

Ordinary image retrieval in a multilingual context: a comparison of two indexing vocabularies

Elaine Ménard

Aslib Proceedings : New Information Perspectives, Vol. 62, n. 4/5, 2010

El artículo presenta un estudio en el que se analiza la recuperación de imágenes en un contexto monolingüe y multilingüe. Según el artículo, la selección de los términos de indización apropiados resulta crucial para la recuperación de las imágenes. La búsqueda de imágenes presenta dos problemas principales: por un lado hay que trasladar la representación visual a una descripción textual (lo cual puede llevar al error y la ambigüedad) y por otro la diversidad lingüística de la red hace que la descripción de la imagen deba corresponder a la del buscador para que este pueda recuperarla. Por lo general, los esquemas de metadatos sugieren la utilización de vocabularios controlados para la descripción de imágenes: listas de encabezamientos de materias, tesauros, diferentes sistemas de clasificación, listas de autoridades o diccionarios visuales; todos ellos suelen estar únicamente en inglés. Es por ello que los vocabularios no controlados pueden constituir una buena alternativa, ya que los sistemas que utilizan este tipo de vocabularios pueden estar escritos en una lengua o en una combinación de varias. Para llevar a cabo el estudio se creó en primer lugar una base de datos de imágenes y se le asignaron términos según cuatro formas de indexación, luego se analizaron esos términos. En segundo lugar, se llevó a cabo una simulación de recuperación con 30 imágenes y 60

participantes escogidos al azar, donde se midió la eficacia, la efectividad temporal, la efectividad humana y la satisfacción del buscador con la imagen. Por último, se realizó un cuestionario para recoger datos sobre las impresiones de los participantes. Los resultados del estudio muestran: 1) los términos de indización, tanto en el enfoque del vocabulario controlado como del no controlado, difieren en las perspectivas terminológica, perceptual y estructural; 2) la simulación de recuperación confirmó que la recuperación de imágenes comunes es más efectiva cuando se combinan los vocabularios controlado y no controlado y que es más eficaz en un entorno monolingüe. Algunas de las conclusiones que se obtienen son que la recuperación de las imágenes de objetos comunes deberían estar basadas en atributos pre-iconográficos, que se debería incluir la función del objeto (para aumentar la correcta recuperación) y que una metodología de indexación basada en la combinación de los lenguajes controlado y no controlado puede ser la mejor solución para la eficacia y la satisfacción del buscador, pero no da tan buenos resultados para efectividad.

Resumen elaborado por: Isabel M^a Domingo Montesinos