

MODELAMIENTO DE UN SISTEMA DE RECUPERACIÓN DE IMÁGENES DE RECURSOS ACUÁTICOS, BASADO EN CONTENIDO Y CALIDAD DE LA INFORMACIÓN

Bell Manrique Losada¹
Jaime Alberto Echeverri Arias²
Francisco Javier Moreno³

Recibido: 29/02/2008

Aceptado: 17/06/2008

RESUMEN

En el marco de los Sistemas de Recuperación de Información, se ubican los Sistemas de Recuperación de Imágenes, los cuales permiten generar procesos de búsqueda y almacenamiento de recursos por medio de coincidencias por palabras claves u otros métodos en tiempo real. Este tipo de sistemas pueden usar como índices, contenido visual de las imágenes como color, textura y brillo, y además combinar diferentes atributos para mejorar los procesos de clasificación y relevancia de los resultados del proceso de búsqueda, que se conocen como ‘de la calidad de la información’. Este artículo presenta el modelamiento de un Sistema de Recupe-

-
- 1 MSc. Email: bmanrique@udem.edu.co
Ingeniera de Sistemas de la Universidad Distrital Francisco José de Caldas, Magíster en Ingeniería - Ingeniería de Sistemas de la Universidad Nacional de Colombia -Sede Medellín. Docente tiempo completo, programa Ingeniería de Sistemas de la Universidad de Medellín. Trabaja en las áreas de Ingeniería de Software, Sistemas de Información en la Web e Informática Educativa. Líder del Grupo de Investigación en Informática Educativa -GIIE- de la Universidad de la Amazonia, y miembro del Grupo de Investigación ARKADIUS de la Universidad de Medellín
 - 2 MSc. Email: jaecheverri@udem.edu.co
Ingeniero mecánico y magister en Ingeniería - Ingeniería de Sistemas de la Universidad Nacional de Colombia -Sede Medellín. Docente tiempo completo del programa Ingeniería de Sistemas de la Universidad de Medellín. Trabaja en las áreas de Visión por Computador y Algoritmia. Líder del Grupo de Investigación ARKADIUS de la Universidad de Medellín.
 - 3 MSc. Email: fjmoreno@unal.edu.co
Ingeniero de Sistemas. Especialista en Gestión y Sistemas de Bases de Datos, M.Sc. en Ingeniería de Sistemas de la Universidad Nacional de Colombia -Sede Medellín. Docente de tiempo completo del programa de Ingeniería de Sistemas y Maestría en Ingeniería de Sistemas de la Universidad Nacional de Colombia -Sede Medellín. Trabaja en las áreas de Ingeniería de Software y Bases de Datos. Miembro del Grupo de Investigación en Sistemas e Informática de la Universidad Nacional -Sede Medellín.

ración de Imágenes que combina atributos de la recuperación basada en contenido como color, textura y forma, con la *recuperación basada en la calidad de la información*, como frecuencia de actualización, portabilidad y relevancia, en el contexto de la Colección Digital de Imágenes de Ecosistemas Acuáticos Amazónicos del Grupo CAPREA de la Universidad de la Amazonia.

Palabras Clave: Recuperación de Información, Recuperación de Imágenes, Ingeniería de Software, Visión por Computador

MODELING OF AN AQUATIC RESOURCE IMAGE RETRIEVAL SYSTEM, BASED ON CONTENT AND INFORMATION QUALITY

ABSTRACT

In Information Retrieval Systems, Image Retrieval Systems are located; they allow generating searching and storing processes of resources through coincidences by key words or other real-time methods. This type of systems can use visual content of images (color, texture, and brightness) as indices. In addition, these systems combine different attributes to improve the processes of classification and relevance of the searching process results known as "Information Quality." This paper presents the modeling of a Image Retrieval System which combines attributes of the *content-based retrieval* such as color, texture, and shape, with the Information Quality-based *retrieval* as updating portability and relevance frequency in the context of Digital Image Collection of Aquatic Amazon Ecosystems of Universidad del Amazonia CAPREA Research Group.

Key words: Information Retrieval, Image Retrieval, Software Engineering, Computer Vision

I. INTRODUCCIÓN

Cuando se agrupan o concentran recursos digitales de información, se obtienen colecciones digitales. Cuando estas colecciones se organizan con un sistema descriptivo a través de algún tipo de catalogación y se les asocian facilidades para la búsqueda y uso de la información, se obtienen los sistemas de recuperación y las bibliotecas digitales (Borgman, 1999). Los avances recientes en este tipo de tecnologías demandan potentes herramientas de procesamiento de información. 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 que contiene millones de imágenes, además de otros casos de grandes bases de datos que incluyen bancos satelitales y médicos, donde muchas veces es difícil describir o anotar el contenido de las imágenes.

Diferentes técnicas que trabajan con sistemas de información tradicionales han sido adecuadas por muchas aplicaciones que involucran registros alfanuméricos. Éstos 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 en las imágenes no es explícito y esto no es adecuado para su directa indexación, clasificación y recuperación. Particularmente, las bases de datos de imágenes en gran escala han surgido como uno de los problemas más retadores en el campo de bases de datos científicas.

Así como en muchas áreas de aplicación (comercio, gobierno, académica, medicina) están siendo creadas grandes colecciones de imágenes digitales, existen necesidades de este tipo de aplicaciones para el tratamiento de imágenes de las áreas biológica, geológica y ecológica, particularmente de ecosistemas acuáticos. Normalmente, la única forma de búsqueda en estas colecciones es por medio de palabras clave, indexación o simple-

mente por búsqueda manual; sin embargo, en las bases de datos, bibliotecas digitales y/o sistemas de recuperación de imágenes -SRI-, es muy útil la búsqueda basada en el contenido.

Este trabajo muestra el modelamiento de un SRI para una colección digital de ecosistemas acuáticos amazónicos, la cual ha sido consolidada luego de un proceso de digitalización de fotografías análogas, diagramas, planos e impresiones. Este sistema propone el uso integrado de dos técnicas de recuperación y búsqueda usadas actualmente pero de forma separada, que son: la *recuperación basada en el contenido* de las imágenes, manejando atributos como el color y el brillo; así como la *recuperación basada en la calidad de la información*, con características como frecuencia de actualización, portabilidad y relevancia.

El artículo está organizado así: primero se presenta el marco referencial del trabajo, alrededor de las áreas de sistemas de recuperación de información visual y calidad de la información; el siguiente apartado muestra el análisis y diseño del SRI propuesto, detallando el modelo de calidad, el análisis y diseño del sri y el diseño del prototipo de consulta definido; finalmente se presentan las conclusiones y trabajo futuro.

II. MARCO REFERENCIAL

A. Sistemas de recuperación de información visual

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

resultados, usando características de bajo nivel como color, textura y forma para consultas de imágenes.

Una técnica para cubrir la brecha 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' (Obeid *et al.*, 2001). Primero se forma un corpus de documentos (en este caso, imágenes con un título); luego, por descomposición de valores singulares, el diccionario se correlaciona con las características derivadas de las imágenes. La búsqueda es por correlación oculta de características y títulos. La colección de imágenes consiste de diez categorías semánticas de cinco imágenes cada una. En este trabajo, se usan características intermedias que son características semánticas de bajo nivel y características de la imagen de alto nivel, para clasificar cada imagen de la colección.

Existen numerosos SRI basados en contenido, que incluyen características como color, textura, formas de objetos en la imagen, entre otros, para hacer la búsqueda. La mayoría son sistemas de propósito general y hay pocos estudios en los que éstos se usen en aplicaciones prácticas. Algunos de los más interesantes son QBIC y MARS (Hirata y Kato, 1992). Un resumen de las técnicas y sistemas de recuperación basados en contenido se puede encontrar en Rui *et al.*, (1999).

Extracción de características

Es el aspecto más importante de todo el proceso de automatización de imágenes. Debe abstraer la información que quedará incluida en la base de datos (en los atributos de cada imagen individualmente). A cada imagen se le asocia un vector de características (atributos), que incluye información sobre color, relaciones de color y formas.

1. *Color y relaciones de color.* Una imagen en color se representa en un espacio de color tridimensional. Existen numerosos espacios de repre-

sentación del color. El más conocido es el RGB. En este espacio, cualquier color se representa como un vector tridimensional que indica la proporción de rojo, verde y azul. Un primer atributo de color que se usa es el histograma de color (Brunelli y Mich, 1999). Estadísticamente, un histograma representa la probabilidad de que un píxel de la imagen tome un determinado color (triada de colores).

2. *Formas.* La extracción de las formas más significativas de una imagen es un proceso complejo, que se conoce como segmentación, y se complica aún más si los objetos están superpuestos, como sucede en imágenes reales. Distintos algoritmos de segmentación se pueden encontrar en Sonka *et al.*, (1998).

3. *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 con una matriz de concurrencia, propiedades psicológicas (contraste, regularidad, tosquedad, aspereza), transformadas *wavelet*, entre otras.

4. *Diseño del color.* Se trata de usar conjuntamente la característica de color y las relaciones espaciales. Una aproximación 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. Otras técnicas usan momentos de color sobre regiones o usan una matriz de concurrencia de color (Aranda *et al.*, 2002).

B. Calidad de la información –IQ–

En esta propuesta, cuando se habla de calidad, se refiere a la calidad de la información, donde esta información base son las imágenes. La calidad de una imagen en este tipo de sistemas se puede relacionar con características como: dispositivo, hora y

condiciones ambientales de captura de la imagen, así como resolución y formato de almacenamiento.

En general no existe un consenso a la hora de definir y clasificar las *características de calidad* que debe presentar un sistema y sus componentes, en este caso las imágenes, que son las fuentes de información del SRI. Para determinar las propiedades de calidad a tener en cuenta en el diseño del SRI de ecosistemas acuáticos amazónicos, se tiene en cuenta la terminología usada para productos software, donde una característica de calidad de un componente es un ‘conjunto de propiedades mediante las cuales se evalúa y describe su calidad’. En este trabajo, se denomina criterio a una propiedad de calidad a la que se le puede asignar una métrica, éste es un procedimiento que examina un componente y produce un dato simple, un símbolo (p.e. excelente, sí, no) o un número.

Un modelo de calidad (Bertoa *et al.*, 2003) es el conjunto de características y sub-características y de cómo éstas se relacionan entre sí. El objetivo principal es detectar los criterios que pueden describir la calidad en cada fuente (imagen), teniendo en cuenta la información que suministra la fuente y que facilitan su valoración por el motor de búsqueda. La definición de los criterios fue producto de una revisión de la literatura sobre la calidad asociada con fuentes de información manejadas en la web (Benett y Zhang, 1996) (Naumann y Rolker, 2000).

Selección y definición de criterios de calidad

Hay diferentes propuestas de clasificaciones de criterios de calidad en la literatura. Algunas como la clasificación de criterios propuesta en Kahn *et al.* (2002) y adaptada en Manrique (2006), determinan tres niveles o clases de criterios de calidad: el usuario, ya que él decide si la información recibida es cualitativamente buena o no; el proceso de la consulta, que contiene criterios relacionados con el proceso de acceso a la información; y la fuente, ya que para muchos criterios la fuente en sí misma es el origen de los criterios de calidad.

III SISTEMA DE RECUPERACIÓN DE IMÁGENES

A. Modelo de calidad

Se definió el siguiente modelo de calidad compuesto por un criterio, una descripción de criterio y una métrica asociada, descritos en la tabla 1.

Tabla 1. Modelo de Calidad

CRITERIO	MÉTRICA	DETALLE
USUARIO	Def.	Grado de satisfacción del usuario de acuerdo con sus necesidades iniciales de búsqueda.
	Descr.	Relacionada con las necesidades del usuario. Si la información contenida en la fuente logra satisfacer sus necesidades en términos del tipo y profundidad proporcionado.
	Métr.	No es medible. Depende de: la semántica que ofrezcan las fuentes de información (ontologías locales) y el modelo de dominio (ontología global); y de la retroalimentación del usuario.
FUENTE	Def.	Número de objetos del mundo real representados por cada fuente de información.
	Descr.	Proporciona la información que se necesita.
	Métr.	# Objetos Ontología Global / # Objetos Ontología Fuente
FUENTE	Def.	Porcentaje de atributos de las fuentes que coinciden con los requeridos.
	Descr.	Evidencia de sesgos o imprecisiones en una fuente. Inconsistencias en la información.
	Métr.	# Objetos Ontología Fuente / # Objetos Ontología Consulta
FUENTE	Def.	Memoria de comportamiento desde operaciones pasadas
	Descr.	Ver la confiabilidad de una fuente, relacionada con la reputación de la fuente (memoria de comportamientos pasados), más no como Confianza o Expectativa acerca del comportamiento futuro.
	Métr.	No es medible. Depende de los valores guardados de consultas anteriores cuya valoración por el Usuario haya sido aceptada como relevante. Uso de un metadato adicional que guarde continuamente estos valores.
PROCESO	Def.	Tiempo para recibir el resultado completo de la consulta
	Descr.	Valor entregado por la fuente, o guardado en cada consulta.
	Métr.	Valor entregado por la fuente.

Fuente: elaboración propia.

Para el SRI, una métrica para un criterio es la medida cuantitativa del grado en que una fuente posee cierto criterio, tal que examina una fuente y produce un dato (símbolo: SÍ, NO, F, V o un número). Siguiendo la estructura de una métrica (Naumann, 2000), ésta debe contener: definición, unidad, escala y fuente. Teniendo en cuenta el número de fuentes que maneje el SRI y el número de criterios IQ definidos, el sistema maneja una matriz en donde las columnas representan el número de fuentes asociadas al sistema, y las filas, el número de criterios IQ. Los valores en cada casilla representan el valor asociado con cada criterio para cada fuente, según la métrica determinada. Como producto de esta matriz, se genera un vector asociado para cada fuente de información, con pesos asignados a él teniendo en cuenta el número de criterios definidos.

B. Análisis y diseño del SRI

El análisis y diseño del sistema se llevó a cabo teniendo en cuenta los pasos propuestos por la metodología RUP (IBM, 1997) para el desarrollo de sistemas y/o productos software, teniendo en cuenta que un SRI es un tipo de sistema web, con unas condiciones particulares. A continuación se describen las etapas principales.

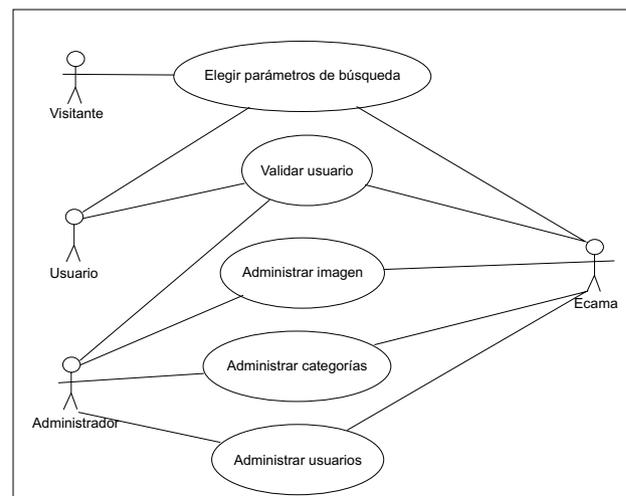
- **Inicio**

Dentro de las actividades ejecutadas en esta etapa, están:

- Estudio de información relevante,
- Identificación de funciones y roles por funciones, así:
 - o Invitado: elige parámetros de búsqueda y guarda imágenes
 - o Usuario: se registra en el sistema, elige parámetros de búsqueda y guarda imágenes, ingresa imágenes y edita metadatos.
 - o Administrador: se registra en el sistema, elige parámetros de búsqueda y guarda

imágenes, ingresa imágenes y edita metadatos; crea categorías y elimina imágenes de categorías, deshabilita usuarios, modifica contraseñas, genera informes del historial de movimientos de los usuarios.

- Identificación de requisitos funcionales y no funcionales
 - o Manejar información de imágenes
 - o Gestionar categorías
 - o Gestionar contenidos
 - o Gestionar informes de visitas de usuarios
 - o Solicitar confirmación de acciones sobre el sistema
 - o Accesibilidad
 - o Gestión de seguridad
 - o Amigabilidad en la GUI (Interfaz Gráfica de Usuario) y facilidad de uso del SRI.
- Modelado de comportamiento. Los casos de uso describen el comportamiento deseado del SRI y cómo es percibido por los visitantes, usuarios y administrador, sin tener que especificar cómo se implementa ese comportamiento. El diagrama de casos de uso se muestra en la figura 1.



Fuente: elaboración propia.

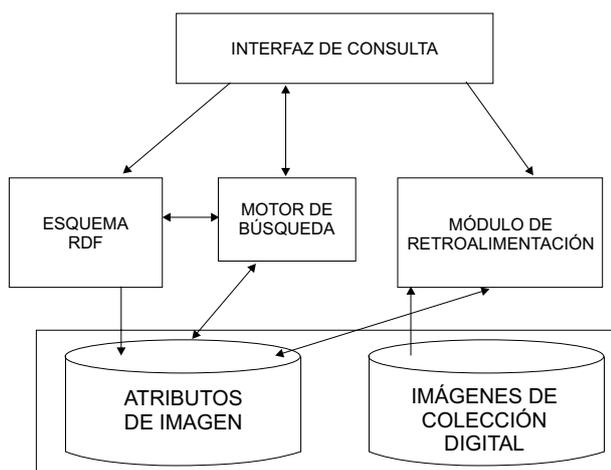
Figura 1. Diagrama de casos de uso general del sistema

• **Elaboracion**

Arquitectura del sistema

La visualización, especificación y diseño de un SRI requiere que sea visto desde varias perspectivas de los diferentes usuarios (visitantes, usuarios internos, administrador).

En la figura 2 se muestra un esquema de la arquitectura del sistema, donde se distinguen tres módulos principales y dos bases de datos. El módulo del nivel superior es la interfaz de consulta o GUI mediante la que el usuario interactúa con el sistema. Los otros dos módulos conforman la parte funcional. Se tiene un componente que lleva a cabo el procesamiento de las imágenes, uno que desempeña la función de motor de búsqueda y otro la de retroalimentación. Finalmente, están las bases de datos: la base de atributos contiene los histogramas de las imágenes, así como los atributos o metadatos asociados con ellas; la base de colecciones digitales que agrupa a las bases de datos de las distintas colecciones digitales (inicialmente se trabaja con una: la colección digital de ecosistemas acuáticos amazónicos) y es donde se guardan físicamente las imágenes.



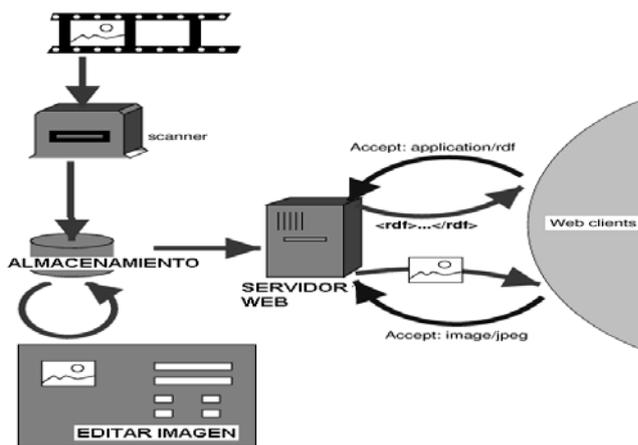
Fuente: elaboración propia.

Figura 2. Arquitectura del sistema

Resource Description Framework **RDF** (Marco de trabajo para la descripción de recursos)

RDF es un lenguaje para representar información acerca de recursos en la World Wide Web; particularmente para representar metadatos de recursos web, tales como título, autor, fecha de modificación, derechos de autor e información de licenciamiento de una página o documento web, pero también para cualquier otro tipo de ‘recurso web’ (que puedan ser recuperados como información acerca de artículos de compra-venta online, preferencias del usuario, entre otros).

La elaboración del SRI se soporta en el uso de RDF para la representación y catalogación de los recursos (imágenes de la colección digital), pues RDF puede usarse en este tipo de área de aplicación, así como también: en la recuperación de recursos para proporcionar mejores capacidades a los motores de búsqueda, en catalogación para la descripción de contenido y relaciones de contenido accesibles en un sitio Web particular, en una página, o biblioteca digital, a través de agentes de software inteligentes para facilitar que el conocimiento se comparta e intercambie, entre otros. La figura 3 muestra la arquitectura propuesta por RDF para la descripción de recursos de un sistema.

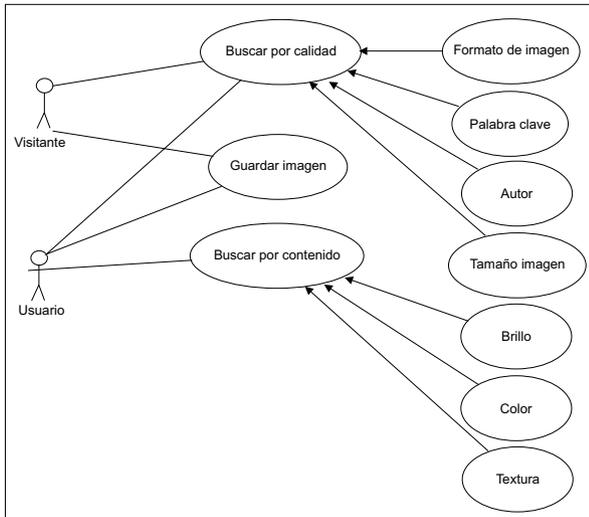


Fuente: W3C, 2002.

Figura 3. Arquitectura del RDF

Dentro de los RDF, los metadatos se separan en tres esquemas (W3C, 2002):

- Esquema de la base de Dublín. Es un esquema general para identificar los trabajos originales, típicamente libros y artículos, pero también películas, imágenes o fotos. Contiene características como: creador, redactor, título, fecha de publicación y editor.
- Esquema técnico. Este esquema captura los datos técnicos sobre la imagen y la cámara fotográfica, tal como el tipo de cámara fotográfica, el tipo de película, la fecha en que la película fue desarrollada y el explorador y el software usados para convertir a formato digital.
- Esquema contenido. Este esquema se usa para clasificar el tema de la imagen por medio de un vocabulario controlado. Este esquema permite recuperar las imágenes, basado en características como retrato, retrato del grupo, paisaje, arquitectura, deporte, animales, etc.



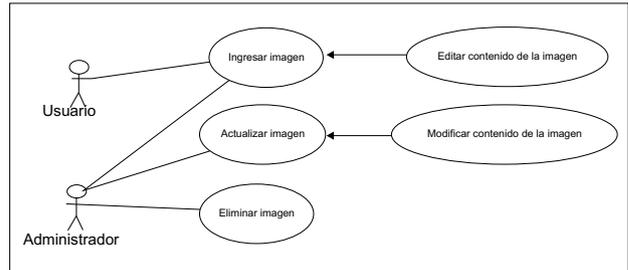
Fuente: elaboración propia.

Figura 4. Caso de uso: elegir parámetros de búsqueda

Modelamiento de comportamiento específico

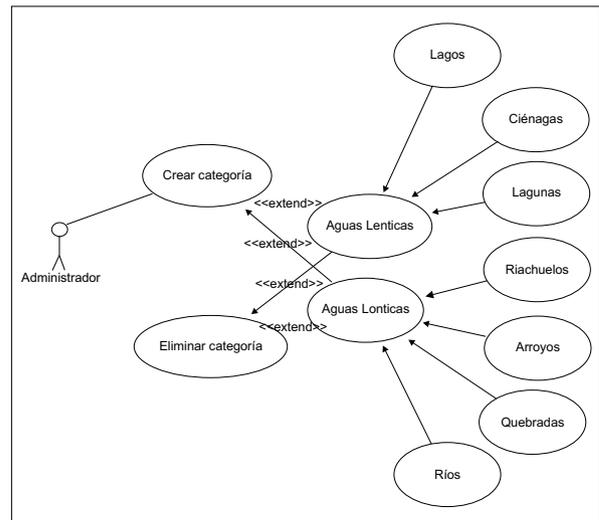
Los diagramas de casos de uso específicos representan la funcionalidad completa del sistema, mostrando su interacción con los agentes externos. Esta representación se hace a través de las relaciones entre los actores (agentes externos) y los casos

de uso (acciones) dentro del sistema. Se pueden visualizar como las funciones más importantes que la aplicación puede hacer o como las opciones presentes en el menú del prototipo. Cada una de ellas aparece en las figuras 4, 5, 6 y 7.



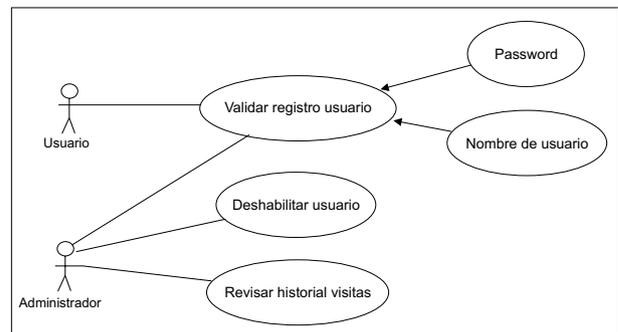
Fuente: elaboración propia.

Figura 5. Caso de uso: administrar imágenes



Fuente: elaboración propia.

Figura 6. Caso de uso: administrar categorías

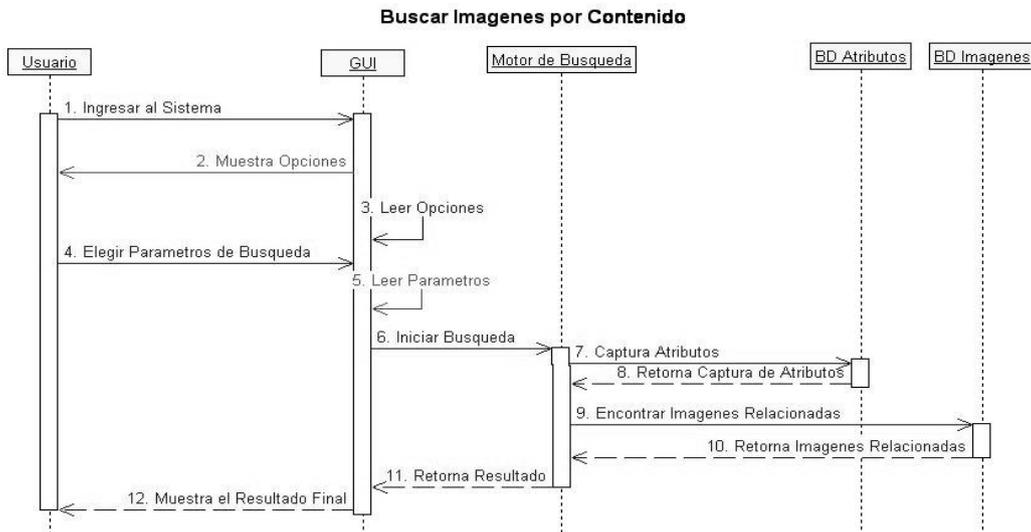


Fuente: elaboración propia.

Figura 7. Caso de uso: administrar usuarios

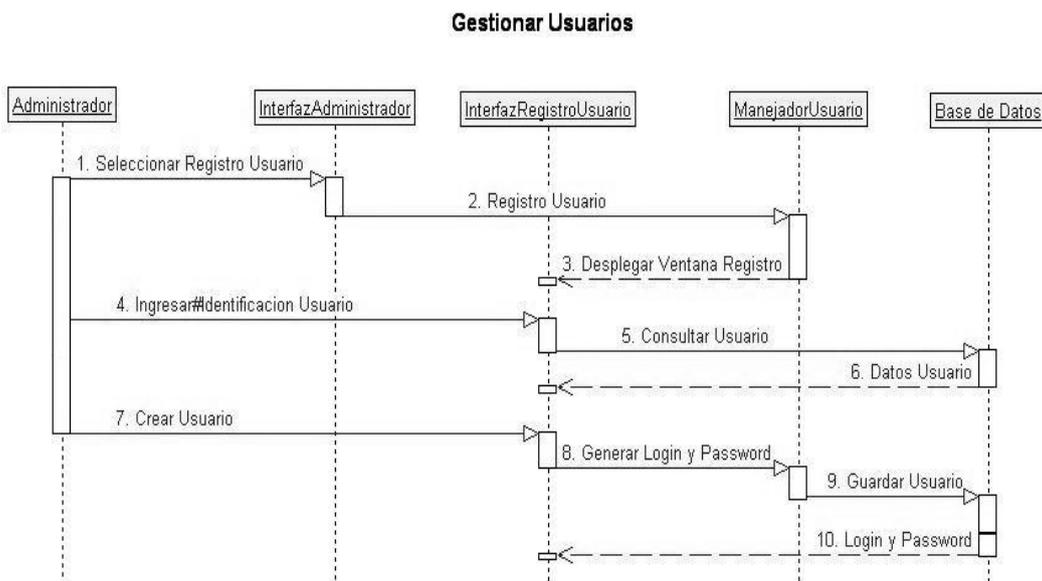
Los diagramas de secuencia representan una interacción entre objetos de manera secuencial en el tiempo. Muestran la participación de objetos en la interacción entre sus “líneas de vida” (desde que se instancian) y los mensajes que ellos organizadamente intercambian en el tiempo. El responsable o actor es quien inicia el ciclo interactuando ini-

cialmente con la interfaz de usuario GUI; luego se inician todos los objetos que intervienen en el funcionamiento del SRI. En este diagrama se comienza a observar el comportamiento del sistema a partir de los eventos generados por los actores. Aquí se interactúa con instancias, no con clases, como se muestra en las figura 8, 9 y 10.



Fuente: elaboración propia.

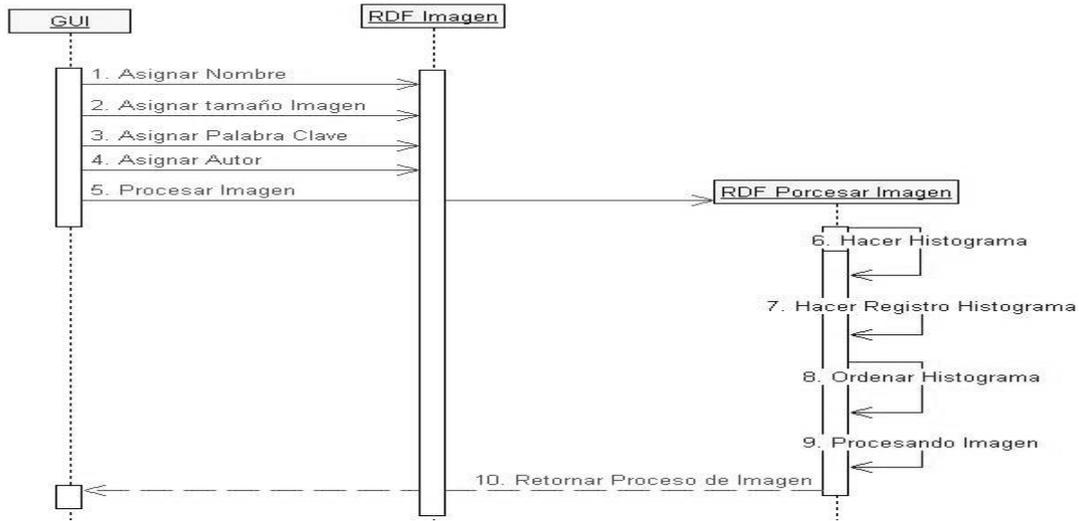
Figura 8. Diagrama de secuencia buscar imágenes por contenido



Fuente: elaboración propia.

Figura 9. Diagrama de secuencia gestionar usuarios

**Procesar una Imagen de Acuerdo al Esquema RDF
Busqueda por Calidad**



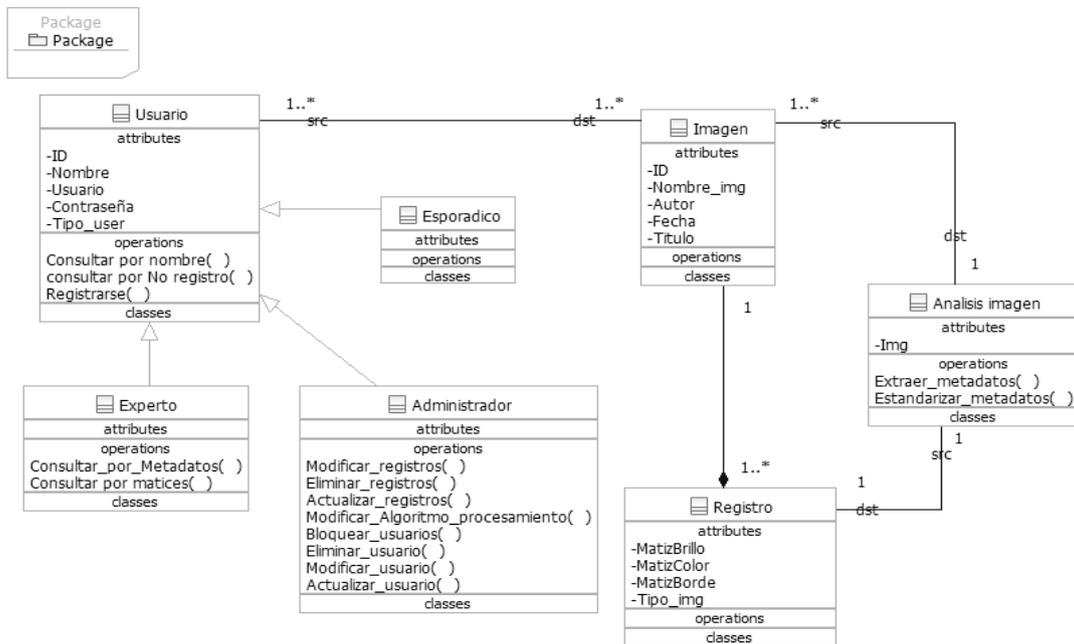
Fuente: elaboración propia.

Figura 10. Diagrama de secuencia buscar por calidad

Modelamiento estructural

La figura 11 muestra el diagrama de clases del SRI propuesto, donde se detalla estructural y

estáticamente cómo se relacionan los diferentes objetos y componentes.

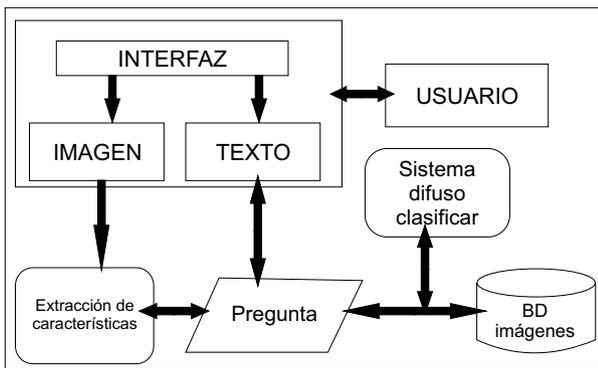


Fuente: elaboración propia.

Figura 11. Diagrama de clases

Diseño del prototipo de consulta

El SRI diseñado consta de un motor de búsqueda; por medio de una interfaz de consulta el usuario experto ingresa las imágenes y los metadatos asociados con cada una de ellas; asimismo, los criterios de búsqueda son ingresados por el usuario normal. Un bosquejo del prototipo se muestra en la figura 12, junto con los componentes del sistema y el proceso de búsqueda en el SRI.



Fuente: elaboración propia.

Figura 12. Modelo del prototipo

Tipos de búsquedas en el prototipo de SRI

- Búsqueda secuencial

La búsqueda secuencial muestra la información más relevante de una imagen aleatoria (título, descripción, formato, tamaño, etc.) de la base de datos y los enlaces para seguir buscando otras imágenes relacionadas, como se muestra en la figura 13.

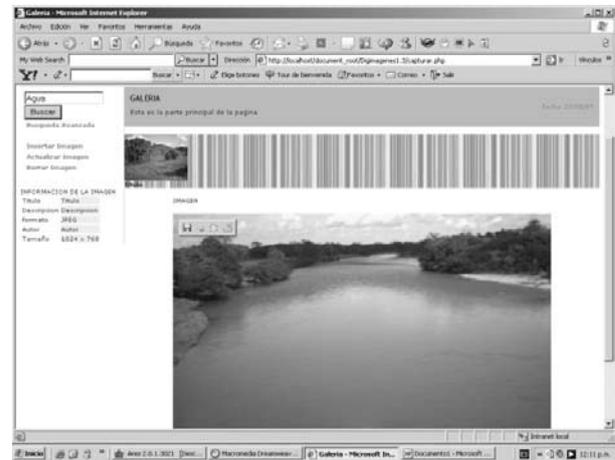
- Búsqueda avanzada

la búsqueda avanzada muestra todos los resultados encontrados en la base de datos para el tipo de imagen que se necesita, teniendo en cuenta los criterios de formato, descripción y autor, como se muestra en la figura 14.

- Búsqueda por contenido

La búsqueda por contenido se ha implementado teniendo en cuenta los criterios de contenido

de la imagen: color y brillo, como aparecen en las figuras 15 y 16.



Fuente: elaboración propia.

Figura 13. Búsqueda secuencial



Fuente: elaboración propia.

Figura 14. Búsqueda avanzada



Fuente: elaboración propia.

Figura 15. Resultado búsqueda por color



Fuente: elaboración propia.

Figura 16. Búsqueda por nivel de brillo

La búsqueda por color solicita al usuario que digite los porcentajes de color que desea en la imagen y el sistema retorna las imágenes con cantidades aproximadas a esos porcentajes dados; lo mismo sucede con la iluminación, se pide un factor de iluminación y el sistema trae los resultados más próximos.

CONCLUSIONES Y TRABAJO FUTURO

En muchas aplicaciones de bases de datos científicas, el contenido de información de las imágenes no es explícito y esto no es adecuado para su indexación directa, clasificación y recuperación. Particularmente, las bases de datos de imágenes en gran escala han surgido como uno de los problemas más retadores en el campo de bases de datos científicas, sobre todo en contextos en los que la mayor cantidad de información asociada con las imágenes de las colecciones está relacionada con su contenido (forma, color, texturas).

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 a una satisfacción mayor del usuario respecto a sus criterios de búsqueda iniciales.

Se ha modelado un SRI, teniendo en cuenta aspectos de la recuperación basada en contenido, y la recuperación basada en la calidad de la información. Este soporte se tuvo en cuenta desde el análisis de requerimientos, hasta el diseño del prototipo del sistema y de las consultas requeridas.

A pesar de no existir un consenso a la hora de definir y clasificar las características de calidad que debe presentar un sistema y sus componentes, en este caso las imágenes, se puede usar el referente para productos software, en donde una característica de calidad de un componente es un conjunto de propiedades mediante las que se evalúa y describe su calidad. Igualmente, marcos de trabajo como el de RDF son adecuados por el esquema de clasificación de atributos para recursos web, como las imágenes.

Es necesario tener en cuenta propuestas como la de RDF para procesar los metadatos asociados con las imágenes y colecciones de imágenes, ya que RDF proporciona interoperabilidad entre los usuarios que intercambian la información máquina-Web, automatiza el procesamiento de los recursos Web, y se puede usar en una variedad de áreas de aplicación. Se está trabajando en la consolidación de la aplicación de los esquemas RDF, particularmente en el contexto de este tipo de recursos web (imágenes), que tiene un comportamiento particular y diferente a otro tipo de recursos.

Aunque existen SRI basados en el contenido, que incluyen alguna o varias de características como color, textura, formas de objetos en la imagen para hacer la búsqueda, la mayoría son sistemas de propósito general y no hay estudios en los que éstos se usen en aplicaciones prácticas para investigar las ventajas e inconvenientes de las distintas opciones.

Se requiere el desarrollo de SRI en contextos específicos que integren, además de técnicas basadas en contenido, técnicas de control de la calidad de la información que se maneja en la colección digital.

Se está desarrollando la fase de implementación, pruebas y puesta en marcha del SRI, que permitirá validar el modelo desarrollado y descrito en este trabajo, y la propuesta de clasificación de imágenes de ecosistemas acuáticos amazónicos. Estos resultados se presentarán en una posterior publicación.

AGRADECIMIENTOS

Los autores agradecen: al Grupo de Investigación CAPREA (Calidad y Preservación de Ecosistemas Acuáticos Amazónicos) de la Universidad de la Amazonia por su colaboración con la clasificación de los ecosistemas y las imágenes de la colección digital; a las Vicerrectorías de Investigaciones de la Universidad de Medellín y la Universidad de la Amazonia, por proporcionar las condiciones y apoyo financiero para la ejecución del presente trabajo.

REFERENCIAS

- 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" Universidad de Málaga, España.*
- BENNETT, J. C. y ZHANG H. (1996). *Worst-case fair weighted fair queueing*. In IEEE INFOCOMM'96, pages 120-128. IEEE.
- BERTOIA, M. F., TROYA, J. M. y VALLECILLO, A. (2003). *Atributos de Calidad para Componentes COTS: Una valoración de la información ofrecida por los vendedores*. <<http://www.lcc.uma.es/~av/Publicaciones/03/TICS03.pdf>> Recuperado el 30 de noviembre de 2007.
- BORGMAN, C. L. (1999). *What are digital libraries, who is building them, and why?* En T. Aparac (Ed.), *Digital libraries: Interdisciplinary concepts, challenges and opportunities*. Proceedings of the Third International Conference on Conceptions of Library and Information Science. Dubrovnik, Croatia. 23-38 pp.
- BRUNELLI, R. y MICH, O. (1999). *On the Use of Histograms for Image Retrieval*. IEEE Conference on Multimedia Computing & Systems. 143-147 pp.
- HIRATA K. y KATO T. (1992). *Query by Visual Example*. Advances in Database Technology EDBT'92. Third Intl. Conf. on Extending Database Technology. <<http://jardzia.ifp.uiuc.edu:8000>> Recuperado el 6 de diciembre de 2007.
- KAHN, B.; LEE, Y.; STRONG, D.; y WANG, R., 2002. *AIMQ: A methodology for information quality assessment*. Information & Management. 40, 133-146 pp.
- 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. 52-62 pp.
- NAUMANN, F., 2000. *Quality-driven Query Planning*. Dissertation PhD. Outline Humboldt-Universität zu Berlin. <<http://www.edbt2000.uni-konstanz.de/phd-workshop/papers/Naumann.ps>> Recuperado el 30 de noviembre de 2007.
- NAUMANN, F. y ROLKER, C. (2000). *Assessment Methods for Information Quality Criteria*. <<http://citeseer.ist.psu.edu/295186.html>> Recuperado el 30 de noviembre de 2007.
- OBEID, M., JEDYNAK, B. & DAOUDI, M. (2001). *Image Indexing & Retrieval Using Intermediate Features*. Francia. <<http://cis.jhu.edu/~bruno/intermediate-features.pdf>> Recuperado el 1 de diciembre de 2007.
- RUI, Y., HUANG, T. & FU, S. (1999). *Image Retrieval: Current Techniques, Promising Directions, and Open Issues*. Journal of Visual Communications and Image Representation. 10, 39-46 pp.
- SONKA, M., HLAVAC, V. & BOYLE, R. (1998). *Image Processing, Analysis and Machine Vision*. PWS Publishing.
- W3C., 2002. *Describir y recuperar recursos usando RDF y HTTP*. <<http://www.w3.org/TR/2002/NOTE-photo-rdf-20020419>> Recuperado el 1 de diciembre de 2007.