

# **IconoTag, un pas vers l'indexation translinguistique d'images**

**Résumé :** On cherche des images efficacement à l'aide du texte. Ce papier décrit une expérience qui permet de démontrer que lorsque l'indexation est effectuée dans une seule langue, c'est possible de générer automatiquement l'indexation de bonne qualité en d'autres langues. Bien que les résultats présentés ici sont préliminaires, ils sont très encourageants.

**Abstract:** Text is still the most efficient way to search for images. This paper describes an experiment that demonstrates how once an image is indexed in a single language, good quality indexing in other languages can be generated automatically. The results presented here, although preliminary, are strong and encouraging.

## **1. Introduction**

Au cours des années, nous avons étudié la question d'indexation bilingue pour l'image fixe et en mouvement (par ex. Turner et Roulier 1999, Turner et Hudon 2002, Da Sylva et Turner 2006). Dans le présent projet, nous étendons l'envergure de ces travaux à l'indexation translinguistique. Il convient de distinguer l'indexation translinguistique de l'indexation bilingue ou multilingue. Lorsque l'on parle d'indexation multilingue, on veut dire que l'objet indexé l'est dans plusieurs langues à la fois, avec correspondances entre les termes dans ces langues. Dans le cas d'indexation bilingue, il s'agit évidemment de deux langues seulement, toujours avec correspondance. L'indexation translinguistique, par contre, s'applique par exemple dans le cas d'objets indexés en plusieurs langues, mais dans l'isolement, sans qu'il y ait nécessairement correspondance entre les langues. Ainsi, pour une même image, par exemple, l'indexation se fera au Japon en japonais, en Australie en anglais, au Maroc en arabe. Autrement dit, il s'agit d'indexation unilingue, mais dans plusieurs langues.

L'approche la plus utile pour la recherche d'images est toujours le texte. En employant cette approche, il faut que la chaîne de caractères de la requête corresponde à celle de l'index inversé de la base de données cherchée pour qu'il y ait repérage. Or, le World Wide Web offre la possibilité de recherches simultanées dans plusieurs bases de données à la fois. Dans le modèle que nous visons, la requête est d'abord « filtrée » par des dictionnaires bilingues, afin d'identifier l'équivalent en d'autres langues, avant de trier les requêtes par langue et les cibler vers des bases de données correspondantes partout au monde. Ainsi, une requête est présentée à chaque base de données dans sa langue d'indexation. Les résultats de toutes les requêtes sont ramassés et présentés à l'utilisateur. L'utilisateur cherche donc dans sa langue, et le système lui répond en offrant des images indexées en plusieurs langues.

Nous avons démontré que la même indexation s'applique indifféremment aux images fixes ou en mouvement lorsque ces dernières sont indexées plan par plan (Turner 1994). Ainsi, nos résultats s'appliqueront tant aux images en mouvement qu'aux images fixes. Finalement, nous croyons que cette approche offrira également une variante intéressante, soit la propagation automatique d'indexation translinguistique d'images fixes et en

mouvement vers d'autres images, une fois qu'une seule image est indexée dans une seule langue. Afin de tester l'approche que nous décrivons ici et d'obtenir des données empiriques pour appuyer ou infirmer notre hypothèse, nous avons construit une petite application web intitulée IconoTag, puis nous avons invité des internautes de neuf groupes linguistiques à participer à notre recherche.

## 2. Méthodologie

Comme son nom le suggère, IconoTag consiste en un site web où l'utilisateur ajoute des étiquettes (*tags*) à des images. Nous avons mis le site en ligne pour le mois de mars 2010 et avons invité des participants à visiter le site et étiqueter les images. Afin d'encourager la participation, nous avons envisagé une tâche facile et rapide à accomplir. Nous avons également décidé de ne pas demander d'informations concernant les participants car dans nos projets précédents, il n'y avait pas de différences significatives entre les réponses d'étudiants ou personnes avec une expérience de travail, entre participants dits visuels et non visuels (Turner 1994) et ce, peu importe le sexe et l'âge.

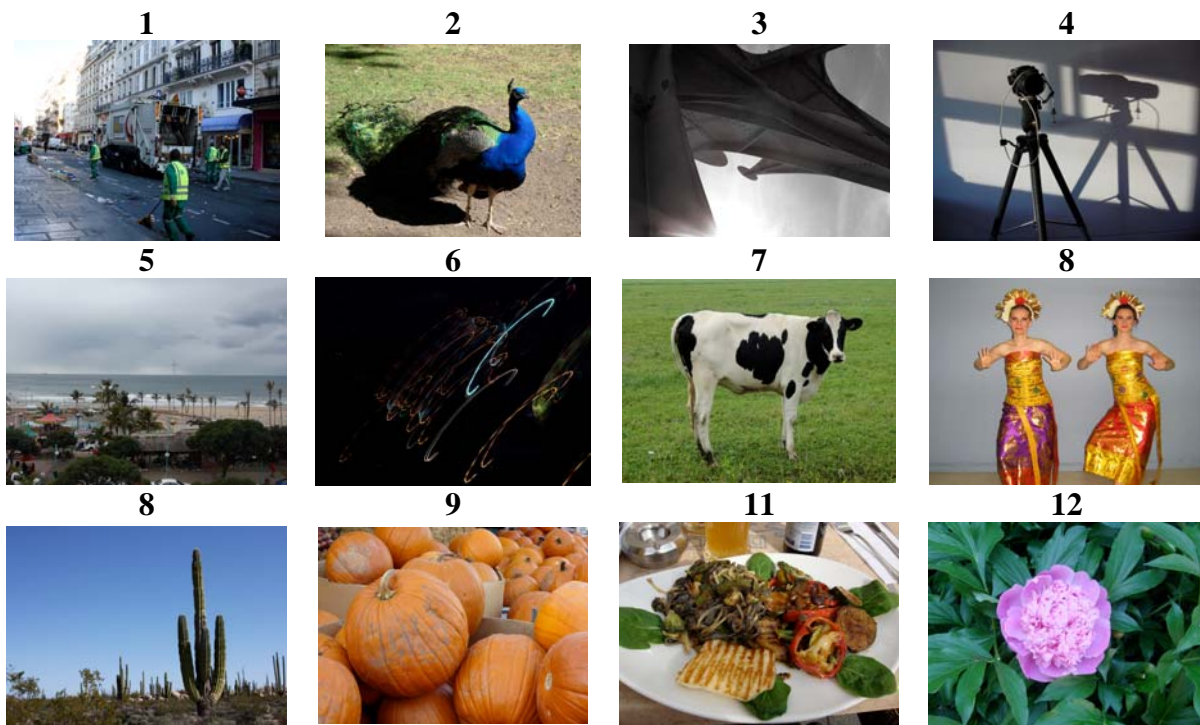


Figure 1. Les images présentées aux participants d'IconoTag.

Nous avons choisi un ensemble de douze images pour inclusion dans IconoTag (figure 1). Nos choix se sont faits sur la base de simplicité ou de complexité de l'image, c.-à-d. selon notre définition (Turner 1994, 7), s'il y a deux objets différents ou moins à nommer, ou s'il y en a plusieurs. Nous cherchions à nous assurer que les objets soient facilement reconnaissables car nous voulions éliciter des réponses simples et spontanées de la part des participants, qu'ils nomment tout simplement les objets visibles dans les photos. Comme contrôle, deux des douze images sont difficilement identifiables, dans l'un cas un gros plan sur un détail d'une sculpture (image 3), dans l'autre cas une image de trajet de lumières (image 6), prise en bougeant l'appareil photo pendant la prise de photo le soir. Finalement, nous n'avons choisi que des images d'orientation paysage, soit l'orientation des écrans d'ordinateur, afin d'assurer l'uniformité de la taille.

IconoTag consistait en neuf sites parallèles, un pour chaque langue. Les langues inclus dans le projet sont les suivantes : français, anglais, allemand, espagnol, portugais, arabe, chinois, suédois et grec. Le site était installé sur un serveur ouvert avec une base de données MySQL et des formulaires codés en PHP pour la collecte de données.

### 3. Résultats

En tout, 999 internautes ont répondu à l'appel en visitant le site d'IconoTag. Parmi eux, 561 ont complété l'exercice et donc fourni des données. Le tableau 1 décrit le nombre de participants par langue, avec le nombre total de termes fournis par chaque groupe et le nombre moyen de termes par plan pour chaque groupe linguistique.

Tableau 1. Nombre de participants par langue

Langue	Nombre de participants	Nombre total de termes	Nombre moyen de termes par plan
français	323	10 126	2,6
chinois	88	1 637	1,6
allemand	11	203	1,5
anglais	127	3 712	2,4
espagnol	11	318	2,4
arabe	0	0	0
grec	0	0	0
portugais	1	10	0,8
suédois	0	0	0
TOTAL	561	16 006	2,4

En tout, 16 006 termes ont été enregistrés, pour une moyenne brute de 2377 termes par image. Pour deux groupes linguistiques, soit le grec et le suédois, il n'y a aucun répondant. Pour le portugais, il n'y a qu'un seul. Il y avait 28 répondants en langue arabe, mais un problème technique dans l'encodage, découvert trop tard dans le processus, a rendu les données inutilisables. Pour le mandarin, l'analyse n'a pas pu être complétée à temps pour l'inclure ici.

Nous offrons ici une première analyse. Pour les langues dont les données peuvent être analysées, on observe que les réponses suivent la distribution de Zipf, soit que la plupart des cellules de la matrice sont vides, la grande majorité des termes uniques sont nommés qu'une seule fois, et que peu de termes sont nommés très souvent. Ce sont ces derniers qui nous intéressent pour l'analyse, car ils sont les candidats les plus probables pour l'indexation des images correspondantes. Le tableau 2 présente les termes les plus souvent nommés. Il s'agit soit de termes complets, soit de fragments de termes plus larges.

Tableau 2. Les quatre termes nommés le plus souvent, pour quatre langues. Les termes sont classés par fréquence.

Image	français (323 répondants) terme, cas	anglais (127 répondants) terme, cas	allemand (11 répondants) terme, cas	espagnol (11 répondants) terme, cas
01	éboueur 188 poubelle 137 camion 115 nettoy* 109	garbage 101 clean 77 street 90 truck 47	Müll 11 Straße 7 rein* 3 Stadt 2	basur* 11 limpi* 5 calle 3
02	paon 288 oiseau 159 plume 56 bleu 55	peacock 127 bird 70 blue 33 feather 26	Pfau 10 Vogel 4 Feder 2 Zoo 2	pavo real 7 ave 3 azul 2
03	métal 150 architectur* 80 sculpture 43 acier 42	structure 44 steel 36 metal 35 architecture 28	Stahl 8 Brücke 3 Architektur 2 Schwarz-Weiß 2	estructura 5 arquitectura 2 cielo 2 metal 2
04	photo 344 appareil 213 ombre 127 trépied 115	camera 116 tripod 89 shadow 54 photo 50	Foto 7 Kamera 6 Stativ 3 Selbstauslöser 2	foto 7 camara 6 sombra 3
05	plage 216 mer 179 palmier 133 bord de mer 55 vacances 55	beach 105 palm 45 ocean 31 arcade 30	Strand 6 Palmen 4 Wolken 4 Vergnügung* 2	playa 6 mar 3 vacaciones 3 palmera 2 turismo 2
06	lumière 184 effet 61 nuit 45 abstra* 36	light 83 photo 21 colo*r 18 night 15	Licht* 4 Nacht* 3 Bewegung 2	luz/luces 7 colores 3 foto* 3 neon* 2
07	vache 281 pré 100 lait 44 campagne 43 animal 33	cow 123 farm 23 field 23 Holstein 23 black 22 white 22 animal 17	Kuh* 6 Bauernhof 2 Holstein 2 Tier 2	vaca 8 animal 4 blanco 2 negro 2
08	danse 308 costume 120 tradition 63 Bali 47	dance 117 costume 44 wom*n 18 Asia* 15	Tanz 9 Indisch 2	bail* 8 danza 4 cultura 3 mujer 2
09	cactus 267 desert 216 ciel 49 arid* 46	cact* 117 desert 89 sky 20 blue 13	Kakt* 8 Wüste 5	desierto 6 cactus 5 azul 2 calor 2
10	citrouille 217 Halloween 102 orange 85 potiron 79	pumpkin 109 harvest 36 Hallowe*en 29 orange 28	Kürbis 7 Halloween 3 Herbst 2 Markt 2 orange 2	calabaza 6 naranja 3 Halloween 2 otoño 2 zapallos 2
11	plat 123 repas 98 assiette 90 légume 86	food 59 vege* 45 meal 33 dinner 27	essen 5 Gericht 3 Fisch 2 grill 2	comida 7 restaurante 3 plato 2

			Restaurant	2		
12	fleur 260	flower 100	Blume	4	flor	5
	pivoine 80	pink 49	Pfingstrose	3	naturaleza	3
	rose 75	peon*	Blüte	2		
	feuille 52	purple 19	lila	2		
			Pflanze	2		

---

L'astérisque indique soit la troncature, soit une ou des lettres manquantes. Dans les cas où il y a le même nombre de cas pour deux termes, ils sont donnés dans l'ordre alphabétique. Dans l'allemand et l'espagnol, lorsqu'il n'y a que deux ou trois termes, c'est que par la suite, il y a d'autres termes mais nommés une seule fois. Pour l'image 4, on voit qu'il y a 344 cas du mot « photo » en français alors qu'il n'y a que 323 répondants pour cette langue. Cela s'explique par le fait que quelques répondants ont donné plus d'un terme contenant ce mot.

#### 4. Discussion et conclusion

Cette première analyse nous permet de constater des taux de correspondance très encourageants. Nous n'avons pas encore effectué une comparaison approfondie. Toutefois, on constate déjà que pour 7 photos sur 12, et pour les 4 langues dont les résultats sont présentés ici, il y a correspondance directe entre le terme en première position. Il s'agit des images 2, 5, 6, 7, 8, 10 et 12. De plus, l'une de ces images, le no. 5, est une image complexe, et une autre, le no. 6, est l'une des images abstraites, ajoutées comme contrôle. Dans trois autres cas (les images 1, 4 et 9), il y a correspondance directe entre trois langues sur quatre pour le terme en première position. Par ailleurs, on voit que lorsqu'on inclut les 4 premiers termes dans l'analyse, le taux de correspondance est très élevé entre ces quatre langues.

Comme on voit dans les tableaux 1 et 2, il y a relativement peu de participants de langues allemande et espagnole. Toutefois, même avec le peu de données pour ces langues figurant au tableau 2, on peut constater des correspondances intéressantes avec le français et l'anglais. Par contre, on observe que l'allemand s'apparente à l'anglais et l'espagnol au français dans la classification des langues, alors que nous souhaitons dans ce projet étudier les correspondances avec des langues autres que celles utilisant l'alphabet romain. Ainsi, c'est regrettable que nous n'avons pas de données en arabe ni en grec, malgré nos efforts. Toutefois, nous comptons bien poursuivre ce travail, et plus tard pourrons présenter d'autres données, dont une comparaison avec les réponses en mandarin. Cette analyse nous offrira un portrait plus complet et des indications plus fiables sur les possibilités pour l'indexation d'images translinguistique et automatisée.

#### 5. Remerciements

Cette recherche a été faite à l'aide de la subvention 410-2005-1213 du Conseil de recherche en sciences humaines du Canada (CRSH). Nous remercions également nos partenaires traducteurs vers les diverses langues employées dans l'étude, messieurs Arnaud d'Alayer, Gabriel Coder et Robin Millette pour leur aide technique, ainsi que tous les participants qui ont contribué à la masse de données dont nous avons fait la collecte.

#### 6. Références

Da Sylva, Lyne et James M. Turner. 2006. Using ancillary text to index web-based multimedia objects. *Literary and linguistic computing* 21, no. 2, 219-228.  
 <<http://llc.oxfordjournals.org/cgi/reprint/21/2/219.pdf>>.

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.

Turner, James M. et Michèle Hudon. 2002. Multilingual metadata for moving image databases: preliminary results. *L'avancement du savoir : élargir les horizons des sciences de l'information, Travaux du 30e congrès annuel de l'Association canadienne des sciences de l'information*, éd. 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.