



Icono Tag. Apoyo empírico para la indización automática de imágenes multilingües

James M Turner

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

Claire Nigay

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

Karine Lespinasse-Sabourault

Alliance Francaise of Shanghai, 上海法语培训中心,
Université Paris 8, Paris, France

Traducción: Alicia Garcia Medina

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

Resumen:

Los usuarios de los servicios de información buscan imágenes para diversos fines, por ejemplo para ilustrar un artículo sobre la comida, para explicar los animales a los niños, o estudiar los vehículos para el transporte. En trabajos anteriores, hemos visto que los hablantes nativos de inglés y francés añaden los correspondientes términos de indización a las mismas imágenes. Por otra parte el inglés y el francés tienen muchas similitudes incluyendo muchas palabras en común y un alfabeto común. Icono Tag era un proyecto que trataba de conocer si el alto nivel de correspondencia también es válido para otros idiomas, especialmente si las raíces de la lengua son completamente diferentes o si el lenguaje utiliza otro alfabeto o ideogramas. Hemos construido una aplicación basada en web para recopilar los datos. A los participantes se les pidió que etiquetaran doce imágenes escogidas y utilizando los criterios establecidos en proyectos anteriores incluyendo sencillez o complejidad de la imagen, el número de objetos que pueden indicarse, la dificultad o facilidad para identificar el contenido, y así sucesivamente.

Dos de las imágenes eran abstractas y se incluyeron como control. El sitio se puso en línea en 2010 y se utilizaron las redes sociales y los contactos personales para reclutar a los participantes para que etiquetaran las doce imágenes en uno de los diez idiomas: árabe, chino, Inglés, francés, alemán, griego, portugués, ruso, español y sueco. Los resultados indican los altos índices de correspondencia lo que corrobora la hipótesis de que una indización automática multilingüe de imágenes es totalmente viable y que los usuarios de

muchas comunidades podrían encontrar imágenes útiles a partir de bases de datos indexadas en cualquier idioma, independientemente del lenguaje que se utilicen para la búsqueda.

Introducción

Un conjunto de nuestros proyectos de investigación han estudiado el tema de la indización de imágenes fijas y móviles. Hemos demostrado que tanto si una imagen es fija como en movimiento, los usuarios utilizan esencialmente las mismas etiquetas para la misma imagen (por ejemplo, Turner 1994, 1995). Nos mostró también que los nativos de habla inglesa o francesa vinculan los mismos términos de indización a las mismas imágenes. Nombran a los objetos que ven en las imágenes y los términos utilizados se traducen directamente en el equivalente de otro idioma. (e.g. Turner et Roulier 1999, Turner and Mathieu 2007). Nuestros estudios mostraron una gran consistencia en los niveles de correspondencia. Esto sugiere que las imágenes pueden ser indizadas en los dos idiomas, y la indización en otro idioma se puede realizar de forma automática. El alto índice de consistencia indica que la indización automática generada sería de la misma calidad que la indización original. Otros estudios referentes a la indización de imágenes fijas en inglés y francés fueron llevados a cabo por Ménard Ménard (2006, 2007).

Proseguimos con nuestros estudios y con otros estudios utilizando los traductores de la web e.g. Hudon, Turner et Devin (2001); Turner and Hudon (2002). Los resultados de estos estudios nos llevaron a creer que con las correspondientes herramientas de Internet, sería posible indizar imágenes fijas y en movimiento en un solo idioma y, a continuación, generar la indización de forma automática en muchos otros idiomas. Sin embargo, la indización multilingüe no es necesariamente una necesidad de los usuarios más elementales de cualquier imagen de una base de datos por lo que no es probable que los que manejan estas bases de datos la pongan en práctica. La necesidad es mayor por parte de los usuarios que visitan la base de datos a través de la web.

Los usuarios de la web escriben una palabra en un motor de búsqueda y el motor entonces comienza la búsqueda para encontrar la imagen correspondiente a la búsqueda requerida. Los motores de búsqueda se podrían programar para filtrar las palabras claves del usuario por medio de vocabularios bilingües que incluyan una serie de idiomas, entonces se reenviará la consulta a las bases de datos indizadas con los diferentes idiomas, entonces se agrupan los resultados y se mandan las imágenes al usuario. Pensamos que este sistema funcionará bien tan pronto como existan suficientes diccionarios bilingües y estén disponibles en la web y además sean de libre acceso y se construya un algoritmo de búsqueda. Una serie de diccionarios bilingües y otras herramientas como vocabularios bilingües o multilingües ya están disponibles en la web, al igual que algunos sitios web de carácter experimental que están tratando de experimentar con la indización multilingüe de las imágenes. En un entorno multilingüe, el problema entre el inglés y el francés es que los idiomas son muy similares, con muchas palabras en común y un con un alfabeto común. ¿Qué sucedería, por el contrario, si no hay correspondencia reconocible entre palabras en lenguas que utilizan una variedad de caracteres o alfabetos?

Hemos demostrado en un estudio anterior (Turner, 1995) que para este tipo de imagen ("normales" en lugar de las imágenes documentales de arte), los usuarios y catalogadores profesionales utilizan los mismos términos para las mismas imágenes. Otro aspecto es el relacionado con las imágenes en movimiento tema éste que se puede encontrar en Lespinasse-Sabourault (2006). Las imágenes de arte son mucho más complejas, debido a los

diferentes niveles de interpretación necesaria para su estudio, y tradicionalmente no se han indizado ni tan siquiera a un nivel básico donde se identifiquen con facilidad los objetos en la imagen (Markey 1986, 1988). Es interesante destacar que últimamente es frecuente describir las imágenes de arte añadiendo términos de indización como los que podemos encontrar en la web y en los museos (Kellogg Smith 2006, Steve 2011).

En este trabajo, se utiliza el término "etiquetas" de manera intercambiable con los "términos de indización" o "palabras clave". Etiquetas o términos de indización pueden estar compuestos por una o varias palabras. La cuestión de si las etiquetas constituyen una indización buena o no, no es un tema para discutir aquí. Otros investigadores estudian este aspecto (por ejemplo, Furner 2007, Kipp 2009) y, como hemos constatado en nuestros propios estudios, parte de la respuesta depende del tipo de materiales que son indizados. Aquí nos centramos en las imágenes fijas que se pueden considerar en su mayoría como imágenes "normales", pero la clasificación es un tanto arbitraria, ya que las mismas imágenes también se pueden considerar como imágenes documentales o incluso imágenes de arte, dependiendo de su contexto.

Metodología

Con el fin de estudiar la cuestión de si las etiquetas que figuran en un idioma pueden traducirse de forma automática en etiquetas útiles para otros idiomas, se recogieron los datos mediante un sitio web llamado IconoTag, que hemos construido en francés. Luego, utilizando la red social Facebook, hemos solicitado ayuda para construir diez versiones idénticas en diez lenguajes diferentes. Los idiomas son el árabe, el chino, el inglés, el francés, el alemán, el griego, el portugués, el ruso, el español y el sueco. El sitio se alojó en un servidor que ofrece una base de datos de SQL back-end para el almacenamiento de los datos recogidos mediante formularios web PHP. La página principal ofrece enlaces a las diez lenguas. En el primer clic (la elección de una lengua), conduce a las miniaturas de las 12 imágenes que se deben marcar (figura 1), así como una breve explicación del proyecto de investigación y las instrucciones para los participantes. Un nuevo clic permite acceder a la siguiente pantalla que ofrece una explicación más detallada del proyecto, información sobre cómo ponerse en contacto con los investigadores, si se desea, información sobre los requisitos de la universidad, sus principios, y un clic de un compromiso para participar en el proyecto.



Figura 1. las doce imágenes escogidas marcadas en cada uno de las 10 lenguas

Una vez que los participantes hacen clic en el botón de compromiso, se pasa a la primera imagen, mucho más grande ahora y, en la misma pantalla, está la opción para marcar las imágenes. Una vez que se ha terminado el etiquetado de la imagen hay que hacer clic en un botón "Enviar" que lleva a una página con la siguiente imagen y una nueva etiqueta, y así sucesivamente hasta las doce imágenes que se deben marcar. Estos datos se incorporan y graban en el back-end de la base de datos, y es en ese momento cuando se pueden exportar y analizar.

Para fomentar la participación, se creó una tarea fácil y que pudiera llevarse a cabo en pocos minutos. Las doce imágenes que utilizamos fueron seleccionadas de una base de datos que reunía una serie de criterios establecidos en proyectos anteriores, incluyendo la sencillez o complejidad de la imagen, el número de cosas que se pueden indizar, facilidad o complejidad a la hora de identificar el contenido. Se incluyeron dos imágenes abstractas (números 3 y 6) pensando incluso que si se daba una elevada tasa de correspondencia entre las etiquetas de las imágenes, incluso los términos más populares se dispersarían ante estas dos imágenes. La imagen 3 presenta un detalle de una escultura estable (a diferencia de móvil) de Alexander Calder, titulada "L'Homme" ("El hombre") instalada en el recinto de la Feria Mundial de Expo67 en Montreal. La imagen recoge 6 puntos geométricos de luz originados al mover una cámara digital cuando se toma una foto en la noche. Elegimos sólo imágenes nítidas de paisaje para reducir el riesgo de que la nitidez se convirtiera en una variable que pudiera provocar diferencias en las respuestas. Elegimos paisajes en orientación vertical, debido a

pantallas de ordenador tienen esta orientación. De esta manera, las imágenes podrían ser tan grandes como fuera posible.

Sobre la base de los resultados anteriores (Turner, 1994), se decidió no recoger información personal con el fin de simplificar aún más la tarea y reducir el tiempo necesario para realizarla. Los estudios previos no mostraron diferencias significativas en la indización proporcionada por los usuarios comunes, ya fueran jóvenes o viejos, hombres o mujeres, con experiencia en el trabajo con imágenes o no, estudiantes o trabajadores, y así sucesivamente.

El sitio fue puesto en línea en marzo de 2010, y se dejó activo durante varios meses, hasta que consideramos que había suficientes datos para analizarlos. Las redes sociales y el contacto personal se utilizaron para reclutar a los participantes. Tenían que elegir un idioma, entonces indizar las doce imágenes. Las instrucciones les pedían que finalizaran la indización una vez que se describían las imágenes adecuadamente. La representación de cada imagen se describe de manera adecuada, con un máximo de cinco etiquetas por imagen.

Resultados y análisis

Se incluyen los resultados parciales de este proyecto expuestos en dos artículos anteriores (Turner y Nigay 2010; Turner, Lespinasse Sabourault y Nigay 2010). A continuación se analizan los resultados completos y definitivos. Puesto que ya se tenía un número bastante grande de participantes para los idiomas de inglés, francés y chino, se llegó a la conclusión de que los términos más importantes se mantienen estables, y por tanto no se hizo ningún otro tipo de análisis sobre estos datos. La Tabla 1 muestra el número de participantes cuyos datos se analizaron para cada idioma.

Tabla 1. Número de participantes para cada lengua

Idioma	Número de Participantes
Francés	323
Inglés	127
Chino	90
Sueco	26
Alemán	17
Español	14
Portugués	7
Árabe	2
Griego	2
Ruso	2
TOTAL	608

Las tablas 2, 3 y 4 muestran los tres primeros términos que figuran en cada uno de los idiomas. Cuando dos o más términos logran la misma puntuación, todos los términos que se hubieran utilizado se ordenan alfabéticamente. Los asteriscos indican que falta una o más letras, a fin de incluir los plurales, variantes ortográficas, y otras palabras compuestas de varias palabras.

Los datos se dividen en tres tablas, debido a las dificultades en su análisis de la representación. Está claro que todavía no han alcanzado la globalización en los programas

informáticos. Para los autores que hablan ni leen chino, se solicitó la ayuda para el análisis de los datos en este idioma del personal de la biblioteca en China. Para el árabe, de hecho hubo un total de 30 participantes, pero un problema técnico con el sistema de recopilación de datos hizo que los datos de 28 de estos participantes no se pudieran utilizar antes de descubrir el problema y ser capaces de arreglarlo, lo que nos dejó solamente los datos de dos participantes. Este hecho fue especialmente lamentable, puesto que el alfabeto árabe utiliza signos no romanos y los datos referentes a esta lengua hubieran sido muy valiosos para lograr los objetivos del proyecto. Los datos de estos dos idiomas se encuentran en tablas separadas.

La Tabla 2 muestra los datos de los restantes ocho idiomas de los que se recogieron los datos. Solamente hemos sido capaces de reclutar a dos participantes rusos y griegos, así que la información es más bien escasa, pero de todos modos se han incluido aquí porque es mejor que nada.

[tabla 2 de gran tamaño, aparece al final en el Apéndice A]

Como es habitual en el análisis de este tipo de datos, las respuestas siguen la distribución de Zipf, es decir, cuando las respuestas se compilan en una hoja de cálculo, la mayoría de las casillas de la matriz están vacías, la mayoría de las etiquetas se dan una sola vez, y unas pocas etiquetas se han repetido muchas veces. Aquellas que se habían repetido muchas veces son los que vamos a estudiar, porque son los términos de indización más útiles. En primer lugar se comparan los tres primeros términos, los que aparecen la mayoría de las veces. Es importante tener en cuenta los tres términos, independientemente de su clasificación, ya que si bien hay variaciones de un idioma a otro en cuanto a la clasificación, nos percatamos que para la mayoría de las imágenes, todos los idiomas o lenguas tienen casi todos los mismos tres primeros términos, o dos de los tres. Al final de la búsqueda, esto significa que un usuario que dio alguno de estos términos puede encontrar la imagen. Sin embargo, con este conjunto de datos, la correspondencia es en realidad mucho más precisa, ya que para las imágenes 2, 4, 5, 6, 7, 8, 9, 10, 11, 12 (10 de las 12 imágenes que corresponde al 83% de las imágenes), todas o casi todas las lenguas tienen el mismo término de indización en la primera posición. Esto es sorprendente, ya que sólo cuatro de estas imágenes son simples, con otras cosas que pueden describir, la clasificación del primer término ofrece una información valiosa sobre cómo las imágenes se perciben e interpretan tanto a nivel fisiológico como bajo las influencias culturales. A continuación nos fijamos en los datos referentes al tipo de imagen. Como había imágenes simples y complejas, el análisis basado en estos criterios es particularmente interesante. Las imágenes simples (un solo objeto importante para describir) corresponden a los números 2, 7, 10 y 12. Las imágenes complejas (varios objetos para describir, los objetos que compiten por la atención, o los objetos más difíciles de identificar o describir) son los correspondientes a los números 1, 4, 5, 8, 9 y 11. Las imágenes abstractas que se incluyeron son las de los números 3 y 6. Esperábamos que la correspondencia entre las lenguas fuera alta para las imágenes simples, más dispersa o totalmente dispersa para las imágenes complejas, y que no existiera correspondencia para las imágenes abstractas. Los resultados actuales son muy alentadores, ya que los tipos de correspondencia son mayores de lo esperado, a veces incluso sorprendentes. Como era de esperar, los participantes de todos los idiomas dieron el mismo término en primer lugar para las imágenes simples. Hay una única excepción a esto: para la imagen 2, Los participantes de habla árabe pusieron en primer lugar el término "pájaro" en lugar de "pavo real" y el pavo real apareció como el segundo término. Sin embargo, la excepción es poco significativa, ya que los datos que se habían utilizado correspondían solamente a dos participantes. Dado que los participantes podían indizar hasta con cinco términos cada imagen, "ave" estuvo en la

primera posición (4 veces) como el elemento principal que escogieron los dos participantes. No esperábamos que las imágenes complejas tuvieran altas tasas de correspondencia. Aquí nos detenemos para analizar en detalle la clasificación. Para la imagen 1, no hay un término único en la primera posición, pero los tres términos en su conjunto abarcan los mismos conceptos (basura, calles, limpieza). Para la imagen 4, "cámara" es el primer término en siete idiomas, y "foto" en 3 (francés, portugués y chino). Para la imagen 5, "playa" es el primero en nueve idiomas, y el "mar" en uno (ruso). Una vez más nos encontramos con una excepción poco significativa porque solo hay datos de dos participantes. Para la imagen 8, todas las lenguas tienen el mismo concepto (la danza) en la primera posición. Para la imagen 9, "cactus" es la primera en 9 idiomas, y "desierto" en 1 (español). "Cactus" estaba unido (una ocurrencia menos) para el primer puesto en español. Los datos del resto de los participantes probablemente favorecerían el hecho de que las posiciones de estos términos fueran más precisas. Para la imagen 11, "comida" es la primera en ocho idiomas, "plato" la encontramos en una ocurrencia (francés), y "cocinar" para otra (chino). Ya que había una cantidad considerable de datos de participantes que utilizaban el francés y el chino, estos resultados reflejan posiblemente interesantes diferencias culturales en la percepción.

Los resultados de las imágenes abstractas, que eran las imágenes control, fueron los más sorprendentes. Ahora nos preguntamos si fuimos ingenuos al esperar una amplia dispersión de los términos. Sin embargo, el hecho de que los participantes intentaron dar el nombre de lo que vieron, y el hecho de que habían visto las mismas cosas les llevó a poner el mismo nombre, en general, este hecho es muy alentador y sirve como un argumento a favor de la idea de que el etiquetado es útil para la indización de imágenes. Para la imagen 3, los términos elegidos en el primer lugar implican los mismos conceptos (estructura, de acero, metal, construcción, escultura). Para la imagen 6, "luz" el término en la primera posición es el mismo en todos los idiomas, excepto para árabe, en el que aparece "fuegos artificiales" como único término dado por los dos participantes. Otros términos que describían esta imagen eran los de "color" y "foto".

A continuación se analizan los resultados de los datos de los participantes de lengua china. Para el chino se llevó a cabo una actividad por separado. Hubo dificultades técnicas en la interpretación de las etiquetas en idioma chino dentro de la base de datos utilizada para la recopilación de los datos, por lo que el análisis no se podía hacer junto con el resto. Además, ninguno de los miembros del equipo del proyecto podía leer los caracteres chinos. Sin embargo, hemos sido capaces de reclutar la ayuda de chino-francés con el personal bilingüe de la biblioteca de la Alianza Francesa de Shanghai (上海法语培训中心图书馆), cuyo trabajo en temas interculturales se ha descrito en Josso y Lespinasse-Sabourault (2008). Este trabajo mostró que el análisis utilizando el francés como lengua intermedia funcionó muy bien para hacer una comparación en bruto.

Debido a los sistemas de escritura diferentes (Zhang et al. 2006), no fue posible utilizar el mismo método de truncamiento de una palabra o sustitución de letras para el análisis de los datos en idioma chino. La Tabla 3 presenta un análisis comparativo de los datos de cada imagen con términos en francés e inglés. Los asistentes de habla china hicieron una versión en francés de los conceptos expresados en chino. A continuación, mostramos las ocurrencias de los términos con la versión en francés de los términos chinos, utilizando el mismo método de truncamiento se había utilizado para analizar los términos en el resto de los idiomas occidentales. Dado que tanto las palabras en francés y en los caracteres chinos representan conceptos, el método, aunque algo torpe, da una imagen razonable de las correspondencias.

Table 3. The most popular tags from Chinese-speaking participants, using French as an intermediate language, and the most popular tags given by French and English speakers.

Image	Chinese (90 participants) term	French from Chinese term, frequency	English from Chinese terms via Google term	French (323 participants) term, frequency	English (127 participants) term, frequency
01	垃圾 清扫 街道 清洁工	ordures 46	garbage	éboueur 188	garbage 101
		nettoy* 42	cleaning	poubelle 137	clean 77
		rue 44	street	camion 115	street 90
		éboueur 29	cleaners	nettoy* 109	truck 47
02	孔雀 动物 蓝色 草地	paon 69	peacock	paon 288	peacock 127
		animal 12	animal	oiseau 159	bird 70
		bleu 9	blue	plume 56	blue 33
		gazon 7	lawn	bleu 55	feather 26
03	钢结构 建筑 飞机 天空	acier 19	steel	métal 150	structure 44
		bâtiment 14	building	architectur* 80	steel 36
		avion 14	aircraft	sculpture 43	metal 35
		ciel 11	sky	acier 42	architecture 28
04	摄影 光 照相机 影子 三角架	photo* 74	photography	photo 344	camera 116
		lumière 71	light	appareil 213	tripod 89
		appareil-photo* 45	camera	ombre 127	shadow 54
		ombre 26	shadow	trépied 115	photo 50
		trépied 30	tripod		
05	海滩 度假 游乐 阴天	plage 64	beach	plage 216	beach 105
		vacances 23	holiday	mer 179	palm 45
		loisir 13	recreation	palmier 133	ocean 31
		ciel gris 12	cloudy day	bord de mer 55 vacances 55	arcade 30
06	光 摄 光影 色彩	lumière 42	light	lumière 184	light 83
		photo* 15	photo	effet 61	photo 21
		éclairage 13	lighting	nuit 45	colo*r 18
		couleur* 11	colour	abstra* 36	night 15
07	奶牛 牛奶 草原 牧场	vache 62	dairy cow	vache 281	cow 123
		lait 11	milk	pré 100	farm 23
		prairie 11	grassland	lait 44	field 23
		pâturage 7	ranch	campagne 43 animal 33	Holstein 23 black 22 white 22 animal 17
08	舞蹈 异域风情 印度 泰国	danse 56	dance	danse 308	dance 117
		exotique 12	exotic	costume 120	costume 44
		Indien* 7	India*	tradition 63	wom*n 18
		Thaïlande 7	Thailand	Bali 47	Asia* 15
09	仙人掌 沙漠 干旱	cactus 49	cactus	cactus 267	cact* 117
		d*sert 21	desert	desert 216	desert 89
		sec* 8	arid	ciel 49 arid* 46	sky 20 blue 13
10	南瓜 万圣节 菜市场 丰收	citrouille 52	pumpkin	citrouille 217	pumpkin 109
		Halloween 14	Halloween	Halloween 102	harvest 36
		marché 14	vegetable market	orange 85	Hallowe*en 29
		moisson 8	bumper harvest	potiron 79	orange 28
11	烹饪 佳肴 食物 餐	cuisine 26	cooking	plat 123	food 59
		plat délicieux 18	cuisine	repas 98	vege* 45
		aliment 11	food	assiette 90	meal 33
		repas 7	meal	légume 86	dinner 27
12	花 紫色 牡丹 极美的	fleur 49	flower	fleur 260	flower 100
		violet 15	purple	pivoine 80	pink 49
		pivoine 10	peony	rose 75	peon* 37
		magnifique 8	very beautiful	feuille 52	purple 19

A partir de este análisis, hemos identificado los caracteres chinos correspondientes a los cuatro términos más populares, y luego los pegamos en Google Translate para obtener una versión en el idioma inglés. Así, podemos ver en el cuadro 3 las relaciones entre los términos de estas tres lenguas, con la columna correspondiente para la versión en francés de los términos chinos. Las dos últimas columnas muestran los datos de los participantes en francés y en inglés, y sirven de ayuda para establecer las comparaciones. Dado que se analizaron los cuatro primeros términos en chino en lugar de los tres primeros en el resto de los idiomas, los datos del cuadro 3 nos han proporcionado una gran riqueza de información.

Los dos participantes de lengua árabe, cuyos datos hemos podido analizar escribieron un solo término para la mayoría de las imágenes. Para tres de las imágenes, se dieron dos términos, y a otras tres otras imágenes no dieron ninguno. La tabla 4 muestra los términos con los que se indicaron.

Tabla 4. Etiquetas de los dos participantes de habla árabe, cuyos datos se podían utilizar

Image	Top term	Freq	Second Term	Freq
1	تنظيف	2	قمامة	2
2	لطيور	4	لطاوس	2
3	<i>no data</i>			
4	صور	3	كاميرا	2
5	شاطئ	2		
6	نارية ألعاب	2		
7	بقرة	3		
8	رقصة	2		
9	صبار	2		
10	<i>no data</i>			
11	غذاء	2		
12	<i>no data</i>			

El árabe es un idioma que ningún miembro del equipo de la investigación habla o lee, pero el funcionamiento las cadenas de caracteres a través de Google Translate nos permitió el poder verificar las equivalencias entre los idiomas que conocemos bien, inglés y francés. Debido a este control cruzado, estamos seguros de que nuestro análisis es el correcto.

Sin embargo, los escasos datos de los participantes de los idiomas con un alfabeto que no sea el alfabeto romano fueron verdaderamente lamentables. A excepción de los datos en chino, de los que se analizaron datos de los 90 participantes, las otras lenguas (árabe, griego y ruso) están escasamente representados. Nuestros esfuerzos para reclutar a los participantes que hablan estas lenguas evidentemente no fue satisfactorio. Sin embargo, los datos obtenidos nos permiten tener una visión útil sobre lo que los resultados podrían haber sido, ya que los

pocos participantes utilizaron en primer lugar los términos también empleados por los participantes del resto de las otras lenguas. Estos términos (que representan los nombres de los objetos en las imágenes) claramente parece ser que se presentan a la mente cuando los participantes ven las imágenes.

Debate

Las dificultades experimentadas con la recopilación de datos es debida en parte a los escasos medios disponibles para llevar a cabo el estudio. Éramos en gran medida dependientes de los voluntarios que nos hicieron las distintas versiones lingüísticas de los sitios, otros que trataron de reclutar a los participantes de diversas comunidades lingüísticas, y a la buena voluntad de aquellos que accedieron a participar. Sin embargo, en los últimos años hemos trabajado en este problema, hemos acumulado experiencia suficiente y hemos visto lo necesario en los datos similares a este tipo de estudios para estar seguros de que nuestros resultados fueran exactos. Tratamos de obtener datos empíricos para dar apoyo teórico para sustentar la hipótesis de la fiabilidad de los sistemas automatizados a través de lenguajes de “indización” para las imágenes “normales”.

En definitiva, se trata de un tipo de indización que se impondrá en un futuro, independientemente de nuestro trabajo. Además, va a poder realizarse sólo por medios automatizados, ya que los costes de hacerlo de otra forma son prohibitivos. La tarea a la hora de etiquetar las imágenes consiste en identificar e indizar los objetos más significativos que se pueden apreciar en una imagen. Hemos demostrado en estudios previos que casi cualquier persona puede indizar los objetos en imágenes de este tipo, así como también lo pueden realizar profesionales especializados ya que algunos términos aparecen con la misma denominación con la que son conocidos por la mayoría de la gente. Una vez que se establecen las condiciones en cualquier idioma, la calidad de la indización en otros idiomas depende a su vez de la calidad de las herramientas de traducción.

A pesar de que el escaso número de participantes que aportaron datos utilizables en los idiomas con alfabetos distintos del alfabeto romano es decepcionante, los pocos datos que nos hicieron llegar son aún importantes y, es razonable, suponer que si hubiéramos tenido más datos, se hubieran producido las mismas concordancias que se produjeron con los otros idiomas. Los datos de nuestros participantes de China son particularmente útiles en este contexto, ya que nos hicieron llegar los datos de un gran número de participantes en este idioma. Dado que los datos en China no muestran ninguna desviación con respecto a los modelos de los datos de las lenguas occidentales, es posible asumir que estamos ante un fenómeno universal que se produce de forma independientemente a la lengua.

Este trabajo contribuirá a la validación de otros enfoques. Uno de estos métodos consiste en transferir los términos de indización y otros metadatos de una imagen a otra una imagen similar, es decir, una imagen con el mismo objeto. Para conseguirlo hay que emplear técnicas de alto nivel, tales como palabras asociadas a las imágenes hechas por personas para que se pueden combinar con otras técnicas de bajo nivel de reconocimiento como son las técnicas utilizadas por la tecnología Photosynth (Aguera y Arcas 2007). Con esta tecnología, mediante la asignación de etiquetas a numerosas fotografías de detalles, posteriormente los resultados se pueden extrapolar a otras de carácter más general de tal forma que muchas de estas fotografías se puedan acoplar a estos modelos más generales. A partir de este supuesto, si un usuario añade etiquetas para identificar una fotografía de un detalle en particular, otras fotos con el mismo detalle pueden heredar sus metadatos. La búsqueda de similitudes

utilizando los diccionarios bilingües y multilingües que hemos descrito aquí, favorece que una fotografía que ha sido indizada por un usuario pueda hacer extensiva esta indización a otras fotografías, independientemente del idioma utilizado lo que favorece la extensión de la indización a una población mucho mayor.

Nuestro trabajo con la indización de fotograma a fotograma de las imágenes en movimiento demuestra que los modelos a aplicar son los mismos, lo que amplía el campo a estudiar. Sin duda hay muchas, muchas otras aplicaciones en las que se puede aplicar este sistema, de modo que no faltará trabajo para los investigadores, especialmente aquellos que conocen a diez o doce lenguas (!), que puedan analizar y estudiar.

Conclusión

Emprendimos el trabajo para demostrar de forma empírica la hipótesis de que los diccionarios en línea bilingües y el resto de herramientas que integran otros sistemas como el traductor de Google puede generar etiquetas en otros idiomas que permitan a los usuarios encontrar imágenes, independientemente del idioma que utilicen para la búsqueda. En otras palabras, es probable que la forma de nombrar los objetos por parte de los participantes en un estudio donde coexisten diversos idiomas también se utilicen los mismos términos que proporcionen las herramientas que se puedan encontrar en la web para generar los términos de forma automática cuando se utiliza un solo idioma para la indización.

A pesar de que estas herramientas de la web son imperfectas, sin embargo son muy utilizadas para la indización y traducen con éxito aquellos términos de indización de imágenes normales siempre y cuando éstas tengan una indización de calidad, es decir cuando los términos en el idioma original sean exactos. Como medida del desarrollo tecnológico que estamos actualmente viviendo, es interesante señalar que este enfoque no podría haberse imaginado en una fecha tan reciente como es la del comienzo de este siglo. Es evidente que el desarrollo continuo de herramientas en la web seguirá ofreciendo un terreno fértil a los investigadores para el estudio de la indización automática de imágenes en múltiples idiomas.

Agradecimientos

Esta investigación se llevó a cabo gracias a la subvención 410-2005-1213 de las Ciencias Sociales y Humanidades del Consejo de Investigación de Canadá. (SSHRC), junto a la participación de los voluntarios. Dedicamos este trabajo final fruto de la subvención a la memoria de Michelle Gauthier, quien fue el principal responsable de la obtención de nuestra subvención. Queremos agradecer a todos nuestros colaboradores que trabajaron como traductores y ayudaron a traducir los diferentes idiomas utilizados en el presente estudio así como a Arnaud d'Alayer, Gabriel Coder and Robin Millette por su ayuda técnica.

Agradecemos a todos y cada uno de los participantes que contribuyeron en la aportación de los datos. Queremos agradecer de forma especial a 潘玥玲 Marion, 陈祎 Camille, 范秉馨 Vincent, 杨珺 Anaïs, al personal de la biblioteca (Alliance Francaise of Shanghai) a la hora de traducir las etiquetas del chino al francés.

Bibliografía

- 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. User tagging of library resources: toward a framework for system evaluation. *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-en.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 » [indexing and retrieval of audiovisual archives using existing textual descriptions]. *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-acs.ca/proceedings/2010/CAIS056_TurnerNigay_Final.pdf>.
- Turner, James M. and Suzanne Mathieu. 2007. Audio description for indexing 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-en.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://www.ifla.org/IV/ifla72/index.htm>>.

