, p. 28.

²²⁵ *Ibid.*, p. 32. Voir ANNEXE H.

²²⁶ Collège de France, « Yann LeCun. Recherche sur l'intelligence artificielle », p. 1.

²²⁷ *Ibid.*, p. 2.

²²⁸ Ian Goodfellow, ..., *L'apprentissage profond*, p. 29.

²²⁹ Collège de France, « Yann LeCun. Recherche sur l'intelligence artificielle », p. 2.

la machine parvient à une « capacité de généralisation », lui permettant de classifier des images qu'elle n'aura pas vues auparavant²³⁰.

Bref, le défi le plus grand de l'IA est sans nul doute l'accomplissement des tâches considérées comme routinières pour les gens. Nous pensons alors à la reconnaissance de la parole et des visages. Pour l'humain, il est simple de reconnaître une voiture ou un chat. Pour l'ordinateur, il est beaucoup plus difficile de repérer les formes²³¹. En effet, l'ordinateur ne va pas voir l'image d'un chat, mais plutôt une série de chiffres et de calculs apparentés à cette dernière²³². Une marge d'erreur existe donc, mais la rapidité de traitement par rapport à l'humain apporte quand même un gain d'efficacité.

Si l'IA fait les manchettes, une autre technique moins connue est aussi à la base de la reconnaissance d'images : la segmentation²³³. Les études portant sur la segmentation d'images mettent en perspective ses avantages et ses modes de fonctionnement à l'aide de modèles²³⁴. Selon Christophe Rosenberger nous devons premièrement améliorer la qualité

²³⁰ *Ibid.*, p. 2.

²³¹ Ian Goodfellow, ..., *L'apprentissage profond*, p. 27. Voir aussi : James M. Turner, *Images en mouvement*, p. 45.

²³² Bharath Ramsundar ..., *TensorFlow pour le deep learning*, p. 123-152. Voir aussi : Jeannita Richard, Éric Lemyre et Louis Faure, « L'intelligence artificielle » [Enregistrement vidéo], 15 juillet 2018, sur le site *Radio-Canada, Découverte*, saison 30, épisode 32, consulté le 20 septembre 2019, < <https://ici.radio-canada.ca/tele/decouverte/site/episodes/409710/intelligence-artificielle> >.

²³³ Jose Dolz, « Vers la segmentation automatique des organes à risque dans le contexte de la prise en charge des tumeurs cérébrales par l'application des technologies de classification de deep learning », Thèse de doctorat (Mathématique), France, Université de Lille 2, 2016, p. 178.

²³⁴ Voir notamment : Philippe Bolon, *et al.*, *Analyse d'images : Filtrage et segmentation. Enseignement de la physique*, France, Éditions MASSON, 1995, p. XXI. N.B. : « Le livre publié en 1995 chez MASSON (EAN13 : 9782225849237) est épuisé. Cette version pdf est une version élaborée à partir de la version préliminaire transmise à l'éditeur. La mise en page est légèrement différente de celle du livre. Malheureusement quelques figures de l'annexe C ont été perdues ».

de l'image pour qu'elle soit facile à traiter²³⁵. Suit alors le processus de segmentation, puis de l'interprétation²³⁶.

Les chercheurs ont élaboré différentes approches afin de comprendre l'image notamment par contour (frontières)²³⁷ et par régions²³⁸. Ces approches découlent d'analyses de bas et de haut-niveaux. Segmenter une image, par exemple en détectant les contours, constitue un traitement de bas-niveau. L'ordinateur traite l'image dans un ensemble sans analyser sa représentation²³⁹. Des chercheurs mentionnent que : « [les] traitements de bas-niveau opèrent plutôt sur des données de nature numérique²⁴⁰ ». En segmentant une image, nous créons des régions, soit des ensembles qui ont tous des propriétés communes²⁴¹. Pour l'analyse de haut-niveau, il s'agit plutôt de reconnaître le sujet de l'image analysée pour ainsi l'étiqueter avec des mots²⁴².

La segmentation consiste donc à faire des sous-ensembles de régions qui auront tous des caractéristiques propres. L'objectif est donc d'échantillonner l'image de manière sémantique²⁴³. Chacune des sections permet une meilleure représentation des

²³⁵ Christophe Rosenberger, « Mise en œuvre d'un système adaptatif de segmentation d'images », Thèse de doctorat (Structure et propriété de la matière), France, Université Rennes 1, 1999, p. 11. Voir aussi : « Les prétraitements les plus utilisés sont les suivants : les transformations ponctuelles de luminance, la réduction du bruit ou lissage [et] le rehaussement de contraste », dans Philippe Bolon, *et al.*, *Analyse d'images ...*, p. 4 et 64-96 et 105.

²³⁶ *Ibid.*, p. 11-12.

²³⁷ Philippe Bolon, *et al.*, *Analyse d'images ...*, p. 3 : « La notion de « frontière » est associée à une variation d'intensité ou à une discontinuité entre les propriétés de deux ensembles connexes de points. L'approche frontière regroupe les techniques de détection de contours. Ces méthodes ne conduisent pas directement à une segmentation de l'image telle qu'on l'a définie précédemment, car les contours obtenus sont rarement connexes, il faut donc procéder à une fermeture des contours si l'on souhaite une partition complète de l'image. En effet, après fermeture de contours, la dualité contours-régions apparaît nettement. Les régions sont définies topologiquement comme l'intérieur d'une ligne fermée ».

²³⁸ *Ibid.*, p. 3 : « La notion de « région » fait référence à des groupements de points ayant des propriétés communes. Les méthodes de l'approche région aboutissent directement à une partition de l'image, chaque pixel étant affecté à une région unique ».

²³⁹ *Ibid.*, p. 1-2.

²⁴⁰ *Ibid.*

²⁴¹ *Ibid.*, p. 2 par exemple: la texture, l'intensités, etc.

²⁴² *Ibid.*, p. 1-2.

²⁴³ Abdelkader Hamadi, « Utilisation du contexte pour l'indexation sémantique ... », 179 p.

composantes²⁴⁴. De ce fait, l'ordinateur sera en mesure de différencier, par exemple, les pixels, la profondeur, la texture, la couleur, etc²⁴⁵. Les méthodes²⁴⁶ de segmentation d'images tentent de copier le système visuel humain et de raisonner de la même manière²⁴⁷.

Une approche de l'IA comme l'apprentissage automatique a pour objectif de reconnaître un animal ou un objet à partir d'une banque prédéfinie alors que la segmentation a d'abord comme objectif de distinguer les différentes zones d'une image avant de procéder à l'analyse. L'hybridation de ces deux approches apporte un degré de précision plus grand, notamment dans le domaine de l'imagerie médicale. L'IA et la segmentation, associées avec un programme de reconnaissance de formes, peuvent donc améliorer le travail des radiologistes et cette approche pourrait aussi s'appliquer dans le domaine des archives.

L'imagerie médicale est un domaine de pointe dont les enjeux permettent d'explorer les défis de la reconnaissance de l'image. La résonance magnétique, la radiographie, l'échographie, etc., ont généré des milliards d'images depuis plusieurs années²⁴⁸. Des chercheurs comme Ismail Ben Ayed²⁴⁹, pensent que l'IA peut solutionner

²⁴⁴ Jose Dolz, « Vers la segmentation ... », p. 23.

²⁴⁵ Christophe Rosenberger, « Mise en œuvre d'un système adaptatif de segmentation d'images », p.11. Voir aussi: Bharath Ramsundar ..., *TensorFlow pour le deep learning*, p. 133. / Jose Dolz, « Vers la segmentation ... », p. 23.

²⁴⁶ *Ibid.*, p. 18 à 43.

²⁴⁷ *Ibid.*, p. 18.

²⁴⁸ Sébastien Langevin, « L'intelligence artificielle au service de l'imagerie médicale », *Actualités du réseau de l'Université du Québec* [En ligne], 24 avril 2018, sur le site de *l'Université du Québec*, consulté le 27 juillet 2019, < <https://www.quebec.ca/reseau/fr/medias/actualites-du-reseau/lintelligence-artificielle-au-service-de-limagerie-medicale> >. Pour en savoir plus: la toute première radiographie fut réalisée en 1895. Dans Bruno Saintôt, « Quelques critères d'utilisation de l'imagerie médicale, conclusion du colloque », Tome 60 de *Laennec*, n°4, 2012, p. 47.

²⁴⁹ « Il est spécialisé dans la conception d'algorithmes servant au traitement, à l'analyse et à l'interprétation des images médicales, notamment d'algorithmes d'apprentissage automatique et d'optimisation, qui permettent de résoudre les problèmes de données massives en imagerie 3D ou 4D ». Dans *Ibid.*

les divers problèmes reliés aux corpus d'images générés par ces techniques. Selon le chercheur, l'IA permet de traiter des milliers d'images rapidement et de repérer efficacement et avec une très grande précision les anomalies²⁵⁰.

Dans la pratique traditionnelle, le technicien effectue manuellement la segmentation des radiographies afin de déceler les anomalies²⁵¹. Cela demande un temps considérable et l'intégration d'une automatisation dans le processus de segmentation permet d'obtenir des résultats rapidement²⁵². Une différence majeure entre la segmentation d'images traditionnelles et celles de l'imagerie médicale est sans aucun doute la qualité des images à la sortie. En effet, la photographie a plus de chance d'être nette et précise comparativement à la radiographie qui demande une technique de prise de vue beaucoup plus difficile donnant une image composée d'un nombre de couches de tissus X où l'intensité entre les couches peut affecter la segmentation²⁵³.

Des techniques d'analyse en imagerie médicale peuvent servir de modèle en histoire. La méthode basée sur les atlas consiste à prendre une image de référence déjà segmentée d'un patient que l'on nomme type. Grâce à cette dernière, il est possible de superposer l'image du patient traité avec l'image type. Il peut en résulter des déformations entre l'atlas et la radiographie du patient. Malheureusement, l'un des inconvénients majeurs de cette méthode est que chaque patient est unique et peut varier pas rapport à l'atlas de base. L'utilisation de plusieurs atlas doit donc être prise en considération afin d'obtenir des résultats adéquats et précis²⁵⁴. Dans l'analyse des photographies historiques,

²⁵⁰ *Ibid.*

²⁵¹ « Dans leur forme brute, les images médicales sont représentées par des tableaux de valeurs représentant des quantités qui montrent le contraste entre les différents types de tissus du corps ». Dans Jose Dolz, « Vers la segmentation ... », p.178.

²⁵² Jose Dolz, « Vers la segmentation ... », p. 1 et p. 178.

²⁵³ *Ibid.*, p. 179.

²⁵⁴ *Ibid.*, p. 179-180.

l'équivalent serait de superposer des portraits afin de comparer l'évolution d'éléments du visage, tel que le sourire ou la pilosité faciale.

Le modèle statistique est plutôt basé sur le modèle d'apprentissage. L'analyse d'une radiographie consiste alors à repérer des points de bases et par la suite d'établir une moyenne en fonction des données recueillies et du modèle d'apprentissage enregistré. Toutefois, certaines problématiques peuvent survenir. En effet, si l'ordinateur n'a pas assez de données d'entrées, il ne sera pas en mesure d'estimer convenablement la valeur moyenne de l'image à analyser. De plus, si l'image du patient est bruitée (flou), il devient plus difficile pour l'ordinateur de repérer les points et d'établir une moyenne convenable²⁵⁵. En photographie historique, l'équivalent serait de définir des géométries types, par exemple pour les bâtiments, que le système tenterait alors de repérer dans d'autres images.

Finalement, la méthode de l'apprentissage automatique pourrait ici apporter quelques solutions. Comme nous l'avons mentionné précédemment, elle consiste en un apprentissage supervisé. Le technicien crée une série d'algorithmes qui va permettre de « rendre autonome » l'ordinateur dans ses analyses²⁵⁶. À partir d'une série d'images types, le système apprend à différencier les anomalies et les formes normales. Dans les photos historiques, ce type d'apprentissage permettrait de mettre l'accent sur des objets précis.

L'expertise médicale en anatomie et en pathologie nécessaire à la création d'algorithmes de qualité²⁵⁷ se transpose aussi en histoire où il faut tirer profit de l'expertise des archivistes et des historiens pour créer les images de base et les processus d'indexation.

²⁵⁵ *Ibid.*, p. 180.

²⁵⁶ *Ibid.*, p. 182-184. Pour plus d'informations sur l'AP, revoir le point concernant l'intelligence artificielle.

²⁵⁷ *Ibid.*, p. 179.

Bref, malgré ces avantages, la segmentation et l'IA ont, elles aussi, leurs inconvénients. Les chercheurs mentionnent que l'intelligence artificielle a comme principal problème son besoin d'assimiler une quantité importante d'images, afin qu'elle ait un apprentissage optimal et qu'elle reconnaisse l'élément qu'elle devra détecter²⁵⁸. La reconnaissance des chiens sur les images pose par exemple le défi de gérer les différentes races de chiens qui ont tous une morphologie propre²⁵⁹.

De plus, l'apprentissage profond, qui découle de l'IA, est dépendant de la qualité de l'image qu'elle analyse²⁶⁰. Si l'image n'est pas claire, il peut en résulter un mauvais repérage et donc une confusion des formes, par exemple, une cheminée avec un pilastre. Pour ce qui est de la segmentation, un des désavantages est que les méthodes sont surtout efficaces sur certains types d'images²⁶¹. En effet, cette dernière ne doit pas être trop bruitée ou texturée²⁶². De plus, lorsqu'elle est faite à la main, cela peut prendre plus de temps et donc n'être en aucun cas efficace. Ces exemples montrent que la difficulté découle non seulement de la qualité de l'image, mais aussi du choix des caractéristiques à extraire de l'image. L'algorithme simple de l'apprentissage automatique prend en compte ces caractéristiques afin d'accomplir la tâche demandée. Il faut prendre en considération des facteurs de variations et d'occurrences dont les valeurs de pixels, les ombres, la position de l'objet, la luminosité, etc²⁶³. D'autres problématiques liées à la déformation de données

²⁵⁸ Substance, « L'intelligence artificielle pour interpréter les images médicales », *Article de recherche, technologies de la santé* [En ligne], 21 octobre 2016, sur le site *Substance, Actualité scientifique et innovation de l'ÉTS*, consulté le 2 mars 2018, <<https://substance.etsmtl.ca/reconnaissance-interpretation-images-medicales>>. Voir aussi: Collège de France, « Yann LeCun. Recherche sur l'intelligence artificielle », p. 2.

²⁵⁹ *Découverte*, saison 30, épisode 32. (6 : 30).

²⁶⁰ Marie-Anne Valiquette, « L'intelligence artificielle et l'imagerie médicale ».

²⁶¹ Christophe Rosenberger, « Mise en œuvre d'un système adaptatif de segmentation d'images », p. 35.

²⁶² *Ibid.*, p. 35.

²⁶³ Ian Goodfellow, ..., « L'apprentissage profond », p. 29 à 31. Voir aussi: Bharath Ramsundar ..., *TensorFlow pour le deep learning*, p. 132.

ou à des biais dans les données d'entraînement qui ne tiennent pas compte de toutes les variantes peuvent aussi amener des malaises et des injustices. La reconnaissance visuelle de Google a notamment connu des ratés. Le programme avait assigné les hommes à la peau noire à l'étiquette des gorilles²⁶⁴. Malgré tout le potentiel de l'AP, il est difficile de savoir sur quelles bases l'application a pris telle ou telle décision, surtout lorsqu'il s'agit de systèmes propriétaires dont le code est fermé. Différentes méthodes de validation sur des corpus variés sont donc nécessaires pour assurer la qualité du traitement.

Malgré leurs faiblesses, il est tout de même possible de tirer profit d'une combinaison des deux approches, comme on l'observe en imagerie médicale. Plusieurs études montrent l'efficacité de la mixité des procédés²⁶⁵. Avec la segmentation automatisée, combinée avec l'apprentissage profond, il est maintenant possible de détecter une anomalie dans les radiographies d'une personne de manière beaucoup plus rapide²⁶⁶. Selon l'état actuel des technologies, un logiciel qui segmenterait les différentes parties d'une photographie pour ensuite en identifier les composantes serait l'outil idéal pour procéder à un classement semi-automatique des images.

3.2 Les défis de la reconnaissance des images sur le terrain

Les avancées dans le milieu médical sont transposables sur le terrain des photographies historiques. Celles-ci, à l'exception des portraits, présentent très souvent des paysages. Ceux-ci comprennent généralement de la végétation ou de la construction

²⁶⁴ Bharath Ramsundar ..., *TensorFlow pour le deep learning*, p. 235.

²⁶⁵ Nous pensons, entre autres à Marie-Anne Valiquette, Jose Dolz ainsi qu'à l'étude menée par les chercheurs June-Goo Lee, Sanghoon Jun, Young-Won Cho, Hyunna Lee, Guk Bae Kim, Joon Beom Seo, et Namkug Kim. (June-Goo Lee, *et al.*, « Deep Learning in Medical Imaging: General Overview », *Korean Journal of Radiology*, n° 4, 2017, p. 570-584. / Jose Dolz, « Vers la segmentation ... », 179 p.).

²⁶⁶ Marie-Anne Valiquette, « L'intelligence artificielle et l'imagerie médicale ». Voir aussi : Collège de France, « Yann LeCun. Recherche sur l'intelligence artificielle », p. 5.

humaine. De ce fait, afin d'avoir une approche plus ciblée vers le cœur de notre recherche, mettons en perspective deux cas orientés vers la reconnaissance de bâtiments. À travers ces exemples, nous verrons les difficultés associées au processus de reconnaissance de formes et par quels moyens il est possible de repérer des éléments ayant des similitudes identifiables.

Commençons par une expérimentation effectuée sur 36 bâtiments faite par Wei Zhang et de Jana Kosecka du Department of Computer Science de l'Université George Mason (Fairfax, Virginie)²⁶⁷. Dans leur rapport, les chercheurs distinguent plusieurs caractéristiques à prendre en considération au moment de la reconnaissance de bâtiments. Il y a notamment les structures répétitives, les occlusions, comme les piétons et les arbres et, les inclusions, comme les points de vue et les perspectives²⁶⁸. Afin d'avoir une reconnaissance optimale, il est essentiel de coordonner les données entre elles. L'objectif principal de leur rapport est : « [...] *we are interested in investigating the efficacy of local image descriptors as building representations, their detection repeatability as a function of viewpoint as well as discrimination capabilities*²⁶⁹ », soit le rendement des descripteurs (points d'association, pixels, etc)²⁷⁰.

Durant l'expérience, la technique de reconnaissance utilisée est un système fonctionnant par vote²⁷¹. La méthode consiste à faire ressortir les points-clés de l'image

²⁶⁷ Wei Zhang et Jana Kosecka, « Experiments in Building Recognition », Department of Computer Science George Mason University Technical Report Series, 9 p.

²⁶⁸ Abdelkader Hamadi, « Utilisation du contexte pour l'indexation sémantique ... », p. 3-4.

²⁶⁹ *Ibid.*, p. 1.

²⁷⁰ Pour plus d'informations, consultez : Chifa Nawal et Abdelmajid Badri, « Combinaison de descripteurs globaux et locaux pour la recherche par le contenu d'informations visuelles », Département Electrique, Faculté des Sciences et Techniques - Mohammedia, Université de Hassan II de Casablanca (Conférence: La 8ème édition de la rencontre nationale des jeunes Chercheurs en Physique), décembre 2015, 6 p.

²⁷¹ Wei Zhang et Jana Kosecka, « Experiments in Building Recognition », p. 4.

analysée et de les associer aux images de référence dans la base de données. Nous pouvons aussi l'appeler modèle de correspondance. Pour ce faire, plusieurs images d'un même bâtiment sont utilisées afin d'avoir de nombreux angles de vue. Le modèle qui a le plus de points correspondants est celui que nous recherchons. La reconnaissance d'image est un processus qui se doit d'être précis et qu'il est important de prendre en considération tous les éléments, comme l'éclairage, qui peuvent jouer sur la précision des résultats²⁷².

Pour le deuxième exemple apporté ici il est question du même principe de reconnaissance de bâtiment. Comparativement aux recherches de Zhang et Kosecka, qui utilisent seulement les images prises à partir du sol, Mark Wolff, Robert T. Collins et Yanxi Liu présentent pour leur part des approches de détection avec autant les prises de vue aériennes qu'au sol²⁷³. Dans ce cas-ci, un treillis est superposé à l'image. Chaque tuile permet de connaître le contexte, la couleur, la texture et sa place sur le bâtiment. L'agencement des tuiles permet donc d'effectuer plus facilement une correspondance entre les divers points d'intérêts comme la texture, les couleurs, les formes, etc²⁷⁴.

L'objectif de la recherche est de voir s'il est possible, avec la technique de reconnaissance de formes, d'associer les vues aériennes de façades de bâtiments avec différents types de vues au sol des mêmes édifices. Les plus grands défis sont, comme mentionné dans la recherche précédente, les occlusions et les inclusions. Afin de limiter les interférences, les chercheurs ont utilisé les différentes vues aériennes et au sol pour

²⁷² *Ibid.*, p. 5: « *In the experiment 53 out of 72 test images are correctly recognized (listed as first candidate), the rest of them are misclassified* ».

²⁷³ Mark Wolff, Robert T. Collins et Yanxi Liu, « *Regularity-Driven Building Facade Matching between Aerial and Street Views* », School of Electrical Engineering and Computer Science (Conference on Computer Vision and Pattern Recognition 2016), The Pennsylvania State University, p. 1591 et 1596: « A quantitative evaluation of our method is carried out on a set of 120 aerial facades. Each facade is visible in 10-15 street images, giving us over 1000 total potential facade matches ».

²⁷⁴ *Ibid.*, p. 1591.

supprimer les occlusions et par la même occasion, améliorer les textures des façades de bâtiments²⁷⁵. Dans les occlusions, il ne faut pas non plus oublier les différentes nuances d'éclairage²⁷⁶. Une problématique importante dans la reconnaissance de bâtiments est leur régularité²⁷⁷. En effet, le programme peut plus facilement confondre un bâtiment avec un autre lorsqu'il y a des similarités. Afin de résoudre le problème, les chercheurs utilisent la méthode du treillis qui génère plusieurs tuiles. Grâce à cette dernière, les chercheurs transforment cette problématique en avantage pour les associations de points entre l'image aérienne et celle au sol. En bref, les résultats de cette expérience montrent bien que l'analyse par méthode de séparation : « *we show that color, shape, and edge-based features combine to form an effective cost function for differentiating between buildings when used within a framework that performs pairwise matching of sample patches summarizing the motif tile of the repeated facade pattern*²⁷⁸ ». Si cette démarche s'avère efficace lorsqu'un grand nombre d'images du même bâtiment est disponible, elle est plus limitée lorsqu'on ne dispose que de quelques photographies historiques, souvent de la même façade, d'un bâtiment disparu.

La reconnaissance d'images / de formes peut certainement s'adapter à différents besoins. En effet, que ce soit pour la reconnaissance des anomalies sur une radiographie ou simplement pour reconnaître un bâtiment par vue au sol ou aérienne, la reconnaissance

²⁷⁵ *Ibid.*, p. 1598.

²⁷⁶ *Ibid.*, p. 1591. Il ne faut pas oublier non plus les autres bâtiments qui peuvent se retrouver aussi dans la photographie soit en partie, soit entièrement. Voir aussi : Abdelkader Hamadi, « Utilisation du contexte pour l'indexation sémantique ... », p. 3-4.

²⁷⁷ *Ibid.*

²⁷⁸ *Ibid.*, p. 1598.

de formes est utilisée dans divers domaines²⁷⁹. Pour les archives photographiques historiques, il serait donc possible de reconnaître les éléments de base d'une photographie afin de perfectionner l'indexation de cette dernière et, ainsi, améliorer sa diffusion.

3.3 Logiciels existants pour la reconnaissance d'objets

Différentes plateformes et logiciels ont été développées pour rendre ces techniques plus accessibles. Dans leur état actuel, pourraient-ils être utilisés dans les centres d'archives? Bien que la technologie ait connu des avancées importantes, nous verrons que quelques composantes doivent être développées pour obtenir une chaîne de traitement complète pouvant être utilisée en archives.

Un des géants dans l'industrie de l'IA est sans contredit Google. En effet, cette entreprise ayant investi beaucoup d'argent depuis de nombreuses années dans la recherche, l'indexation, la voiture intelligente, la publicité ciblée, la reconnaissance de la parole, etc.

Une grande partie de ces développements sont associés à l'arrivée du web 2.0 au tournant des années 2000. Google va en partie utiliser la création et le partage de contenus par les utilisateurs afin d'améliorer ses compétences en matière de technologies de reconnaissance.

²⁷⁹ D'autres recherches touchent plutôt un sujet bien connu aujourd'hui, la reconnaissance des visages. Voir Shiry Ginosar, *et al.*, « A Century of Portraits: A Visual Historical Record of American High School Yearbooks », *International Conference on Computer Vision*, 7 p. Les chercheurs Shiry Ginosar, Kate Rakelly, Sarah Sachs, Brian Yin et Alexei A. Efros analysent l'importance de la photographie dans la complémentarité des écrits des historiens actuels. Le corpus est constitué de portraits de finissants et ils sont analysés grâce à un système de codage. Toutefois, un élément à prendre en considération concernant ce type de recherche est que la photographie scolaire est, de manière générale, standardisée et définie. Ce qui amène une paramétrie parfaite et donc une facilité dans le repérage pour les ordinateurs.

L'entreprise a su se hisser, au tournant du siècle, au niveau mondial en matière de reconnaissance d'objets. Enregistrée le 4 septembre 1998, son nom fut inspiré par le nombre gogolplex (1 suivi de 100 zéros) car les deux créateurs Larry Page et Sergey Brin désiraient, dès ses débuts, montrer la puissance et la capacité du moteur de recherche. Au fil du temps, Google est devenu un des leaders dans le domaine de la recherche sur le web. Sa place est si importante qu'en 2013, lorsque le système a cessé de fonctionner durant quelques minutes, le trafic web aurait diminué de 40%²⁸⁰.

La base de l'expertise de Google en matière de reconnaissance se retrouve dans son système TensorFlow :

TensorFlow™ is an open source software library for high performance numerical computation. Its flexible architecture allows easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices. Originally developed by researchers and engineers from the Google Brain team within Google's AI organization, it comes with strong support for machine learning and deep learning and the flexible numerical computation core is used across many other scientific domains²⁸¹.

Diffusé en accès libre depuis 2015, TensorFlow est utilisé, entre autres, pour améliorer la vision par ordinateur. La librairie liée au programme est notamment utilisée par Intel²⁸², AirBnb, Twitter, etc²⁸³. Le partage d'expertises permet ainsi d'améliorer les performances des algorithmes. Selon les chercheurs Matthieu Montaigu et Kasra Mansouri: « *The use cases for which TensorFlow is mostly known are image recognition,*

²⁸⁰ Radio-Canada, « 11 choses que vous ignorez peut-être sur Google », *Radio-Canada techno* [En ligne], 4 septembre 2018, sur le site de ici.radio-canada.ca, consulté le 4 décembre 2019, <<https://ici.radio-canada.ca/nouvelle/1121706/google-20-ans-histoire-larry-page-sergey-brin>>.

²⁸¹ TensorFlow, *An end-to-end open source machine learning platform* [site Web], 2019, consulté le 4 décembre 2019, <<https://www.tensorflow.org/>>.

²⁸² Intel Software, *AI Frameworks Frameworks* [site Web], 2019, consulté le 4 décembre 2019, <<https://software.intel.com/en-us/frameworks/tensorflow>>. voir aussi : Vivek Rane et AG Ramesh, « TensorFlow* Optimizations for the Intel® Xeon® Scalable Processor », *Intel AI Blog* [En ligne], 3 août 2018, sur le site intel.ai, consulté le 4 décembre 2019, <<https://www.intel.ai/tensorflow-optimizations-intel-xeon-scalable-processor/#gs.CxZXtywx>>.

²⁸³ Stephane Nachez, « Classifier des images avec TensorFlow ». Voir aussi : TensorFlow, *Case studies* [site Web], 2019, consulté le 11 décembre 2019, <<https://www.tensorflow.org/about/case-studies>>.

*natural language processing and speech to text analysis*²⁸⁴». Dans le cas de la détection d'objets, plusieurs niveaux sont disponibles grâce au modèle proposé par TensorFlow. En effet, « *Image classification models are designed to extract features from the whole image and classify the image as a whole*²⁸⁵».

Bref, Google a rapidement su mettre de l'avant la technologie générée grâce à l'IA. Avec le libre droit d'usage de l'algorithme TensorFlow et l'élaboration d'un programme de reconnaissance d'image adapté pour les photographies d'archives, nous serions en mesure de numériser et d'indexer efficacement les documents. Le programme proposé par Google est une bonne base pour qui veut se lancer dans l'aventure de la reconnaissance d'image. TensorFlow offre plusieurs possibilités dont déjà un corpus analysé et défini par le système.

Il est certain que les technologies de reconnaissance de formes évoluent aussi rapidement que les heures sur notre cadran. Toutefois, il est important de prendre conscience des possibilités qui nous sont proposées en cette période X pour le consommateur. Bien sûr, l'algorithme de Google est une option très intéressante en matière de performance en IA. Toutefois, nous nous devons de connaître quelles seraient les autres possibilités pour un centre d'archives régionales qui veut améliorer ou créer un système numérique.

²⁸⁴ Matthieu Montaigu et Kasra Mansouri, « How to use TensorFlow and its related open source contributions to build a tailor-made product recognition model », *Artefact* [En ligne], octobre 2018, sur le site *artefact.com*, consulté le 4 décembre 2019, < <https://www.artefact.com/news/tensorflow-and-open-source-made-product-recognition-model/> >.

²⁸⁵ *Ibid.* Pour en savoir plus: Bharath Ramsundar, ..., *TensorFlow pour le deep learning*, p. 17. Et Thomas Lebrun, « Classifier des images avec TensorFlow », *INFINITE BLOGS* [En ligne], 8 octobre 2018, sur le site *blogs.infinitesquare.com*, consulté le 15 décembre 2019, < <https://blogs.infinitesquare.com/posts/divers/classifier-des-images-avec-tensorflow> >.

Un grand système tel que TensorFlow peut sembler inaccessible aux plus petites institutions. Toutefois, plusieurs plateformes mettent de l'avant des technologies de reconnaissance qui sont à la portée de tous ou presque. En effet, de nombreuses solutions y sont proposées à petits prix et même gratuitement. Parmi « les 10 meilleurs outils de reconnaissance d'images » en 2019 identifiés par Talkwalker, on constate que l'accent est souvent mis sur le repérage de logos, de visages et d'images similaires à d'autres, question de protéger les marques et les œuvres de création²⁸⁶. L'indexation selon des taxonomies ne fait pas encore partie des préoccupations des développeurs. La technologie est tellement répandue qu'on la retrouve sur les smartphones. Google Lens permet d'identifier un objet ou une image à partir d'une recherche sur internet²⁸⁷, ce qui laisse entrevoir des applications à faible coût pour une institution qui accepterait de s'intégrer à l'écosystème Google.

Un autre géant du logiciel n'est toutefois pas à négliger : Adobe²⁸⁸. Logiciel payant, ce dernier permet, en plus de retoucher et de partager les images, de les classer automatiquement. En effet, « [g]râce à l'apprentissage automatique optimisé par Adobe Sensei, Lightroom est capable d'identifier des personnes et des éléments sur vos photos, pour y associer des mots-clés qui peuvent ensuite faire l'objet de recherches²⁸⁹ ». La démarche s'appuyant sur la folksonomie plutôt que sur la taxonomie, des ajustements seraient requis pour répondre aux besoins des archives.

²⁸⁶ François Georges, « Les 10 meilleurs outils de reconnaissance d'images », *Talkwalker* [En ligne], 10 avril 2019, sur le site [talkwalker.com/fr/blog/](https://www.talkwalker.com/fr/blog/), consulté le 15 décembre 2019, <<https://www.talkwalker.com/fr/blog/outils-reconnaissance-images>>.

²⁸⁷ Google, *Google Lens* [site Web], 2020, consulté le 13 janvier 2020, <<https://lens.google.com/>>.

²⁸⁸ Adobe, *Adobe* [site Web], 2020, consulté le 13 janvier 2020, <https://www.adobe.com/ca_fr/products/photoshop-lightroom.html>.

²⁸⁹ *Ibid.*

Une dernière application pour la reconnaissance d'images mérite qu'on s'y attarde : Image Analysis Toolset²⁹⁰. Lancée en 2020, alors que terminait la rédaction de ce mémoire, cette application gratuite propose plusieurs possibilités d'intérêt dont l'identification des éléments principaux de l'image comme les animaux, les objets, les bâtiments, les paysages, etc. Le logiciel permet surtout d'effectuer une indexation automatique et de sortir les résultats en fichiers textes, ce qui est prometteur pour l'intégration dans un catalogue. De plus, afin de répondre aux préoccupations courantes de certains usagers, l'application analyse les images similaires sur le web et peut ainsi repérer les logos ou les visages. Développé par Sylvio Marano à partir des bibliothèques TensorFlow, cette nouvelle boîte à outils démontre le potentiel encore à venir de ces applications, notamment pour les adapter à des fins d'identification et d'indexation archivistique.

3.4 Conclusion

La technologie de reconnaissance de l'image est donc arrivée au point où elle permet de reconnaître la végétation, les bâtiments, les formes humaines, les objets. Nous n'avons toutefois pas réussi à identifier d'études dans lesquelles cette technologie est adaptée aux besoins des centres d'archives, soit de générer des identificateurs qui puissent s'arrimer aux normes de classement, que ce soit la typologie du thésaurus AAT²⁹¹ ou celle du logiciel de classement néerlandais *Iconclass*²⁹². Une telle intégration permettrait d'uniformiser les descriptions des corpus iconographiques des petits centres d'archives et

²⁹⁰ Google Play, *Image Analysis Toolset (IAT)* [site Web], 2020, consulté le 3 juin 2020, <https://play.google.com/store/apps/details?id=tk.silviomarano.imageanalysisistoolset&hl=fr_CA>.

²⁹¹ The Getty Institute, *Art & Architecture Thesaurus® Online* [site Web], 2017, consulté le 3 avril 2018, <<https://www.getty.edu/research/tools/vocabularies/aat/>>.

²⁹² Iconclass, *Home Iconclass* [site Web].

ainsi faciliter l'interopérabilité entre les différents systèmes de classement des différentes institutions. Si le respect des normes n'est pas intégré à la démarche d'indexation, le résultat générera simplement de nouveaux silos plus imposants, mais impossibles à relier.

Le défi d'intégrer l'informatique dans les systèmes de classement des archives est bien présent depuis de nombreuses années²⁹³. En effet, dès les années 2000, plusieurs chercheurs ont remis en question les théories et la pratique archivistique avec l'avancée massive du web. 2.0 dans la sphère du partage d'informations²⁹⁴. En plus du défi d'intégration, il y a aussi le défi de temps et d'avancée scientifique. La rapidité avec laquelle évolue la technologie liée à l'apprentissage profond montre à quel point une technologie révolutionnaire en 2018 peut devenir désuète en 2019²⁹⁵. Les principes de base de reconnaissance des images sont cependant bien établis et le défi pour les centres d'archives sera de miser sur une technologie qui permettra de poursuivre le développement en fonction de ses besoins plutôt que de résoudre un problème précis tel que le repérage de propriété intellectuelle (logos, œuvres d'art).

²⁹³ Carol Couture, « Le concept de document d'archives ... », p. 3. Voir aussi : Carol Couture et Marcel Lajeunesse, *L'archivistique à l'ère du numérique. ...*, p. 185 à 250.

²⁹⁴ Carol Couture et Marcel Lajeunesse, *L'archivistique à l'ère du numérique...*, p. 187.

²⁹⁵ Bharath Ramsundar, ..., *TensorFlow pour le deep learning*, p. 18.

Conclusion : de la reconnaissance à l'indexation

L'amélioration de l'indexation grâce à la reconnaissance des images est reliée aux différents enjeux et défis entourant l'intégration du numérique dans les centres d'archives régionaux. En effet, comme le mentionne Sophie Roy : « Plusieurs fonctions archivistiques doivent être adaptées pour prendre en compte ces nouveaux environnements numériques en réseau²⁹⁶ ». Que ce soit par l'analyse, la planification, l'organisation, l'acquisition, le traitement ou même la diffusion, chacune de ces fonctions évolue au fur et à mesure que les technologies entrent dans notre quotidien²⁹⁷. Une de celles qui constitue une part importante dans ce modèle d'évolution est sans contredit la numérisation. Permettant autant la diffusion que la préservation des originaux, elle est la porte d'entrée de l'archive dans le monde numérique²⁹⁸. L'ajout des technologies de reconnaissance de formes et d'indexation automatisée à cette étape favorise une augmentation plus rapide de leur diffusion, mais aussi, améliore la démocratisation du savoir. Elles établissent une complémentarité entre le travail des archivistes et la mise en ligne des sources. L'intégration d'une liste de mots génériques automatique permet d'effectuer un premier classement hiérarchique des thèmes pour ensuite compléter l'information avec des éléments beaucoup plus spécifiques à l'image.

Bref, après avoir passé en revue les modes actuels d'indexation, les fondements de ces dernières et les logiciels de reconnaissance d'image, nous pouvons maintenant

²⁹⁶ Sophie Roy, « La diffusion d'archives à l'aide des technologies du Web 2.0 ... », p. 1.

²⁹⁷ Association des Archivistes du Québec, *Archivistes ...* Voir aussi : Carol Couture et Marcel Lajeunesse, *L'archivistique à l'ère du numérique. ...*, p. 120 : Grandes fonctions de l'archivistique : création, acquisition, évaluation, classification, description, diffusion et préservation de l'information.

²⁹⁸ Sophie Roy, « La diffusion d'archives à l'aide des technologies du Web 2.0... », p. 1.

répondre à notre question : est-il possible d'intégrer ces technologies dans le processus de numérisation et d'indexation des archives photographiques?

La reconnaissance de formes consiste à reconnaître et classer les formes transmises à l'aide de données brutes et de les catégoriser en fonction des données préalablement entrées dans l'ordinateur²⁹⁹. Pour une analyse d'image historique de paysage urbain quelconque, la technologie actuelle serait selon nous adéquate. En effet, si le programme est déjà entraîné (par apprentissage automatique) avec des données d'indexation de bases qui sont utilisées par l'archiviste, le programme pourra donc générer une série de concepts généraux associées à l'image. Il faut préciser dans le cas de cet exemple qu'il y a de fortes probabilités que l'archiviste reçoive une seule photographie d'un lieu donné. À défaut d'un jeu de données d'entraînement pour ce lieu, le logiciel ne pourra qu'analyser son aspect général (contenu global de la photo) et ce sera à l'archiviste d'ajouter plus de précision (identification du lieu, des gens, etc.). L'organisation des images à l'aide d'une indexation adéquate et automatisée maximise et facilite ainsi le travail des archivistes en générant des données qui peuvent s'inscrire dans des systèmes taxonomiques.

Nous avons déjà pu observer que les modes d'indexation des images en usage dans différentes institutions sont très génériques ou très pointus, le travail humain ne permettant pas de traiter de grands corpus de manière uniforme. Un classement hiérarchique et sémantique est important pour retrouver l'image et la diffuser ou bien pour l'indexer correctement dans la base de données en ligne. De ce fait, l'intégration d'un premier niveau d'indexation à la suite de la numérisation permet de suivre adéquatement des règles

²⁹⁹ Bharath Ramsundar, ..., *TensorFlow pour le deep learning*, p. 132.

définies et uniformes pour chaque image numérisée. Quand il est question de règles définies, l'AAT défini par l'Institut Getty ou aux travaux du RCIP pour les institutions canadiennes viennent orienter la gestion des collections muséales³⁰⁰ et à standardiser le premier niveau d'indexation. Les logiciels effectuant un premier niveau de saisie permettent de donner un regard uniforme autant pour les chercheurs que pour les archivistes. L'intégration d'un programme de reconnaissance et d'un logiciel d'indexation éviterait des problèmes de repérage³⁰¹. Nous parlons ici bien sûr du classement dans les bases de données en ligne de l'institution. Avec un classement adéquat, il sera plus facile pour un chercheur de repérer, donc de retrouver les images qu'il recherche. Une mauvaise indexation peut mener à des recherches infructueuses, notamment à ne pas découvrir des images pertinentes. Il devient donc important que le choix des mots soit uniforme, exhaustif et constant afin de s'adapter aux besoins des usagers et des archivistes.

Toutefois, il est bon de voir l'envers de la médaille quand il est question de nouvelles technologies. En effet, cette étape ne peut être unique dans le processus de numérisation et d'indexation. Ce premier niveau peut générer des mots généraux comme rue, bâtiment, arbre, etc. Il ne peut donner les noms des rues, des bâtiments, des commerces, des personnes, etc., à moins d'être configuré spécifiquement pour les caractéristiques d'un milieu donné. L'intervention humaine reste donc nécessaire afin de s'assurer une indexation précise et plus détaillée. De plus, les changements ainsi que les

³⁰⁰ RCIP : Réseau canadien d'information sur le patrimoine. Dans Gouvernement du Canada, *Réseau canadien d'information sur le patrimoine* [site Web], 22 janvier 2020, consulté le 9 février 2020, <<https://www.canada.ca/fr/reseau-information-patrimoine.html>>. Le RCIP est un organisme issu du Ministère du Patrimoine canadien. Leur rôle est « [d'appuyer] les musées canadiens dans leur travail de documentation et de gestion de leurs collections, ainsi que dans la diffusion de l'information sur leurs collections afin que ces renseignements soient accessibles aujourd'hui et dans l'avenir ».

³⁰¹ James M. Turner, *Images en mouvement*, p. x.

mises à jour toujours constants en matière de technologie demandent des investissements à long terme pour les centres. Il devient alors très important de rester connecté aux nouveautés et de toujours suivre et d'entretenir la plateforme pour éviter qu'elle ne devienne obsolète. Finalement, il ne faut pas non plus oublier que même si l'aspect humain dans le processus d'identification demeure d'une grande importance, l'utilisation d'un programme permet, pour sa part, de traiter un plus grand corpus d'images avec un aspect plus uniforme. En traitant plus rapidement les images, il devient donc possible d'accélérer le processus de mise en ligne et donc, d'augmenter la diffusion des ressources.

Vers une nouvelle chaîne de traitement³⁰²

Il faut savoir que l'indexation n'est pas uniforme d'une base de données à l'autre ainsi que d'une institution à l'autre. En effet, comme nous avons vu précédemment, chaque base de données, qu'elle soit d'une grande ou d'une petite institution, répond aux besoins qui sont propre au milieu. Le choix des éléments mis de l'avant dans les fiches descriptives découle des politiques de l'institution. Leurs chaînes de traitement comportent néanmoins des points communs. Il y a l'acquisition, le tri des documents (choisir ceux que l'on conserve ou pas), établir la cote, compléter les descriptions, numériser, effectuer la mise en enveloppe, ranger le document, mettre en ligne et diffuser³⁰³. Dans un même fonds d'archives, d'une institution à l'autre, plusieurs techniciens, dont des stagiaires, peuvent s'y attarder au fil des traitements donnant des résultats très variés dans les termes génériques employés. Avec l'ajout de la technologie,

³⁰² Voir ANNEXE I.

³⁰³ Tiré du Guide pour le traitement des archives iconographiques de la Société d'histoire de Sherbrooke.

peu importe qui effectuera la chaîne de traitement, les premiers échelons de la taxonomie sera plus constante et mènera à une standardisation des données.

Le processus de numérisation des archives photographiques a maintenant atteint un niveau qui répond aux bonnes pratiques dans la plupart des institutions. Les fichiers qui découlent de cette étape, incluant les métadonnées de base (fonds, cote, date, titre) permettent alors de passer au traitement des images par le logiciel de reconnaissance de formes et d'indexation³⁰⁴. Les données en sortie seront alors regroupées en termes généraux : rues, arbres, bâtiments, etc. Avec ces regroupements, il est alors possible d'effectuer des traitements en grands lots. En effet, l'archiviste ou le technicien peut créer des regroupements uniformes en fonction des besoins de classement du fonds traité. Par exemple, nous pourrions sortir toutes les photographies avec des personnes dans un même fonds ou bien toutes les images contenant un même édifice. L'ajout de termes spécifiques devient alors plus facile et efficace. Un des avantages de la reconnaissance de forme que nous avons perçu au cours de notre recherche est la quantité de données qu'il est possible d'en extraire. En effet, l'œil humain ne va pas donner la même importance d'un indexeur à l'autre. Par exemple, pour l'un, les éléments du paysage peuvent sembler plus importants que l'événement qu'ils représentent. L'expertise spécifique d'une personne peut mener à un déséquilibre des descripteurs pour deux images semblables traitées par deux personnes différentes. La reconnaissance permet donc de sortir tous les termes généraux d'une image et d'éviter l'oubli d'un concept qui pourrait être important pour un chercheur.

En effectuant un premier regroupement général, il est possible de développer des stratégies de traitement manuel efficaces en fonction des caractéristiques ressorties lors

³⁰⁴ Voir ANNEXE J

de la première analyse automatisée. Une fois l'analyse des regroupements faite, c'est là que l'archiviste va mettre à profit l'expertise humaine réelle et plus poussée. Il s'agit alors d'attribuer des tâches d'identification à des gens spécialisés dans plusieurs domaines. Par exemple, si la caractéristique à identifier sont les bâtiments, les personnes qui ont une connaissance approfondie du domaine pourront alors prendre en charge l'ajout de contenu d'indexation. Prenons à titre d'exemple une image d'un homme politique sherbrookoïse tel que Daniel McManamy. Nous pouvons identifier l'homme en question, mais aussi les différents aspects entourant sa vie. Nous pouvons ajouter maire, conseiller, rue (toponymie), etc. Finalement, quand nous ajoutons des mots d'indexation plus spécifiques, nous pouvons y définir des toponymes significatifs aux lieux, aux bâtiments, aux personnes, etc. Les liens historiques seront plus évidents mais aussi, ils seront plus faciles à comprendre.

Une autre possibilité liée à l'ajout de termes d'indexation serait d'établir une collaboration avec d'autres centres d'archives. En effet, les spécialistes ne se retrouvent pas tous dans le même milieu. Dans une région, plusieurs institutions peuvent faire équipe afin d'établir des partages de connaissances en lien avec les images liées entre elles. En établissant cet appui entre chaque organisation, il devient donc plus facile d'agrandir le corpus de concepts et aussi, les divers liens entre les collections et les fonds d'institutions diverses. Cette collaboration entre centres peut permettre des liens à long terme.

Finalement, afin d'avoir encore plus de précision dans l'indexation de groupe de photos, nous pouvons prendre en compte la folksonomie, soit le rôle du public. Demander de l'aide au public permet de créer des liens entre la communauté et l'institution. Avec l'aide de passionnés, il sera possible d'obtenir ce que nous appelons de la toponymie

informelle, soit le nom de lieux-dits désignés de manière plus populaire. Étant des connaissances informelles, elles peuvent apporter des connaissances anecdotiques qui vont renforcer l'identification de l'image, son contenu historique mais aussi, la rendre plus accessible à un public élargi³⁰⁵. Grâce à l'apport du public, nous pouvons aussi découvrir des bâtiments, des lieux ou même des personnages moins connus dans l'histoire.

Malgré les nombreux avantages qu'apporte l'ajout de la folksonomie au processus d'indexation, il faut garder en tête qu'elle n'est pas parfaite. En effet, l'intégration du public dans la chaîne de traitement est un domaine en soi. La plupart du temps, ce sont des passionnés et non des spécialistes, il faut donc prendre certaines identifications avec parcimonie et toujours vérifier ce qui est ajouté. Il ne faut pas non plus oublier la surabondance de données qui peut être produite si plusieurs répondent à l'appel. Bref, si l'institution décide d'entreprendre les démarches vers le chemin de la collaboration avec le public, elle se doit de le faire adéquatement et y investir les ressources nécessaires à son succès.

Nous venons de voir l'apport considérable que peut apporter l'ajout des technologies de reconnaissance de formes et d'indexation dans le processus de numérisation pour les centres d'archives. Nous pouvons conclure à ce titre que l'ajout des technologies faciliterait le travail des archivistes et des techniciens dans le processus de numérisation des images. Les retombées à long terme de ces ajouts peuvent devenir très nombreux pour les chercheurs qui utiliseront la base de données par la suite. Nous nous

³⁰⁵ Sophie Roy, « La diffusion d'archives à l'aide des technologies du Web 2.0 ... », p. 2.

posons donc la question : qu'en est-il pour les chercheurs qui utiliseront les résultats de ces nouvelles démarches?

L'ajout de précision dans l'identification des ressources apporte bon nombre d'avantages pour un chercheur qui utilise la plateforme. En effet, elles permettent d'établir un modèle hiérarchique des connaissances. Nous pouvons donc, en tant que chercheurs, choisir de faire les recherches du plus large au plus spécifique. Trouver de l'anecdotique avec l'apport de la folksonomie, mais aussi de l'information formelle ajoutée par des spécialistes. Nous pouvons aussi donner notre apport en ajoutant des mots d'indexation qui pourraient avoir une pertinence pour le centre. Une identification plus approfondie permet de trouver des liens qui peuvent être inconnus pour le chercheur lorsqu'il effectue sa recherche.

L'avenir de la recherche numérique

Nous croyons avoir démontré qu'une standardisation de l'indexation permet une uniformisation des contenus. Un catalogage précis et standardisé dans les institutions permet aux chercheurs d'effectuer leurs recherches de manière autonome. En effet, avec les liens normalisés qui se créent entre les centres d'archives, il devient plus facile d'établir des requêtes transversales dans plusieurs collections, mais aussi d'un centre à l'autre. Un autre avantage à long terme est sans contredit la diminution du temps d'apprentissage d'une plateforme à l'autre.

Bien sûr, les centres d'archives régionaux n'ont pas besoin d'attendre que les grands centres emboîtent le pas en matière d'intégration du numérique. En effet, plusieurs possibilités s'offrent à eux. Il y a notamment des programmes de numérisation mis sur pied qui favorisent la création de consortiums entre plusieurs petits centres régionaux. Ces

derniers peuvent amener à traiter un grand volume de documents, mais aussi à faciliter les requêtes recouvrant plusieurs dépôts en même temps.

La création des outils de consultation informatisés s'est souvent faite sans l'apport des historiens. Notre bibliographie, dans laquelle on retrouve peu de publications par des historiens, reflète cette situation. Pendant que nous avons travaillé à ce mémoire, des spécialistes de la bibliothéconomie ont réalisé de nouvelles recherches, dont Laure Amélie Guitard³⁰⁶ et Jean-Philippe Moreux³⁰⁷. La première propose d'améliorer l'accès thématique aux archives patrimoniales par l'automatisation de l'indexation. Elle compare entre autres les termes employés par les archivistes pour indexer les documents et ceux employés par les usagers dans leurs recherches. Le second relève les mêmes problématiques que nous avons souligné par rapport au manque d'uniformité dans l'indexation des documents et propose aussi de valoriser les ressources iconographiques par l'apport des technologies d'intelligence artificielle. Au-delà des enjeux très concrets que nous avons soulevé dans ce mémoire, les historiens doivent participer aux échanges en cours. Nous vivons dans un monde numérique en constante évolution et il est important d'agir, de s'informer et de voir les possibilités pour ainsi aller de l'avant.

³⁰⁶ Laure Amélie Guitard, « Vocabulaire employé pour l'accès thématique aux documents d'archives patrimoniales : étude linguistique exploratoire de termes de recherche, de description, d'indexation », Thèse de doctorat (science de l'information), Montréal, Université de Montréal, 2018, 328 p.

³⁰⁷ Jean-Philippe Moreux. « Recherche d'images dans les bibliothèques numériques patrimoniales et expérimentation de techniques d'apprentissage profond », *Documentation et bibliothèques*, vol. 65, n°2, avril-juin 2019, p.5–27.

BIBLIOGRAPHIE

Ouvrages :

- BONI, Alexandre, et Nicolas STEMART. *Google, Le guide complet*. Paris, Éditions Micro Application, 2008, 508 p.
- BOLON, Philippe, *et al.* *Analyse d'images : Filtrage et segmentation. Enseignement de la physique*. France, Éditions MASSON, 1995, 447 p.
- Comité de planification sur les normes de description. *Règles pour la description des documents d'archives*. Ottawa, Bureau canadien des archivistes, 2008, 2^e éd. (1990), 737 p.
- CHARBONNEAU, Normand, et Mario ROBERT dirs. *La gestion des archives photographiques*. Sainte-Foy, Presses de l'Université du Québec, 2011, 306 p.
- COUTURE, Carol, et Marcel LAJEUNESSE. *L'archiviste à l'ère du numérique, Les éléments fondamentaux de la discipline*. Québec, Presses de l'Université du Québec, 2014, 278 p.
- COUTURE, Carol, et Jean-Yves ROUSSEAU. *Les archives au XXe siècle. Une réponse aux besoins de l'administration et de la recherche*. Montréal, Université de Montréal, Secrétariat général et Service des archives, 1982, VI-491 p.
- CHALLONER, Jack. *L'intelligence artificielle, un guide d'initiation au futur de l'informatique et de la robotique*. Paris, Éditions Pearson Éducation France, Focus Sciences, 2003, 72 p.
- CORNUÉJOLS, Antoine, et Laurent MICLET. *Apprentissage artificielle. Concepts et algorithmes*. Paris, Éditions Eyrolles, 2010, 2^e éd. (2003), 804 p.
- Conseil international des archives. *Déclaration universelle sur les archives*. Paris, 36e sessions de la conférence générale de l'UNESCO, 10 novembre 2011, 1 p.
- CHAUMIER, Jacques. *Travail et méthodes du documentaliste, connaissance du problème, applications pratiques*. Paris, ESF, 1996, 5e éd. 106-54 p.
- DE MIRANDA, Luis dir. *Intelligence artificielle et robotique en 30 secondes, 50 avancées majeures, expliquées en moins d'une minute*. Montréal, Édition Hurtubise, 2019, 159 p.
- DELPORTE, Christian, Laurent GERVEREAU et Denis MARÉCHAL, dirs. *Quelle est la place des images en histoire?* Paris, Éditions Nouveau monde, 2008, 480 p.
- FERRO, Marc. *Cinéma, une vision de l'histoire*. France, Éditions Du Chêne, 2003, 168 p.
- GERVEREAU, Laurent. *Les images qui mentent, histoire du visuel au XXe siècle*. Paris, Éditions du Seuil, 2000, 458 p.
- GANASCIA, Jean-Gabriel. *Intelligence artificielle vers une domination programmée?* Paris, Éditions Le Cavalier Bleu, 2017, 2^e éd. (2007), 215 p.
- GOODFELLOW, Ian, Yoshua BENGIO et Aaron COURVILLE. *L'apprentissage profond*. France, Éditions Massot et Quantmetry, 2018, 768 p.

HASKELL, Francis. *L'historien et les images*. Trad. de l'anglais par Alain Tachet et Louis Évrard. Paris, Gallimard, 1995, 781 p.

HUDON, Michèle. *Le Thésaurus : Conception, élaboration, gestion*. Montréal, Les Éditions ASTED, 1994, 220 p.

KESTEMAN, Jean-Pierre, 2000-2002 - *Histoire de Sherbrooke*, Sherbrooke, GGC Éditions, 4 vol. (coll. Patrimoine), t. 1 : *De l'âge de l'eau à l'ère de la vapeur, 1802-1866*, 2000, 353 p. / t. 2 : *De l'âge de la vapeur à l'ère de l'électricité, 1867-1896*, 2001, 280 p. / t. 3 : *La ville de l'électricité et du tramway, 1897-1929*, 2002, 292 p. / t. 4 : *De la ville ouvrière à la métropole universitaire, 1930-2002*, 2002, 486 p.

KESTEMAN, Jean-Pierre. *Guide historique du vieux Sherbrooke*. Sherbrooke, Société d'histoire de Sherbrooke, 2001, 2^e éd. (1985), 269 p.

LE DEUFF, Olivier. *Du tag au like, La pratique des folksonomies pour améliorer ses méthodes d'organisation de l'information*. France, Éditions FYP, 2012, 159 p.

M. TURNER, James. *Images en mouvement, stockage. Repérage. Indexation*. Sainte-Foy, Presses de l'Université du Québec, 1998, 102 p.

MILLIGAN, Ian. *History in the age of abundance? How the web is transforming historical research*. Montreal & Kingston, McGill-Queen's University Press, 2019, 310 p.

PETERSEN, Toni, dir. *Art & Architecture Thesaurus*. New York, Oxford University Press, 1994, 2^e éd. (1990), 3182 p.

RAMSUNDAR, Bharath, et Reza BOSAGH ZADEH. *TensorFlow pour le deep learning, La regression linéaire à l'apprentissage par renforcement*. Paris, Éditions First interactive, 2018, 245 p.

SHS. *Guide pour le traitement des archives iconographiques*. Sherbrooke, SHS, 1987-1988 (révisé en 2011), 45 p.

Articles périodiques et journaux :

ALMEIDA, Rodrigo, *et al.*. « Interactions et métadonnées riches pour les bibliothèques numérisées ». *Document numérique*, vol. 9, n° 2 (2006), p. 83-109.

ALAIN, Jean-Marc. « La convergence : enjeux et défis ». *Documentation et bibliothèques*, vol. 47, n° 2 (2001), p.71-74.

ABOUT, Ilsen, et Clément CHÉROUX (2002, novembre). « L'histoire par la photographie ». *Études photographiques* [En ligne]. N° 10, sur le site *études photographiques*. Consulté le 8 mars 2017. <http://etudesphotographiques.revues.org/261>

AUBIN, Danielle. « La mondialisation et la diffusion des archives : entre continuité et rupture ». *Archives*, vol. 31, n° 3 (1999-2000), p. 7-19.

BORGÈS DA SILVA, Roxane. « Taxonomie et typologie : comment rapprocher les cadres conceptuels des différentes écoles de pensée ? ». *Santé Publique*, vol. 26, n°1 (2014), p. 87-88.

- BORGÈS DA SILVA, Roxane. « Taxonomie et typologie : est-ce vraiment des synonymes ? ». *Santé Publique*, vol. 25, n°5 (2013), p. 633-637.
- BESSON, Rémy. « Regard intermédial sur le devenir numérique des archives ». *Intermédialités: histoire et théorie des arts, des lettres et des techniques*, n°23 (2014), 18 para.
- BLAIR, Rolande. « La gestion des documents électroniques : le point de vue de la Bibliothèque nationale du Canada ». *Documentation et bibliothèques*, vol. 42, n° 3 (1996), p. 127-133.
- BOYER, Deborah, Robert CHEETHAM et Mary JOHNSON. « Using GIS to Manage Philadelphia's Archival Photographs ». *The American Archivist*, vol. 74, n° 2 (2011), p. 652-663. Disponible en ligne: <http://dx.doi.org/10.17723/aarc.74.2.028511q0360x327r>
- BELLVER, Julien (10 Avril 2012). « Facebook se paye Instagram pour 1 milliard de dollars ». *Pure média* [En ligne]. Sur le site *oxap.com*. Consulté le 20 avril 2019. <https://www.oxap.com/actu/facebook-se-paye-instagram-pour-1-milliard-d-euros/440310>
- BIERSDORFER, J. D. (21 juin 2018). « Moving On From Picasa ». *The New York Times* [En ligne]. Sur le site *nytimes.com*. Consulté le 20 avril 2019. <https://www.nytimes.com/2018/06/21/technology/personaltech/picasa-replace-photos.html>
- BROUSSEAU, François. « La victoire de la ferraille Deep Blue remporte le match contre Kasparov ». *Le Devoir* (Montréal), vol. LXXXVIII, n°104 (12 mai 1997), p. A1 et A8.
- Comité de planification sur les normes de description. « L'indexation en archivistique, rapport du groupe de travail sur l'indexation par sujet ». *Bureau canadien des archivistes*, n° 4 (1992), 154 p.
- CHAMPEAU, Guillaume (23 septembre 2009). « Picasa reconnaît les visages et tag automatiquement vos photos ». *Numerama* [En ligne]. Sur le site *numerama.com*. Consulté le 20 avril 2019. <https://www.numerama.com/magazine/14036-picasa-reconnait-les-visages-et-tag-automatiquement-vos-photos.html>
- CHARBONNEAU, Normand, et Mario ROBERT. « Indexation ». Dans Normand Charbonneau et Mario Robert dirs., *La gestion des archives photographiques*, Sainte-Foy, Presses de l'Université du Québec, 2011, p. 153-173.
- COUTURE, Carol. « Le concept de document d'archives à l'aube du troisième millénaire ». *Archives*, vol. 27, n° 4 (1996), p. 3-19.
- CLOSE, Susan. « Reading Photographs: Playing the Visual Literacy Game ». *CV Photo* 52, n° 52 (2000), p. 5-6.
- COTTE, Michel. « Les outils numériques au service de l'histoire des techniques ». *Revue d'histoire des techniques*, vol. 1, n° 2 (2012), p. 12-27.
- CHARLIN, Laurent. « Intelligence artificielle : une mine d'or pour les entreprises ». *HEC MTL Gestion*, vol. 42, n° 1 (2017), p. 76-79.

- Collège de France (2018). « Yann LeCun. Recherche sur l'intelligence artificielle ». *Collège de France*. Sur le site du *Collège de France* [En ligne]. Consulté le 22 janvier 2018. https://www.college-de-france.fr/media/yann-lecun/UPL4485925235409209505_Intelligence_Artificielle_Y._LeCun.pdf
- DASCAL, Marcelo. « Culture numérique. Enjeux pragmatiques et philosophiques ». *Diogène*, vol. 3, n° 211 (2005), p. 26-47.
- DURETZ, Marlène (1 avril 2013). « Phonéographie. Qui "instagramme" ses photos et quels sont les sujets fédérateurs de cette application ? ». *Le monde* [En ligne]. Sur le site *Lemonde.fr*. Consulté le 20 avril 2019. https://www.lemonde.fr/vous/article/2013/04/01/phoneographie_3151452_3238.html
- Facebook (9 avril 2012). « Facebook to Acquire Instagram ». *Facebook Newsroom* [En ligne]. Sur le site *newsroom.fb.com*. Consulté le 20 avril 2019. <https://newsroom.fb.com/news/2012/04/facebook-to-acquire-instagram/>
- GAGNON-ARGUIN, Louise. « Études d'usagers ». Dans Normand Charbonneau et Mario Robert dirs., *La gestion des archives photographiques*, Sainte-Foy, Presses de l'Université du Québec, 2011, p.221-235.
- GUITARD, Laure Amélie. « Indexation par sujet en archivistique et en bibliothéconomie : du pareil au même? ». *Documentation et bibliothèques*, vol. 59, n° 4 (octobre-décembre 2013), p. 201-212.
- GERON, Tomio (2014). « A Look Back At Yahoo's Flickr Acquisition For Lessons Today ». *Techcrunch* [En ligne]. Sur le site *techcrunch.com*. Consulté le 10 avril 2019. <http://tcrn.ch/1qC8IPe>
- Google (2016). « Qu'advient-il de Picasa, de Picasa Albums Web et de l'API Picasa Albums Web? ». *Support google* [En ligne]. Sur le site *support.google.com*. Consulté le 20 avril 2019. <https://support.google.com/picasa/answer/6383491?hl=fr>
- Google (12 février 2016). « Moving on from Picasa ». *Picasa Blog* [En ligne]. Sur le site *googlephotos.blogspot.com*. Consulté le 20 avril 2019. <http://googlephotos.blogspot.com/2016/02/moving-on-from-picasa.html>
- GEORGES, François (10 avril 2019). « Les 10 meilleurs outils de reconnaissance d'images ». *Talkwalker* [En ligne]. Sur le site *talkwalker.com/fr/blog/*. Consulté le 15 décembre 2019. <https://www.talkwalker.com/fr/blog/outils-reconnaissance-images>
- GREVELINK, Evelyn (18 décembre 2017). « A Closer Look at Object Detection, Recognition and Tracking ». *Intel Software, Developer zone* [En ligne]. Sur le site *software.intel.com*. Consulté le 1 octobre 2019. <https://software.intel.com/en-us/articles/a-closer-look-at-object-detection-recognition-and-tracking>
- HEIMBURGER, Franziska, et Émilien RUIZ. « Faire de l'histoire à l'ère numérique : retours d'expériences ». *Revue d'histoire moderne et contemporaine*, vol. 5, n° 58 (2011), p. 70-89.
- HUDON, Michèle. « Indexation et langages documentaires dans les milieux archivistes à l'ère des nouvelles technologies de l'information ». *Archives*, vol. 29, n°1 (1997-1998), p. 75-98.

HILL, Simon (16 juin 2015). « Google Photos: Should You Be Worried about Privacy? ». *Android Authority*. [En ligne]. Sur le site *androidauthority*. Consulté le 28 mai 2019. <https://www.androidauthority.com/google-photos-worried-privacy-616339/>

HUANG, Thomas S. « Computer Vision: Evolution and Promise ». Dans Carlo E. Vandoni, éd., *19th CERN School of Computing (conference), Organisation européenne pour la recherche nucléaire, CERN european organization for nuclear research*, Egmond aan Zee, The Netherlands, Éditions Geneva, n°8 (8-21 septembre 1996), p. 21-25.

LEMAY, Yvon, et Anne KLEIN. « La diffusion des archives ou les 12 travaux des archivistes à l'ère du numérique ». *Les Cahiers du numérique*, vol. 8, n° 3 (2012), p. 15-48.

LE DEUFF, Olivier (2006). « Folksonomies. Les usagers indexent le web ». *Bulletin des bibliothèques de France (BBF)*. N° 4, sur le site *BBF*. Consulté le 28 février 2018. <http://bbf.enssib.fr/consulter/bbf-2006-04-0066-002>

LÉVESQUE, Michel. « L'indexation: Luxe ou nécessité? ». *Archives*, vol. 33, n° 1 (2001-2002), p. 17 à 45.

LAUSSON, Julien (15 février 2016). « Google tire un trait sur Picasa, son service d'albums photos ». *Numerama* [En ligne]. Sur le site *numerama.com*. Consulté le 20 avril 2019. <https://www.numerama.com/tech/145584-google-tire-un-trait-sur-picasa-son-service-dalbums-photos.html>

LECUN, Yann, Yoshua BENGIO et Geoffrey HINTON. « L'apprentissage en profondeur ». *Nature*, vol. 521 (28 mai 2015), p. 436 – 444.

LEBRUN, Thomas (8 octobre 2018). « Classifier des images avec TensorFlow ». *INFINITE BLOGS* [En ligne]. Sur le site *blogs.infinitesquare.com*. Consulté le 15 décembre 2019. <https://blogs.infinitesquare.com/posts/divers/classifier-des-images-avec-tensorflow>

LANGÉVIN, Sébastien (24 avril 2018). « L'intelligence artificielle au service de l'imagerie médicale ». *Actualités du réseau de l'Université du Québec* [En ligne]. Sur le site de *l'Université du Québec*. Consulté le 27 juillet 2019. <https://www.quebec.ca/reseau/fr/medias/actualites-du-reseau/lintelligence-artificielle-au-service-de-limagerie-medicale>

LEE, June-Goo, *et al.* « Deep Learning in Medical Imaging: General Overview ». *Korean Journal of Radiology*, n° 4 (2017), p. 570-584.

MOREUX, Jean-Philippe. « Recherche d'images dans les bibliothèques numériques patrimoniales et expérimentation de techniques d'apprentissage profond ». *Documentation et bibliothèques*, vol. 65, n°2 (avril-juin 2019), p.5–27.

MONTAIGU, Matthieu, et Kasra MANSOURI (2018, octobre). « How to use TensorFlow and its related open source contributions to build a tailor-made product recognition model ». *Artefact* [En ligne]. Sur le site *artefact.com*. Consulté le 4 décembre 2019. <https://www.artefact.com/news/tensorflow-and-open-source-made-product-recognition-model/>

NACHEZ, Stéphane (2 juillet 2017). « Classifier des images avec TensorFlow ». *ActuIA* [En ligne]. Sur le site *actuia.com*. Consulté le 1 octobre 2019.

<https://www.actuia.com/tutoriel/classifier-images-tensorflow/>

PIROLI, Fabrice. « Pratiques d'indexation sociale et démarches de veille informationnelle ». *Études de communication*, n° 36 (2011), p. 53-66. Disponible en ligne :

https://journals.openedition.org/edc/2615#xd_co_f=MmI5YWJIOTdkMjY1MTVIM2FhMzE1NTgxMjI1MTg2Njg=~

Philadelphie (AFP). « Garry Kasparov remporte la 5 e partie contre « Deep Blue » ». *Le Devoir* (Montréal), vol. LXXXVI, n°40 (17 février 1996), p. A5.

RONCERO, Leticia (14 décembre 2018). « Photography and Licensing: Q&A with Creative Commons CEO Ryan Merkley ». *Flickr blog* [En ligne]. Sur le site *blog.flickr*. Consulté le 15 avril 2019.

<https://blog.flickr.net/en/2018/12/14/photography-and-licensing-qa-with-creative-commons-ceo-ryan-merkley/>

RONCERO, Leticia (16 janvier 2019). « Celebrating 11 years of The Commons ». *Flickr blog* [En ligne]. Sur le site *blog.flickr*. Consulté le 15 avril 2019.

<http://blog.flickr.net/en/2019/01/16/celebrating-the-11th-anniversary-of-flickr-commons/>

RANE, Vivek, et AG RAMESH (3 août 2018). « TensorFlow* Optimizations for the Intel® Xeon® Scalable Processor ». *Intel AI Blog* [En ligne]. Sur le site *intel.ai*.

Consulté le 4 décembre 2019. <https://www.intel.ai/tensorflow-optimizations-intel-xeon-scalable-processor/#gs.CxZXtywx>

Radio-Canada (4 septembre 2018). « 11 choses que vous ignorez peut-être sur Google ». *Radio-Canada techno* [En ligne]. Sur le site de *ici.radio-canada.ca*. Consulté le 4

décembre 2019. <https://ici.radio-canada.ca/nouvelle/1121706/google-20-ans-histoire-larry-page-sergey-brin>

ROTH, Matthew (20 avril 2018). « Together is where photographers belong! ». *Flickr Blog* [En ligne]. Sur le site *blog.flickr*. Consulté le 15 avril 2019.

<http://blog.flickr.net/en/2018/04/20/together-smugmug-flickr/>

SCALLON-CHOUINARD, Pascal (5 juin 2013). « Développer les sciences humaines numériques au Québec ». *Histoire Engagée* [En ligne]. Sur le site *Histoireengage.ca*.

Consulté le 20 janvier 2017. <http://histoireengagee.ca/?p=2899>

SÉCAIL-TRAQUES, Claire. « Image, représentations, mémoire de l'événement ». *Vingtième Siècle, revue d'histoire*, n°72 (2001), p. 136-137.

SAINTÔT, Bruno. « Quelques critères d'utilisation de l'imagerie médicale, conclusion du colloque ». Tome 60 de *Laennec*, n°4 (2012), p. 47-50.

Substance (21 octobre 2016). « L'intelligence artificielle pour interpréter les images médicales ». *Article de recherche, technologies de la santé* [En ligne]. Sur le site *Substance, Actualité scientifique et innovation de l'ÉTS*. Consulté le 2 mars 2018.

<https://substance.etsmtl.ca/reconnaissance-interpretation-images-medicales>

TONY, Gill (2008). « Metadata and the Web, In Introduction to Metadata: Pathways to Digital Information, edited by Murtha Baca ». *Online Edition (Version 3.0)* [En ligne]. Sur le site *Getty éducation*. Consulté le 20 janvier 2017.

http://www.getty.edu/research/publications/electronic_publications/intrometadata/metadata.html

VALIQUETTE, Marie-Anne (6 février 2018). « L'intelligence artificielle et l'imagerie médicale ». *Article de recherche, technologies de la santé*, sur le site *Substance* [En ligne]. Sur le site *Actualité scientifique et innovation de l'ÉTS*. Consulté le 2 mars 2018. <https://substance.etsmtl.ca/lintelligence-artificielle-imagerie-medecale#respond>

« Text Retrieval with New Wrinkles : Documensa's Edibase SGML ». *XML.COM* [En ligne]. 10 janvier 1998, sur le site *XML.com*. Consulté le 13 décembre 2019.

<https://www.xml.com/pub/a/SeyboldReport/ip020516.html>

Mémoires et thèses :

DOLZ, Jose. « Vers la segmentation automatique des organes à risque dans le contexte de la prise en charge des tumeurs cérébrales par l'application des technologies de classification de deep learning ». Thèse de doctorat, Université de Lille 2, (Mathématique) 2016, 238 p.

GINOSAR, Shiry, *et al.* « A Century of Portraits: A Visual Historical Record of American High School Yearbooks », *International Conference on Computer Vision*, 7 p.

GUITARD, Laure Amélie. « Vocabulaire employé pour l'accès thématique aux documents d'archives patrimoniaux : étude linguistique exploratoire de termes de recherche, de description, d'indexation ». Thèse de doctorat, Université de Montréal, (science de l'information) 2018, 328 p.

HAMADI, Abdelkader. « Utilisation du contexte pour l'indexation sémantique des images et vidéos ». Thèse de doctorat, Université Joseph Fourier Grenoble -1-, (Informatique) 2014, 179 p.

MICHON, Philippe. « Vers une nouvelle architecture de l'information historique : L'impact du Web sémantique sur l'organisation du Répertoire du patrimoine culturel du Québec ». Mémoire de maîtrise, Université de Sherbrooke, (histoire) 2016, 134 p.

NAWAL, Chifa, et Abdelmajid BADRI. « Combinaison de descripteurs globaux et locaux pour la recherche par le contenu d'informations visuelles ». Département Electrique, Faculté des Sciences et Techniques - Mohammedia, Université de Hassan II de Casablanca (Conférence: La 8ème édition de la rencontre nationale des jeunes Chercheurs en Physique), décembre 2015, 6 p.

ROY, Sophie. « La diffusion d'archives à l'aide des technologies du Web 2.0 : le cas du Musée McCord ». Mémoire de maîtrise, Université de Montréal, (Sciences de l'information) 2012, 158 p.

ROSENBERGER, Christophe. « Mise en œuvre d'un système adaptatif de segmentation d'images ». Thèse de doctorat, Université Rennes 1, (Structure et propriété de la matière) 1999, 170 p.

WOLFF, Mark, Robert T. COLLINS et Yanxi LIU. «Regularity-Driven Building Facade Matching between Aerial and Street Views ». School of Electrical Engineering and Computer Science (Conference on Computer Vision and Pattern Recognition 2016), The Pennsylvania State University, p. 1591.

X EISEN, Erica (15 novembre 2018). « Who's Behind That Beard? Historians are using facial recognition software to identify people in Civil War photographs ». *Slate* [En ligne]. Sur le site *slate.com*. Consulté le 4 novembre 2020.
< <https://slate.com/technology/2018/11/civil-war-photo-sleuth-facial-recognition.html> >.

ZHANG, Wei, et Jana KOSECKA. « Experiments in Building Recognition ». Department of Computer Science George Mason University Technical Report Series, 9 p.

Article dans un acte de Colloque :

CHOMINOT, Marie. « L'image photographique, une source pour écrire l'histoire de la guerre d'Algérie/ guerre de libération ? ». Dans REMAOUN, Hassan, et Mahamed BENSALAD, dirs. *Image, Histoire, Mémoire. Les représentations iconographiques en Algérie et au Maghreb : Actes de colloque international organisé par la CRASC*, Algérie, CRASC, 2007, 153 p.

Encyclopédies, dictionnaires et site web:

Archives Montréal (2018). *La section des archives* [site Web]. Consulté le 25 juillet 2018. <<http://archivesdemontreal.com/la-section-des-archives/>

Adobe (2020). *Adobe* [Site web]. Consulté le 13 janvier 2020.
https://www.adobe.com/ca_fr/products/photoshop-lightroom.html

Association des Archivistes du Québec (2016). *Archivistes* [site Web]. Consulté le 20 février 2019. <https://archivistes.qc.ca/devenir-archiviste/>

Bibliothèque et Archives Nationales du Québec (2019). *BANQ Archives* [site Web]. Consulté le 28 mars 2019.
http://www.banq.qc.ca/archives/archivistique_gestion/services_partenaires/serv_archives_privees_agrees/

Bibliothèques de l'Université Cornell (2000-20003). *De la théorie à la pratique. Didacticiel d'imagerie numérique* [site Web]. Consulté le 1 octobre 2018.
<http://preservationtutorial.library.cornell.edu/tutorial-french/tutorial-french.pdf>

DeepMind (2020). *AlphaGo* [site Web]. Consulté le 26 avril 2020.
<https://deepmind.com/research/case-studies/alphago-the-story-so-far>

Flickr (2019). *Flickr, Site officiel* [site Web]. Consulté le 10 avril 2019.
<https://www.flickr.com/>

Gouvernement du Canada (2020). *Normes principales pour les musées canadiens* [site Web]. Consulté le 1 octobre 2018. <https://www.canada.ca/fr/reseau-information-patrimoine/services/normes-documentation-collections/guide-rcip-normes-musees/normes-principales-musees-canadiens.html>

Gouvernement du Canada (22 janvier 2020). *Réseau canadien d'information sur le patrimoine* [site Web]. Consulté le 9 février 2020. <https://www.canada.ca/fr/reseau-information-patrimoine.html>

Gouvernement du Québec (2013). *Office québécoise de la langue française* [site Web]. Consulté le 30 mars 2019. https://www.oqlf.gouv.qc.ca/ressources/bibliotheque/dictionnaires/terminologie_blogue/folksonomie.html

Google (2019). *Google, Site officiel* [site Web]. Consulté le 21 avril 2019. <https://www.google.com/photos/about/?hl=fr>

Google Cloud (2019). *Produits d'IA et de machine learning* [site Web]. Consulté le 15 décembre 2019. <https://cloud.google.com/vision/?hl=fr>

Google (2020). *Google Lens* [site Web]. Consulté le 13 janvier 2020. <https://lens.google.com/>

Google Play (2020). *Image Analysis Toolset (IAT)* [site Web]. Consulté le 3 juin 2020. https://play.google.com/store/apps/details?id=tk.silviomarano.imageanalysisistoolset&hl=fr_CA

The Getty Research Institute (2018). *About the Research Institute* [site Web]. Consulté le 1 octobre 2018. <http://www.getty.edu/research/institute/>

Iconclass (2012). *Home Iconclass* [site Web]. Consulté le 18 juillet 2018. <http://www.iconclass.nl/home>

Intel Software (2019). *AI Frameworks Frameworks* [site Web]. Consulté le 4 décembre 2019. <https://software.intel.com/en-us/frameworks/tensorflow>

IBM (2019). *Visual Recognition* [site Web]. Consulté le 15 décembre 2019. <https://www.ibm.com/watson/services/visual-recognition/>

Idealab (2019). *Idealab, site officiel* [site Web]. Consulté le 15 avril 2019. <https://www.idealab.com/>

Larousse (2019). *Dictionnaire Larousse* [En ligne]. Consulté le 20 janvier 2019. <https://www.larousse.fr/dictionnaires/francais>

Musée McCord (2018). *À propos des collections et de l'offre numérique en ligne* [site Web]. Consulté le 25 juillet 201. <http://collections.musee-mccord.qc.ca/fr/clefs/collections/>

Mon McCord (2019). *Mon McCord, site officiel* [site Web]. Consulté 21 avril 2019. <http://collections.musee-mccord.qc.ca/scripts/user.php?Lang=2>

Office québécois de la langue française (2012). *Le grand dictionnaire terminologique (GDT)* [site Web]. Consulté le 1 octobre 2019. <http://gdt.oqlf.gouv.qc.ca/index.aspx>

Picasa (2016). *Picasa, site officiel* [site Web]. Consulté le 20 avril 2019.

<https://picasa.google.fr/>

Publication Québec (2019). *LégisQuébec source officielle* [site Web]. Consulté le 28 mars 2019. <http://legisquebec.gouv.qc.ca/fr/ShowDoc/cr/A-21.1,%20r.%201/>

La Société d'histoire de Sherbrooke (2020). *Histoire Sherbrooke, à propos de la Société d'histoire de Sherbrooke* [site Web]. Consulté le 15 février 2019.

http://www.histoiresherbrooke.org/a_propos_de_la_shs

Smugsmug (2019). *Smugsmug, Site officiel* [site Web]. Consulté le 15 avril 2019.

<https://www.smugmug.com/>

Studio Line (2020). *Studioline* [site Web]. Consulté le 13 janvier 2020.

<http://studioline.net/fr/>

TensorFlow (2019). *An end-to-end open source machine learning platform* [site Web].

Consulté le 4 décembre 2019. <https://www.tensorflow.org/>

Émission de télévision

RICHARD, Jeannita, Éric LEMYRE et Louis FAURE (15 juillet 2018). « L'intelligence artificielle » [Enregistrement vidéo], sur le site *Radio-Canada, Découverte*. Saison 30, épisode 32. Consulté le 20 septembre 2019. [https://ici.radio-](https://ici.radio-canada.ca/tele/decouverte/site/episodes/409710/intelligence-artificielle)

[canada.ca/tele/decouverte/site/episodes/409710/intelligence-artificielle](https://ici.radio-canada.ca/tele/decouverte/site/episodes/409710/intelligence-artificielle)

ANNEXE A

Thèmes abordés	
1. Constructions diverses	1 Édifices publics 2 Édifices militaires 3 Édifices religieux 4 Édifices commerciaux et industriels 5 Résidences 6 Immeubles à logements multiples, places 7 Édifices et industries non-identifiés 8 Sites archéologiques
2. Événements	9 Événements religieux 10 Événements sportifs et loisirs 11 Événements artistiques et culturels 12 Événements sociaux 13 Événements militaires 14 Événements économiques 15 Événements politiques 16 Événements judiciaires
3. Paysages	17 Paysages urbains 18 Paysages naturels 19 Paysages / sinistres 20 Paysages : organisation et peuplement
4. Personnages	21 Individus (seuls) 22 Couples 23 Groupes
5. Transports et communications	24 Automobiles et camions, motoneiges 25 Tramways 26 Navigation 27 Ponts 28 Trains et chemins de fer (Comprend les gares) 29 Routes 30 Journaux, imprimeries de journaux, édifices de journal 31 Etc.
6. Divers	32 Hors région 33 Œuvres d'artistes 34 Images pieuses

	35	Etc.
7. Monuments et plaques	36 37 38 39 40	Monuments commémoratifs Plaques commémoratives; panneaux d'interprétation Croix diverses Armoiries Etc.
8. Équipements divers	41 42 43 44	Service d'incendie, de police, de la G.R.C. Autres (équipement électrique, hydro-électrique, inventions) Équipement militaire Pièces de musée
9. Cartes, plans et cadastres	45 46 47	Plans de la ville et photographies aériennes Cadastres Cartes
10. Animaux		
11. Scènes humoristiques		
12. Documents écrits	48 49 50 51	Archives Publicités Statuts et règlements Autres

Source : SHS, *Guide pour le traitement des archives iconographiques*, p.24 à 30.

ANNEXE B

Recherche avancée

Rechercher

Partout	<input type="text"/>
NO	<input type="text"/>
TITRE_DU_FONDS	<input type="text"/>
COTE_DU_FONDS	<input type="text"/>
LOCALISATION	<input type="text"/>
CODE_SUPPORT	<input type="text"/>
NO_DOCUMENT	<input type="text"/>
CONSULTATION	<input type="text"/>
SUJET_PRINCIPAL	(séparés par des virgules)
RESPONSABILITE	<input type="text"/>
DATE_CREATION	<input type="text"/>
COULEUR	<input type="text"/>
ETAT_DOCUMENT	<input type="text"/> et <input type="text"/>
DESCRIPTION	<input type="text"/>
LIEUX	<input type="text"/>
SUJETS	<input type="text"/>
PERSONNAGES	<input type="text"/>
NOTES	<input type="text"/>
NB_COPIES	résultats avec des photos ? <input type="checkbox"/>
NO_DU_NEGATIF	<input type="text"/>
RENVOIS	<input type="text"/>

[Rechercher dans une autre catégorie](#)

Source : Société d'histoire de Sherbrooke, *Histoiresherbrooke* [En ligne], 2018, consulté le 18 juillet 2018, <<http://histoiresherbrooke.ca/search.php?advanced>>.

ANNEXE C

Résultats obtenus dans la base de données en ligne

ACCUEIL | À PROPOS DE LA SHS | DEVEZ-VOUS ÊTRE MEMBRE | COMMENT NOUS JOINDRE | PANIER


RECHERCHE Recherche avancée

RÉSULTAT DE LA RECHERCHE

Archives (21)
Audiovisuels (9)
Bibliothèque (4)
Cartes (14)
Images (903)
Objets (23)
Périodiques (10)
Sonores (4)

COLLECTION : Images (903)


Fiche #10011



No : 10011
Titre du fonds : CLOVIS ROY
Cote du fonds : IP306
Localisation : R
Code support : CPM
No document : 31D 6

Consultation : L
Sujet principal : WELLINGTON STREET, SHERBROOKE, P.Q.
Responsabilité : VALENTINES AND SONS
Date creation : 1910
Couleur : OUI
Etat document : TRES BON, MAIS TRACE D'OBLITERATION SUR LE RECTO
Description : VUE SUR LA RUE WELLINGTON NORD A PARTIR DE KING AU DEBUT DU SIECLE. ON VOIT DES CHARRETTES TIREES PAR DES CHEVAUX. DES HOMMES ET DES FEMMES EN HABIT DE VILLE D'EPOQUE. SUR LA DEVANTURE D'UN MAGASIN. ON PEUT LIRE : "LAPAS BROS. POOL ROOM". PLUSIEURS DEVANTURES POSSEDENT DES AUVENTS EN TISSUS.
Lieux : SHERBROOKE; RUE WELLINGTON NORD
Sujets : PAYSAGES URBAINS; RUES
Notes : TIRE DE L'ALBUM 2. PHOTO 2.19. CARTE ANNOTEE AU VERSO. ON RETROUVE LA MEME PHOTO EN IP460 PN 31D 12.

Fiche #10014



No : 10014
Titre du fonds : CLOVIS ROY
Cote du fonds : IP306
Localisation : R
Code support : CPM
No document : 31D 9

Consultation : L
Sujet principal : WELLINGTON ST. LOOKING NORTH, SHERBROOKE, P.Q.

Source : Société d'histoire de Sherbrooke, *Histoiresherbrooke* [En ligne], 2018, consulté le 18 juillet 2018,

<[http://histoiresherbrooke.ca/list_sheets.php?category_uid=7&search_input\[\]=wellington&search_rubrique\[\]=0](http://histoiresherbrooke.ca/list_sheets.php?category_uid=7&search_input[]=wellington&search_rubrique[]=0)>.

ANNEXE D

Recherche détaillée

<p>RECHERCHE RAPIDE Aide ?</p> <p>Mots-clés : <input style="width: 100%;" type="text"/> <input type="button" value="Aller"/></p> <p><input type="checkbox"/> Inclure images des partenaires</p> <p>RECHERCHE AVANCÉE ▶ (McCord seulement)</p> <p>EXPLORER ▶ La collection en ligne</p> <ul style="list-style-type: none"> Collections Artistes Objets Périodes Régions géo-culturelles Carte géo-culturelle Ensembles thématiques (McCord et partenaires) Fonds <p>JEUX ▶</p> <p>CIRCUITS THÉMATIQUES ▶</p> <p>MON MCCORD ▶</p> <p>ACHETER UNE REPRODUCTION ▶</p>	<p>Collections : Photographie</p> <p>Mots-clés : <input style="width: 100%;" type="text"/></p> <p>Objet : <input style="width: 100%;" type="text" value="Tous les objets"/></p> <p>Catégorie : <input style="width: 100%;" type="text" value="Toutes les catégories"/></p> <p>Sujet : <input style="width: 100%;" type="text" value="Tous les sujets"/></p> <p>Artiste : <input style="width: 100%;" type="text" value="All Artists/Tous les artistes"/></p> <p>Période : <input style="width: 100%;" type="text" value="Toutes les périodes"/></p> <p>Entre <input style="width: 50px;" type="text"/> et <input style="width: 50px;" type="text"/></p> <p>Province : <input style="width: 100%;" type="text" value="(Aucune sélection)"/></p> <p>Ordre de tri : <input style="width: 100%;" type="text" value="Pertinence"/></p> <p style="text-align: center;"><input type="button" value="Poursuivre"/></p> <p style="text-align: center;"><input type="button" value="Nouvelle recherche"/></p>
--	---

Source : Musée McCord, *Collections et recherche* [En ligne], 2018, consulté le 25 juillet 2018, <http://collections.musee-mccord.qc.ca/scripts/advanced_search.php?Lang=2&department=0>.

ANNEXE E

Résultats de l'expérimentation de la base de données AAT

Termes	Résultats	Ajouts
History	Architectural History/ Industrial History/ Local History/ Urban History	
Scene	Street Scenes	Arbre/ arbre couché/ jour/ abattage/ entretien.
Event	Events (activities)	Abattage/ entretien.
Historic event	Historic Districts/ Historic Quarters/ Historic Sites/ Local Historic Districts	
Place	Inhabited Places / Market Places/ Place	Carré Strathcona
Region	Inhabited Regions/ Local regions/ Regions (Administrative Divisions)/ Regions (Geographic)	Estrie/ Québec/ Canada/ Cantons-de-l'Est,
City	Central Business Districts/ Cities/ Inner cities/ Urban Beautification/ Urban Planning	Sherbrooke
Street	Arterial Streets/ Main Streets/ Streetcar Systems/ Streets	Wellington/ Wellington Nord/ Frontenac/ Factory Street/ rue de la Facterie/ King's Highway/ Chemin du Roi/ tramways/ rails/ ligne/ ligne Belt/ urbaine.

Building	Administrative Building/ Building (Structure)/ Commercial Building/ Communications Building/ Government Office Building/ Institutional Building/ Landmark Building/ Municipal Building / Office Building/ Societies' Building/ Stores/ Transportation Building/ Universities (Building)	Quebec Central Railway Building/ Édifice Odell/ Édifice Whiting/ Édifice Bell Telephone/ quincaillerie Luke & Mitchell, J. S. Mitchell/ H. C. Wilson/ Magasin general/ Hôtel de ville/ Palace of illusion/ magasin de chaussures Wiggett/ Banque de Montréal/ Bell Telephone/ Faculté de Droit UdeS/ Université de Sherbrooke/ Musée des beaux- arts/ Petit Théâtre de Sherbrooke/ Red School House/ Chapelle méthodiste/ station de pompiers/ Bellevue Hotel.
Companies	Companies/ Railroads (Administrative)	Quebec Central Railway Building/ Sherbrooke Railway & Power/ Sherbrooke Street Railway.

ANNEXE F

Publication sur la plateforme Facebook

Carré Stratcohna, Rue Wellington, Rue Frontenac, marché, square, centre ville, Bâtiments Odell, Lucke et Mitchel, tramway, 20e siècle, arbre, Québec central railway, entretien urbain, électricité, parque, accident

J'aime · Répondre · 2 sem · Modifié

Sherbrooke, Rue Wellington, Rue Factory, Rue Frontenac, Arbre, Individu, Personne, Rue, Rail, Tramway, Électricité, Commerce, Accident, Édifice Odell, Quebec Central Railway Building, Palais de Justice, Hôtel de Ville, Centre-Ville, Lampadaire, Réverbère, Voiture, Magasin

J'aime · Répondre · 1 sem

Carré Stratcohna, Rue Wellington ...

J'aime · Répondre · 2 sem

Centre ville

J'aime · Répondre · 2 sem

Michel? Veux tu aider une jeune étudiante? 😊😊

J'aime · Répondre · 2 sem

1940 ou koi les voitures ont vraiment l'air vieilles 😊

J'aime · Répondre · 2 sem

Wellington, Frontenac, Brûlerie de café, hôtel de ville.

J'aime · Répondre · 2 sem

Pourquoi l'arbre est à terre? Il y a eu une tempête ?

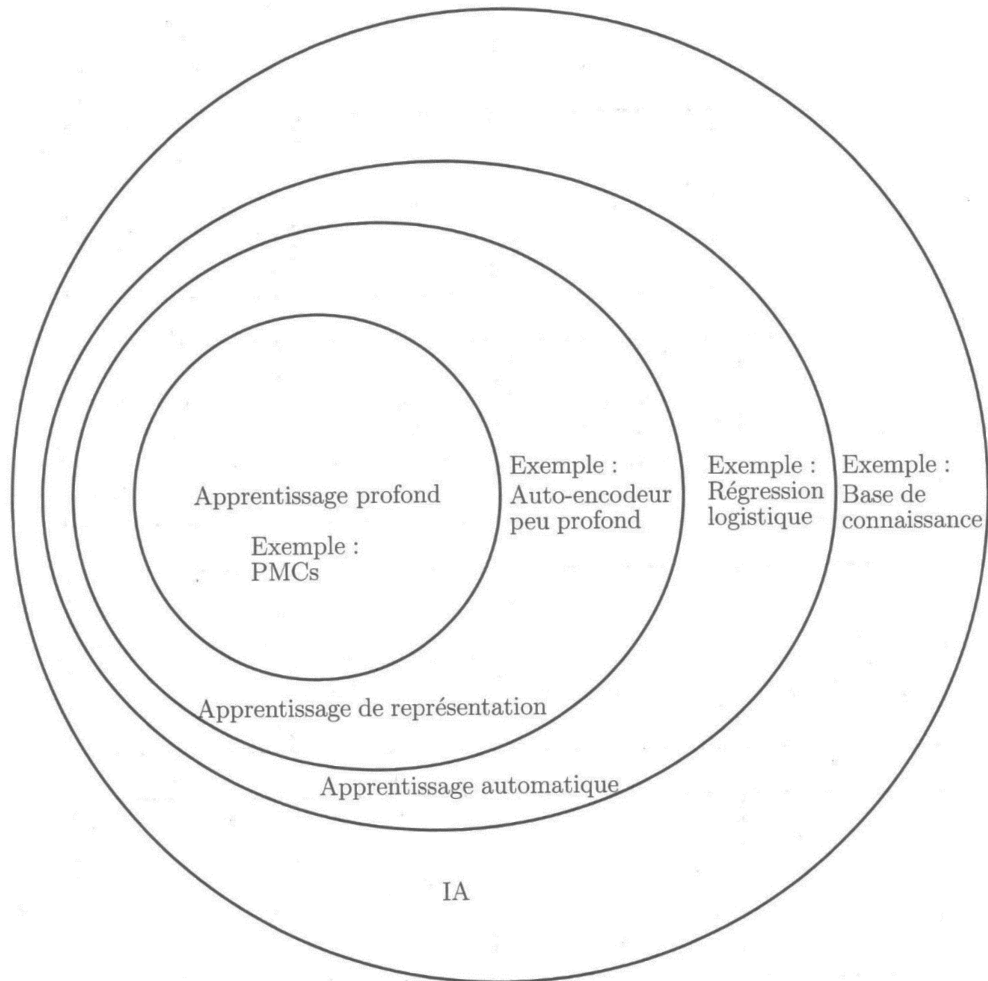
J'aime · Répondre · 2 sem

Rue Wellington, (année de la photo), (tempête ou événement quelconque qui a fait que l'arbre est tombé/a été abattu), tramway

J'aime · Répondre · 2 sem

ANNEXE G

Les apprentissages liés à l'intelligence artificielle



Source : Ian Goodfellow, ..., *L'apprentissage profond*, p. 35.

ANNEXE H

Analyse de l'image par couches

CHAPITRE 1

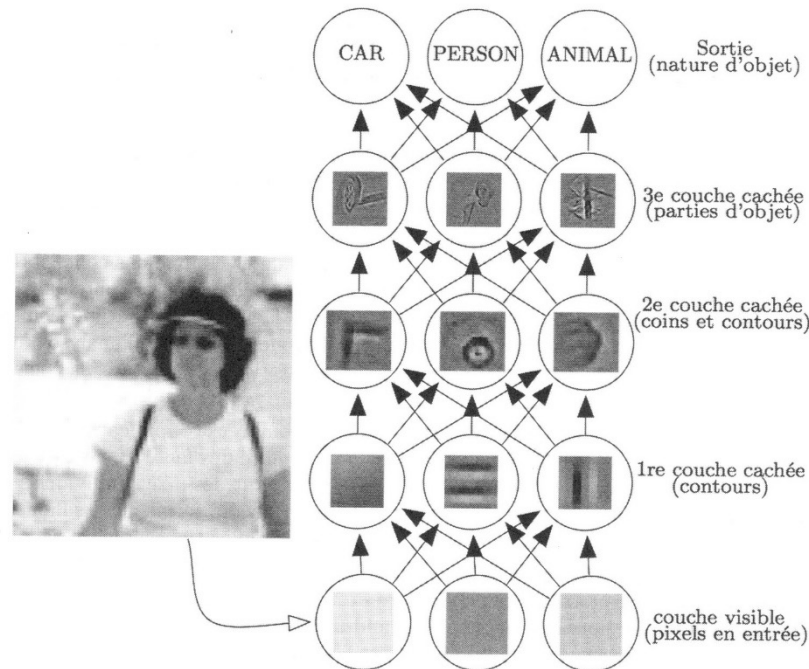
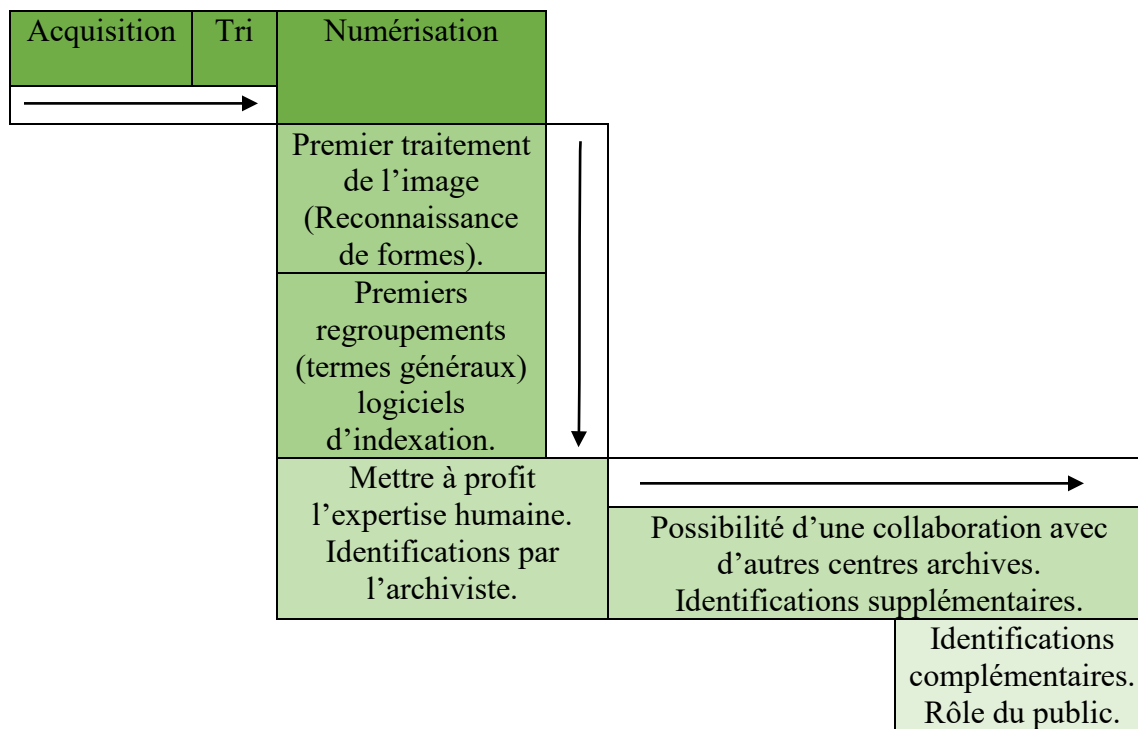


Figure 1.2 : Illustration d'un modèle d'apprentissage profond. Il est difficile pour un ordinateur de comprendre la signification des données d'entrée sensorielles brutes, telles que cette image représentée sous la forme d'une collection de valeurs de pixels. La transformation d'un ensemble de pixels en une identité d'objet est très compliquée. L'apprentissage ou l'évaluation de cette transformation semble insurmontable si elle est abordée directement. L'apprentissage profond résout cette difficulté en brisant la transformation complexe désirée en une série de transformations simples imbriquées, décrite chacune par une couche différente du modèle. L'entrée est présentée à la **couche visible**, ainsi nommée car elle contient les variables que nous pouvons observer. Puis une série de **couches cachées** extrait des caractéristiques de plus en plus abstraites de l'image. Ces couches sont appelées "cachées" parce que leurs valeurs ne sont pas fournies dans les données ; au lieu de cela, le modèle doit déterminer lui-même quels concepts sont utiles pour expliquer les relations dans les données observées, et inférer ces valeurs. Les images ici sont des visualisations du type de caractéristique représenté par chaque unité cachée. En fonction des pixels, la première couche peut facilement identifier les arêtes, en comparant la luminosité des pixels voisins. Compte tenu de la description des arêtes de la première couche cachée, la deuxième couche cachée peut aisément rechercher des coins et des contours étendus, reconnaissables comme des collections d'arêtes. Étant donnée la deuxième couche cachée décrivant l'image en termes d'angles et de contours, la troisième couche cachée peut détecter des parties complètes d'objets spécifiques, en trouvant des collections spécifiques de contours et d'angles. *In fine*, cette description de l'image, en termes de parties d'objets qu'elle contient, peut être utilisée pour reconnaître les objets présents dans cette image. Images reproduites avec l'autorisation de Zeiler et Fergus (2014).

ANNEXE I

Chaîne de traitement



ANNEXE J

Reconnaissance de formes

Voici une image représentant un Rafale au décollage :



```
python classify_image.py --model_dir=tmp --image_file=imagespersonnelles/rafale_dec
```

Voici le résultat obtenu :

```
aircraft carrier, carrier, flattop, attack aircraft carrier (score = 0.71103)
warplane, military plane (score = 0.08506)
projectile, missile (score = 0.04319)
missile (score = 0.03181)
wing (score = 0.00225)
```

Source : Stéphane Nachez, « Classifier des images avec Tensorflow ».