

Recuperación Semántica de Imágenes y Datos de Fuentes Heterogéneas

María Constanza Pabón B.
Departamento de Ciencias e Ingeniería de la Computación
Pontificia Universidad Javeriana – Seccional Cali

Diciembre de 2010

Resumen: En el campo médico, la información requerida para procesos de ayuda diagnóstica, enseñanza, e investigación, supone integrar datos de diversas fuentes y formatos; entre ellos imágenes médicas y datos clínicos del paciente. Una forma de lograrlo es aplicar un modelo de integración basado en un mediador, en el que se define un modelo de datos global sobre el cual se realizan las consultas. Este reporte resume el trabajo de revisión de literatura sobre recuperación de imágenes, integración de datos de fuentes heterogéneas, y modelos de datos conceptuales; e introduce la posibilidad de proponer de un lenguaje conceptual de consulta, aplicado en primera instancia al dominio médico, que tome ventaja de la experticia del usuario, para proveerle la posibilidad de construir consultas complejas.

1. Introducción

En los últimos años gran cantidad de información ha sido capturada en imágenes. En particular, las imágenes médicas permiten a los especialistas conocer la condición del paciente a partir de su interpretación y análisis. Día a día se generan imágenes médicas en grandes volúmenes, información que podría ser utilizada para apoyar, aún de mejor manera, el diagnóstico, así como también para el aprendizaje y la investigación. Para ello, se requieren herramientas eficientes que permitan recuperar las imágenes que son relevantes para un objetivo dado. Sin embargo, centrar las consultas solamente en las características de la imagen (de bajo nivel o semánticas) no permite al usuario dar respuesta a preguntas complejas que requieren relacionar la información proveniente de la imagen con datos de otras fuentes. De acuerdo con Rubin [1] el diagnóstico médico en radiología sigue tres pasos: primero la percepción de la imagen, seguido de la interpretación de la misma, y finalmente la decisión y recomendaciones. En los dos últimos pasos el especialista tiene en cuenta, además de la imagen, la historia clínica y la información del paciente. Con base en ello, y aplicando heurísticas propias de su disciplina, toma una decisión sobre el caso. Normalmente la información relacionada con la imagen se encuentra en diversas fuentes de datos (Ej. historias clínicas, reportes de diagnóstico, datos del paciente, metadatos de la imagen, ontologías) que cuentan con mecanismos de almacenamiento de diversos tipos (bases de datos relacionales, documentos RDF, DICOM, XML, etc.). Se hace entonces necesaria la integración de fuentes de datos heterogéneas, de manera que permita a los usuarios recuperar información relacionando de manera transparente los datos de diversas fuentes. Una de las arquitecturas usadas para la integración de datos se basa en un mediador, ente en el que se define un modelo global de los datos, y un mapeo entre este y los

esquemas de datos de las fuentes de datos locales. Sobre el modelo de datos global se pueden expresar consultas que al ejecutarse recuperen datos de las diversas fuentes. Los modelos de datos conceptuales, que constituyen una opción para la representación del modelo de datos global, pueden soportar el desarrollo de lenguajes de consulta, de alto nivel semántico, y cercanos al usuario.

Este reporte resume el trabajo de revisión de literatura de sobre los temas mencionados. La sección 2, presenta métodos de recuperación de imágenes. La sección 3, hace una breve descripción de los aspectos relevantes de integración de datos, particularmente de los enfoques basados en mediación. La sección 4 introduce los modelos de datos conceptuales y presenta algunos lenguajes de consulta que operan sobre estos modelos. Finalmente, en la sección 5 se esboza, de manera preliminar, la posibilidad de proponer un lenguaje conceptual de consulta, aplicado en primera instancia al dominio médico, que permita a los médicos realizar consultas que integren diversas fuentes de información sobre sus pacientes, y tome ventaja de la experticia de los usuarios en el tema. Esto con el fin de proveer una herramienta de consulta para el proyecto *Sistema de gestión y recuperación de imágenes médicas enriquecidas semánticamente*, el cual está llevando a cabo el grupo de investigación GEDI, de la Universidad del Valle.

2. Recuperación de Imágenes

La recuperación de imágenes es diferente de la recuperación de información en texto. El contenido de una imagen no es explícito, no existe una terminología o una sintaxis para su descripción [2]. El texto por su parte, tiene una estructura sintáctica y un vocabulario que asocia un significado. El elemento mínimo de un texto, el alfabeto, forma palabras fácilmente identificables en un proceso automático. Si bien las palabras en si mismas no definen totalmente la semántica de un texto, son parte fundamental de ella. Luego, la búsqueda con palabras clave tiene inherentemente semántica. Por otro lado, las técnicas de análisis de imágenes disponibles solo son capaces de identificar características visuales de bajo nivel como color, forma, textura, y relaciones espaciales. Pero la identificación de contenido semántico de alto nivel a partir de las características de bajo nivel es un problema no resuelto. Las imágenes contienen gran cantidad de información cuya semántica aún no puede ser capturada por métodos automáticos, a este hecho se le ha denominado la *Brecha Semántica (Semantic Gap)*. La información capturada en imágenes crece cada día, y constituye una fuente importante de información y conocimiento. Para acceder a ello, es necesario desarrollar estrategias que reduzcan la brecha semántica y permitan a los usuarios recuperar imágenes que satisfagan sus necesidades. Algunas de estas estrategias se pueden basar en el desarrollo de mejores métodos de consulta, que provean una semántica de alto nivel.

2.1. Métodos de Consulta de Imágenes

Los métodos de consulta pueden jugar un rol importante en la disminución de la brecha semántica en sistemas de recuperación de imágenes [3]. Todo método de consulta de imágenes requiere del análisis e indexación de las imágenes sobre las cuales se realiza la consulta, de un mecanismo mediante el cual el usuario describa lo que busca, y de un proceso de recuperación que sea capaz de seleccionar las imágenes más relevantes para la consulta. A continuación se presenta una síntesis de los métodos comúnmente usados en estas tres etapas.

2.1.1. Anotación de las imágenes

Comúnmente, los sistemas de recuperación de imágenes trabajan sobre un repositorio de imágenes previamente anotadas, con el fin de hacer la recuperación más eficiente en términos de tiempo de

respuesta. En general, es posible identificar dos tipos de anotaciones: anotaciones basadas en el contenido de la imagen, y anotaciones basadas en texto. Las del primer tipo [4], consisten de descriptores numéricos que representan características visuales de las imágenes, como el color, la textura, la forma y la ubicación espacial. Se calculan descriptores, de toda la imagen ó de segmentos de la misma, aplicando modelos estadísticos o matemáticos. Los segmentos se definen usando algoritmos de segmentación. Las anotaciones con texto, el segundo tipo, surgen debido a que la semántica de las imágenes se expresa, de mejor forma, con palabras clave o conceptos que describen su contenido o el de una región de interés (ROI por sus siglas en inglés).

2.1.2. Tipos de Consulta de Imágenes

Khapli y Bhalachandra [5] identifican tres tipos de consulta usadas para recuperación de imágenes: Consulta con imagen de ejemplo (*Query by image example*), Consulta por características de la imagen (*Query by image feature*), y Consulta con Texto (*Query by text*). Para solventar las desventajas de cada tipo [6][7][8], muchos sistemas los combinan implementando consultas híbridas.

En las consultas basadas en imagen de ejemplo, el usuario provee una imagen o selecciona una ROI, que en su criterio es representativa de lo que busca. Dependiendo de la forma de recuperación implementada, el sistema retornará las imágenes con características visuales o semánticas similares. Este tipo de consulta ha sido ampliamente aplicada en sistemas CBIR (*Content Based Image Retrieval*), pero tiene dos desventajas principales. La primera, consiste en que en la mayoría de los casos, es difícil para los usuarios proveer una imagen apropiada para iniciar la búsqueda. La segunda, es la dificultad de capturar la intención del usuario, es decir que el sistema pueda identificar las características que lo llevaron a escogerla como ejemplo. QBIC [9], I²Cnet [10], y CervigramFinder [11] son sistemas que usan consulta con imagen de ejemplo.

En consultas basadas en características de la imagen, el usuario provee las características visuales representativas de la imagen que busca. Una manera de hacerlo es mediante un boceto en el cual representa regiones con formas, colores, y texturas. Chalechale et al. [12] proponen APAI un método que mide la similitud entre bocetos en blanco y negro, y objetos complejos de imágenes en color. Assfalg et al. [6], por su parte, presentan un paradigma de interacción en el cual el usuario crea un boceto en 3 dimensiones, y especifica la consulta con una “foto” del ambiente creado, capturada desde algún punto de vista específico. De esta manera, el boceto incluye la relación espacial de los objetos.

En las consultas con texto, el usuario especifica su búsqueda mediante palabras clave (descriptores en texto) o lenguaje natural (texto libre). En algunos casos, el texto se restringe con alguna sintaxis o vocabulario definido. Möller y Sintek [13] denominan *Consulta por Concepto* al caso especial en el cual los términos de la consulta se restringen a los definidos en una ontología. En este caso, el sistema puede permitir al usuario recorrer la ontología para seleccionar los términos que va a incluir en la consulta, o digitar la consulta y luego verificar que los términos se encuentren en la ontología. Algunas aplicaciones usan técnicas de refinamiento para desambiguar los términos. Las consultas en lenguaje natural, pueden expresarse como una oración o una historia. Barnard y Forsyth [14] proponen una asociación estadística entre las imágenes y el texto, que permite seleccionar las imágenes que mejor describen un texto. En SPE (*Story Picturing Engine*) [15] el usuario digita una historia, ésta se procesa, y se identifican las palabras clave que se usan para seleccionar las imágenes.

Las consultas híbridas son usadas frecuentemente, ejemplo de ello son los sistemas Paragrab [16], Shoebox [17], SPIRS [11], e IRMA [18]. Paragrab usa técnicas de procesamiento de lenguaje natural y refinamiento, y combina metadatos con características visuales para recuperar las imágenes. Shoebox ofrece dos métodos de consulta, uno basado en texto y reconocimiento del habla, y el otro basado en características visuales. SPIRS (*Spine Pathology and Image Retrieval System*) combina los tres tipos de consulta. Por su parte, el proyecto IRMA (*Image Retrieval in Medical Applications*) trata de proveer soporte para el diagnóstico médico realizado con base en imágenes radiológicas, para lo cual implementa consultas híbridas, con bosquejos e imágenes de ejemplo.

2.1.3. Proceso de Recuperación de Imágenes

Los procesos de recuperación de las imágenes dependen de la forma como estas están anotadas y del tipo de consulta. Las consultas con texto, pueden recuperar imágenes anotadas con texto o con descriptores de bajo nivel. El primer caso, es una búsqueda textual. En el segundo caso [3][19], el sistema cuenta con descriptores visuales de imágenes representativas de los conceptos que soporta en las consultas. El resultado se basa en la similitud entre los descriptores visuales de los conceptos y los de las imágenes del repositorio. De esta forma se plantea una solución al hecho de que las anotaciones con texto, de una imagen, siempre son incompletas, puesto que solo capturan las características que son relevantes bajo algunos criterios.

A su vez, las consultas con imagen de ejemplo o bocetos, pueden recuperar imágenes anotadas con descriptores de bajo nivel o con texto. Si las imágenes del repositorio están anotadas con descriptores visuales [20], el resultado se elige con base en una medida similitud entre los descriptores de la imagen ejemplo y las del repositorio. Por otro lado, si las imágenes del repositorio están anotadas con texto [21], se trata de identificar conceptos u objetos que aparecen en la imagen ejemplo, a través de la similitud entre regiones de la imagen y descriptores visuales asociados a los conceptos. Una vez los conceptos son identificados, se hace una búsqueda textual. De esta manera, es posible consultar imágenes de diversos formatos a partir de una imagen de ejemplo.

En todos los casos, debido a que los resultados no son precisos, estos tienen asociado un valor que especifica su relevancia. Sólo las imágenes con mayor relevancia (mayor probabilidad de satisfacer la consulta) se presentan al usuario. Para ello, se pueden seleccionar las imágenes de mayor rango (*top K-rank*) o permitir que el usuario defina un mínimo valor aceptable (*threshold*).

2.2. Otras Estrategias Usadas en la Recuperación de Imágenes

Los tipos de consulta y procesos de recuperación de imágenes se complementan, en muchos casos, con otras estrategias que buscan para mejorar la precisión de los resultados, tales como el refinamiento de consultas (*Query Refinement*), y el *relevance feedback*. Comúnmente, estas estrategias se basan en el aprendizaje a partir de la interacción con el usuario, capturando información que permita identificar la intención de su búsqueda, con lo cual se mejoran los resultados de las consultas [4].

El propósito de las estrategias de refinamiento de consultas es ayudar al usuario a representar su consulta de manera más precisa. Se encuentran diversas formas de refinamiento, entre ellas: sugerir palabras similares a las que el usuario está digitando (*spelling suggestions*) [16], sugerir expansión semántica presentando sinónimos, hipónimos, hiperónimos, y términos relacionados [16], ofrecer explicación que justifique la selección de las imágenes para el resultado [18] y, en dominios

específicos, identificar conceptos muy generales para solicitar al usuario agregar mayor detalle [13]. En estos casos, se usan vocabularios (*lexicons*), taxonomías, u ontologías. Sin embargo, no siempre los resultados son los esperados. Una investigación empírica hecha por Chai et al. [22], en la cual el sistema propone al usuario una lista de términos relacionados que puede agregar a la consulta, muestra como resultado, que los usuarios tienden a incluir términos relevantes e irrelevantes en proporciones similares. Por lo tanto, los resultados de las consultas no mejoran.

El *Relevance feedback* (RF) [23] es un método en el cual sistema solicita al usuario retroalimentación sobre la relevancia de las imágenes que recupera la consulta. El usuario clasifica cada resultado como positivo o negativo, o elige los resultados positivos, o le asigna un grado de relevancia a cada imagen. RF captura la información y aprende de ella usando diversas técnicas [24], tales como: SVN (*Support Vector Networks*), ANN (*Artificial Neural Networks*), y aprendizaje supervisado

Para mejorar la anotación y la recuperación de las imágenes, se han usado ontologías y tesauros. Estrategias de refinamiento de consultas se han desarrollado usando ontologías para expandir los términos (i.e. MeSH [25]), o para explicar y justificar resultados. En anotación [26], automática o manual, con ontologías se establece un vocabulario compartido. En anotación automática, la imagen se segmenta, los descriptores de características visuales se determinan para cada segmento, se calcula su similitud con los descriptores de los conceptos en la ontología, y se anota la imagen con aquellos conceptos de mayor similitud. Algunos de estos sistemas de anotación automática usan las relaciones espaciales y la información de contexto para refinar la anotación [27][28][13].

En la formulación de las consultas [7], el usuario selecciona de la ontología los conceptos que incluye en la búsqueda. De esta forma, se recuperan imágenes anotadas con texto, o se usan descriptores de características visuales representativos de los conceptos para buscar imágenes similares. En lenguajes de consulta [29] las ontologías se han empleado para mantener el vocabulario y la descripción semántica de los conceptos que el usuario puede incluir en la consulta. En recuperación de imágenes [13] se han usado ontologías de conocimiento del dominio de la aplicación, ontologías para anotación (incluyen atributos usados para anotar la imagen, i.e. relaciones espaciales), y ontologías para análisis (ontologías de objetos u ontologías visuales, que facilitan asociar los conceptos del dominio de la aplicación con propiedades visuales y relaciones espaciales que se pueden detectar con base en características visuales de bajo nivel).

2.3. Lenguajes de Consulta para Recuperación de Imágenes

De acuerdo con Liu et al. [3][30], los lenguajes de consulta diseñados para la recuperación de imágenes pueden ayudar a reducir la brecha semántica. Los lenguajes de consulta son una forma de mitigar problemas encontrados en los CBIR (*Content Based Image Retrieval*), tales como la ambigüedad en el contenido de la imagen o la necesidad de inferir la intención de búsqueda del usuario [3]. Lenguajes como PICQUERY+ [31][32], FOQL [33], OQUEL [3][7], LUCENE [34], y SQL/MM se han desarrollado (o extendido) con este propósito.

PICQUERY+ [31][32] ofrece operaciones de manipulación, reconocimiento de patrones, espaciales o geométricos, y funciones definidas por el usuario. Usa una consulta tabular que permite establecer las condiciones de búsqueda, que se evalúan de manera exacta o relajada, y seleccionar las bases de datos y los métodos de procesamiento y visualización. La consulta puede incluir predicados temporales, evolutivos y de tipo *where*. El lenguaje es independiente del dominio, pero la interfaz para la consulta se debe desarrollar a la medida de cada aplicación.

FOQL [33] es una extensión del estándar OQL de ODMG (*Object Data Management Group*). Su consulta tiene la estructura SELECT-FROM-WHERE, sus predicados pueden tener un peso asociado, y se pueden evaluar con valores *booleanos* o *fuzzy*. El usuario puede nombrar las consultas para su uso posterior.

OQUEL [3][7] es un lenguaje de propósito general, desarrollado para el sistema ICON. ICON combina consultas con imágenes ejemplo (o contraejemplo), bocetos, propiedades de la imagen (i.e. ajuste de la cámara, color), y texto. Las consultas con texto se expresan en OQUEL, lenguaje basado en una ontología que incluye conceptos de alto y bajo nivel. Los últimos se refieren a características de la imagen y a relaciones que se evalúan a partir de los descriptores de la imagen, i.e. por similitud de características. A partir de la detección de conceptos de bajo nivel se calcula la probabilidad de la presencia de un concepto de alto nivel. Los autores ofrecen un prototipo para la recuperación de imágenes fotográficas de paisajes, cuyas categorías visuales están basadas en un descriptor de color.

Pein et al. [34] extienden LUCENE, un lenguaje de consulta de documentos de texto, y le agregan la capacidad para tratar consultas híbridas, que permiten integrar la semántica de alto nivel con las características visuales de bajo nivel. Para ello, le agregan operadores fuzzy y una gramática de dos niveles que permite extender el lenguaje. En el primer nivel la gramática es fija, y permite especificar búsquedas en la meta-data. En el segundo nivel, el lenguaje puede ser extendido con *plug-ins* que se pueden usar para definir, por ejemplo, métodos para calcular vectores de características y sus funciones de similitud.

Los lenguajes de consulta para multimedia son también un referente importante, ya que incluyen imágenes entre los datos que manejan. Múltiples lenguajes de consulta para datos multimedia se han propuesto, Cao et al. [36], proponen UMQL y hacen referencia a 16 lenguajes, clasificándolos en tres categorías: lenguajes con enfoque lógico o funcional (ej. EVA), extensiones de SQL (ej. MOQL, FOQL, VQL, VideoSQL, PSQL, SpatialSQL, SEQL, ESQL), y lenguajes especializados (MDQL, PICQUERY+, MMQL, CVQL). ISO/IEC 13249, SQL/MM, estandariza las extensiones SQL para manejo de datos multimedia, incluyendo texto (*full text*), imágenes (*still images*) y datos espaciales (*spatial data*); también define las extensiones para minería de datos. *Still image* permite almacenar imágenes en diferentes formatos (JPEG, TIFF, GIF, etc.), acompañadas por datos como formato, dimensiones, espacio de color, y descriptores visuales. El estándar propone para este tipo de datos, métodos para escalar, cortar, rotar, crear una miniatura, y crear descriptores de la imagen. Con algunas variaciones, SQL/MM está implementado en [35] PostgreSQL-IE, DB2 Image extender, Informix Excalibur/Image Foundation Datablade, y Oracle Intermedia. De otra parte, MPQF (*MPEG Query Format*) [37] propone un estándar para el acceso unificado a repositorios multimedia distribuidos, es un lenguaje basado en XML y hace parte del estándar MPEG-7.

3. Recuperación de datos de fuentes heterogéneas

La Integración de Datos busca proveer a los usuarios una vista unificada de datos que se encuentran en fuentes diversas y heterogéneas. Esta vista unificada puede ser materializada o virtual, materializada cuando se replican los datos de las diversas fuentes en un repositorio central, virtual cuando los datos se acceden en sus fuentes locales. Diferentes modelos y arquitecturas se han propuesto para resolver el problema de integración de datos de fuentes heterogéneas, desde el punto de vista técnico los principales enfoques son [38]: *Data Translation*, *Query Translation* e *Information Linkage*. En el primero, *Data Translation* (traducción de datos, también llamado *Data*

Exchange), los datos de las diferentes fuentes se copian a un repositorio central, modificándolos para encajarlos en un esquema unificado (*target schema*). Este enfoque es usado en la creación de bodegas de datos. En *Query Translation* (traducción de consultas), los datos permanecen en sus respectivas bases de datos. La consulta del usuario se traduce en consultas que son procesadas en cada una de las fuentes de datos, cuyas respuestas se integran. Finalmente, en *Information Linkage*, la integración se logra usando referencias entre los datos de las diversas fuentes.

Los sistemas que integran datos usando *Query Translation*, o sistemas basados en mediación, se implementan sobre una arquitectura que usualmente incluye la interfaz de usuario, un mediador, *wrappers*, y las fuentes de datos. La interfaz de usuario captura la consulta, la expresa en un lenguaje y la entrega al mediador. El mediador contiene el modelo de datos global y los mapeos entre éste y los esquemas de datos locales de las fuentes que integra. Con estos modelos, global y locales, se procesa la consulta, generando subconsultas que procesan las respectivas fuentes de datos. Cuando el lenguaje de las subconsultas es diferente al lenguaje que procesa una fuente de datos, se necesita un *wrapper*, que haga la traducción entre ambos lenguajes. Finalmente, las subconsultas se ejecutan en cada fuente de datos, el mediador recibe los resultados, los integra y los entrega al usuario. La Figura 1, muestra una arquitectura genérica de un integrador basado en mediación, adaptada de [39].

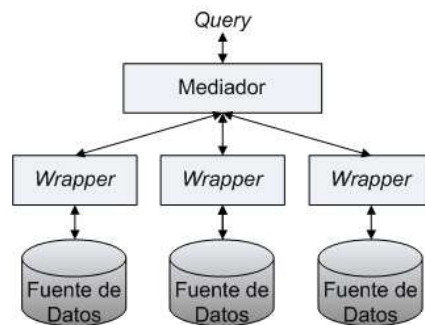


Figura 1. Arquitectura genérica de un integrador basado en mediación.

En términos generales, existen dos formas de abordar el problema de definir el mapeo entre los esquemas de datos locales de las fuentes y el modelo global [40]: *Global-as-View* (GAV) o *Local-as-View* (LAV). En el primero, el modelo global se define a partir de los esquemas locales, y cada uno de sus elementos mapea a una vista que combina datos de los esquemas locales. En el segundo, LAV, el modelo global es independiente de los esquemas locales, y cada elemento de los esquemas locales mapea a una vista del modelo global. Este último, facilita la extensibilidad del sistema, pues es posible agregar nuevas fuentes locales sin alterar el modelo global.

Entre los enfoques basados en mediación, los modelos de integración semántica [41] de datos se caracterizan por construir el modelo global que no incluye detalles de implementación, solo incluye los elementos necesarios para representar la semántica del dominio. Varios métodos de modelado de conocimiento y de datos cumplen esta característica, entre ellos las ontologías y los modelos de datos conceptuales. En el caso de las ontologías las representaciones más usadas son las asociadas con la Web Semántica, RDF y OWL, con SPARQL como lenguaje de consulta. Algunas aplicaciones generan una vista en RDF de un esquema relacional y permiten realizar consultas SPARQL sobre él [42], entre ellas: Virtuoso, DartGrid, SPASQL, Squirrel, RelationalOWL y D2RQ. La importancia de estas aplicaciones se debe a la necesidad de publicar en la web una gran cantidad de datos que se tiene en bases de datos relacionales.

Del uso de modelos de datos conceptuales se trata en la sección 4. Los modelos más usados son [41] el modelo ER (entidad-relación), el ORM (*Object Role Modeling*) y FCO-IM.

3.1. Integración de datos en el dominio médico

En el dominio médico se han desarrollado algunas soluciones que integran semánticamente fuentes de datos heterogéneas, entre ellas, MEDICO [43], NIF [44] (*Neuroscience Information Framework*), ACGT [38] (*Advancing Clinico-genomic Trials on Cancer*), y NeuroBank [45].

MEDICO [46-48], [43] hace parte del proyecto THESEUS, cuyo propósito es construir un motor de recuperación de imágenes médicas. Las imágenes se consultan con base en el contenido semántico y en descriptores de bajo nivel. La arquitectura de MEDICO incluye una jerarquía de ontologías, cuyos conceptos se usan para anotar las imágenes, dotándolas así de semántica. Las consultas se hacen a través de imágenes de ejemplo, conceptos (palabras de la ontología), o lenguaje natural. Como resultado, la interfaz construye una consulta en SPARQL. Con base en esta consulta es posible recuperar imágenes en diferentes formatos, datos del paciente, y metadatos almacenados en formato DICOM o RDF. Las consultas se ejecutan a través de *wrappers* y tres capas de mediación: *Query/Inference/Analysis*, *Index/Knowledge Base*, y *Data Access*.

El proyecto NIF [44][49][50] tiene como objetivo brindar un acceso integrado a diversos recursos sobre neurociencia, disponibles en la web. Para ello, cuenta con una ontología de dominio, resultado de una mezcla de ontologías existentes y nuevas. Con base en esta ontología, el usuario puede hacer consultas basadas en conceptos para recuperar datos de sitios web, bases de datos relacionales, documentos XML y documentos en texto. A partir de los conceptos usados en la consulta de usuario, la aplicación crea consultas en los lenguajes o formatos soportados por cada una de las fuentes de datos. La aplicación soporta también consultas más complejas, que incluyen condiciones verificables con la ontología (ej. ser subclase-de) o condiciones sobre los datos. La aplicación provee un *framework* que guía al usuario durante la formulación de la consulta.

Por su parte, ACGT [38] [11] busca proveer acceso integrado a bases de datos clínicos, genéticos e imágenes, a través de una infraestructura *grid*, con el fin de dar soporte a estudios sobre cáncer. Para ello, cuenta con un mediador, que funciona como un servicio, que provee herramientas de descubrimiento de conocimiento y que, a su vez, es un cliente de los *wrappers* que dan acceso a las bases de datos. Para abordar los problemas de semántica presentes al integrar las fuentes de datos, usan una ontología sobre cáncer (*Master Ontology*) que permite la homogeneización, y usa SPARQL como lenguaje de consulta.

NeuroBank [45] [51] es un proyecto en desarrollo, que busca crear un sistema flexible para administrar, analizar y compartir el conocimiento sobre circuitos neuronales (neuronas y sus conexiones). Para ello se propone integrar información sobre el sistema nervioso provista por diversas entidades, y hacerla disponible para la comunidad de investigadores en esta área. La integración se basa en un modelo conceptual NeuroDM, y usa NeuroQL como lenguaje de consulta.

4. Modelos de Datos Conceptuales y sus Lenguajes de Consulta

Un modelo de datos [52] es un conjunto de herramientas que permiten representar las entidades del mundo real y sus relaciones; permiten estructurar y describir los datos, mantenerlos, y recuperarlos o consultarlos [53]. Un modelo de datos [54] incluye un conjunto de tipos de estructuras de datos, un conjunto de operadores o reglas de inferencia, y un conjunto de reglas de integridad. Los

modelos de datos se diferencian por los detalles semánticos que pueden expresar, identificándose cuatro niveles [55] [56]: físico, lógico, conceptual, y externo (de programación/presentación). El modelo físico describe la representación de los datos en los medios físicos (discos o memoria); en este nivel se definen registros, formatos, archivos, e índices. Los cambios en el modelo físico no impactan la lógica de la aplicación, pero si su rendimiento. El modelo lógico contiene una especificación de los datos que se van a almacenar; el modelo lógico más usado es el Relacional. El modelo conceptual representa el conocimiento específico de un dominio, usando conceptos que le son familiares a los expertos del mismo [57]. Este nivel se creó con el fin de facilitar la comunicación entre los usuarios y los diseñadores de sistemas de información. Entre los modelos conceptuales más conocidos están el Modelo Entidad Relación [58] y ORM [59]. Finalmente, el modelo externo, de programación/presentación, describe las vistas que se presentan al usuario de acuerdo con la actividad que él realiza. A nivel de programación están los objetos que implementan la lógica de la aplicación, los *web services*, o estructuras de memoria en las que se capturan los datos para presentarlos en la interfaz del usuario; a nivel de presentación, se tienen las interfaces de usuario que permiten expresar consultas.

Entre otros propósitos, un modelo de datos es la base para el desarrollo de lenguajes de consulta y manipulación de datos [54]. En el nivel externo, las interfaces generalmente restringen la expresividad de la consulta y no involucran operaciones complejas. Por su parte, los lenguajes de consulta en el nivel lógico son más expresivos, puesto que permiten definir consultas complejas; sin embargo para usuarios no-técnicos son difíciles de usar, debido a que la formulación de una consulta implica conocer la estructura de almacenamiento de los datos. En este nivel, el lenguaje de consulta más usado es SQL. Los Lenguajes de Consulta en el Nivel Conceptual pueden permitir al usuario expresar consultas con mayor nivel de abstracción, de forma similar a como lo hace en su dominio y con términos específicos del mismo [60]. Sin embargo, pocos lenguajes han sido propuestos en este nivel.

4.1. Referencia a Algunos Lenguajes Conceptuales de Consulta

Un Lenguaje en el nivel conceptual puede tener dos roles: la definición del modelo de datos, o la recuperación de datos. Estos últimos son llamados Lenguajes Conceptuales de Consulta, y permiten formular consultas con base en el modelo conceptual de los datos [57]. Los lenguajes conceptuales de consulta [61] se han propuesto con dos objetivos, el primero, formalizar las restricciones sobre los conceptos del modelo, formulando estas restricciones como una combinación de consulta y aserción sobre los resultados. El segundo objetivo, es proveer un lenguaje de alto nivel y fácil uso, para que los usuarios puedan, de forma sencilla, formular consultas complejas, obviando las desventajas que se presentan por la formulación de las consultas con base en los modelos lógico o físico.

En los últimos diez años, algunos lenguajes de consulta conceptuales han sido propuestos [62-64], [41], [65], [57]. En esta sección se hace una breve descripción de estos lenguajes, que incluye una referencia al modelo conceptual sobre el cual operan y a la estructura de sus expresiones, presentando un ejemplo.

CUDL (*Conceptual Universal Database Language*) [62] basado en FDB (*Frame Database Model*). FDB es cercano al modelo Entidad-Relación, y soporta tipos de datos compuestos. El objetivo de CUDL es proveer la posibilidad de manipular, directamente, tipos de datos compuestos y la expresión de restricciones a nivel de los subcampos. CUDL es un lenguaje declarativo, de alto

nivel, configurable para diferentes idiomas, y que no exige al usuario conocer los detalles técnicos del almacenamiento de los datos. Las sentencias en CUDL permiten definir, controlar, manipular y recuperar datos. Por ejemplo, la consulta “Recupere los títulos de los videos” se expresa en CUDL: *#Find data when entity = 'videos' and tag = 'title'.*

Constellation Query Language [63] permite definir modelos ORM (*Object Role Modeling*) y expresar consultas sobre los mismos. Los modelos y consultas se expresan en texto, debido a que los autores encontraron que esto mejora, con respecto al uso de modelos gráficos, la participación de los expertos de negocio en la definición del modelo conceptual. El modelo semántico constituye el vocabulario de las consultas, y está conformado por conceptos, hechos, instancias, restricciones y unidades. Este vocabulario puede definirse en cualquier idioma. El lenguaje de consulta está en proceso de definición, por lo cual no se proveen detalles del mismo, aunque si algunos ejemplos, como: *¿Person has givenName 'Daniel'?*

NeuroQL [64] es un lenguaje de consulta de dominio específico para datos de neurociencia, desarrollado siguiendo las metodologías [66] DSC-DM (*Domain Specific Conceptual Data Model*) y DSC-QL (*Domain Specific Conceptual Query Languages*), para modelado de datos conceptual y para la construcción de lenguajes de consulta de dominio específico, respectivamente. DSC-DM tiene una estructura similar al modelo Entidad Relación Extendido (EER) y permite la definición de funciones de usuario. DSC-QL se basa en los conceptos, relaciones y funciones de usuario definidas en el DSC-DM del dominio específico; e incluye características de OQL como expresiones *dot-path* y superclases; es independiente del DBMS y soporta tipos de datos y funciones específicas de las neurociencias (neurona, soma, connect, neighbors,...). NeuroQL fue implementado en el sistema NeuroBank, en el cual las consultas NeuroQL se traducen a expresiones SQL, OQL, o Xquery de acuerdo con el tipo de DBMS en el cual se encuentran los datos que se requieren. Una consulta NeuroQL tiene la estructura (*<result_list>*) [*<condition_list>*]; *<result_list>* es la lista de objetos a recuperar, y *<condition_list>* la lista de criterios que deben cumplir esos objetos. Como ejemplo, la consulta “*Find all neurons have molecule '5HT' and connect to another neuron, whose connid is '1'*”, se expresa en NeuroQL así: (*Neuron n*) [*n.hasMolecule.Molecule.name='5HT', n.connect.Connection.connid='1'*].

SCQL (*Semantically Complete Query Language*) [41] basado en SCM (*Semantically Complete Model*). En SCM solo una asociación es permitida entre dos conceptos, esto garantiza que el modelo sea semánticamente completo. SCQL expresa las consultas en términos de los conceptos, sin incluir las relaciones. Con ello se busca mantener la estructura del lenguaje simple, sin afectar el poder expresivo, ya que elimina factores de complejidad como el uso de nombres propios de las asociaciones (relaciones), elementos complicados en la estructura de la consulta, y asociaciones sin sentido. Una expresión de consulta en SCQL es un árbol de operaciones. Las operaciones de las hojas son selecciones, las operaciones intermedias pueden ser de composición (análogo a los *join* del algebra relacional), unión, diferencia o transformación. Como ejemplo, la consulta “traiga los números de teléfono de todas las personas que no tienen un alto nivel en alguna habilidad” se formula así: (*(Person, Phone) minus (Person, Skill, (Skill level = "High")*)).

COQL (*Concept-Oriented Query Language*) [65] es un lenguaje de representación y recuperación de datos en el modelo CODM (*Concept-Oriented Data Model*). Un modelo de datos orientado a conceptos, se caracteriza por su estructura jerárquica y multidimensional, que no distingue entre objetos y sus características. Todo concepto es caracterizado por un conjunto de dimensiones, cada una de las cuales hace referencia a otros conceptos. La semántica de un ítem depende de su posición

en la estructura, por ello el lenguaje de consulta enfatiza las conexiones entre conceptos. COQL es usado para la construcción de conceptos en el modelo de datos, por ejemplo: *ResultCube = CUBE (Cities city, Banks bank) WHERE (bank.name STARTSWITH 'A')*.

ConQuer (*Conceptual Query*) [57] está basado en ORM (*Object Role Modeling*). La consulta ConQuer se expresa como un árbol de predicados que conecta objetos. La interfaz visual presenta una lista de objetos que el usuario arrastra hacia la ventana de la consulta, para construir el árbol, donde también se especifican las condiciones y las operaciones sobre los datos. Los ítems que deben ser desplegados tienen una marca de aprobación. La semántica del lenguaje está especificada de dos formas, con base en operadores relacionales y con base en *bag comprehension*. ConQuer fue construido para la herramienta ActiveQuery, en la cual se proponía, inicialmente, la verbalización de la consulta en inglés. La consulta “Liste el id y el salario de los empleados que tienen un salario superior a \$90000, y hablan más de un idioma o manejan un carro”, se expresa en ConQuer así:

```

Employee
  +-- has Salary > 90000
  +-- either speaks count (Language) > 1
    |
    +-- or-- drives Car
      +-- has Color 'red'

```

5. Un lenguaje de consulta en el nivel conceptual

Los trabajos de integración de datos que se han realizado en el dominio médico muestran una tendencia a realizar las consultas del usuario por medio de conceptos o palabras clave, siendo la excepción MEDICO, en el cual se están desarrollando interfaces con lenguaje natural. Las consultas basadas en conceptos limitan la expresividad de lo que el usuario puede formular porque no es posible establecer relaciones entre estos conceptos o imponer condiciones sobre los mismos. Por otro lado, lenguajes como SPARQL o SQL proveen esa expresividad, pero ésta no puede ser explotada por el usuario de la aplicación, porque expresión de la consulta se basa en la estructura de los datos, por lo que la formulación de consultas en estos lenguajes exige conocer y entender esta estructura. Además, se genera cierta complejidad en la expresión la consulta cuando ésta esta basada en la estructura de los datos. Por ejemplo, en el caso de las consultas basadas en triplas, para recuperar datos en el modelo RDF, es necesario usar variables para recuperar datos, o para hacer *joins* entre las triplas. En el caso de SQL, un ejemplo de ello es la necesidad de hacer explícitos tanto los *joins* entre las tablas. Esta complejidad, difícilmente, la gestiona el usuario de una aplicación, aun cuando sea un experto en el dominio de la misma, lo cual ha generado la necesidad de simplificar la consulta especificada por el usuario confinándola a un grupo de palabras clave. De esta forma, las aplicaciones pierden la posibilidad de tomar ventaja del conocimiento del usuario experto, que podría permitir especificar mejor sus consultas para acceder a los datos.

Considerando la necesidad planteada en el dominio médico, de enriquecer la recuperación de imágenes médicas con información de otras fuentes de datos, y los enfoques de integración vigentes, se propone como alternativa una solución de integración de datos basada en un mediador. El mediador, a diferencia de las propuestas actuales, se basa en un lenguaje simple y de alto nivel de abstracción, dependiente del modelo conceptual de los datos, e independiente de su estructura. Este lenguaje estará basado en una vista del modelo conceptual de datos global que integra la información de las diversas fuentes. La vista del modelo conceptual ofrece un modelo simplificado, exento de los detalles propios del modelo conceptual, porque, un usuario experto, no requiere que,

explícitamente, estos detalles se representen en la consulta. El lenguaje permitirá acceder a diferentes tipos de datos y ofrecer diversos tipos de consulta: *Query By Example*, para recuperar datos multimedia, con base en una imagen de ejemplo, una ROI (*Region of Interest*), o con los descriptores de bajo nivel de la imagen. Y conservará, al menos, el mismo poder expresivo del álgebra relacional, la cual es base de la ejecución de consultas en SQL y SPARQL, de esta manera, se estaría garantizando la posibilidad de traducir las consultas a estos lenguajes para su ejecución en las fuentes de datos.

Agradecimientos: A la Doctora Martha Millán quien ha dirigido este trabajo, por su apoyo constante, y la acertada guía que me brinda. Y a la Pontificia Universidad Javeriana, quien apoya el desarrollo de estos estudios.

6. Referencias

- [1] D. L. Rubin, "Informatics Methods to Enable Patient-centered Radiology," *Academic Radiology*, vol. 16, n° 5, págs. 524-534, 2009.
- [2] D. L. Rubin, P. Mongkolwat, V. Kleper, K. Supekar, y D. Channin, "Medical imaging on the Semantic Web: Annotation and image markup," in *AAAI Spring Symposium Series, Semantic Scientific Knowledge Integration*, 2008.
- [3] C. Town, "Ontology based Visual Information Processing," University of Cambridge, 2004.
- [4] O. Karam, A. Hamad, y M. Attia, "Exploring the Semantic Gap in Content-Based Image Retrieval: with application to Lung CT," presented at the GVIP 05 Conf., Cairo, Egypt, 2005.
- [5] V. Khapli y A. Bhalachandra, "CBIR system for biomedical images: challenges and open issues," in *IET Intl. Conf. on Wireless, Mobile and Multimedia Networks*, págs. 85-88, 2008.
- [6] J. Assfalg, A. Del Bimbo, y P. Pala, "Three-dimensional interfaces for querying by example in content-based image retrieval," *IEEE Trans. on Visualization and Computer Graphics*, págs. 305-318, 2002.
- [7] C. Town y D. Sinclair, "Language-based querying of image collections on the basis of an extensible ontology," *Image and Vision Computing*, vol. 22, n° 3, págs. 251-267, Mar. 2004.
- [8] C. Wang, F. Jing, L. Zhang, y H. Zhang, "Scalable search-based image annotation of personal images," in *Proc. of the 8th ACM international workshop on Multimedia information retrieval*, págs. 269-278, 2006.
- [9] M. Flickner et al., "Query by Image and Video Content: The QBIC System," *Computer*, vol. 28, n° 9, págs. 23-32, 1995.
- [10] S. C. Orphanoudakis, C. E. Chronaki, y D. Vamvaka, "I2Cnet: Content-based similarity search in geographically distributed repositories of medical images," *Computerized Medical Imaging and Graphics*, vol. 20, n° 4, págs. 193-207, 1996.
- [11] R. Long, S. Antani, T. Deserno, y G. Thoma, "Content-Based Image Retrieval in Medicine: Retrospective Assessment, State of the Art, and Future Directions," *Intl. Journal of Healthcare Information Systems and Informatics*, vol. 4, n° 1, Mar. 2009.
- [12] A. Chalechale, G. Naghdy, y A. Mertins, "Sketch-based image matching using angular partitioning," *IEEE Trans. on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 35, n° 1, págs. 28-41, 2005.
- [13] M. Möller y M. Sintek, "A Generic Framework for Semantic Medical Image Retrieval," in *Proc. of the 1st International Workshop on Knowledge Acquisition from Multimedia Content KAMC'07*, vol. 253, 2007.
- [14] K. Barnard y D. Forsyth, "Learning the semantics of words and pictures," in *Proc. 8th IEEE Intl. Conf. on Computer Vision*, vol. 2, págs. 408-415 vol.2, 2001.

- [15] D. Joshi, J. Z. Wang, y J. Li, "The Story Picturing Engine---a system for automatic text illustration," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 2, n° 1, págs. 68-89, 2006.
- [16] D. Joshi et al., "Paragrab: A comprehensive architecture for web image management and multimodal querying," *Proc. of Very Large Databases, Seoul, Korea*, pág. 1166, 2006.
- [17] T. J. Mills, D. Pye, D. Sinclair, y K. R. Wood, "Shoebox: A Digital Photo Management System," *AT&T RESEARCH*, vol. 2000, 2000.
- [18] T. M. Deserno et al., "Extended query refinement for medical image retrieval," *Journal of Digital Imaging: The Official Journal of the Society for Computer Applications in Radiology*, vol. 21, n° 3, págs. 280-289, Sep. 2008.
- [19] V. E. Ogle y M. Stonebraker, "Chabot: Retrieval from a Relational Database of Images," *Computer*, vol. 28, n° 9, págs. 40-48, 1995.
- [20] B. Shah, R. Benton, Z. Wu, y V. Raghavan, "Automatic and Semi-Automatic Techniques for Image Annotation," in *Semantic-Based Visual Information Retrieval*, IRM Press, 2007, págs. 112-134.
- [21] M. Möller, M. Sintek, P. Buitelaar, S. Mukherjee, y X. S. Zhou, "Medical image understanding through the integration of cross-modal object recognition with formal domain knowledge," in *Proceedings of Healthinf 2008*, Madeira, Portugal: , 2008.
- [22] J. Y. Chai, C. Zhang, y R. Jin, "An empirical investigation of user term feedback in text-based targeted image search," *ACM Transactions on Information Systems (TOIS)*, vol. 25, n° 1, pág. 3, 2007.
- [23] X. S. Zhou y T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia Systems*, vol. 8, n° 6, págs. 536-544, 2003.
- [24] V. P. S. Rallabandi y S. K. Sett, "Knowledge-based image retrieval system," *Know.-Based Syst.*, vol. 21, n° 2, págs. 89-100, 2008.
- [25] M. C. Díaz-Galiano, M. Martín-Valdivia, y L. A. Ureña-López, "Query expansion with a medical ontology to improve a multimodal information retrieval system," *Comput. Biol. Med.*, vol. 39, n° 4, págs. 396-403, 2009.
- [26] L. Hollink y M. Worring, "Building a visual ontology for video retrieval," Hilton, Singapore, págs. 479-482, 2005.
- [27] S. Dasiopoulou, C. Doulaverakis, V. Mezaris, I. Kompatsiaris, y M. G. Strintzis, "An Ontology-Based Framework for Semantic Image Analysis and Retrieval," in *Semantic-based visual information retrieval*, IRM Press, 2007, págs. 269-293.
- [28] V. Mezaris, I. Kompatsiaris, y M. G. Strintzis, "An Ontology Approach to Object-Based Image Retrieval," in *Proc. IEEE Int. Conf. on Image Processing (ICIP03)*, págs. 511--514, 2003.
- [29] C. Town y D. Sinclair, "Ontological query language for content based image retrieval," in *Content-Based Access of Image and Video Libraries, 2001. (CBAIVL 2001). IEEE Workshop on*, págs. 75-80, 2001.
- [30] Y. Liu, D. Zhang, G. Lu, y W. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, n° 1, págs. 282, 262, Ene. 2007.
- [31] T. Joseph y A. F. Cárdenas, "Picquery: a high level query language for pictorial database management," in *Research foundations in object-oriented and semantic database systems*, Prentice-Hall, Inc., 1990, págs. 283-302.
- [32] A. F. Cardenas, I. T. Jeong, R. Barker, R. K. Taira, y C. M. Breant, "The Knowledge-Based Object-Oriented PICQUERY+ Language," *IEEE Trans. on Knowl. and Data Eng.*, vol. 5, n° 4, págs. 644-657, 1993.
- [33] S. Nepal, M. V. Ramakrishna, y J. A. Thom, "A fuzzy system for content based image retrieval," *Proc. 2 nd IEEE International Conference on Intelligent Processing Systems*, 1998.

- [34] R. Pein, J. Lu, y R. Wolfgang, "An extensible query language for content based image retrieval based on Lucene," in *8th IEEE Intl. Conf. on Computer and Information Technology, CIT 2008*, págs. 179-184, 2008.
- [35] D. Guliato, E. V. Melo, R. M. Rangayyan, y R. C. Soares, "POSTGRESQL-IE: An Image-handling Extension for PostgreSQL," *Journal of Digital Imaging*, vol. 22, n°. 2, págs. 149-165, 2008.
- [36] Z. Cao, Z. Wu, y Y. Wang, "UMQL: A Unified Multimedia Query Language," in *Signal-Image Technologies and Internet-Based System, International IEEE Conference on*, vol. 0, págs. 109-115, 2007.
- [37] M. Doller, R. Tous, M. Gruhne, K. Yoon, M. Sano, y I. Burnett, "The MPEG Query Format: Unifying Access to Multimedia Retrieval Systems," *Multimedia, IEEE*, vol. 15, n°. 4, págs. 82-95, 2008.
- [38] L. Martín et al., "Ontology Based Integration of Distributed and Heterogeneous Data Sources in ACGT," in *Proceedings of the First International Conference on Health Informatics, HEALTHINF 2008*, 2008.
- [39] J. D. Ullman, "Information Integration Using Logical Views," *Proc. of the 6th Int. Conf. on Database Theory ICDT'97*, vol. 1186, págs. 19-40, 1997.
- [40] D. Calvanese, G. De Giacomo, y M. Lenzerini, "Description Logics for Information Integration," in *Computational Logic: Logic Programming and Beyond*, vol. 2408, Springer, 2002.
- [41] V. V. Ovchinnikov, "Architecture of a semantic data integration system based on a semantically complete model and a semantically complete query language," *Programming and Computer Software*, vol. 32, n°. 4, págs. 228-242, 2006.
- [42] C. Bizer y R. Cyganiak, "D2RQ - Lessons Learned," Ene-2009. [Online]. Available: <http://www.w3.org/2007/03/RdfRDB/papers/d2rq-positionpaper/>. [Accessed: 11-Oct-2010].
- [43] D. Sonntag y M. Möller, "Unifying Semantic Annotation and Querying in Biomedical Images Repositories," in *Proceedings of the International Conference on Knowledge Management and Information Sharing (KMIS)*, 2009.
- [44] A. Gupta et al., "Federated Access to Heterogeneous Information Resources in the Neuroscience Information Framework (NIF)," *Neuroinformatics*, Sep. 2008.
- [45] R. Calin-Jageman et al., "Development of NeuronBank: A Federation of Customizable Knowledge Bases of Neuronal Circuitry," in *Services, 2007 IEEE Congress on*, págs. 114-121, 2007.
- [46] M. Möller y M. Sintek, "A Generic Framework for Semantic Medical Image Retrieval," in *Proceedings of the Knowledge Acquisition from Multimedia Content (KAMC) Workshop*, 2007.
- [47] M. Möller, S. Regel, y M. Sintek, "RadSem: Semantic Annotation and Retrieval for Medical Images," *The Semantic Web: Research and Applications*, vol. 5554, págs. 21-35, 2009.
- [48] M. Moller, N. Vyas, M. Sintek, S. Regel, y S. Mukherjee, "Visual Query Construction for Cross-Modal Semantic Retrieval of Medical Information," in *Proc. of the Malaysian Joint Conference on Artificial Intelligence MJCA*, 2009.
- [49] L. Marengo, Y. Li, M. Martone, P. Sternberg, G. Shepherd, y P. Miller, "Issues in the Design of a Pilot Concept-Based Query Interface for the Neuroinformatics Information Framework," *Neuroinformatics*, vol. 6, n°. 3, págs. 229-239, 2008.
- [50] A. Gupta, C. Condit, y X. Qian, "BioDB: An ontology-enhanced information system for heterogeneous biological information," *Data & Knowledge Engineering*, Jul. 2010.
- [51] Hao Tian, Yanchao Wang, Hong Yang, R. Sunderraman, P.S. Katz, y Ying Zhu, "A Novel Neuron Data Model with Domain Specific Query Language," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, págs.

- 6068-6071, 2005.
- [52] A. Silberschatz, H. F. Korth, y S. Sudarshan, "Data models," *ACM Computing Surveys*, vol. 28, n° 1, págs. 105-108, 1996.
 - [53] R. Angles y C. Gutierrez, "Survey of graph database models," *ACM Computing Surveys*, vol. 40, n° 1, págs. 1-39, 2008.
 - [54] E. F. Codd, "Data models in database management," presented at the the 1980 workshop, Pingree Park, Colorado, United States, págs. 112-114, 1980.
 - [55] J. A. Blakeley, D. Campbell, S. Muralidhar, y A. Nori, "The ADO.NET entity framework: making the conceptual level real," *ACM SIGMOD Record*, vol. 35, n° 4, pág. 32, 2006.
 - [56] C. Fankam, S. Jean, G. Pierra, L. Bellatreche, y Y. Ameur, "Towards Connecting Database Applications to Ontologies," presented at the Advances in Databases, Knowledge, and Data Applications, 2009. DBKDA '09. First International Conference on, págs. 131-137, 2009.
 - [57] T. Halpin, "Conceptual Queries," *Database Newsletter*, vol. 26, n° 2, Abr-1998.
 - [58] P. P. Chen, "The entity-relationship model--toward a unified view of data," *ACM Transactions on Database Systems*, vol. 1, n° 1, págs. 9-36, 1976.
 - [59] G. Bronts, S. Brouwer, C. Martens, y H. Proper, "A Unifying Object Role Modelling Approach," *Information Systems*, vol. 20, n° 3, págs. 213-235, 1995.
 - [60] M. Lawley y R. Topor, "A Query Language for EER Schemas," *ADC'94 Proc. of the 5th Australian Database Conf., Global Publications Service*, págs. 292--304, 1994.
 - [61] V. Owei, "Development of a Conceptual Query Language: Adopting the UserCentered Methodology," *The Computer Journal*, vol. 46, n° 6, págs. 602--624, 2003.
 - [62] N. N. Karanikolas, "Conceptual universal database language," presented at the the 2009 Euro American Conference, Prague, Czech Republic, págs. 1-5, 2009.
 - [63] C. Heath, "The Constellation Query Language," in *On the Move to Meaningful Internet Systems: OTM 2009 Workshops*, vol. 5872, 2009, págs. 682-691.
 - [64] H. Tian, R. Sunderraman, R. Calin-Jageman, H. Yang, Z. Ying, y P. Katz, Eds., "NeuroQL: A Domain-Specific Query Language for Neuroscience Data," in *Current Trends in Database Technology – EDBT 2006*, vol. 4254, Springer Berlin Heidelberg, 2006.
 - [65] A. Savinov, *Principles of the Concept-Oriented Data Model*. Institute of Mathematics and Computer Science, Academy of Sciences of Moldova, 2004, pág. 54.
 - [66] H. Tian y R. Sunderraman, "A Domain-Specific Conceptual Data Modeling and Querying Methodology," presented at the 1st Intl. Conf. on Information Systems, Technology and Management, 2007.