



IconoTag, données empiriques pour soutenir le concept d'indexation multilingue automatique d'images

Claire Nigay

École de bibliothéconomie et des sciences de l'information
Université de Montréal, Montréal, Québec, Canada

James M Turner

École de bibliothéconomie et des sciences de l'information
Université de Montréal, Montréal, Québec, Canada

Karine Lespinasse-Sabourault

Alliance française de Shanghai, 上海法语培训中心,
Université Paris 8

Meeting:

161 — "The media is the message!" The convergence of media in rapidly changing societies from a user perspective as well as the demand for preservation — Audiovisual and Multimedia Section with Preservation and Conservation

Résumé

Les utilisateurs de services d'information cherchent des images à diverses fins, par exemple pour illustrer un article sur la nourriture, présenter des animaux auprès d'enfants, ou encore étudier des véhicules. Lors de projets antérieurs, nous avons trouvé que les participants dont la langue maternelle est le français ou l'anglais écrivaient des termes d'indexation correspondant à ceux de l'autre langue. Toutefois, il existe beaucoup de similarités entre ces deux langues, par exemple l'alphabet romain et beaucoup de mots en commun. IconoTag était un projet qui cherchait à vérifier si autant de correspondances se manifesteraient entre d'autres langues, surtout celles dont les racines sont différentes ou qui emploient d'autres alphabets ou encore des idéogrammes. Afin d'effectuer une collecte de données, nous avons construit un site web. Les participants devaient ajouter des étiquettes (tags) à des images, lesquelles étaient choisies selon des critères établis lors de travaux précédents, par exemple la simplicité ou la complexité de l'image, le nombre d'objets qu'on peut nommer, la facilité ou la difficulté d'identifier le contenu, et ainsi de suite. L'ensemble comportait deux images abstraites, incluses comme contrôle. Nous avons installé le site web en 2010. À l'aide de réseaux sociaux et de contacts personnels, nous avons recruté des participants qui devaient ajouter des étiquettes en l'une de dix langues, soit l'allemand, l'anglais, l'arabe, le chinois, l'espagnol, le français, le grec, le portugais, le russe ou le suédois. Le taux de correspondance entre les langues était élevé, ainsi validant la notion que l'indexation multilingue automatique d'images en environnement réseauté est tout à fait faisable. Les utilisateurs de diverses communautés réussiraient donc à trouver des images, peu importe la langue d'indexation de la base de données.

Introduction

Dans le cadre de plusieurs projets de recherche, nous avons étudié l'indexation d'images fixes et en mouvement. Nous avons démontré que les utilisateurs écrivent une même étiquette, que l'image soit fixe ou en mouvement (par exemple Turner 1994, 1995). Dans d'autres projets, nous avons démontré que les utilisateurs écrivent les termes correspondants, que l'utilisateur ait le français ou l'anglais comme langue maternelle. Ils nomment les objets visibles dans les images, et les noms donnés se traduisent directement à un nom équivalent dans l'autre langue (par exemple Turner et Roulier 1999, Turner and Mathieu 2007). Nos résultats de recherche font état de beaucoup de constance dans les taux de correspondance, ce qui suggère que l'on peut indexer une image dans une seule langue, puis générer automatiquement l'indexation dans l'autre langue. Ces taux de correspondance indiquent que l'indexation générée automatiquement sera de la même qualité que l'indexation dans la langue source. Ménard (2006, 2007) a réalisé d'autres études traitant de l'indexation d'images fixes en français et anglais.

Nous avons poursuivi nos recherches en effectuant d'autres études, à l'aide de traducteurs web (Hudon, Turner et Devin 2001; Turner and Hudon 2002). Les résultats indiquent qu'à l'aide d'outils web, il serait possible de générer automatiquement, à partir d'images indexées en une seule langue, de l'indexation en beaucoup d'autres langues. Toutefois, l'indexation multilingue ne faisant pas nécessairement partie des besoins des usagers principaux de telle ou telle base de données d'images, il est peu probable que ceux qui gèrent ces bases de données l'effectuent. Le besoin se manifeste plutôt chez les visiteurs internautes de la base de données.

Les internautes saisissent des mots-clés dans un moteur de recherche, puis le moteur leur trouve des images. On pourrait programmer des moteurs de recherche pour acheminer les mots-clés à des dictionnaires bilingues en de nombreuses langues. Par la suite, la requête pourrait être acheminée à des bases de données indexées en ces langues, les résultats ramassés et acheminés vers l'utilisateur. Nous sommes d'avis qu'un tel système pourrait bien fonctionner, dès que les dictionnaires bilingues nécessaires sont disponibles sur le web et qu'un algorithme de recherche est formulé et installé. En fait, de nombreux dictionnaires et autres outils de gestion de vocabulaire bilingues et multilingues sont déjà disponibles sur le web. Il existe également quelques sites expérimentaux avec tentatives d'indexation multilingue d'images.

Dans le contexte du présent projet, notre objectif est de démontrer empiriquement soit que cette approche marcherait bien, soit qu'elle ne marcherait pas. En environnement multilingue, le problème entre le français et l'anglais est la similarité des langues, avec bien des mots en commun et un alphabet commun. On se demande ce qui arriverait lorsqu'il n'y a pas ce genre de correspondance entre les langues.

Lors d'un projet antérieur (Turner 1995), nous avons démontré que pour ce type d'image (images « ordinaires » par opposition aux images documentaires ou artistiques) et pour les mêmes images, les utilisateurs donnent les mêmes termes d'indexation que les indexeurs professionnels. On peut trouver une discussion des questions entourant les images en mouvement de type documentaire dans Lespinasse-Sabourault (2006). Les images artistiques sont beaucoup plus complexes, à cause des divers niveaux d'interprétation utiles pour les étudier. La pratique jusqu'à présent est de ne pas les indexer au niveau primaire, soit de nommer les objets qu'on y voit (Markey 1986, 1988). Il est toutefois intéressant d'observer

que ce genre d'activité se trouve maintenant sur le web et dans les musées (Kellogg Smith 2006, Steve 2011).

Dans le présent article, nous utilisons le terme « étiquette » (*tag*) comme synonyme de « terme d'indexation » ou « mot-clé ». Les étiquettes ou termes d'indexation peuvent comporter un seul mot ou plusieurs mots. Nous n'aborderons pas ici la question de la qualité d'indexation reliée aux étiquettes, mais d'autres chercheurs étudient cette question (par exemple Furner 2007, Kipp 2009). Au cours des années, nous avons trouvé qu'une partie de la réponse à cette question dépend du type d'information étiquetée ou indexée. Dans le présent contexte, nous avons affaire à des images fixes que nous pouvons considérer comme des images « ordinaires ». Soulignons quand même que cette catégorisation est plutôt arbitraire, puisque les mêmes images pourraient être considérées comme documentaires ou artistiques, selon le contexte.

Méthodologie

Afin d'étudier la question de la traduction automatique d'étiquettes d'une langue à plusieurs autres langues, nous avons créé un site web, IconoTag, pour la collecte de données. Nous avons fait la conception en français. Par la suite, nous avons recruté de l'aide en utilisant le site de réseautage social Facebook, afin d'aboutir à dix versions parallèles du site. Les langues du projet sont : l'allemand, l'anglais, l'arabe, le chinois, l'espagnol, le français, le grec, le portugais, le russe et le suédois. Nous avons installé le site sur un serveur avec une base de données SQL afin d'héberger les données saisies à l'aide d'un formulaire PHP. La page d'accueil du site consistait en liens vers les dix langues. Ce premier clic menait à des images de douze images à étiqueter (figure 1), ainsi que de brèves explications sur le projet et des instructions.



Figure 1. Les douze images à étiqueter en l'une de dix langues.

Un clic vers la page suivante offrait de plus amples explications sur le projet, les coordonnées des chercheurs, de l'information sur le certificat d'éthique de l'université, et enfin un clic pour signifier le consentement du participant.

Ce dernier clic menait à la première image, laquelle paraissait alors beaucoup plus grande. Sur la même page on trouvait le formulaire pour saisir les étiquettes correspondant à l'image. Une fois les étiquettes saisies, un clic sur le bouton intitulé « Soumettre » menait à la prochaine image, et ainsi de suite jusqu'à ce que toutes les images soient étiquetées. Les données étaient enregistrées dans une base de données, et pouvaient alors être exportées et analysées.

Afin d'encourager la participation, nous avons créé une tâche rapide et facile. Nous avons choisi les images selon des critères établis lors de projets précédents, dont la simplicité ou la complexité de l'image, le nombre d'objets à nommer, et la facilité ou la difficulté à identifier les contenus. Nous avons inclus deux images abstraites (les numéros 3 et 6) comme contrôle, afin de valider les données. Nous croyions que même si les taux de correspondance entre étiquettes seraient élevés pour les images simples, les termes nommés le plus souvent seraient dispersés pour ces deux images. L'image 3 présente un détail d'une sculpture (un *stable*, par opposition à un mobile) d'Alexander Calder intitulé *L'Homme*, installée sur le site de l'exposition universelle Expo67 à Montréal. L'image 6 présente les tracés de lumière enregistrés en bougeant un appareil-photo numérique lors de la prise de vue le soir. Nous avons préféré l'orientation paysage à l'orientation portrait pour mieux se conformer à la forme des écrans d'ordinateur et pour assurer que les dimensions des images soient ainsi les plus importantes possible.

Sur la base de résultats de recherche antérieurs (Turner 1994), nous avons décidé de ne pas demander de renseignements personnels, afin de simplifier la tâche et de réduire le temps nécessaire pour la compléter. Les études antérieures ne révélaient aucune différence significative dans l'indexation, que les participants soient jeunes ou vieux, hommes ou femmes, expérimentés dans l'utilisation d'images ou non, étudiants ou travailleurs, et ainsi de suite.

Le site était installé sur le web en mars 2010, puis fermé plusieurs mois plus tard, une fois que nous croyions avoir assez de données. Le réseautage social et les contacts personnels étaient employés pour le recrutement de participants. Il fallait choisir une langue, puis étiqueter les douze images. Selon les instructions, il s'agissait d'arrêter d'ajouter des étiquettes dès que le participant pensait qu'il y en ait suffisamment pour décrire adéquatement l'image, jusqu'à concurrence de cinq étiquettes par image.

Résultats et analyse

Dans deux articles précédents, nous avons donné des résultats préliminaires de cette recherche (Turner et Nigay 2010; Turner, Lespinasse-Sabourault and Nigay 2010). Dans le présent papier, nous donnons les résultats finaux du projet. Puisque nous avons déjà un nombre important de participants lors de l'analyse préliminaire pour le français, l'anglais et le chinois, nous avons conclu qu'il y avait déjà assez de stabilité dans les termes nommés le plus souvent. Ainsi, nous n'avons pas analysé davantage de données pour ces langues. Le tableau 1 présente le nombre de participants pour chaque langue.

Tableau 1. Nombre de participants pour chaque langue.

Langue	Nombre de participants
français	323
anglais	127
chinois	90
suédois	26
allemand	17
espagnol	14
portugais	7
arabe	2
grec	2
russe	2
TOTAL	608

Les tableaux 2, 3 et 4 présentent les termes nommés le plus souvent dans chacune des langues. Lorsque deux termes ou plus sont nommés le même nombre de fois, nous donnons tous ces termes, dans l'ordre alphabétique. Les astérisques indiquent une ou plus lettres manquantes, afin d'inclure les formes plurales, des variations dans l'orthographe et des mots composés de plusieurs éléments.

Les données sont divisées en trois tableaux, à cause des difficultés dans l'analyse et la représentation des diverses langues. Il est évident que la globalisation n'est pas encore complétée en ce qui concerne les logiciels ! Or, les auteurs ne parlent ni ne lisent le chinois. Ainsi, nous avons cherché de l'aide pour l'analyse auprès du personnel d'une bibliothèque en Chine. Pour l'arabe, 30 participants ont contribué aux données, mais un problème technique dans le mécanisme de collecte a rendu inutilisables les données provenant de 28 de ces participants, avant qu'on ne découvre le problème et qu'on réussisse à le régler. Il ne restait alors que des données de deux participants. Cela est particulièrement malheureux, puisque l'arabe utilise un alphabet autre que romain, un des aspects que nous espérions étudier dans le cadre de ce projet. Les données pour le chinois et l'arabe sont présentées dans un tableau séparé pour chaque langue.

Le tableau 2 présente les données pour les huit autres langues dans lesquelles des participants ont contribué aux données. Trop grand pour présenter ici, le tableau se trouve à la fin, en Annexe A. Nous n'avons réussi à recruter que deux participants pour le russe et deux pour le grec. Ainsi, les données pour ces langues sont maigres.

La distribution de Zipf est typique de ce genre de données, et on la retrouve dans ce contexte. Ainsi, en compilant les réponses dans un chiffrier, la plupart des cellules de la matrice sont vides, la plupart des étiquettes ne sont données qu'une fois, et quelques étiquettes paraissent souvent. Ce sont ces dernières qui nous intéressent, parce que les plus utiles pour l'indexation des images correspondantes.

Nous comparons d'abord les trois termes donnés le plus souvent pour chaque image. Il est utile de les prendre en considération indépendamment de leur rang. Bien qu'il y ait de la variation dans le rang d'une langue à l'autre, nous observons que pour la plupart des images, toutes les langues ou presque ont les trois mêmes termes en vedette ou encore deux sur trois. Cela signifie que celui qui cherche une image va la trouver en saisissant n'importe quel de ces trois termes. Nous observons que la correspondance est même plus précise avec cet ensemble de données : pour les images 2, 4, 5, 6, 7, 8, 9, 10, 11, 12 (soit 10 des 12 images ou 83 pour cent), le terme nommé le plus souvent est le même dans toutes les langues ou

presque. Cela est surprenant, puisque quatre images seulement sont classifiées comme simples, ayant un seul objet évident à nommer. Les participants ont toutefois trouvé d'autres choses à nommer. Ainsi, le terme au premier rang suggère de l'information importante, d'un point de vue physiologique et culturel, concernant la perception et l'interprétation des images.

Nous traitons ensuite les données selon le type d'image, soit simple, complexe ou abstrait. Cette catégorisation offre une analyse particulièrement intéressante. Les images simples (un seul objet significatif à nommer) sont les numéros 2, 7, 10 et 12. Les images complexes (plusieurs objets à nommer, des objets qui sollicitent l'attention en concurrence les uns avec les autres, ou encore des objets plus difficiles à identifier ou à nommer) sont les numéros 1, 4, 5, 8, 9 et 11. Les images abstraites, images de contrôle, sont les numéros 3 et 6. Nous nous attendions à ce que la correspondance entre les langues soit élevée pour les images simples, plus dispersée ou complètement dispersée pour les images complexes, et qu'il n'y ait que peu ou pas de correspondance pour les images abstraites.

Les résultats obtenus sont mieux qu'attendus, ce qui est très encourageant. Les taux de correspondance sont parfois même étonnamment élevés. Comme prévu, les participants de toutes les langues pour lesquelles nous avons des données pour l'image ont donné le même terme au premier rang pour les images simples. Il n'y a qu'une exception : pour l'image 2, pour les arabophones, « oiseau » est au premier rang, alors que « paon » est au deuxième. L'exception est insignifiante toutefois, puisque nous avons des données utilisables de deux participants seulement. Puisque les participants pouvaient donner jusqu'à cinq étiquettes pour chaque image, « oiseau » se trouve au premier rang (4 cas) en tant qu'élément de divers termes donnés par les deux participants.

Nous n'avions pas prévu de si hauts taux de correspondance pour les images complexes. Regardons de plus près maintenant. Pour l'image 1, aucun terme ne se démarque au premier rang. Toutefois, pris ensemble, les trois termes nommés le plus souvent correspondent aux mêmes concepts (ordures, rue, nettoyage). Pour l'image 4, « appareil-photo » est au premier rang dans 7 langues, « photo » pour les trois autres (le français, le portugais, le chinois). Pour l'image 5, « plage » est au premier rang pour 9 langues, « mer » pour 1 (le russe). Encore ici, l'exception est insignifiante parce que deux participants seulement ont contribué aux données. Pour l'image 8, toutes les langues ont le même concept (danse) au premier rang. Pour l'image 9, « cactus » arrive en premier pour 9 langues, « désert » pour 1 (l'espagnol). Si on avait plus de données, les rangs pour ces termes seraient sans doute plus précis. Pour l'image 11, « nourriture » est au premier rang pour 8 langues, « plat » pour 1 (le français) et « cuisine » pour 1 (le chinois). Puisque nous avons amplement de données pour le français et le chinois, il est possible que ces résultats reflètent des différences culturelles de perception.

Ce sont les images abstraites qui étonnent le plus, au point où nous nous demandons si nous étions tout simplement naïfs en pensant qu'il y aurait une dispersion large parmi les étiquettes. Les participants ont essayé de nommer ce qu'ils voyaient, ils voyaient les mêmes choses et ils ont donné les mêmes noms aux contenus, de façon générale. Cela constitue un argument assez éloquent pour appuyer la notion que l'étiquetage marche bien pour l'indexation de l'image « ordinaire », peut-être même pour l'image abstraite. Pour l'image 3, les termes nommés le plus souvent couvrent les mêmes concepts (structure, acier, métal, bâtiment, sculpture). Pour l'image 6, « lumière » est au premier rang pour toutes les langues sauf l'arabe. Pour cette image, les deux arabophones ont écrit « feu d'artifice » comme seule étiquette. D'autres termes offerts dans les autres langues pour cette image sont « couleur » et

« photo ».

Nous abordons ensuite l'analyse des données chinoises, une activité à part. Des difficultés techniques dans l'interprétation des étiquettes provenant de la base de données avec laquelle nous avons effectué la collecte empêchaient l'analyse des données avec celles provenant d'autres langues. Par ailleurs, aucun membre de l'équipe de recherche ne lit l'écriture chinoise. Néanmoins, nous avons pu recruter l'aide de membres du personnel bilingue de la bibliothèque de l'Alliance française de Shanghai (上海法语培训中心图书馆). On peut trouver des informations sur leur travail sur les questions interculturelles dans Josso and Lespinasse-Sabourault (2008). Ce travail a fait la démonstration que l'analyse, en employant le français comme langue intermédiaire, fonctionne assez bien.

Les différences entre les systèmes d'écriture (Zhang et al. 2006) rendaient impossible pour les données chinoises la méthode utilisée pour la troncature des mots ou pour la substitution de lettres dans les autres langues. Le tableau 3 présente l'analyse des données chinoises, en comparant les données chinoises pour chaque image avec les termes ramassés pour le français et l'anglais. Dans un premier temps, le personnel bilingue de la bibliothèque de l'Alliance française a créé une version française des concepts exprimés en langue chinoise. Dans un deuxième temps, nous avons compté les occurrences des termes de cette version française des données, en utilisant la même méthode de troncature employée pour les langues occidentales. Puisqu'il s'agit des mêmes concepts représentés soit par idéogrammes, soit par mots, la méthode, bien qu'un peu malhabile, offre un portrait assez fidèle des correspondances.

Tableau 3. Les étiquettes données le plus souvent par les participants sinophones, utilisant le français comme langue intermédiaire, et celles données le plus souvent par les francophones et anglophones.

Image	Chinois (90 participants) terme	Français à partir de chinois terme, fréquence	Anglais à partir de chinois via Google terme	Français (323 participants) terme, fréquence	Anglais (127 participants) terme, fréquence
01	垃圾	ordures 46	garbage	éboueur 188	garbage 101
	清扫	nettoy* 42	cleaning	poubelle 137	clean 77
	街道	rue 44	street	camion 115	street 90
	清洁工	éboueur 29	cleaners	nettoy* 109	truck 47
02	孔雀	paon 69	peacock	paon 288	peacock 127
	动物	animal 12	animal	oiseau 159	bird 70
	蓝色	bleu 9	blue	plume 56	blue 33
	草地	gazon 7	lawn	bleu 55	feather 26
03	钢结构	acier 19	steel	métal 150	structure 44
	建筑	bâtiment 14	building	architectur* 80	steel 36
	飞机	avion 14	aircraft	sculpture 43	metal 35
	天空	ciel 11	sky	acier 42	architecture 28
04	摄影	photo* 74	photography	photo 344	camera 116
	光	lumière 71	light	appareil 213	tripod 89
	照相机	appareil-photo* 45	camera	ombre 127	shadow 54
	影子	ombre 26	shadow	trépied 115	photo 50
05	三角架	trépied 30	tripod		
	海滩	plage 64	beach	plage 216	beach 105
	度假	vacances 23	holiday	mer 179	palm 45
	游乐	loisir 13	recreation	palmier 133	ocean 31
阴天	ciel gris 12	cloudy day	bord de mer 55	arcade 30	
			vacances 55		

06	光 摄 光影 色彩	lumière photo* éclairage couleur*	42 15 13 11	light photo lighting colour	lumière effet nuit abstra*	184 61 45 36	light photo colo*r night	83 21 18 15
07	奶牛 牛奶 草原 牧场	vache lait prairie pâturage	62 11 11 7	dairy cow milk grassland ranch	vache pré lait campagne animal	281 100 44 43 33	cow farm field Holstein black white animal	123 23 23 23 22 22 17
08	舞蹈 异域风情 印度 泰国	danse exotique Indien* Thaïlande	56 12 7 7	dance exotic India* Thailand	danse costume tradition Bali	308 120 63 47	dance costume wom*n Asia*	117 44 18 15
09	仙人掌 沙漠 干旱	cactus d*sert sec*	49 21 8	cactus desert arid	cactus desert ciel arid*	267 216 49 46	cact* desert sky blue	117 89 20 13
10	南瓜 万圣节 菜市场 丰收	citrouille Halloween marché moisson	52 14 14 8	pumpkin Halloween vegetable market bumper harvest	citrouille Halloween orange potiron	217 102 85 79	pumpkin harvest Hallowe*en orange	109 36 29 28
11	烹饪 佳肴 食物 餐	cuisine plat délicieux aliment repas	26 18 11 7	cooking cuisine food meal	plat repas assiette légume	123 98 90 86	food vege* meal dinner	59 45 33 27
12	花 紫色 牡丹 极美的	fleur violet pivoine magnifique	49 15 10 8	flower purple peony very beautiful	fleur pivoine rose feuille	260 80 75 52	flower pink peon* purple	100 49 37 19

Cette analyse nous a permis d'identifier les idéogrammes chinois correspondant aux 4 termes nommés le plus souvent. Comme triangulation, nous avons passé chaque terme par Google Traduction afin d'obtenir une version anglaise des données. Ainsi, on voit dans le tableau 3 les relations entre les termes dans les trois langues. Pour le chinois, le compte est fait à partir de la version française des termes créée par les employés de la bibliothèque. Les deux dernières colonnes du tableau présentent les données provenant des participants francophones et anglophones, afin de faciliter la comparaison. Pour la langue chinoise, nous avons analysé les quatre termes donnés le plus souvent plutôt que les trois premiers termes pour les autres langues. Ainsi, nous n'ajustons pas le tableau 3 pour nous conformer à la présentation des autres langues, car les données du tableau 3 sont un peu plus riches.

Les deux participants arabophones dont on a pu analyser les données ont fourni un seul terme pour la plupart des images. Pour trois images, ils ont donné deux termes, et pour trois autres images, aucun. Les données sont présentées dans le tableau 4.

Tableau 4. Étiquettes provenant des deux participants arabophones dont les données n'étaient pas corrompues.

Image	Terme au 1er rang	Fréq	Terme au 2e rang	Fréq
1	تنظيف	2	تقمام	2
2	لطيور	4	لطاوس	2
3	<i>pas de données</i>			
4	صور	3	كاميرا	2
5	شاطئ	2		
6	نارية ألعاب	2		
7	بقرة	3		
8	رقصة	2		
9	صبار	2		
10	<i>pas de données</i>			
11	غذاء	2		
12	<i>pas de données</i>			

Les membres de l'équipe de recherche ne parlent ni ne lisent l'arabe non plus. Cependant, on a pu confirmer les résultats par triangulation en passant les chaînes de caractères par Google Traduction. Ainsi, on a pu vérifier les équivalents dans deux langues que nous connaissons bien, le français et l'anglais. Nous sommes donc confiants de la validité de l'analyse.

Il demeure que le peu de données provenant de participants parlant des langues écrites avec un alphabet autre que romain est vraiment regrettable. Mis à part le chinois, pour lequel nous avons pu analyser des données provenant de 90 participants, les autres langues de cette catégorie (arabe, grec, russe) ne sont pas bien représentées dans notre ensemble de données. De toute évidence, nos efforts pour recruter des participants qui parlent l'une de ces langues n'ont pas porté fruit. Néanmoins, le peu de données que nous avons pu collecter offre un coup d'oeil utile sur les résultats potentiels. Les quelques participants ont quand même nommé les mêmes termes que ceux nommés le plus souvent dans d'autres langues. Ces termes, lesquels représentent les noms d'objets que l'on voit dans les images, semblent bien être ceux qui viennent à l'esprit lorsque les participants regardent ces images.

Discussion

Les difficultés de collecte de données sont en partie causées par les moyens limités disponibles pour effectuer cette recherche. Nous dépendions en grande partie des bénévoles qui ont créé les versions du site web en diverses langues, d'autres bénévoles qui ont essayé de recruter des participants des diverses communautés linguistiques, ainsi que d'autres qui étaient nos participants. Malgré ces contraintes, l'expérience que nous avons gagnée au cours

des années avec ce genre de données nous permet d'être confiants de la justesse de nos résultats. Nous cherchions des données empiriques pour soutenir la question théorique de la robustesse de l'indexation multilingue automatisée pour des images « ordinaires ».

Ultimement, il s'agit d'un type d'indexation qu'on verra de toute façon, avec le temps, en faisant abstraction de notre travail. Par ailleurs, ce sera certainement automatisé, puisque les coûts pour l'effectuer autrement sont trop élevés. La tâche d'étiqueter les images est de nommer les objets significatifs que l'on voit dans l'image. Dans nos travaux antérieurs, nous avons démontré que presque n'importe qui peut nommer les objets dans les images de ce type aussi bien que le font les indexeurs professionnels, et que quelques termes seulement sont nommés par la plupart des gens. Une fois qu'on établit les termes dans une langue, la qualité de l'indexation dans les autres langues dépend de la qualité des outils de traduction.

Malgré le peu de données obtenues dans des langues avec un alphabet autre que l'alphabet romain, ces données sont néanmoins pertinentes. Il est raisonnable d'imaginer que d'autres données éventuelles suivraient les mêmes tendances que celles provenant d'autres langues. Dans ce contexte, les données provenant des participants sinophones sont particulièrement significatives, à cause du grand nombre de participants. Puisque les données chinoises suivent exactement les mêmes tendances que celles provenant de langues occidentales, on peut probablement conclure qu'il s'agit d'un phénomène universel, lequel sera vrai pour à peu près n'importe quelle langue.

Cette recherche contribue à valider d'autres approches, dont l'une cherche à propager les termes d'indexation et d'autres métadonnées d'une image à une image similaire, donc d'un objet similaire. Afin d'effectuer une telle manoeuvre, des techniques de haut niveau, par exemple des termes créés par des êtres humains, peuvent être combinées avec des techniques de bas niveau, par exemple la reconnaissance d'objets telle qu'employée par la technologie Photosynth (Aguera y Arcas 2007). En employant une telle stratégie, et en reliant un grand nombre de photos de détails d'un modèle général, de multiples photos trouvent leur place dans la zone appropriée du modèle. À partir de cette situation, lorsqu'un utilisateur ajoute des étiquettes pour identifier une image d'un détail particulier, d'autres photos du même détail peuvent recevoir automatiquement les mêmes métadonnées. En ajoutant une autre étape, celle de filtrer les métadonnées à travers des dictionnaires bilingues et multilingues comme nous l'avons fait dans le cadre de notre recherche, les images qui reçoivent les métadonnées par propagation pourraient aussi les recevoir dans de multiples autres langues. Ainsi, l'image peut être cherchée par une communauté d'utilisateurs beaucoup plus large.

Nos recherches sur l'indexation plan par plan de l'image en mouvement indiquent que les mêmes tendances s'appliquent. Ainsi, le champ de recherches utiles à effectuer devient plus large encore. On peut imaginer beaucoup d'autres utilisations de ces connaissances. Ainsi, l'offre de sujets de recherche est riche, surtout pour les chercheurs qui parlent dix ou douze langues (!).

Conclusion

Nous cherchions à obtenir des données empiriques pour soutenir le concept de l'indexation multilingue automatique d'images par la traduction automatique d'étiquettes par des dictionnaires bilingues et multilingues. Ainsi, l'utilisateur pourrait chercher des images en saisissant des noms d'objets en n'importe quelle langue. Dit autrement, les noms probables pour un objet que donneront les participants provenant d'une variété de groupes linguistiques

sont les mêmes que générera automatiquement un outil web lorsqu'on saisit un nom d'objet dans une seule langue. Les outils sont imparfaits mais ils sont largement utilisés par les internautes. Dès que le terme à l'origine est juste, la traduction automatique dans d'autres langues du terme d'indexation pour les images « ordinaires » est juste également. Ce phénomène offre une mesure des progrès technologiques actuels. On n'aurait pu imaginer une telle approche même aussi récemment que le début du présent siècle. De toute évidence, le développement d'outils web va continuer à offrir aux chercheurs un terrain fertile pour l'étude de l'indexation multilingue automatique d'images.

Remerciements

Nous avons effectué une partie de cette recherche à l'aide de la subvention 410-2005-1213 du Conseil de recherche en sciences sociales et humaines du Canada (CRSH). Ce rapport final du projet est dédié à la mémoire de Michelle Gauthier, largement responsable de l'obtention de cette subvention. Nous avons bénéficié également de beaucoup d'aide de la part de bénévoles. Nous remercions nos partenaires traducteurs pour la création des diverses versions du site, ainsi qu'Arnaud d'Alayer, Gabriel Coder et Robin Millette pour leur soutien technique. Nous sommes reconnaissants envers chaque participant qui a contribué aux données. Nous remercions particulièrement 潘玥玲 Marion, 陈祎 Camille, 范秉馨 Vincent et 杨珺 Anaïs, personnel de la bibliothèque de l'Alliance française de Shanghai pour leur version française des concepts exprimés en langue chinoise.

Références

- Aguera y Arcas, Blaise. 2007. Blaise Aguera y Arcas demos Photosynth [video]. *Ted Talks*. <http://www.ted.com/index.php/talks/blaise_aguera_y_arcas_demos_photosynth.html>.
- Furner, Jonathan. 2007. L'indexation des ressources des bibliothèques par les usagers : vers un modèle d'évaluation. *World Library and Information Congress (IFLA) : 73rd IFLA General Conference and Council, 19-23 August 2007, Durban, South Africa, Proceedings*. <<http://archive.ifla.org/IV/ifla73/papers/157-Furner-trans-fr.pdf>>.
- Hudon, Michèle, James Turner et Yves Devin. 2001. Description et indexation des collections d'images en mouvement : résultats d'une enquête. *Documentation et bibliothèques* 47, no. 1, 5-12.
- Josso, Guillaume, and Karine Lespinasse-Sabourault. 2008. Towards the empowerment of the reader: a new approach in a traditional public library: cross-cultural challenge in the French Language centre — Hongkou Spare-Time College. *SILF 2008, Shanghai*, 265-276.
- Kellogg Smith, Martha. 2006. Viewer tagging in art museums: comparisons to concepts and vocabularies of art museum visitors. In *Advances in classification research, Vol. 17: Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop (Austin, TX, November 4, 2006)*, ed. Jonathan Furner and Joseph T. Tennis.
- Kipp, Margaret E.I. 2009. Exploring inter tagger consistency measures [poster]. In *20th Annual SIG/CR Classification Research Workshop, American Society for Information Science and Technology, Vancouver, BC, 6-11 November 2009*. <<http://eprints.rclis.org/17218/1/sigcrposter2009.pdf>>.
- Lespinasse-Sabourault, K. 2006. « L'Heure de Vérité » de la linguistique informatique, Institut National de l'Audiovisuel ». *Cahiers de l'IUT de l'université Paris 5 - René Descartes*, numéro spécial.
- Markey, Karen. 1988. Access to iconographical research collections. *Library Trends* 27, no. 2, 154-174.

- Markey, Karen. 1986. *Subject access to visual resources collections: a model for computer construction of thematic catalogs*. New York: Greenwood Press.
- Ménard, Elaine. 2007. Indexing and retrieving images in a multilingual world. *Knowledge Organization* 34, no. 2, 91-100.
- Ménard, Elaine. 2006. La recherche d'information multilingue. *Documentation et bibliothèques* 52, no. 4, 255-261.
- Steve: the museum social tagging project. 2011. <<http://www.steve.museum/>>.
- Turner, James M., Karine Lespinasse and Claire Nigay. 2010. IconoTag, an experiment in multilingual picture indexing. *City life and libraries: the proceedings of the fifth Shanghai (Hangzhou) International Library Forum (SILF), August 24-27, 2010*. Shanghai: Shanghai Scientific and Technological Publishing House, 55-62.
- Turner, James M. et Claire Nigay. 2010. IconoTag, un pas vers l'indexation translinguistique d'images. 38e congrès de l'ACSI/CAIS, Congrès des sciences sociales et humaines. Concordia University, Montréal, 2010.06.04. <http://www.cais-acsi.ca/proceedings/2010/CAIS056_TurnerNigay_Final.pdf>.
- Turner, James M. et Suzanne Mathieu. 2007. Audiovision et indexation de films. *World Library and Information Congress (IFLA) : 73rd IFLA General Conference and Council, 19-23 August 2007, Durban, South Africa, Proceedings*. <http://archive.ifla.org/IV/ifla73/papers/157-Turner_Mathieu-trans-fr.pdf>.
- Turner, James M. and Michèle Hudon. 2002. Multilingual metadata for moving image databases: preliminary results. *Advancing Knowledge: Expanding Horizons for Information Science, Proceedings of the 30th annual conference of the Canadian Association for Information Science*, ed. Lynne C. Howarth, Christopher Cronin, Anna T. Slawek. Toronto: Faculty of Information Studies, 34-45.
- Turner, James M. et Jean-François Roulier. 1999. La description d'images fixes et en mouvement par deux groupes linguistiques, anglophone et francophone, au Québec. *Documentation et bibliothèques* 45, no. 1 (janvier-mars), 17-22.
- Turner, James M. 1995. Comparing user-assigned terms with indexer-assigned terms for storage and retrieval of moving images: research results. *Proceedings of the 58th ASIS Annual Meeting, Chicago, Illinois, October 9-12, 1995*, vol. 32, 9-12.
- Turner, James Ian Marc. 1994. Determining the subject content of still and moving image documents for storage and retrieval: an experimental investigation. PhD thesis, University of Toronto. 300 p.
- Zhang Zhixiong, Li Sa, Wu Zhengxin, and Lin Ying. 2006. Towards constructing a Chinese information extraction system to support innovations in library services. *World Library and Information Congress : 72nd IFLA General Conference and Council, 20-24 August 2006, Seoul, Korea*. <http://archive.ifla.org/IV/ifla72/papers/097-Zhixiong_Sa_Zhengxin_Ying-en.pdf>.

Annexe A: Tableau 2. Les trois termes les plus fréquents, pour huit langues, avec la fréquence de chaque terme.

	Anglais (127 participants)		Allemand (17 participants)		Suédois (26 participants)		Français (323 participants)		Espagnol (14 participants)		Portugais (7 participants)		Russe (2 participants)		Grec (2 participants)	
<i>Image</i>	terme, fréquence		terme, fréquence		terme, fréquence		terme, fréquence		terme, fréquence		terme, fréquence		terme, fréquence		terme, fréquence	
01	garbage clean street	101 77 90	Müll* Straß* Stadt*	20 15 6	gat* renhållning sop*	19 8 8	éboueur poubelle camion	188 137 115	basur* limpi* calle camión	19 6 5 5	limpeza rua lixo	7 7 5	мусоровоз	2	Σκουπιδιάρικο	3
02	peacock bird blue feather	127 70 33 26	Pfau* Vogel Zoo Tier	20 5 3 3	påfågel fågel	22 13	paon oiseau plume bleu	288 159 56 55	pavo real ave azul	10 5 3	pavão ave beleza cores penas	8 2 2 2 2	павлин птица	2 2	<i>no data</i>	
03	structure steel metal	44 36 35	Stahl* *Konstruktion Träger	11 8 6	Konstruktion Byggnad Metall	10 6 5	métal architectur* sculpture	150 80 43	estructura arquitectura cielo metal	6 4 4 4	estrutura* metal ponte	3 2 2	<i>no data</i>		<i>no data</i>	
04	camera tripod shadow	116 89 54	*Kamera Foto* Stativ	12 10 7	Kamera Stativ Skugg	28 18 9	photo appareil trépied	344 213 115	camara foto sombra tripode	10 7 4 4	foto* máquina tripé	10 4 3	фотоаппарат	4	Κάμερα	2
05	beach palm ocean arcade	105 45 31 30	Strand* *Wolk Palmen Meer	13 6 4 4	Strand Palme Hav	22 12 9	plage mer palmier	216 179 133	playa palmera* mar vacaciones	9 4 3 3	praia mar palm	8 4 3	море	2	Παραλία	2
06	light photo color	83 21 18	Lichter* Nacht* Feuer*	8 6 4	Ljus* Neon Foto*	15 9 6	lumière effet nuit	184 61 45	luc* colores foto neon	7 3 3 3	luz arte festa 3fogo de artifício	4 2 2 2	<i>no data</i>		<i>no data</i>	
07	cow farm field Holstein black white	123 23 23 23 22 22	Kuh* Land* Weide*	11 4 3	Ko Kviga Svart*	19 6 6	vache pré lait	281 100 44	vaca animal pasto prado	12 5 3 3	vaca pasto leite* prado	6 3 2 2	<i>no data</i>		<i>no data</i>	
08	dance costume wom*n	117 44 18	Tanz* Asiatisch Frauen	16 5 2	Dans* Dräkt Asiatisk	21 6 4	danse costume tradition	308 120 63	bail* danza* cultura mujer* tradicional	12 6 3 3 3	dança bailarinas	9 2	Танец* экзот	5 2	<i>no data</i>	
09	cact* desert sky	117 89 20	Kakt* Wüste* Landschaft Himmel	13 8 2 2	Kaktus Öken Landskap	19 12 8	cactus desert ciel	267 216 49	desierto cactus azul cielo	9 8 2 2	cact* deserto	7 3	Кактус	2	<i>no data</i>	
10	pumpkin harvest Hallowe*en	109 36 29	Kürbis* Herbst Halloween	13 7 5	Pump* Orange Grönsak Halloween Höst	29 7 4 4 4	citrouille Halloween orange	217 102 85	calabaza naranja Halloween	10 5 3	abóbora feira	10 2	Тыква	2	<i>no data</i>	

