

# RECUPERACIÓN DE IMÁGENES EN LA WEB: SISTEMA PROTOTIPO BASADO EN CONTENIDO Y MANEJO DE CALIDAD

---

Bell Manrique Losada\*  
Jaime Alberto Echeverri Arias\*\*

Recibido: 01/03/2007

Aceptado 30/04/2007

## RESUMEN

El presente artículo muestra una revisión de las investigaciones realizadas acerca de la recuperación de imágenes en colecciones digitales en la web, basado en su contenido y en el manejo de la calidad de la información. Teniendo en cuenta este estado del arte, se presenta una propuesta metodológica para el diseño e implementación de un prototipo de sistema de recuperación de imágenes que utilice estas dos técnicas de búsqueda: recuperación basada en contenido y recuperación manejando calidad de la información y, por último, se muestran los avances en el desarrollo de la propuesta.

**PALABRAS CLAVE:** Recuperación de imágenes, colección digital, técnicas de búsqueda, calidad de la información

## ABSTRACT:

This article presents a compilation of the researches carried on the image recovery from web digital collections, based upon their contents and data quality management. Using this state-of-the-art study as a support, the article presents a methodological proposal to design and implement a system prototype in image recovery that uses these two search techniques; and finally the article presents the progress achieved in developing this proposal.

**KEY WORDS:** Image recovery, digital collection, search techniques, data quality.

---

\* Ingeniera de sistemas Universidad Distrital FJC- Universidad de la Amazonia, Magister en Ingeniería de Sistemas Universidad Nacional de Colombia Sede Medellín, Profesora de Tiempo Completo, Programa de Ingeniería de Sistemas, Universidad de la Amazonia. e-mail: bellm@uniamazonia.edu.co

\*\* Ingeniero, Magister en Ingeniería de Sistemas, Universidad Nacional de Colombia Sede Medellín, Profesor Tiempo Completo, Universidad de Medellín, e-mail: jaecheverri@udem.edu.co

## INTRODUCCIÓN

Internet está basado en un paradigma de búsqueda que hace difícil recuperar datos desde múltiples sitios y en su mayoría provee aplicaciones como motores de búsqueda y meta-buscadores que permiten a los usuarios buscar la información que necesitan, por medio de una búsqueda basada generalmente en recuperación de información sintáctica de textos. Debido a esto, se ha desarrollado una serie de propuestas que tratan de integrar un conjunto de diferentes fuentes especializadas, para lo cual extraen, filtran y representan eficientemente la información obtenida de la web, pero la mayoría están enfocadas principalmente a la cantidad de información recuperada y a la calidad de la consulta, medida ésta con criterios de tiempo y costos de ejecución para encontrar planes óptimos, como lo muestran los trabajos de Chen *et al.*, (1998) y Ambite, (1999).

Las colecciones de imágenes digitales son un tipo de sistema de información en la web, en donde se integra un conjunto de fuentes especializadas (fuentes de imágenes) y se implementan estrategias de procesamiento de consultas enfocadas hacia minimizar el tiempo de respuesta. A diferencia de los sistemas de información tradicional en donde los tiempos y costos de ejecución de consultas y generación de respuestas son el punto central, en este tipo de sistemas en la web el principal factor de eficiencia es la *calidad de la información* (IQ) de los resultados de la consulta, como lo muestran diferentes investigaciones (Naumann, 2000).

Por otro lado, en una colección digital los usuarios requieren definir estas consultas basándose en el contenido de las imágenes, normalmente utilizando técnicas de descripción de su contenido para realizar el análisis, además de considerar criterios relacionados con la calidad de la información que contienen, y la calidad de las imágenes en sí. Esto genera oportunidades de acción para profundizar en la temática de determinar el contenido de las imágenes, de forma que sea posible la clasificación

de objetos con base en este contenido y en la calidad de la información que representan.

De esta forma, este trabajo presenta los resultados preliminares logrados alrededor de la tarea de proponer un prototipo de sistema de recuperación de imágenes, basado en contenido y en un modelo de calidad adecuado, que permita la ejecución de un proceso de búsqueda de imágenes en una colección digital en la web. Estos resultados preliminares se enmarcan en la exploración bibliográfica, determinación de antecedentes y definición de un estado del arte de este tópico de investigación y, por último, en la definición de una propuesta metodológica y su respectivo avance en ejecución.

## DESCRIPCIÓN DEL PROBLEMA

En sistemas web, como las colecciones de imágenes digitales, el principal factor de eficiencia que tienen en cuenta las diferentes estrategias de procesamiento de consultas no es el tiempo de respuesta, como lo es en sistemas de información tradicional, sino la calidad de la información (IQ) de los resultados. Diferentes investigaciones se han acercado a este tema con la exploración de criterios de calidad para responder consultas de usuario en sistemas de información en la web (Naumann, 2000).

Las aproximaciones convencionales para clasificar las características visuales en términos de descripciones textuales han demostrado ser inadecuadas para indexar imágenes (Leung, 1999). De hecho, la menor riqueza expresiva del texto, respecto a las características visuales, no permite explotar de forma plena las habilidades de la memoria del ser humano, y los resultados de una consulta pueden no ser relevantes a las expectativas del usuario. Esta es la razón por la que, actualmente, las tendencias se encaminan a utilizar contenidos visuales como descriptores. Las limitaciones de una aproximación basada en descripciones textuales y la oportunidad de apoyar la labor de los diseñadores sugiere incorporar al sistema la

capacidad de incluir un sistema de consulta y recuperación basado en descriptores sintácticos y/o semánticos que permitan una aproximación mayor al contenido de las imágenes. Este proporcionará un método de exploración y recuperación de las imágenes en función de, por ejemplo, el motivo de partida que ofrezca el diseñador. Es por esto que se hace necesario utilizar otros descriptores de mayor nivel expresivo.

En una primera aproximación, pueden ser derivados del análisis estructural de las imágenes, basándose en la utilización de elementos como el color, la forma y la textura que ya se obtienen en la etapa de análisis de la imagen, o bien, agrupaciones u organizaciones de éstos, de mayor entidad. En un segundo nivel, la exploración se realiza a partir de la elección de un criterio, cuyo resultado es la obtención de una secuencia de coincidencias que, para su discriminación, pueden ir acompañadas de un valor numérico que indica la presencia y cercanía del criterio en cada una de las imágenes mostradas (Melchor y Valiente, 2001).

Los métodos de búsqueda basados en descriptores son extremadamente eficientes (Stone, 1998). Es por este motivo que esta es la forma más utilizada de enfocar las operaciones de búsqueda en el ámbito de las imágenes. Pero hay dos problemas principales en la aproximación basada en descriptores. Por un lado, el determinar qué descriptores hay que utilizar, y por otro lado, la representación de una determinada base de datos en forma de descriptores.

Si el usuario realiza de forma natural las consultas basándose en el contenido de las imágenes, se pueden utilizar técnicas de descripción del contenido de las mismas para hacer el análisis de las imágenes. Estas aplicaciones son difíciles de llevar a cabo por las técnicas clásicas de reconocimiento de objetos utilizadas en visión por computador, por la complejidad de la tarea de encontrar objetos generales en contextos abiertos. Así, es necesario

profundizar en la temática de determinar el contenido de las imágenes, de forma que sea posible la clasificación de objetos. Este es uno de los ejes centrales de este trabajo.

La calidad de la información es un tema importante que tiene mucha consideración, y es tópico de investigaciones sobre captura y modelamiento de la información, y no es la excepción en este tipo de ambientes de tratamiento y recuperación de imágenes. Sin embargo, pocos trabajos tratan de aplicar esa calidad de la información al proceso de planificación de consultas sobre la web (Naumann *et al.*, 2001), y mucho menos cuando se habla de bases de datos de imágenes o colecciones de imágenes digitales en Internet.

En una base de datos tradicional, la planificación de consultas considera un conjunto de criterios basados en tiempos y costos de ejecución; y últimamente se están tratando criterios relacionados con la calidad de la información como completitud, frecuencia de actualización o exactitud. En estos sistemas se encuentran eficientemente los resultados de consultas con máxima calidad con respecto a los criterios IQ. Cuando se habla de bases de datos de imágenes, o colecciones de imágenes digitales, son pocos los trabajos que se han ocupado de la planificación de consultas con manejo de criterios de calidad de la información, relacionados con esos descriptores básicos y típicos de las imágenes, como forma, textura, entre otros.

Este tipo de sistemas normalmente se han enfocado hacia medir la recuperación de su información (imágenes) con criterios de calidad relacionados con minimalidad, costos de ejecución y tiempos de respuesta de sus consultas, y ha recibido poca atención el tratamiento de otros criterios (metadatos) que se relacionan con la calidad de la información de las respuestas, esto es, aspectos como la relevancia de acuerdo con las necesidades iniciales del usuario y con sus criterios de búsqueda particulares.

## ESTADO DEL ARTE

A continuación se discuten las propuestas relacionadas con el control de calidad en sistemas de búsqueda en la web, en cuyo campo de acción se relacionan varios aspectos como lo son: la optimización de consultas tradicional, planificación de consultas en la Web, calidad de la información en planificación de consultas Web; y posteriormente se muestran los trabajos desarrollados alrededor de la recuperación de imágenes basada en contenido.

## CONTROL DE CALIDAD EN SISTEMAS DE BÚSQUEDA EN LA WEB

### Optimización de consultas tradicional

En la literatura de bases de datos, la optimización de consultas ha sido ampliamente estudiada. Un optimizador de consultas intenta encontrar la forma algebraica más eficiente de una consulta y escoger métodos específicos para implementar cada operación de procesamiento de datos. La investigación desarrollada en esta sub-área por Chu y Hurley (1982) se enfoca hacia la optimización de la consulta basada en criterios de eficiencia: minimización de tiempos y costos de ejecución.

### Calidad del proceso de planificación en sistemas de información en la Web.

Los siguientes trabajos enfrentan el problema de la planificación de consultas y se acercan al tratamiento de la calidad de los planes, en términos de selección de las fuentes, y eficiencia, en términos de costos computacionales de ejecución de las consultas, sin tener en cuenta calidad de la información de las respuestas encontradas.

Gran número de proyectos ha desarrollado propuestas en la planificación de consultas en mediadores. Por ejemplo, el Information Manifold (Levy *et al.*, 1996) y el TSIMMIS (Hammer *et al.*, 1995)

enfocan la planificación hacia la optimización basada en costos, donde primero un conjunto de planes recuperables son encontrados y luego se optimiza cada uno independientemente. El proyecto GARLIC (Tork Roth *et al.*, 1996; Roth y Schwarz, 1997) considera la optimización de costos para mediadores y evaluación de sub-consultas de las fuentes de información. El sistema SAGE (Knoblock, 1996), considera la calidad del plan, soportándola con la propuesta de intervención entre planificación y ejecución. El sistema OCCAM (Kwok y Weld, 1996) es un planificador para recuperación de información en dominios distribuidos y heterogéneos que se enfoca principalmente en el problema de la selección de las fuentes relevantes para la consulta, más no del procesamiento de la consulta como tal ni de la medición de la calidad de la información.

Un marco de trabajo más relacionado con el problema y que provee buenos resultados se presenta en Ambite (1999) donde se propone el paradigma de la planificación por reescritura-PbR-, que combina la selección de las fuentes y la optimización de la consulta basada en costos. Esta propuesta tiene tres ejes de acción: la calidad del plan de la consulta, la generación de un plan inicial, y las reglas de reescritura del plan; el corazón del proceso de planificación consiste en la aplicación iterativa de un conjunto de reglas de reescritura de un plan hasta que sea encontrado uno de aceptable calidad.

En Ives, 2002, se propone y evalúa un conjunto de técnicas para procesamiento de consultas que se adaptan a su medio de ejecución, que permite al procesador de la consulta reaccionar a las condiciones cambiantes o a conocimiento que va creciendo en tiempo de ejecución. En comparación con otros trabajos, la calidad del procesamiento de la consulta mejora por la consideración de cambios en tiempos de ejecución y por las técnicas adaptativas que permiten mayor rapidez, sin embargo, igualmente la calidad la relaciona únicamente con eficiencia.

## Calidad de la información en planificación de consultas

En Knoblock *et al.*, 2000, se propone el Sistema ARIADNE basado en una arquitectura que hace eficiente integrar múltiples fuentes por medio de métodos para mapearlas en una representación uniforme. Esta propuesta se construye basada en técnicas de representación de conocimiento, aprendizaje de máquina y planificación automatizada. Este trabajo está más enfocado hacia los esquemas de representación y herramientas de modelamiento de las fuentes, que hacia el procesamiento de la consulta como tal, en donde igualmente la calidad de los planes se mide con métricas basadas en costos.

En Naumann, 2000, se investiga la exploración de criterios de calidad de la información IQ para responder consultas de usuarios en SIBM, y discute qué criterios de IQ son necesarios, cómo pueden ser adquiridos y cómo pueden ser usados para mejorar la calidad de los resultados de la consulta. Este autor plantea el hecho de la importancia que tiene la calidad de la información en los sistemas distribuidos a gran escala, luego de algunos trabajos desarrollados en el área, y enfatiza en la ausencia de investigaciones que apliquen razonamiento sobre la calidad de la información en el área de la planificación de consultas sobre la web. Desarrolla un sistema que encuentra resultados de consulta con alta calidad basado en criterios de calidad definidos, aplicado sobre un meta-motor de búsqueda que usa motores de búsqueda existentes como sus fuentes de información. Como criterios de calidad para este dominio, incluye completitud y frecuencia de actualización, entre otros.

Chen *et al.* (1998), presentan una investigación sobre la calidad del procesamiento de consultas en la WWW, debido a muchos factores tales como impredecible tiempo de respuesta, resultados irrelevantes y datos no actualizados, y propone un método para el procesamiento de consultas controlando calidad en este ambiente web. Introduce

parámetros de calidad que los usuarios pueden especificar cuando se introducen las consultas, al igual que funciones que son usadas para evaluar la bondad de estos parámetros y algoritmos de programación, planificación y ejecución.

## TÉCNICAS DE RECUPERACIÓN DE IMÁGENES -RI- EN COLECCIONES DIGITALES

Los avances recientes en tecnologías de computación y comunicación están tomando herramientas de procesamiento de información para lograr sus objetivos. En los últimos años ha habido una sobre acumulación de datos digitales tales como imágenes, video y audio. Internet es un excelente ejemplo de bases de datos distribuidas conteniendo varios millones de imágenes. Otros casos de grandes bases de datos de imágenes incluyen bancos satelitales y médicos, donde es muchas veces difícil describir o anotar el contenido de las imágenes.

Técnicas que trabajan con sistemas de información tradicional han sido adecuadas por muchas aplicaciones involucrando registros alfanuméricos. Ellos pueden ser ordenados, indexados y buscados por coincidencia de patrones en una forma sencilla. Sin embargo, en muchas aplicaciones de bases de datos científicas, el contenido de información de imágenes no es explícito, y esto no es fácilmente adecuado para la directa indexación, clasificación y recuperación. Particularmente, las bases de datos de imágenes en gran escala han surgido como el problema más retador en el campo de bases de datos científicas.

Los sistemas de recuperación de información visual están relacionados con almacenamiento eficiente y recuperación de registros. En general, son útiles solamente si pueden recuperar coincidencias aceptables en tiempo real. Adicional a palabras clave asignadas por un humano, los sistemas RI pueden usar el contenido visual de las imágenes como índices (color, textura y forma). Recientemente, varios sistemas combinan atributos heterogéneos

para mejorar la discriminación y clasificación de resultados. Mientras estos sistemas usan características de bajo nivel como color, textura y forma para consultas de imágenes, los usuarios usualmente tienen una noción más abstracta de qué los satisfará usando características de bajo nivel para corresponder a abstracciones de alto nivel.

Una técnica interesante para cubrir la brecha que existe entre las descripciones textuales y pictóricas para explorar información al nivel de los documentos es tomada de la recuperación de información, llamada Análisis Semántico 'Latente' y trabajada por Obeid *et al.* (2001); primero es formado un corpus de documentos (en este caso, imágenes con un título); luego, por descomposición de valores singulares, el diccionario es correlacionado con las características derivadas de las imágenes. La búsqueda es por correlación oculta de características y títulos. La colección de imágenes consta de diez categorías semánticas de cinco imágenes cada una. En este trabajo, se utilizan para relacionar cada imagen de la colección características intermedias que son características semánticas de bajo nivel y características de la imagen de alto nivel. Es decir, pueden ser utilizadas para producir conceptos de alto nivel y pueden ser aprendidas desde una base de datos de anotaciones pequeñas.

Existen numerosos sistemas de recuperación de imágenes basados en el contenido, los cuales incluyen alguna o varias características como color, textura, formas de objetos en la imagen, entre otros, para hacer la búsqueda. La mayoría son sistemas de propósito general y faltan estudios en los que éstos se usen en aplicaciones prácticas para investigar las ventajas e inconvenientes de las distintas opciones. Algunos de los más interesantes son QBIC y MARS. Un interesante resumen de las técnicas y sistemas de recuperación basados en el contenido, así como abundante bibliografía sobre el tema, puede encontrarse en Rui *et al.* (1999).

## PROPUESTA METODOLÓGICA

Teniendo en cuenta la literatura revisada en el estado del arte, se propone aplicar dos técnicas que han sido trabajadas aisladamente, que son: recuperación de imágenes basada en contenido, y recuperación basada en la calidad de la información.

La recuperación de imágenes basada en contenido es un mecanismo que permite la extracción automática de información a partir de una imagen, lo cual genera que se puedan ejecutar consultas más complejas. En esta técnica, las imágenes son indexadas y recuperadas por su propio contenido visual como el color o la textura, además de la utilización de anotaciones textuales y taxonomías establecidas, que por sí solas se muestran como insuficientes e inadecuadas por problemas relacionados con la cantidad de trabajo que requiere asociar un texto a una imagen, y por la subjetividad que implica expresar el gran contenido de una imagen (Aranda *et al.*, 2002). La recuperación de imágenes basada en el contenido se fundamenta en 3 aspectos: la extracción de características visuales, la indexación multidimensional (para facilitar una búsqueda rápida) y el diseño del sistema de recuperación. Las características visuales se clasifican en generales y específicas, las cuales dependen de la aplicación. Las características visuales generales que son usadas en la mayoría de las aplicaciones son:

- **Color:** Es una de las características más usadas. Es relativamente robusta a las variaciones del fondo e independiente del tamaño y orientación de la imagen. La técnicas más usadas como representación del color son el histograma de color, el histograma de color acumulado, los momentos de color y los conjuntos de color.
- **Textura:** Se refiere a patrones visuales homogéneos formados por diversos colores o intensidades. Es una propiedad innata de prácticamente todas las superficies, como nubes,

árboles, pelo o ladrillos. Las características de textura se suelen representar usando una matriz de concurrencia, propiedades psicológicas (contraste, regularidad, tosquedad, aspereza...), transformadas wavelet...

- Formas de objetos en la imagen: Algunas aplicaciones requieren que la representación de la forma sea invariante a traslación, rotación y escalado, mientras que otras no. En general, las representaciones de la forma se dividen en dos categorías: las basadas en contornos y las basadas en regiones. Las primeras usan sólo el contorno exterior de la forma, mientras que las últimas usan la región de la forma completa. Se han desarrollado numerosos métodos para ambas categorías, pero los más representativos son los descriptores de Fourier (transformada de Fourier del contorno) para la primera, y los momentos invariantes (momentos basados en regiones que sean invariantes a transformaciones) para la segunda.
- Diseño del color: Se trata de usar conjuntamente la característica de color y las relaciones espaciales. Una aproximación sencilla es dividir la imagen en bloques y extraer las características de color de cada bloque. Otra aproximación es segmentar la imagen en regiones con características de color destacadas y luego almacenar el conjunto de características de color y la posición de cada región. Su desventaja es la problemática que supone la segmentación de una imagen. Otras técnicas son usar momentos de color sobre regiones, usar una matriz de concurrencia de color.

El valor de cualquier sistema de información está condicionado no solo por la cantidad de información contenida, sino también por su calidad y, al mismo tiempo, por su 'findability' o facilidad para encontrar dicha información, cualidad que naturalmente disminuirá conforme aumente el tamaño del sistema. Es por esta razón que la investigación en técnicas de acceso y recuperación

de información ha experimentado su mayor auge con el advenimiento de un sistema de información de gigantesco volumen y crecimiento exponencial, como la World Wide Web o WWW.

En esta evolución de técnicas de recuperación de información se ha llegado a hablar entonces de recuperación de información e integración de información basada en la calidad de información. Se han desarrollado, por lo tanto, modelos de planificación de consultas en este tipo de sistemas como el encontrado en Manrique, (2006). Este tipo de trabajos tienen como objetivo determinar lo que realmente implica tener en cuenta la calidad de la información en el proceso de planificar una consulta en sistemas de información en la web. Esta técnica de clasificación y búsqueda se ha trabajado en forma aislada con respecto a nuevas técnicas de indexación y recuperación, sobre todo en los sistemas de recuperación de imágenes.

En este sentido, se propone el diseño e implementación de un prototipo de sistema de recuperación de imágenes -SRI-, basado en contenidos y en un modelo de calidad adecuado, que permita la ejecución de un proceso de búsqueda de imágenes en una colección digital en la web.

Para cumplir con este objetivo, se ha definido una metodología enmarcada dentro de tres fases.

La fase teórica que pretende lograr los objetivos relacionados con profundización de la temática de determinación del contenido de las imágenes, de técnicas de descripción del contenido para realizar el análisis de la colección y de definición del modelo de calidad a aplicar en el SRI. Las actividades específicas propuestas en esta fase son:

- Exploración de trabajos relacionados con la determinación de contenido de las imágenes y técnicas utilizadas para desarrollar SRI basadas en contenido.
- Definición de criterios de calidad de la información, métricas para cada uno y la forma en la que pueden ser implementados en el SRI.

- Definición de un modelo de calidad que contenga los criterios, la descripción detallada y métrica de cada uno.

La fase de estructuración que pretende lograr los objetivos relacionados con la determinación de los contenidos y categorías base de la colección digital y el planteamiento del modelo del SRI. Las actividades específicas propuestas en esta fase, teniendo en cuenta la descripción de contenido de las imágenes de la colección y el modelo de calidad definido, son:

- Diseño del modelo de diseño del SRI.
- Planteamiento del prototipo del SRI.
- Implementación de un prototipo del modelo.
- Determinación y definición de la colección bibliográfica
- Recolección de información fotográfica por medio de salidas de campo a ecosistemas seleccionados para la colección.

La fase de validación, que por medio de la implementación de un prototipo del SRI, pretende validar el modelo propuesto y comparación con otros del área de estudio, así:

- Validación del prototipo con la ejecución de pruebas y comparación con otros modelos y sistemas en el área de estudio.
- Análisis y determinación de resultados de la aplicación del prototipo propuesto.

## LOGROS Y TRABAJO EN CURSO

En el desarrollo de la primera fase metodológica propuesta para lograr el objetivo planteado, se han logrado las siguientes conclusiones:

Se han revisado trabajos representativos acerca de técnicas de recuperación de imágenes, específicamente las basadas en descriptores semánticos y en contenido, al igual que los trabajos desarrollados

alrededor de calidad de la información en sistemas de procesamiento o planificación de consultas en la web. Esta revisión se caracteriza por la serie de limitaciones encontradas en los métodos de recuperación de información basada en descripciones textuales, y en la oportunidad de apoyar la labor de los diseñadores incorporando a este tipo de sistemas la capacidad de incluir un sistema de consulta y recuperación basado en descriptores sintácticos y/o semánticos que permitan una aproximación mayor al contenido de las imágenes y una satisfacción mayor del usuario respecto a sus criterios de búsqueda iniciales.

Es necesario desarrollar sistemas de recuperación de imágenes en contextos específicos que integren además de técnicas basadas en contenido, técnicas de control de calidad de la información que se maneja en la colección digital.

Por las condiciones cambiantes de la web y las necesidades de búsqueda de los usuarios, sobre todo en colecciones digitales de imágenes, es oportuno proponer soluciones para otro tipo de usuarios, profesionales y con necesidades específicas de consultas de imágenes, que permita encontrar resultados más cercanos a sus necesidades iniciales y más relevantes de acuerdo con sus contextos profesionales y de trabajo.

Como avance de la propuesta metodológica planteada, se ha logrado terminar, casi en su totalidad, la fase teórica y se está iniciando con la ejecución de las actividades definidas en la fase de estructuración.

Entre los resultados logrados hasta el momento en la primera fase, la exploración teórica de trabajos relacionados con la calidad de la información y con técnicas de recuperación basadas en contenido, se incluyen dentro del estado del arte de este artículo.

El modelo de calidad que se definió para el SRI fue el estructurado en Bertoa *et al.*, (2003) y aplicado y definido por Manrique (2006), compuesto por un

conjunto de características y sub-características y de cómo se relacionan entre sí, reflejados como criterios que pueden describir la IQ de cada elemento de información, en este caso cada imagen, teniendo en cuenta la información que suministran y que facilita su valoración para el proceso de consulta.

Como contexto específico de aplicación, se definió la colección digital de imágenes de ecosistemas acuáticos, utilizando la base pictórica que tiene el grupo de investigación CAPREA (Calidad y Preservación de Ecosistemas Acuáticos Amazónicos) de la Universidad de la Amazonia -Florencia Caquetá.

## BIBLIOGRAFÍA

- AMBITE, J. L. 1999. *Planning by Rewriting*. PhD Thesis, University of Southern California.
- ARANDA, M. C., GALINDO, J. & URRUTIA, A. 2002. Museos Digitales en Internet: Modelo EER Difuso y Recuperación de Imágenes Basada en su Contenido. IV Congreso "Turismo y Tecnologías de la Información y las Comunicaciones". En: <http://bibemp2.us.es/turismo/turismonet1/economia%20del%20turismo/turismo%20cultural/museos%20digitales%20en%20Internet.pdf>
- BERTOIA, M. F., TROYA, J. M. & VALLECILLO, A. 2003. Atributos de Calidad para Componentes COTS: Una valoración de la información ofrecida por los vendedores. Dpto. Lenguajes y Ciencias de la Computación. Universidad de Málaga. España, 2003.
- CHEN, Y., ZHU, Q. & WANG, N. 1998. *Query processing with quality control in the World Wide Web*. World Wide Web.
- CHU, W. & HURLEY, P. 1982. *Optimal query processing for distributed database systems*. IEEE Transactions on Computers, Septiembre.
- HAMMER, J., GARCIA-MOLINA, H., IRELAND, K., PAKONSTANTINO, Y., ULLMAN, J. & WIDOM, J. 1995. *Information translation, mediation, and mosaic-based browsing in the TSIMMIS system*. En: Proceedings of the ACM SIGMOD International Conference on Management of Data, San Jose, California.
- IVES, Z. 2002. *Efficient Query Processing for Data Integration*. PhD Thesis, University of Washington.
- KNOBLOCK, C. A. 1996. Building a planner for information gathering: A report from the trenches. En: Proceedings of the Third International Conference on Artificial Intelligence Planning Systems.
- KNOBLOCK, C. A., MINTON, S., AMBITE, J. L., ASHISH, N., MUSLEA, I., PHILPOT, A. G. & TEJADA, S. 2000. *The ARIADNE approach to Web-Based Information Integration*. En: International Journal of Cooperative Information Systems. Originalmente publicado en AAAI'98.
- (KWOK, C. T. & WELD, D. S. 1996. *Planning to gather information*. En Proceedings of the Thirteenth National Conference on Artificial Intelligence, Portland, OR.
- LEUNG, C. 1999. Image and Vision Computing, pp. 463-464.
- LEVY, A. Y., RAJARAMAN, A. & ORDILLE, J. J. 1996. *Query-Answering algorithms for information agents*. En 13<sup>th</sup> AAAI National Conference on Artificial Intelligence, Portland, Oregon.
- MANRIQUE, B. 2006. Modelo de Planificación de Consultas con Manejo de Calidad de la Información en Sistemas de Integración de Información. Escuela de Sistemas, Universidad Nacional de Colombia - Medellín, Colombia.

- MELCHOR, M. A. & VALIENTE, J. M. 2001. Bases de datos para Multimedia: Recuperación por Contenido. Departamento de Informática de Sistemas y Computadores. Universidad Politécnica de Valencia. Valencia, España.
- NAUMANN, F. 2000. *Quality-driven Query Planning*. Dissertation Outline Humboldt-Universität zu Berlin.
- NAUMANN, F., LESER, U. & FREYTAG, J. C. 2001. *Quality-driven Integration of Heterogeneous Information Systems*. Humboldt-Universität zu Berlin.
- OBEID, M., JEDYNAK, B. & DAOUDI, M. 2001. Image Indexing & Retrieval Using Intermediate Features. France. En: <http://cis.jhu.edu/~bruno/intermediate-features.pdf>
- ROTH, M. & SCHWARZ, P. 1997. Don't Scrap It, Wrap it! A *Wrapper Architecture for Legacy Data Sources*. Proc. VLDB Conference.
- RUI, Y., HUANG, T. & FU, S. 1999. Image Retrieval: Current Techniques, Promising Directions, and Open Issues. *Journal of Visual Communications and Image Representation*, 10, pp. 39-46.
- STONE, H. 1998. Content-based image retrieval - Research issues, *Multimedia Technology for Applications*.
- TORK ROTH, M., ARYA, M., HAAS, L. M., CAREY, M. J., CODY, W., FAGIN, R., SCHWARZ, P. M., THOMAS, J. & WIMMERS, E. L. 1996. The Garlic project. *SIGMOD Record (ACM Special Interest Group on Management of Data)*.